

Lecture Notes in Civil Engineering

Adel Francis
Edmond Miresco
Silvio Melhado *Editors*

Advances in Information Technology in Civil and Building Engineering

Proceedings of ICCCBE 2024, Volume 2,
Simulation and Automation



Series Editors

Marco di Prisco, *Politecnico di Milano, Milano, Italy*

Sheng-Hong Chen, *School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, China*

Ioannis Vayas, *Institute of Steel Structures, National Technical University of Athens, Athens, Greece*

Sanjay Kumar Shukla, *School of Engineering, Edith Cowan University, Joondalup, Australia*

Anuj Sharma, *Iowa State University, Ames, USA*

Nagesh Kumar, *Department of Civil Engineering, Indian Institute of Science Bangalore, Bengaluru, India*

Chien Ming Wang, *School of Civil Engineering, The University of Queensland, Brisbane, Australia*

Zhen-Dong Cui, *China University of Mining and Technology, Xuzhou, China*

Xinzheng Lu, *Department of Civil Engineering, Tsinghua University, Beijing, China*

Lecture Notes in Civil Engineering (LNCE) publishes the latest developments in Civil Engineering—quickly, informally and in top quality. Though original research reported in proceedings and post-proceedings represents the core of LNCE, edited volumes of exceptionally high quality and interest may also be considered for publication. Volumes published in LNCE embrace all aspects and subfields of, as well as new challenges in, Civil Engineering. Topics in the series include:

- Construction and Structural Mechanics
- Building Materials
- Concrete, Steel and Timber Structures
- Geotechnical Engineering
- Earthquake Engineering
- Coastal Engineering
- Ocean and Offshore Engineering; Ships and Floating Structures
- Hydraulics, Hydrology and Water Resources Engineering
- Environmental Engineering and Sustainability
- Structural Health and Monitoring
- Surveying and Geographical Information Systems
- Indoor Environments
- Transportation and Traffic
- Risk Analysis
- Safety and Security

To submit a proposal or request further information, please contact the appropriate Springer Editor:

- Pierpaolo Riva at pierpaolo.riva@springer.com (Europe and Americas);
- Swati Meherishi at swati.meherishi@springer.com (Asia—except China, Australia, and New Zealand);
- Wayne Hu at wayne.hu@springer.com (China).

All books in the series now indexed by Scopus and EI Compendex database!

Adel Francis · Edmond Miresco · Silvio Melhado
Editors

Advances in Information Technology in Civil and Building Engineering

Proceedings of ICCCBE 2024,
Volume 2, Simulation and Automation

Editors

Adel Francis
Département de génie de la Construction
École de Technologie Supérieure
Montreal, QC, Canada

Edmond Miresco
Département de génie de la Construction
École de Technologie Supérieure
Montreal, QC, Canada

Silvio Melhado
Département de génie de la Construction
École de Technologie Supérieure
Montreal, QC, Canada

ISSN 2366-2557

ISSN 2366-2565 (electronic)

Lecture Notes in Civil Engineering

ISBN 978-3-031-87363-8

ISBN 978-3-031-87364-5 (eBook)

<https://doi.org/10.1007/978-3-031-87364-5>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Contents

Big Data, Data Mining, Machine Learning and Artificial Intelligence

Transparency and Trust: Evaluating XAI for Critical Infrastructure Systems – A Comprehensive Analysis	3
<i>Jens Wala</i>	
Quantum Computing in Civil Engineering: Potentials and Limitations	15
<i>Joern Ploennigs, Markus Berger, Martin Mevissen, and Kay Smarsly</i>	
A Contribution to Process-Oriented Graph Data Management for Structural Data Exchange	29
<i>Marcel Heiß, Christian-Dominik Thiele, and Uwe Rüppel</i>	
Machine Learning and Internet of Things for Construction Air Pollution Monitoring and Prediction	44
<i>Ruichuan Zhang, Irene Paek, Samuel Park, and Jenna Krall</i>	
Exploring the Potential of BIM Models for Deriving Synthetic Training Data for Machine Learning Applications	54
<i>Simon K. Hoeng, Friedrich Eder, Marc Schmailzl, and Mathias Obergriesser</i>	
Natural Language Communication with Sensor Data Through a LLM-Integrated Protocol: A Case Study	64
<i>Fanglai Jia, Arianna Fonsati, and Kjartan Gudmundsson</i>	
Integrating Machine Learning for Cyber Risk Analysis in Construction 4.0	76
<i>Dongchi Yao and Borja García de Soto</i>	
Automated Hierarchical Object Clustering for Multi-Image Documentation: A Comprehensive Image Analysis Pipeline	92
<i>Michael Disser, Maximilian Gehring, and Uwe Rüppel</i>	
Image Segmentation for Enhanced Visualization and Accessibility of Historical Municipal Development Plans	107
<i>Phillip Schönfelder, Husan Duski, Jonas Maibaum, and Markus König</i>	
Optimization of BIM Collaboration Format Data Analysis Through Advanced Classification and Information Extraction	119
<i>Benedikt Kandler, Marcel Heiß, and Uwe Rüppel</i>	

Optimizing Truck Allocation in Open-Pit Mining Using a Deep Reinforcement Learning Policy	133
<i>Mohsen Hatami and Ian Flood</i>	
An Image-Based Approach for Construction Site Monitoring and Documentation Using Machine Learning	148
<i>Matthias W. Glunz, Florian Steinbach, Lina Wedekind, Martina Mellenthin Filardo, and Jürgen Melzner</i>	
Enhancing Information Extraction from German Regulatory Documents Using Deep Learning	159
<i>Sherief Ali and Markus König</i>	
Evaluating the Capabilities of Machine Learning Surrogate Modeling to Support High-Resolution Building Performance Simulation	169
<i>Elin Markarian, Seif Qiblawi, Shivram Krishnan, Anagha Divakaran, Albert Thomas, and Elie Azar</i>	
Data-Driven Predimensioning: Applying Graph Neural Networks to Reinforced Concrete Design	181
<i>Nils Schäfer, Jan-Friedrich Köhle, and Joaquín Díaz</i>	
Neural Network Based Defect Classification in Real-Field Rail Inspection Data Augmented by Simulated Defects	191
<i>Olm Georg</i>	
Federated Learning in Infrastructure Predictive Models: A Case Study of Utah's Culverts	205
<i>Pouria Mohammadi, Abbas Rashidi, and Sadegh Asgari</i>	
Dimensionality Reduction in Structural Reliability Analysis	217
<i>I.-Tung Yang and Jonathan Aloysius Budiman</i>	
An Artificial Intelligent Design System for Shear Wall Structures with Large Language Model Controlling Generation and Optimization	227
<i>Si-zhong Qin, Hong Guan, Wen-jie Liao, Yi Gu, Zhe Zheng, Hong-jing Xue, and Xin-zheng Lu</i>	
Efficient and Rapid Optimal Chiller Loading Using a Simplified Two-Stage Algorithm	238
<i>Mohamed S. Kandil, Gary Chang, and J. J. McArthur</i>	
Gray-Box Modeling of an Energy Recovery Ventilator Using Streamed Data ...	252
<i>Neda Adimi and J. J. McArthur</i>	

Near Real-Time Steel Rust Recognition Using Transformer-Based Convolutional Neural Networks	264
Shahinuzzaman, Po-Han Chen, and Luh-Maan Chang	
The Cost of Modelling Silicon-Based Materials with Machine Learning Potential	275
Karim Zongo, Laurent Karim Béland, and Claudiane Ouellet-Plamondon	
Autonomous Crack Segmentation Based on Segment Anything Model	283
Ghodsiyeh Rostami, Po-Han Chen, and Yang Wang	
Building the Blueprint for AI-Powered Compliance Checking: Analyzing ChatGPT-4 & Gemini by Question Category in Engineering Regulations	294
Entesar Al Nama and Maqsood Mahmud	
Data Mining and Feature Selection in Large Hospital Central Heating Plant	311
Marjan FatehiJananloo, Helen Stopps, and J. J. McArthur	
Automation and Robotics for Construction	
Automated Inspection in Mechanical and Electrical Construction	
Leveraging OpenVSLAM and Instance Segmentation	325
Wei Wei, Yujie Lu, Lijian Zhong, Yufan Chen, Ruihan Bai, and Yijun Lin	
Multi-Agent-Based Swarm Gas Source Localization Using Nano Aerial Robots	337
Jan Stührenberg, Felix S. Häusler, Patrick P. Neumann, Kosmas Dragos, and Kay Smarsly	
Autonomous Navigation of Quadruped Robots for Monitoring and Inspection of Civil Infrastructure	347
Aditya Tandon, Jan Stührenberg, Kosmas Dragos, and Kay Smarsly	
Investigating the Barriers to the Adoption of 3D Printing Technology in the Turkish Construction Industry	361
Leva Latifilkhechi, Saman Aminbaksh, and Emre Caner Akcay	
Electrodermal Activity as an Indicator of Cognitive Load for Human-Robot Collaboration in Construction	374
Seulbi Lee and Bogyeong Lee	
Development of Automatic Steel Deck Cutting System Using a Multi-Axis Robot for Power Plant Construction Work	383
Tsuyoshi Yamada, Naoya Shinoda, Kazuyuki Takeuchi, and Munechika Masaki	

Toward Future-Proof Asset Information Management	398
<i>Marlena Block, Alexander Buttgeret, and Markus König</i>	
Evaluating Automated Floorplan Generation: Benchmark on Residential Buildings	408
<i>Abdullah Elsafty and Timo Hartmann</i>	
Multi-Objective Optimization of a Sustainable LC3 Mortar for 3D Printing	421
<i>Willy Jin, Claudiane Ouellet-Plamondon, and Jean-François Caron</i>	
An Assessment of the Usage of Collaborative Robotics in the South African Construction Industry	429
<i>Ayanda Boya, Clinton Aigbavboa, and Opeoluwa Akinradewo</i>	
Analysis, Simulation and Sensing	
Development of an Electronic Guidance Device “eGuiDev” Based on Wireless Communication Technology for Pedestrian Navigation	443
<i>Abduaziz Juraboev, Abdulrahman Othman, Rüdiger Kern, Joaquín Díaz, and Uwe Rüppel</i>	
Simulation-Based Predictive Analytics for Construction Safety Incidents	455
<i>Abbey Dale Abellanoza, Hamidreza Golabchi, Stephen Hague, Simaan AbouRizk, and Yasser Mohamed</i>	
Formation Process Simulation Method Based on the Real Construction Data for Vibro Stone Columns	464
<i>Daoxiang Chen, Peng Lin, Zhiwei Zhang, and Guo Li</i>	
Prototyping a Wireless Hardware System for Automated Collection and Transmission of Audio and Kinematic Data at Construction Jobsites	473
<i>Seyedeh Zahra Golazad, Farzad Ordubadi, Abbas Mohammadi, Armin Tajalli, Abbas Rashidi, and Sadegh Asgari</i>	
Interactive Light and Sound-Sensitive Mortar for Architectural Environments	489
<i>Karina Hwang Arcolezi, Claudiane Ouellet-Plamondon, and Bora Ung</i>	
Automated FEM-Based Determination of Thermal Diffusivity and Temperature Profile for Rockfall Prediction	498
<i>Jaka Dujc and Mateja Jemec Auflič</i>	

Virtual Reality and Augmented Reality (VR/AR)

Augmented Reality Technologies in Education and Training: A Pathway to Enhancing the Built Environment	509
<i>Opeoluwa Akinradewo, Clinton Aigbayboa, Douglas Aghimien, Devon Rocha, Samuel Adekunle, and John Aliu</i>	
Architectural Design Revolution: Enhancing Student Learning Through VR ...	520
<i>Jean-Pierre Basson and Ashvin Manga</i>	
Ink or Pixels: Exploring the Impact of Hand-Drawing Skills on Learning Abilities in a Tech-Savvy Generation Z	533
<i>Jean-Pierre Basson and Riëtte Kotzé</i>	
VR-Based Computer Gaming Application in Digital Rehearsing of Mobile Crane Construction Operations	546
<i>Tan Qu and Marwan Shaban</i>	
Pre-design Stage: Computerizing Architect-Client Communication Using Virtual Reality and Artificial Intelligence	562
<i>Omar Bagasi, Nawari Nawari, and Fazil Najafi</i>	
Implementation of VDC and Virtual Reality to Meet the Deadline and Budget of Work. Case Study: UTEC CAMPUS BARRANCO—FASE 02	571
<i>Luis L. Espinoza Cauti and Enrique A. Juarez Aquino</i>	
Evaluating Impact of Smart Technologies Within Construction Site and Office Operations Among South African Contractors	585
<i>Johannes S. Mchunu, Irufka C. Anugwo, and Jachike C. Anugwo</i>	
Virtual Reality Experience and Motion Sickness in Construction Human-Robot Collaboration Learning	599
<i>Adetayo Onososen, Innocent Musonda, and Thembani Moyo</i>	

Technologies and Smart Data Management

Integrating Privacy-Preserving Occupancy Estimation Using Thermal Camera with a Digital Twin Platform	613
<i>Benjamin Rafatian, Pezhman Sharafdin, and Ali Motamed</i>	
Automated Registration of Ground 3D Point Cloud Data for Individual Buildings	626
<i>Huaiyuan Weng, Chul Min Yeum, Derek T. Robinson, and Bruce Macvicar</i>	

Automated Stakeout Survey for Efficient Earthwork Construction Project Control	640
<i>Sida Wang, Qingying Wu, Ming Lu, and Chen Wang</i>	
Exploring Critical Success Factors for Cloud Data Management in the South African Construction Industry	652
<i>Wanda Buhle Mpingana, Opeoluwa Akinradewo, and Clinton Aigbavboa</i>	
Author Index	669

Big Data, Data Mining, Machine Learning and Artificial Intelligence



Transparency and Trust: Evaluating XAI for Critical Infrastructure Systems – A Comprehensive Analysis

Jens Wala^(✉)

Technical University of Darmstadt, Darmstadt, Germany
wala@kritis.tu-darmstadt.de

Abstract. This paper explores the potential benefits and challenges of implementing Artificial Intelligence (AI) technologies in critical infrastructure systems, particularly Explainable Artificial Intelligence (XAI). As the shortage of skilled workers continues to grow, automating monitoring and control duties in critical infrastructures becomes increasingly necessary. However, the lack of transparency, trust, and explainability in traditional AI approaches creates hesitation among stakeholders. This paper evaluates the regulatory requirements for AI systems in critical infrastructures, analyzes how XAI can enhance resilience in these systems, and explores the potential design of an XAI-based autonomous system for critical infrastructure. Regulations from the European Union, Canada, and the United States are analyzed to evaluate the overall regulatory framework for AI applications in critical infrastructure. It also discusses how AI could optimize the performance of different critical infrastructure systems by predicting potential failures, reducing downtime, and extending the lifespan of infrastructure components. However, the lack of transparency in traditional AI techniques becomes a significant concern in fields where understanding the rationale behind decisions is crucial for ethical, legal, and practical reasons. The paper concludes by discussing how XAI can foster trust in AI systems through greater transparency and enhance the resilience of critical infrastructure systems.

Keywords: XAI · Trust · Transparency · AI Regulation

1 Introduction

Critical infrastructures form a complex socio-technical system with extensive monitoring and control requirements [1]. Supervisory Control and Data Acquisition (SCADA) systems are used to control and monitor elements of critical infrastructures [2]. Systems of this type enable rule-based automation by defining different actions depending on the input variables collected by industrial sensors [3]. The increasing shortage of skilled workers necessitates the automation of monitoring and control duties in critical infrastructures [4]. Many control center tasks are already highly automated, requiring experienced personnel only for specific interactions with the systems when action must be taken [5–7]. Changing policies and regulations concerning safety and security requires

an ever-growing effort in monitoring and record-keeping duties, which contributes to the problem [7].

This paper delves into the requirements, potential risks, and advantages of implementing Artificial Intelligence (AI) technologies in critical infrastructure systems. To achieve this, I have analyzed the regulations from the European Union, Canada, and the United States to evaluate the overall regulatory framework for AI applications in critical infrastructure. An exploration of how Explainable Artificial Intelligence (XAI) technologies can enhance the resilience of critical infrastructure systems and foster trust in AI systems through greater transparency was also conducted.

To address the ongoing shortage of skilled workers, it is necessary to utilize AI technologies. This is particularly important in critical infrastructure systems, where transparency, trust, and explainability are crucial when utilizing AI models for monitoring and control. To set the stage for further research, this paper addresses the following questions:

- RQ 1: What are the regulatory requirements for using AI systems in critical infrastructures?
- RQ 2: How can XAI enhance resilience in critical infrastructure systems?
- RQ 3: How could an XAI-based autonomous system for critical infrastructure look like?

2 AI in Critical Infrastructure: Efficiency vs. Transparency

2.1 Demand for AI in Managing Critical Infrastructure Systems

With the increasing complexity and scale of critical infrastructure systems, such as power grids, water supply networks, and transportation systems, the demand for skilled workers who can effectively monitor and manage these systems has escalated [4]. However, there is a growing shortage of such skilled professionals [4]. This gap can be significantly bridged by integrating AI technologies [8]. With their ability to process and analyze vast amounts of data quickly, AI systems can effectively perform tasks that normally require human intervention [8]. They can monitor system performance, detect anomalies, and even predict potential failures before they occur [9]. This could reduce the reliance on a large workforce while ensuring that the critical systems are under constant surveillance for any signs of malfunction or inefficiency, which could lead to a higher system resiliency.

AI technologies bring a new dimension of efficiency to the operation and management of various industries [8]. By leveraging machine learning algorithms and predictive analytics, AI can optimize the performance of these systems. For example, AI can predict demand patterns in power grids and adjust supply accordingly, reducing energy wastage [10]. In water management systems, it could optimize the distribution and predict potential leakages or contamination events [10]. Furthermore, AI-driven predictive maintenance can significantly reduce downtime and extend the lifespan of infrastructure components by scheduling repairs and maintenance activities before breakdowns occur [9]. Because systems used in critical infrastructure are considered critical, stakeholders remain hesitant to implement AI into their workflow due to the models' missing explainability [10].

2.2 The Transparency Gap in Traditional AI Approaches

One of the core issues with many traditional AI systems, especially those based on deep learning and complex neural networks, is their ‘black box’ nature [11, 12]. In these systems, the decision-making process is opaque; while the inputs and outputs are observable, the internal mechanisms that lead to a particular decision are not transparent [12]. This lack of explainability becomes a significant concern in fields where understanding the rationale behind decisions is crucial for ethical, legal, and practical reasons [12].

The issue of explainability is a persistent challenge in traditional AI approaches, which are often characterized by a high degree of complexity that further impedes interpretability [12]. Interpretability refers to the ease with which a human can comprehend the reasoning behind the AI system’s decision-making process [12, 13]. The high complexity of AI models, which often involve intricate algorithms and large datasets, makes it challenging for domain experts to interpret how the system arrived at a particular conclusion [12]. This lack of interpretability is not only a technical issue but also a barrier to the responsible deployment of AI, as it obstructs the assessment of the system’s resilience, reliability, and bias [10, 12]. Therefore, it is crucial to develop novel AI models with enhanced interpretability [12].

The transparency gap in traditional AI approaches thus poses significant hurdles for deployment in areas where human understanding is critical. Without explainability and interpretability, these AI systems can become unreliable in critical decision-making scenarios [14]. This gap can significantly hinder the deployment of AI applications to be put into use in critical infrastructure systems as regulations evolve and end-user trust in AI systems is low [12].

3 Regulatory Framework for AI in High-Risk-Sectors

3.1 European Union Artificial Intelligence Act

The use of artificial intelligence in the context of critical infrastructure is discussed in the European Union’s regulatory framework, the so-called EU Artificial Intelligence Act (AIA), which was proposed in 2021 [15]. It categorizes the use of AI according to the use case in which it is applied [15]. For that, it proposes four general use cases of AI depending on the risk assessment of usage: Prohibited AI practices, high-risk AI systems, limited risk application, and low/minimal risk application. According to the EU, prohibited AI practices pose an unacceptable risk [16]. Examples of unacceptable AI practices include the use of manipulative “subliminal techniques”, exploitation of vulnerable groups, social scoring, and real-time remote identification in public spaces on a persistent scale for law enforcement purposes [10].

According to Title III Article 6, high-risk AI systems include systems that are intended to be used as a safety component of products falling under the veil of the Union Harmonization legislation [17] (CE-seal) and systems deployed in specific areas, which include the management and operation of critical infrastructures [18]. Title III Chapter 2, Articles 8–15 outline requirements for Compliance, Risk management, data and data governance, technical documentation, record-keeping, as well as transparency,

human oversight, accuracy, robustness, and cybersecurity when leveraging AI systems in high-risk scenarios [18].

Another important aspect of Regulation is the European Union’s General Data Protection Regulation (GDPR) [19, 20]. Especially for users, Recital 71 of the GDPR requires the “right to an explanation” by which users can request a justification for the decision of an algorithm or model [21].

3.2 US Regulation on AI

The United States of America doesn’t have a unified regulation on AI applications [22]. In 2022, the White House proposed a blueprint for an “AI Bill of Rights”, which shall regulate individual rights regarding information when interacting with AI systems, combat discrimination, and data privacy [23]. This document, however, is not legally binding [24]. Several sector-specific regulations exist that govern the use of AI in different sectors, such as healthcare (HIPAA [25]) or finance (Gramm-Leach-Bliley Act [26]).

The National Institute of Standards and Technology (NIST) plays a key role in developing standards and guidelines for AI technologies, focusing on aspects like reliability, robustness, and trustworthiness [27]. In January 2023, they released the final version of the AI Risk Management Framework (AI RMF), which builds on the OECD Framework for AI system classification [28]. It defines a risk-based approach for classifying systems that use artificial intelligence and what risks are tolerable in those systems [28]. It specifies guidelines for when an AI system can be considered trustworthy [28]. For an AI-powered system to be trustworthy, according to the AI RMF, a system must be “valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed” [28, p. 12]. It is important to note that the AI RMF has a guideline character and is not legally binding [28].

Generally, the United States AI regulation can be summarized as a collection of risk-based approaches across highly distributed federal agencies and sectors [27].

3.3 Canadian AI and Data Act

The Canadian government has proposed the AI and Data Act (AIDA) as part of the Digital Charter Implementation Act, 2022 (Bill C-27) [29]. This act aims to establish a framework for the development, design, and deployment of AI systems that may impact Canadian citizens [29]. It seeks to regulate “high-impact systems” by improving accountability and transparency and ensuring safe and non-discriminatory operations [29]. AIDA builds upon existing legal frameworks for consumer protection, human rights, and criminal law and addresses specific regulatory requirements for handling AI usage [30]. To achieve this, a risk-based approach is proposed, considering existing international proposals and regulations like the EU-AIA and the Principles for AI by the OECD and the US National Institute of Standards and Technology’s Risk Management Framework [30].

The AIDA considers various key factors in determining if a system is considered “high-impact” [30]. A high-impact system is an AI system that has the potential to cause significant harm to individuals, society, or the environment if it fails or operates

erroneously [30]. For example, the Companion Document specifically names systems critical to health and safety [30].

The regulatory requirements contained in the AIDA require “appropriate measures” to be put in place to identify, assess, and mitigate risks associated with using AI during the whole lifecycle of the product [30]. These requirements include human oversight and monitoring, transparency, fairness and equity, safety, accountability, validity, and robustness [30].

3.4 Regulatory Commonalities, Differences, and Implications for High-Risk AI Systems

All three regions emphasize a risk-based approach to AI regulation. The EU’s Artificial Intelligence Act (AIA), the US’s AI Risk Management Framework (AI RMF), and Canada’s AI and Data Act (AIDA) all classify AI systems based on the level of risk they pose, ranging from high-risk to low-risk applications.

Each framework categorizes AI systems according to their potential impact and risks. For example, the EU AIA categorizes AI applications into prohibited practices, high-risk systems, limited-risk applications, and low/minimal-risk applications. Similarly, Canada’s AIDA focuses on “high-impact systems”, and the US’s AI RMF defines criteria for trustworthy AI systems. All frameworks pay special attention to AI systems that are considered high-risk, especially those used in critical infrastructure and impacting public safety and well-being.

All three approaches require a certain degree of interpretability and explainability in the use of models when applied in high-risk scenarios such as critical infrastructures.

Where the Canadian AIDA and the EU AIA have a legal enforcement mechanism, the US-AI-RMF is only a guideline for developing AI systems and is, therefore, not legally binding for the developers.

Compared to the AIDA and AIA, which are highly centralized, top-down regulations, the US guidelines and regulations are sectorally segregated from each other, which leads to a bottom-up view and complicates the regulatory process.

The examined regulatory frameworks require a certain degree of explainability and transparency to increase trust, ensure liability, and assign responsibilities for the development and deployment of critical systems, such as critical infrastructure control systems.

4 XAI in Critical Infrastructure Operation

4.1 Fundamentals of XAI

Each of the examined regulatory frameworks name transparency and interpretability as crucial aspects of successfully implementing AI technologies in critical infrastructure sectors [10]. Currently, there are efforts to develop more transparent AI models and techniques, such as Explainable AI (XAI), which aim to make AI decisions more understandable to humans [12]. The ultimate goal is to create AI systems that are powerful, efficient, transparent, and interpretable, which would foster trust and accountability in AI-driven

decision-making processes [30]. There are several methods to achieve explainability in machine learning models that use different approaches in demystifying decisions made by machine learning and AI models.

4.2 XAI Technologies and Methodologies

Intrinsically explainable machine learning models have a certain degree of explainability built directly into the model itself. They sacrifice some degree of model performance to allow for greater interpretability and transparency in their decision-making to circumvent the need for complex external explanation methods [9].

Linear regression models are often used because of their simplicity and interpretability. They predict the outcome as a weighted sum of the input features. The weights (or coefficients) assigned to each feature directly represent the relationship between that feature and the target variable. This makes it relatively easy to understand which features are most important and how changes in those features will affect the predicted outcome [11].

Generalized Additive Models (GAMs) extend linear models by allowing non-linear functions of the features while maintaining overall interpretability. In a GAM, the predicted outcome is the sum of smooth functions of the input features. Each feature has its own smooth function (often splines), which defines how it affects the target variable. This allows for more complex, non-linear relationships than linear regressors, but in a way that's still interpretable: you can plot these smooth functions to see exactly how each feature is affecting the prediction. GAMs are widely used in statistics for their flexibility and interpretability, especially in scenarios where understanding the impact of each feature is crucial [31].

Another kind of inherently explainable model are decision trees. They split the data into branches to make predictions, making their decision-making process very transparent. Each node in the tree represents a decision based on a single feature, leading to a clear path from the root to a leaf that embodies the prediction. When combined into forests (as in Random Forests), individual trees' predictions are aggregated, which can complicate interpretability slightly, but the importance of each feature can still be quantified by measuring how much it decreases the impurity in the nodes [11].

Depending on the kind of problem, the performance of intrinsically explainable models may not suffice. In that case, more complex models can be used. The problem with those is that, with growing complexity, transparency and interpretability decrease. Such models are often called “black boxes” as the decisions made by those models are not understandable. Various techniques and methodologies have been developed to make those kinds of models more explainable.

Feature importance is a post-hoc, model-agnostic way of determining a feature's impact on the output. Shapley Additive explanations (SHAP) values are based on the concept of Shapley Values from cooperative game theory [11]. For a given prediction, SHAP values explain the impact of each feature by shifting the dataset from the original prediction. SHAP can be used to generate consistent explanations and provide global as well as local interpretability [14].

Local interpretable model agnostic explanations (LIME) is a technique that explains individual model predictions by approximating the model locally with an interpretable

model [9]. It enables a high amount of local explainability but often lacks a global scope due to the limited performance of the chosen approximation model [9]. Compared to SHAP, it offers a faster computation time at the cost of scope of explainability.

Both approaches simplify the base model for the sake of explanation. The approaches often fail to capture the nuances of a given model. Those negatives must be considered when choosing a model for use in critical infrastructure systems.

4.3 Enhancing System Resilience with XAI

Incorporating XAI into critical infrastructure systems has the potential to improve their resilience, especially in the current climate of workforce challenges characterized by a shortage of skilled workers, high fluctuation, and dynamic operational environments [10].

The dynamic nature of critical infrastructure necessitates rapid response to unexpected changes or failures. XAI could provide immediate insights into why certain behaviors or anomalies occur within the system, enabling faster diagnosis and implement corrective measures. By leveraging the enhanced data analysis and forecasting capabilities of AI technologies, predictive maintenance measures could be taken to ensure the systems they control and monitor remain operational. This rapid problem-solving capability is essential for maintaining continuous operation and minimizing the impact of disruptions, contributing to system resilience.

Due to high workforce fluctuation and the widespread shortage of skilled workers, knowledge loss is a significant concern [10]. XAI could help enable knowledge transfer by providing new workers with essential skills and support in their work by providing insights into the complex inner workings of the systems. By making the decision-making processes of the AI more transparent, XAI could lower the threshold of expertise required to operate and manage sophisticated systems, thereby mitigating the risks associated with workforce limitations.

Critical infrastructure systems function in ever-changing environments with multiple potential points of failure [1]. These systems and their operators must be able to respond to various states of failure, including those that have never occurred before. AI systems could have the capability to adapt to new scenarios. By providing XAI with historical and current data, it could learn to adjust to new situations and alert operators of impending changes. This ability to adapt and provide advance notice can be extremely beneficial for decision-making.

4.4 Trust and Public Perception of AI Systems

The general populace is often distrustful of AI systems due to a lack of understanding. It remains unclear to them how such a system reaches the given decision and on what kinds of data this decision came to be. Especially in critical systems this could be a hurdle for adoption by workers. XAI could solve this problem by making AI's decisions and processes transparent and understandable [32].

Transparency is also critical in enhancing public trust. It's important for users to know when an AI is being used and for what purpose. This knowledge helps demystify the

technology rather than obfuscating the decision-making process. By being clear about when AI is being used in operational conditions and outlining its goals and limitations, trust in the technology [10]. Kraft, Zweig [33] have suggested that different levels of transparency should be enforced depending on the scope of AI applications to establish trust. More transparency is required for AI applications that have a more significant potential impact on the population in the case of misjudgment and where decisions are more centralized.

The shift to more explainability is crucial for the broader acceptance and ethical integration of AI technologies into critical infrastructure control systems.

4.5 Specific Requirements for Implementing XAI in Critical Infrastructure

To comply with upcoming regulations such as the EU-AIA or AIDA, developers creating models for critical infrastructure monitoring and control systems must ensure that their models are interpretable and transparent. This is because failures in such systems can have severe consequences for society. Therefore, it is important to focus on making AI model decisions understandable to human operators, who should be able to override certain decisions made by the system. Ideally, implementing XAI in critical infrastructures should involve using inherently explainable models, as many of these systems are in use for a long period of time.

To give the human operator an overview of the decisions the AI proposes for the physical system, a human-machine interface (HMI) must be implemented to allow for direct control and feedback for the AI. This feedback can then be used as context for the AI to make better decisions in future occurrences.

To manage data that is being collected over the lifecycle of the CI-System, a digital twin can be used, providing a way to aggregate data from different data sources and providing a context for an AI system to make decisions upon. A delay can be implemented before certain decisions are implemented to allow the operator to interfere with the decision before it is executed in the physical environment.

Figure 1 shows a system architecture that complies with these requirements. In the *physical space* (green), data is aggregated and sent to a *digital twin* of an infrastructure. This data can then be used by technologies in the *digital space* (blue). There, an AI-Framework analyses that data and can, with the context given during the setup, feed a decision-making model that was trained on the context as well as own training data. The context is also extended from the data in the data analysis step. The decision maker then generates inputs for the digital twin to be implemented in the physical space as well as feeds a user interface with the explanations of how that decision came to be. Depending on the decision's severity, an operator may decide to override the AI-based system's decision. Here, an operator can also give manual contextual inputs to the data foundation for the model. Finally, the commands given by either manual or autonomous input will be executed in the physical space by actuators and other control systems.

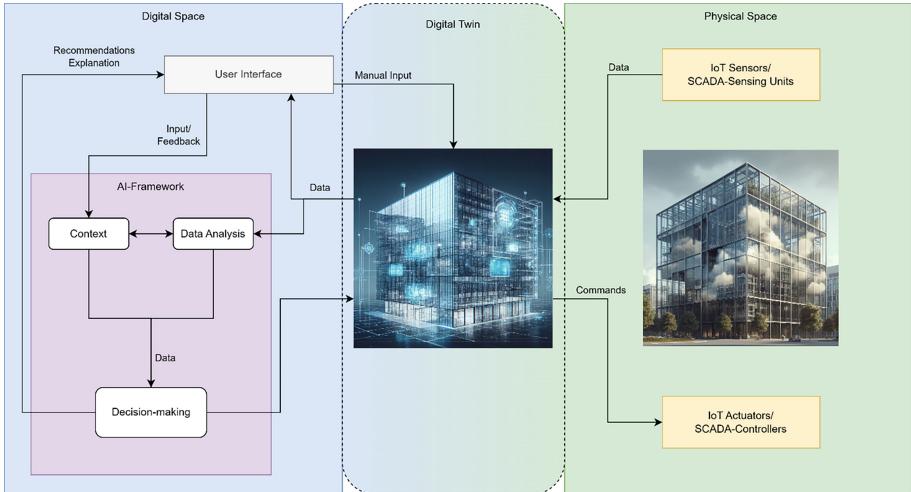


Fig. 1. A system architecture for autonomous monitoring and control of critical infrastructure

5 Discussion

This paper aims to explore the necessity and implications of Artificial Intelligence usage in the context of critical infrastructure monitoring and control systems. The regulatory base frameworks governing AI applications in the European Union, Canada, and the United States have been analyzed and concluded that although there are differences, many similarities exist as well. A risk-based approach is being taken when looking at AI usage in critical, high-risk systems, which necessitates the use of technologies from the explainable AI domain. A more detailed analysis of the regulations could have been conducted, including a detailed analysis of additional policies and regulations like the EU's general data protection regulation for example, but would have exceeded the scope of this paper.

Furthermore, general methodologies, techniques, and technologies from the domain of explainable AI were examined. The analysis of technologies doesn't include all available methods of explaining black boxes and omits many intrinsically explainable models, as a certain robustness is needed for use in critical infrastructure systems. The focus was, therefore, set on numerical models. Further analysis of the input data, as well as the application scenario, needs to be taken into consideration when choosing a model. Explainability is necessary not only for compliance with the requirements of the analyzed regulatory frameworks but also for making decisions that are understandable to the operators and public, thus building trust in the application. The use of black box AI systems would not suffice in high-risk scenarios such as critical infrastructures, where reliability and resilience are particularly important, and biases in the model could potentially have severe consequences for the safety and security of not only the system itself but also the general population.

A system schema that meets the requirements formulated in this paper has been proposed. The schema has been deliberately kept abstract in order to represent a complete

system of what a critical infrastructure control system might look like, illustrating the core components of the architecture. Future work should focus on the implementation of the specific systems, in particular the decision-making framework and the underlying model(s).

6 Conclusion and Future Work

This paper examines the use and necessity of AI in controlling critical infrastructure systems. Due to the growing challenges in infrastructure operation, such as a shortage of skilled workers and high fluctuation, the implementation of autonomous systems seems necessary to ensure the reliability and resilience of critical infrastructure. However, traditional AI approaches lack the explainability required, making them unsuitable for use within critical systems that may harm the general population in case of malfunction.

In turn, this paper analyzed the regulatory frameworks of the EU, Canada, and the US to gauge the feasibility of AI usage in critical or high-risk systems. The comparison of those frameworks showed that many of them share common goals and have similar regulatory implementation, accountability mechanisms, and focus on transparency.

Fundamental methods and technologies have been discussed on how to overcome the transparency gap that traditional AI approaches pose, as well as the resilience gains that could be achieved by leveraging XAI technologies.

Ultimately, specific requirements for an XAI system in critical infrastructure operation and a system implementing those were proposed.

Future works in the field of XAI in critical infrastructure could focus on the specific analysis of different AI models and their compliance with regulatory frameworks.

Newer developments in post-hoc explainability focus on maintaining model performance. Methods like SHAP or LIME could be used to build a compliant framework, but attention must be given that operators may not be able to understand the explanations given by those more complex and abstract explanation methods.

References

1. Fekete, A.: Kritische Infrastruktur und Versorgung der Bevölkerung: Klimawandel, Epidemien, digitale Transformation und das Risikomanagement. in *essentials*. Springer Berlin Heidelberg, Berlin, Heidelberg (2022). <https://doi.org/10.1007/978-3-662-65047-9>
2. Baker, T., et al.: A secure fog-based platform for SCADA-based IoT critical infrastructure. *Softw. Pract. Exp.* **50**(5), 503–518 (2020). <https://doi.org/10.1002/spe.2688>
3. Grilo, A.M., Chen, J., Diaz, M., Garrido, D., Casaca, A.: An Integrated WSAN and SCADA System for Monitoring a Critical Infrastructure. *IEEE Trans. Ind. Inform.* **10**(3), 1755–1764 (2014). <https://doi.org/10.1109/TII.2014.2322818>
4. Alberto: Fünf fundamentale Sicherheitsprobleme für kritische Infrastrukturen. [Online]. Available: <https://ap-verlag.de/fuenf-fundamentale-sicherheitsprobleme-fuer-kritische-infrastrukturen/66965/>. Accessed 06 June 2023.
5. “What is SCADA?” Inductive Automation [Online]. Available: <http://www.inductiveautomation.com/resources/article/what-is-scada>. Accessed 14 Apr 2023
6. “Water Treatment Plant - Remote Automation & Control System” [Online]. Available: <https://ctisupply.vn/en/water-treatment-plant-automation-control-system/>. Accessed 05 June 2023

7. Plank, S.: Wie KRITIS-Institutionen physische Sicherheitsmaßnahmen zur Gefahrenabwehr umsetzen können. kes, no. Februar 2022 [Online]. Available: https://www.kes.info/fileadmin/kundenbereich/Dokumente/Specials/2022/Special_Kritis_inkl._Leseprobe_.pdf. Accessed 06 June 2023
8. Alufaisan, Y., Marusich, L.R., Bakdash, J.Z., Zhou, Y., Kantarcioglu, M.: Does Explainable Artificial Intelligence Improve Human Decision-Making? (2021)
9. Burkart, N., Brajovic, D., Huber, M.F.: Explainable AI: introducing trust and comprehensibility to AI engineering. *Autom.*, **70**(9), 787–792 (Sep. 2022). <https://doi.org/10.1515/automat-2022-0013>
10. Laplante, P., Amaba, B.: Artificial Intelligence in Critical Infrastructure Systems. *Computer* **54**(10), 14–24 (2021). <https://doi.org/10.1109/MC.2021.3055892>
11. Gianfagna, L., Di Cecco, A.: Explainable AI with Python. Springer International Publishing, Cham (2021). <https://doi.org/10.1007/978-3-030-68640-6>
12. Andrae, K.: Erklärbare Künstliche Intelligenz - Tools und Methoden für erklärbare KI-Modelle. Jan. 18, 2021 [Online]. Available: <https://phaidra.fhstp.ac.at/detail/o:4486.pdf>. Accessed: 26 July 2023
13. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* **6**, 52138–52160 (2018). <https://doi.org/10.1109/ACCESS.2018.2870052>
14. Kale, A., Nguyen, T., Harris, F.C., Jr., Li, C., Zhang, J., Ma, X.: Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intell.* **5**(1), 139–162 (2023). https://doi.org/10.1162/dint_a_00119
15. “EU AI Act: first regulation on artificial intelligence | News | European Parliament.” [Online]. Available: <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. Accessed 16 June 2023
16. “Art. 5 AI-Regulation (Proposal)” [Online]. Available: https://lexparency.org/eu/52021PC0206/ART_5/. Accessed 23 August 2023
17. “Harmonised Standards.” [Online]. Available: https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en. Accessed 23 Aug 2023
18. “ANNEX III AI-Regulation (Proposal)—HIGH-RISK AI SYSTEMS REFERRED TO IN ARTICLE 6(2)” [Online]. Available: https://lexparency.org/eu/52021PC0206/ANX_III/. Accessed 10 Aug 2023
19. Saeed, W., Omlin, C.: Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowl. Based Syst.* **263**, 110273, Mar. 2023. <https://doi.org/10.1016/j.knosys.2023.110273>
20. “General Data Protection Regulation (GDPR)—Official Legal Text,” General Data Protection Regulation (GDPR) [Online]. Available: <https://gdpr-info.eu/>. Accessed 12 Dec 2023
21. “Recital 71 - Profiling,” General Data Protection Regulation (GDPR) [Online]. Available: <https://gdpr-info.eu/recitals/no-71/>. Accessed 12 Dec 2023
22. “US state-by-state AI legislation snapshot,” Bryan Cave Leighton Paisner - US state-by-state AI legislation snapshot [Online]. Available: <https://www.bclplaw.com/en-US/events-insights-news/2023-state-by-state-artificial-intelligence-legislation-snapshot.html>. Accessed 12 Dec 2023
23. The White House: Blueprint for an AI Bill of Rights [Online]. Available: <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>
24. The White House: Blueprint for an AI Bill of Rights | OSTP. The White House [Online]. Available: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>. Accessed 12 Dec 2023
25. O. for C. Rights (OCR): Health Information Privacy [Online]. Available: <https://www.hhs.gov/hipaa/index.html>. Accessed 12 Dec 2023

26. "Financial Services and AI: Regulatory Developments | Loeb & Loeb LLP" [Online]. Available: <https://www.loeb.com/en/insights/publications/2023/06/financial-services-and-ai-regulatory-developments>. Accessed 12 Dec 2023
27. The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment. Brookings [Online]. Available: <https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/>. Accessed 12 Dec 2023
28. Tabassi, E.: Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology (U.S.), Gaithersburg, MD, NIST AI 100-1 (Jan. 2023). <https://doi.org/10.6028/NIST.AI.100-1>
29. Innovation, Science and Economic Development Canada: Artificial Intelligence and Data Act [Online]. Available: <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act>. Accessed 08 Dec 2023
30. Innovation, Science and Economic Development Canada, "The Artificial Intelligence and Data Act (AIDA)—Companion document." Accessed: Dec. 08, 2023. [Online]. Available: <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>
31. Zschech, P., Weinzierl, S., Hambauer, N., Zilker, S., Kraus, M.: GAM(e) changer or not? An evaluation of interpretable machine learning models based on additive model constraints. <https://doi.org/10.48550/arXiv.2204.09123>. Accessed 19 Apr 2022
32. Fuchs, D.: Künstliche Intelligenz—Verstehbarkeit & Transparenz (AK13). self, Vienna, 0xc1aa5576_0x003d5f4a, 2023. <https://doi.org/10.1553/ITA-pb-2022-01>
33. Krafft, T., Zweig, K.A.: Transparenz und Nachvollziehbarkeit Algorithmenbasierter Entscheidungsprozesse. verbraucherzentrale Bundesverband, Jan. 22, 2019 [Online]. Available: https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf. Accessed 05 Jan 2024



Quantum Computing in Civil Engineering: Potentials and Limitations

Joern Ploennigs¹(✉), Markus Berger¹, Martin Mevissen², and Kay Smarsly³

¹ University of Rostock, Rostock, Germany

{Joern.Ploennigs,Markus.Berger}@uni-rostock.de

² IBM Research Europe, Dublin, Ireland

martmevi@ie.ibm.com

³ Hamburg University of Technology, Hamburg, Germany

Kay.Smarsly@tuhh.de

Abstract. Quantum computing is a new computational paradigm with the potential to solve certain computationally challenging problems much faster than traditional approaches. Civil engineering encompasses many computationally challenging problems, which leads to the question of how well quantum computing is suitable for solving civil engineering problems and how much impact and implications to the field of civil engineering can be expected when deploying quantum computing for solving these problems. To address these questions, we will, in this paper, first introduce the fundamentals of quantum computing. Thereupon, we will analyze the problem classes to elucidate where quantum computing holds the potential to outperform traditional computers and, focusing on the limitations, where quantum computing is not considered the most suitable solution. Finally, we will review common complex computation use cases in civil engineering and evaluate the potential and the limitations of being improved by quantum computing.

Keywords: Quantum computing · Civil engineering

1 Introduction

Civil engineering is a complex area with many challenging problems in design, construction, and operation, which leads to the adaptation of many digital technologies from computer aided design to construction robotics and machine learning, aiming to solve the underlying complex computational problems. The digitization of the industry is pushing the boundaries of what can be planned, analyzed, and optimized on current conventional computers, entailing a continuous need for new approaches.

Quantum computing (QC) is one of these new approaches. It may allow us to push some of the boundaries, as QC promises to compute numerous complex problems significantly faster than conventional computers. QC is explored in different industries [8, 25], but has not yet gotten much attention in civil engineering. The absence in civil engineering raises questions about the suitability of QC for addressing the computational challenges in civil engineering and the extent of its anticipated impact on the field. In

this paper, we will review promising QC methods that may be applied to current problems in civil engineering.

Therefore, we will first introduce the fundamental principles of quantum computing. We will delve into an analysis of scenarios where QC may demonstrate an advantage over traditional computers and identify areas where its application may be far in the future. With this context in mind, we will assess prevalent complex computational problems within civil engineering, determining problem classes that qualify for enhancement through QC.

2 Introduction to Quantum Computing

2.1 Fundamentals

Traditional computers are based on processing and storing binary data in the form of bits that are either 0 or 1. A modern CPU is specialized on processing the binary data, usually in blocks of 64 bits, by applying logical operators (AND, OR, XOR, NOT), numerical operators (addition, subtraction, multiplication and division), or specialized operators (encoding, encryption, etc.). Well established means have been proposed to represent any kind of data in binary code, such as numbers, texts, to images, and videos. The main assumptions behind binary encoding and processing are that (i) data is encodeable in binary 0 or 1 representation and (ii) all operators are strictly deterministic.

Quantum computing, by contrast, does not follow either of these assumptions, but uses a model of computation centered on *quantum bits* or *qubits* [69]. First, a qubit can be in a complex linear combination of both basis states $|0\rangle$ and $|1\rangle$, thanks to the principles of quantum *superposition*, which allows a single qubit to represent multiple states in a Hilbert space.

Secondly, quantum computers use the quantum mechanical effect of *entanglement*, where the state of one qubit becomes linked or correlated with the state of another qubit, even when separated by large distances. With those entangled qubits it is possible to create quantum gates that implement specific quantum operations as exemplified in the next section.

Third, due to the superposition principle (famously known as the Schrodinger's cat paradox), a qubit will collapse from its superposition state once measured and is identifiable only as one of the binary ground states. So, instead of being able to read the complex state between $|0\rangle$ and $|1\rangle$ directly, it is only possible to randomly sample binary states as either 0 or 1. Therefore, the qubit setup, computation and measurement operation needs to be repeated many times to retrieve a probability distribution representing the quantum state.

Figure 1 summarizes the process of loading data from a traditional computer into a quantum computer, applying quantum algorithms, and subsequently iterating through result retrieval. Several architectures are currently being explored to realize universal gate-based quantum computers at scale. These include superconducting quantum computing, trapped ion systems, linear optical quantum computing and others. For example, IBM has published a development roadmap for its superconducting quantum computing systems for the next 10 years [30].

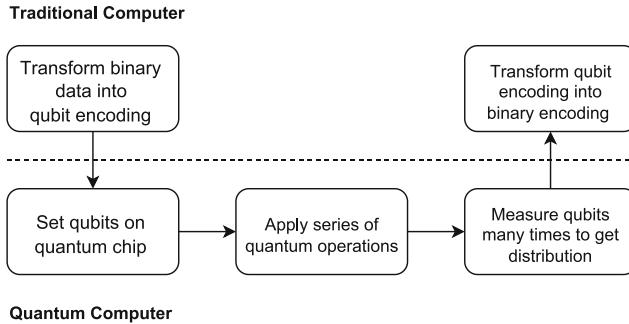


Fig. 1. Process for combining traditional computers with quantum computers.

2.2 Introduction to Quantum Circuits

In this section we want to discuss the universal quantum gates that form quantum circuits to illustrate the complexity of quantum computation. To visualize the states of one qubit, we use the Bloch sphere model [80]. On it the states of $|0\rangle$ or $|1\rangle$ are positioned at opposite poles. Points off the poles represents a superposition of qubit states.

Quantum gates alter the states of the qubits on the surface of this sphere. Notable among these are one-qubit gates like the Hadamard gate (H), which initiates superposition of the initial states, and the Phase gate (S), which rotates the qubit state on the sphere through complex space [60]. With these gates we can build a quantum circuit to negate a single qubit from $|0\rangle$ to $|1\rangle$ as shown in Fig. 2.

Another essential gate is the CNOT (Controlled-NOT-Gate) gate that operates on two qubits. It allows us to manipulate the state of one qubit based on the state of a second qubit. Specifically, the value of the second qubit (target) is either retained ($|00\rangle \rightarrow |00\rangle$; $|01\rangle \rightarrow |01\rangle$) or negated ($|10\rangle \rightarrow |11\rangle$; $|11\rangle \rightarrow |10\rangle$) depending on whether the first qubit (control) is $|0\rangle$ or $|1\rangle$, respectively. If the control is in a state of superposition, this results in the qubits getting entangled, as the state of the target is entirely dependent on how the superposition of the control resolves. Fig. 3 shows an example.

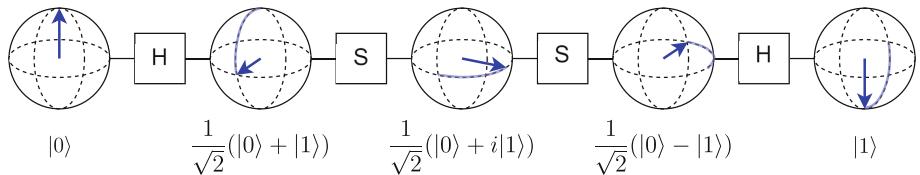


Fig. 2. A quantum circuit with a series of Hadamard gates (H) and Phase gate (S) to move a qubit from a pure $|0\rangle$ to a pure $|1\rangle$ state, shown as vectors on a Bloch sphere.

The CNOT gate and all one-qubit gates like the Hadamard and Phase gate form the set of universal gates that allow us to implement all other quantum gates and quantum algorithms. The resulting state of the target qubits at the end of the operation is measured multiple times, as discussed before.

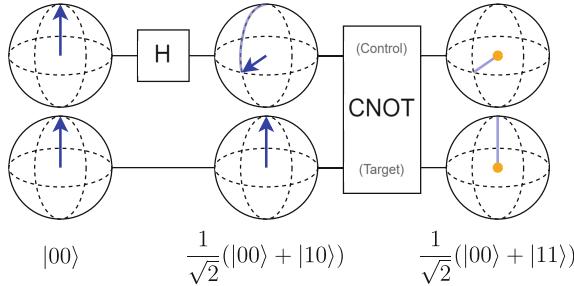


Fig. 3. A quantum circuit with two qubits. The Hadamard gate (H) places the control qubit of the CNOT gate into a superposition state, resulting in the qubits becoming fully entangled.

3 Challenges

There is a lot of ongoing fundamental research in quantum computing, and many challenges exist that limit its application at scale in practice. In this section, we aim to address some of the most prominent challenges, while dispelling common misconceptions about the current readiness of QC.

Error handling: Compared to classical computers, there are many sources of interference in quantum computing hardware, also called noise. This noise constrains the number of operational qubits. Current quantum computers are, therefore, *noisy quantum computers* [68]. As a result, one of the most prominent challenges in scaling up quantum computers is handling these errors affecting the actual quantum states and qubits. Broadly, we distinguish two approaches: Error mitigation and Error correction [12].

Error mitigation uses the outputs of quantum circuits to reduce the effect of noise in estimating expectation values and is an approach to improve the performance of current noisy quantum computer in the short term. Two state-of-the-art techniques for error mitigation are Probabilistic Error Cancellation [78] and Zero-Noise Extrapolation [39]. Those approaches estimate the noise and introduce additional computational overhead with limited scalability.

The long-term research targets Fault-Tolerant universal Quantum Computers (FTQC) by using error correction. *Error correction* is based on built-in redundancies, that is, on encoding each logical qubit through multiple redundant physical qubits [67] and running error correction code for detecting and correcting errors. The redundancy, while critical for error correction, poses a significant challenge to the potential scalability of FTQC, as solving practical problems may require millions of qubits [14]. Thus, it is expected to take a number of years before first FTQC can be realised. Therefore, both approaches feature a trade-off between the additional overheads introduced by running the error mitigation or correction method on the one hand and the reduction in noise achieved on the other.

Quantum Compilation: Over the past few decades, we have become accustomed to the convenience of enhancing our computational performance by either upgrading hardware components or adding more machines due to standardized hardware, like the IBM PC, and operating systems, such as Linux, Windows, or iOS. This familiarity has led to an expectation that transitioning to quantum computing should be equally

straightforward, allowing us to run our existing applications on quantum computers for vastly improved performance.

As shown in the introduction, quantum computers operate on fundamentally different principles than traditional computers, in terms of problem representation and solution methods. Consequently, we cannot simply cross-compile existing code to be run on a quantum computer in the same manner as we would do with a new CPU. Instead, each problem must be specifically adapted and implemented for quantum computation. Special programming languages for QC, such as IBM Qiskit, have been developed over the past few years [40]. Nevertheless, the development of efficient quantum compilers remains a significant challenge, with many approaches under investigation on how to optimize exact or approximate quantum compilation [57, 58]. As a result, quantum computing programming languages, while resembling modern high-level programming languages, often operate at a very low level, close to the quantum gates themselves, and require a profound understanding of the underlying quantum computing models. Thankfully, experts develop problem specific frameworks, e.g. for quantum chemistry, machine learning or optimization problems that lower the entry barrier to utilize quantum computers for those use cases.

The development of GPU computing offers some parallels that can provide valuable insights. While modern PCs are equipped with GPUs boasting hundreds, if not thousands, of cores—far more than traditional CPUs—GPUs have not replaced CPUs because many computational problems in computing are not *embarrassingly parallelizable* and, thus, well-suited for GPUs. Instead, most problems are highly sequential, allowing the faster CPUs to solve them much more efficiently. Parallelizable problems also need to be specially implemented for the GPU in dedicated programming languages or frameworks. For instance, NVIDIA introduced CUDA in 2007 [50] with support for many programming languages. Yet, its applications in civil engineering software remain limited [73]. Instead, it took several years until the area of neural networks took advantage of it to scale deep learning, which now also has a disruptive impact on civil engineering.

Quantum advantage: Most existing software runs efficiently on traditional computers, both from a computational and from an energy perspective [44]. For certain specific problem classes quantum computers have the potential to outperform classical computers in the near term. This is called *quantum advantage*, when we are able to demonstrate that the quantum computer can outperform classical computers and find problem instances where this speedup is useful. Early quantum algorithms, such as Deutsch–Jozsa [22], Shor [72], and Grover [33], were ground setting for this research area, but, have limited direct applicability on their own. The current focus is on finding problems where noisy quantum computers may provide computational advantage over classical computers considering that most algorithms require error correction, for example the widely re-used Shor algorithm for integer factorization. Indeed, there have been recent results demonstrating the utility of quantum computing for simulating the time evolution of 2D transverse-field Ising models [49]. There is a lot of ongoing research at the moment, in order to find further challenging problems, where noisy quantum computers—along with techniques like error mitigation—are demonstrated to be a useful tool.

4 Use Cases

4.1 Use Cases for Quantum Computing

In this section, we will explore use cases that are expected to exhibit a quantum advantage in the near future and share similarities with problems encountered in civil engineering. The use cases encompass: (i) *simulating* natural physical processes, (ii) *mathematical and machine learning* problems that process data with complex structure, and (iii) solving *optimization* problems as well as search.

Simulating nature represents one significant application area for quantum computers, particularly for systems operating (at or close to) the quantum level, such as simulating Hamiltonian dynamics or preparation of ground states. There is an increasing body of work that studies the properties of many body quantum systems with the help of quantum computers, e.g. [47]. Current exploration in the field includes applications in chemistry, such as the study of molecular structures [62] or chemical reactions [13]. Quantum computers facilitate the investigation of complex materials, including superconductors and novel compounds, allowing for more efficient comprehension of the properties and potential applications compared to classical methods. Leveraging the precision in simulating quantum mechanics, quantum computers are well-suited for providing more accurate predictions of molecular behavior, which is promising for future developments in drug discovery, materials design, and catalyst development [17]. A particularly encouraging recent result has been the demonstration of the utility of quantum computing for simulating the time evolution of 2D transverse-field Ising models [49].

Certain **Mathematical and Machine Learning** problems that process data with complex structure are expected to benefit from approaches based on quantum computing. In particular, demonstrating exponential speed-up is expected for quantum machine learning methods. While theoretic development of quantum machine learning algorithms goes back several decades, there has been a growing body of work on quantum machine learning approaches explored on current noisy quantum computers.

A promising class of quantum machine learning approaches are quantum kernel methods for classification problems, one of the most fundamental problems in machine learning. In a recent result, it was shown that there are specific classification tasks where a quantum kernel method with only classical access to data provides an exponential speed-up over classical machine learning algorithms [55]. It is an ongoing direction of research to develop supervised quantum machine learning algorithms for broader sets classification tasks.

Another notable group of quantum algorithms with potential applications in machine learning includes quantum phase estimation (QPE) [24] and the underlying quantum Fourier transform (QFT) [34], which is similar to the classical discrete Fourier transform and performs a Fourier transformation of the amplitudes but scales in $O(n^3)$ whereby the best-known classical algorithm requires $2^{O(n^{1/3})}$. QFT finds widespread usage in various other algorithms, including Shor's algorithm for integer factoring [72], the Harrow-Hassidim-Lloyd (HHL) algorithm for solving linear equations [36], and quantum gradient estimation [45]. One application area of Shor's algorithm is cryptography, where the algorithm enables efficient factorization of integers to find the prime factors.

The advancements made in solving specific mathematical matrix problems using QAOA have found applications in enhancing traditional machine learning models, including linear regression [20], clustering [31, 42], reinforcement learning [23, 52], support vector machines [38], and active learning [63]. Additionally, there is ongoing exploration of quantum algorithms for neural networks [48, 70]. Furthermore, quantum computing has been investigated for its potential to improve classical machine learning algorithms, particularly in the context of handling tensor and dot products in higher dimensions, thereby reducing computation time [19, 27]. In all these cases, quantum computing addresses specific aspects of traditional algorithms.

A sub-category of the aforementioned problems includes classical data analysis and statistics problems, which can be advanced through machine learning algorithms and, also, by more classically adapted methods, such as quantum principal component analysis [56] or quantum clustering [3]. Again, quantum advantage has not been shown to be practical for any of these models, yet, and achieving it on current noisy quantum computers is challenging due to the necessity of conducting numerous measurement runs and the stochastic nature of the objective function.

Optimization is another class of problems where quantum computing may be demonstrated to be useful, even if no exponential speed-up over classical computing is expected. However, various problems may also derive benefits from quantum computing [1]. Several gate-based optimization algorithms have been developed, involving the simulation of the evolution of a system through a sequence of quantum operators. Two common methods include the variational quantum eigensolver (VQE) [66] and the quantum approximate optimization algorithm (QAOA). Both methods are meta-heuristics for solving combinatorial optimization problems that can leverage gate-based quantum computers and potentially outperform purely classical heuristic algorithms [29]. The approaches mentioned above find application in logistical problems [5, 79], such as optimizing traffic networks to reduce congestion and enhance efficiency [81] or formulating and solving routing problems [37]. Similarly, the methods are employed for optimizing energy distribution networks [4]. A common feature among these applications is the utilization of a graph representation [32], where the problems are modeled as a Max-Cut problem.

Finding an exact solution to the Max-Cut problem is known to be NP-hard [46]. The objective of the Max-Cut problem is to divide the set of graph vertices into two subsets in a way that the sum of the weights of the edges connecting one subset to the other is maximized. Solving Max-Cut with QAOA has seen some focus since it can be formulated as an unconstrained quadratic unconstrained binary optimization (QUBO) problem and has wide number of applications. The approach with QAOA is based on the following mapping:

$$\underset{x}{\operatorname{argmin}} \sum_{i,j} w_{i,j} x_i x_j \stackrel{\substack{\text{mapped} \\ \text{to QAOA}}}{\rightarrow} H = \sum_{i,j} w_{i,j} \sigma_i \sigma_j$$

where $w_{i,j}$ are the weights for the binary assignment x_i and x_j . To solve the problem using QAOA, we construct the cost Hamiltonian H by mapping the binary assignment

variables x_k onto the eigenvalues of the Pauli Z operator σ_k . By employing the mapping, we obtain a representation on the right that closely resembles the original problem. It is expected, albeit unproven, that a quantum computer can solve this problem more efficiently than a classical one [71].

4.2 Use Cases in Civil Engineering

Based on this discussion of use cases for quantum computers we can derive some insights on which use cases in civil engineering may be most appropriate to benefit from quantum computers. We derive the mapping shown in Fig. 4 with good (bold), some (dotted) and limited (none) applicability as explained below.

Simulations are a common approach in civil engineering to evaluate design variants or to understand real-world phenomena. Some of those approaches may benefit from quantum computing.

Urban and traffic planning is using simulations specifically for modeling people and traffic flows and congestion within transportation networks, encompassing interactions among vehicles, infrastructure, and traffic signals. Traffic simulations rely on computational models that account for a multitude of variables, ranging from vehicle behavior and road conditions to traffic patterns [77]. These simulation models are less suited for quantum.

Structural analysis and design is commonly employing either rigid body models or finite element analysis (FEA) to simulate and analyze the behavior of structures under various conditions, including stress, heat transfer, and fluid flow. FEA, in particular, demands substantial computational resources as it involves dividing a structure into numerous small interacting elements to approximate the overall system behavior [82]. *Geotechnical engineering* problems are another application area for FEA. Here, engineers simulate soil-structure interactions, slope stability analysis, and foundation design [18]. *Seismic analysis and design* combines aspects of both structural and geotechnical engineering and focuses on designing structures capable of withstanding seismic forces, entailing complex analyses of how structures respond to earthquakes [43].

Hydraulic engineering often employs computational fluid dynamics (CFD) models to analyze the behavior of rivers, dams, coastal areas, and other hydraulic structures. The models, like FEA, break down problems into small interacting fluid (or gas) elements, but emphasize the solution of intricate fluid dynamics equations such as turbulence modeling, sediment transport, and interactions with structures [6].

Environmental Impact Assessment is looking into simulating the environmental impact of constructions often applying a combination of CFD or FEA methods [35].

The FEA and CFD problems fundamentally involve simulating the underlying physics in the form of higher-order differential equations, without relying on specific quantum mechanical effects for which quantum computers are well-suited. Instead, only quantum algorithms that enhance the efficiency of solving differential equations are explored [11, 54]. They break them down into systems of linear equations solvable with quantum algorithms, as the HHL [36].

A few applications to civil engineering problems can be found. Ajagekar and You [4] show a heat exchanger network synthesis problem, where multiple cooling units need to

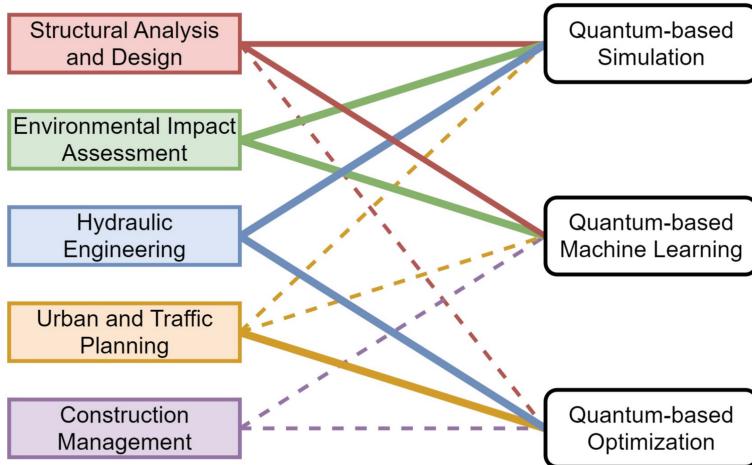


Fig. 4. An overview of engineering disciplines associated with the potential applicability of quantum computing. Lines highlight good (bold), some (dotted), or limited (none) applicability.

be balanced. Quantum approaches are utilized to simulate the behaviour and to identify the optimal solution.

Various **Mathematical and Machine Learning** problems exist in civil engineering, primarily in operation scenarios, but also in design and construction.

Structural analysis and design is using methods for *structural health monitoring* to record and analyze sensor data for assessing structural conditions of infrastructure, including the detection of defects or damage, aiming to support predictive maintenance and life-cycle management [28, 74]. *Predictive maintenance* is a related topic that focuses on analyzing system performance to predict maintenance actions [15], applicable not only to structures but to various types of assets.

Environmental Impact Assessment is another field using prediction models to analyze performance measures in general for a system like their energy consumption [2]. *Performance-based design* can conceptually use the models to evaluate different design options under various conditions or environmental impacts. If the prediction models are trained from simulation models, we speak of *surrogate models* [75].

Hydraulic and Energy Engineering is commonly targeting *demand prediction* of consumers in energy or water grids, or the traffic flow of people and cars in *Urban and Traffic Planning*. The objectives are similar to the simulation scenarios discussed earlier but rely on data-driven models instead of simulation models [53].

The specific machine learning models employed depend on the use case and data type. Typically, the data is recorded through sensors and stored in the form of time series. Regression models, such as linear regression or support vector machines, are often sufficient for prediction [65]. Quantum computing variants exist for these models, as discussed earlier [20, 38], but their necessity is debatable, as time series problems can be decomposed and trained individually at larger scale [64]. For problems with a graph structure, such as energy, water or traffic networks, Graph Neural Networks may be more appropriate [7], although more performant quantum variants are currently lacking.

In cases involving image or video data sets, deep learning models are commonly used, as the models tend to benefit more from GPU processing than quantum computation do [16, 21].

Optimization in civil engineering involves finding the best arrangement, combination, or configuration of elements within a discrete set of choices, considering various constraints and objectives.

Construction Management deals with efficiently assigning resources (e.g. labor, equipment, materials) in allocation and scheduling problems and schedule construction activities [41]. These problems can often be modelled as a QUBO problem [51].

Urban and traffic planning addresses the design of optimal traffic networks [76], which is an extension of the logistic routing problem discussed earlier [5, 79]. However, there has not been a discussion yet on the evaluation of the impact of the algorithms on network design. Conceptually related is the *site location selection* problem, which refers to determining the most suitable locations for infrastructure facilities, such as schools, hospitals, waste management sites, or transportation hubs. The determinations take into account factors, such as accessibility, environmental impact, population distribution, and cost [61]. The problems can often be modeled as coverage problems [26], and thus we may be able to map the problems onto a QUBO problem, which, in turn, might be solved by QAOA. However, in practice, the number of constraints and weights is often higher, as demonstrated by Farahani et al. [26].

Hydraulic engineering and distribution networks planning problems are prevalent in civil engineering, spanning various forms from power to fresh and wastewater networks. Early work explores the use of quantum computers for optimizing energy distribution networks [4]. Similar approaches may apply to other network types by mapping the problems to graph Max-Cut problems.

Environmental Impact Assessment is usually more looking in using optimization for improving energy efficiency. Here the optimization approaches are either control problems or scheduling problems.

Structural analysis and design is optimizing the arrangement and sizing of structural components (beams, columns) that ensure structural integrity while minimizing material usage [9, 59]. Optimization often requires solving complex mathematical problems, such as nonlinear programming or genetic algorithms. Considering the problem as a rigid body problem with a graph representation may allow to map it to QADA. For example, Wang et al. [79] discuss an application for a topology optimization problem called the Messerschmitt-Bolkow-Blohm (MBB) problem [10] to find the stiffest design of a desired volume fraction. Traditionally, such design problems were solved by engineers through experience or trial and error.

5 Conclusion

In this paper, we have discussed the current state of quantum computing and potential applications in civil engineering. We first provided an overview of the fundamental concepts of quantum computing, highlighting its distinctions from traditional binary computing. Subsequently, we have addressed the current challenges associated with scaling quantum computing, with a specific focus on algorithms and problem classes

that can be studied with current, noisy quantum computers. We then have reviewed the areas where quantum computing has the potential to outperform traditional computing, namely in simulating quantum mechanics, quantum machine learning approaches, and classes of optimization problems. For each area, we have also introduced the relevant use cases in civil engineering and have discussed the potential of applying QC within them. The preliminary conclusions are summarized as follows:

- **Simulation:** Among the most computationally demanding tasks are FEA and CFD problems, where quantum computing holds the potential to enhance performance in solving differential equations. Here, the discretization into linear equation systems is an important step.
- **Machine Learning:** Quantum computing shows the largest promise in the speed-up that may be achieved through quantum kernel methods for a variety of classification tasks. In addition, advancements in machine learning involve algorithmic steps that may have the potential to perform better on a quantum computer. To benefit from these future possibilities, it is recommended to keep investigating the applicability of these classification models in civil engineering.
- **Optimization:** Quantum computing has already been explored for a number of optimization problems arising across different application domains. There is a lot of ongoing research in the speed up that may be achieved for particular algorithms. In civil engineering this means to identify relevant problem-mappings to solutions like Max-Cut or QUBO.

Quantum computing has the potential to catalyze a revolutionary paradigm shift similar to the transformative impact observed with GPU computing, currently reshaping the landscape of science through the empowerment of deep learning. The innovations in quantum computing will not arise primarily from accelerating existing code; instead, innovations are expected to derive from the ability to address an entirely new set of challenges using specialized software applicable across a broad spectrum of domains.

References

1. Abbas, A., Ambainis, A., Augustino, B., Bartschi, A., Buhrman, H., et al.: Quantum optimization: potential, challenges, and the path forward (2023). arXiv e-prints, art. [arXiv:2312.02279](https://arxiv.org/abs/2312.02279)
2. Ahmed, A., Korres, N.E., Ploennigs, J., Elhadi, H., Menzel, K.: Mining building performance data for energy-efficient operation. *Adv. Eng. Inform.* **25**(2), 341–354 (2011)
3. Aïmeur, E., Brassard, G., Gambs, S.: Quantum clustering algorithms. In *ICML*, pp. 1–8 (2007)
4. Ajagekar, A., You, F.: Quantum computing for energy systems optimization: challenges and opportunities. *Energy* **179**, 76–89 (2019)
5. Ajagekar, A., Humble, T., You, F.: Quantum computing based hybrid solution strategies for large-scale discrete-continuous optimization problems. *Comput. Chem. Eng.* **132**, 106630 (2020)
6. Anderson, D., Tannehill, J.C., Pletcher, R.H., Munipalli, R., Shankar, V.: Computational Fluid Mechanics and Heat Transfer (2020)
7. Ba, A., O'Donncha, F., Ploennigs, J., Azmat, M.: Efficient extraction of insights at the edges of distributed systems. In *IEEE BigData* (2023)

8. Bayerstadler, A., et al.: Industry quantum computing applications. *EPJ Quant. Technol.* **8**(1), 25 (2021)
9. Belegundu, A.D., Chandrupatla, T.R.: Optimization Concepts and Applications in Engineering (2019)
10. Bendsoe, M.P., Sigmund, O.: Topology Optimization: Theory, Methods, and Applications (2003)
11. Berry, D.W.: High-order quantum algorithm for solving linear differential equations. *J. Phys. A: Math. Theor.* **47**(10), 105301 (2014)
12. Cai, Z., et al.: Quantum error mitigation. *Rev. Mod. Phys.* **95**(4), 045005 (2023)
13. Cao, Y., Romero, J., Olson, J.P., Degroote, M., Johnson, P.D., Kieferova, M., Kivlichan, I.D., Menke, T., Peropadre, B., Sawaya, N.P., et al.: Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**(19) (2019)
14. Chatterjee, A., Stevenson, P., De Franceschi, S., Morello, A., de Leon, N.P., Kuemmeth, F.: Semiconductor qubits in practice. *Nature Rev. Phys.* **3**(3), 157–177 (2021)
15. Cheng, J.C., Chen, W., Chen, K., Wang, Q.: Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Autom. Construct.* **112** (2020)
16. Combarro, E.F., Gonzalez-Castillo, S., Di Meglio, A.: A Practical Guide to Quantum Machine Learning and Quantum Optimization: Hands-on Approach to Modern Quantum Algorithms (2023)
17. Daley, A.J., et al.: Practical quantum advantage in quantum simulation. *Nature* **607**(7920), 667–676 (2022)
18. Das, B.M.: Principles of Geotechnical Engineering (2021)
19. Dasari, V.R., Im, M.S., Beshaj, L.: Solving machine learning optimization problems using quantum computers. In: Disruptive Technologies in Information Sciences IV, vol. 11419, pp. 60–69 (2020)
20. Date, P., Potok, T.: Adiabatic quantum linear regression. *Sci. Rep.* **11**(1), 21905 (2021)
21. Date, P., Schuman, C., Patton, R., Potok, T.: A classical-quantum hybrid approach for unsupervised probabilistic machine learning. In: Future of Information and Communication Conference (FICC), pp. 98–117 (2020)
22. Deutsch, D.: Quantum theory, the church–turing principle and the universal quantum computer. In: Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, vol. 400, no. 1818, pp. 97–117 (1985)
23. Dong, D., Chen, C., Li, H., Tarn, T.-J.: Quantum reinforcement learning. *IEEE Trans. Syst., Man, Cybern., Part B (Cybernetics)* **38**(5), 1207–1220 (2008)
24. Dorner, U., et al.: Optimal quantum phase estimation. *Phys. Rev. Lett.* **102**(4), 040403 (2009)
25. Egger, D.J., et al.: Quantum computing for finance: state-of-the-art and future prospects. *IEEE Trans. Quant. Eng.* **1**, 1–24 (2020)
26. Farahani, R.Z., Asgari, N., Heidari, N., Hosseiniinia, M., Goh, M.: Covering problems in facility location: a review. *Comput. Ind. Eng.* **62**(1), 368–407 (2012)
27. Fastovets, D.V., Bogdanov, Y.I., Bantysh, B.I., Lukichev, V.: Machine learning methods in quantum computing theory. In: International Conference on Micro- And Nano-Electronics 2018 (2019). <https://doi.org/10.11117/12.2522427>
28. Flah, M., Nunez, I., Ben Chaabene, W., Nehdi M.L.: Machine learning algorithms in civil structural health monitoring: a systematic review. *Arch. Comput. Methods Eng.* **28** (2021)
29. Fuchs, F.G., Lye, K.O., Møll Nilsen, H., Stasik, A.J., Sartor, G.: Constraint preserving mixers for the quantum approximate optimization algorithm. *Algorithms* **15**(6), 202 (2022)
30. Gambetta, J.: The hardware and software for the era of quantum utility is here. IBM Research Blog (Dezember 2023) (2023)
31. Gao, L.-Z., Lu, C.-Y., Guo, G.-D., Zhang, X., Lin, S.: Quantum k-nearest neighbors classification algorithm based on mahalanobis distance. *Front. Phys.* **10**, 1047466 (2022)

32. Goodrich, T.D., Sullivan, B.D., Humble, T.S.: Optimizing adiabatic quantum program compilation using a graph-theoretic framework. *Quant. Inf. Process.* **17**, 1–26 (2018)
33. Grover, L.K.: A fast quantum mechanical algorithm for database search. In: An ACM Symposium on Theory of Computing, pp. 212–219 (1996)
34. Hales, L., Hallgren, S.: An improved quantum Fourier transform algorithm and applications. In: An Symposium on Foundations of Computer Science, pp. 515–525 (2000)
35. Harish, V., Kumar, A.: A review on modeling and simulation of building energy systems. *Renew. Sustain. Energy Rev.* **56**, 1272–1292 (2016). ISSN 1364-0321
36. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.* **103**(15), 150502 (2009)
37. Harwood, S., Gambella, C., Trenev, D., Simonetto, A., Bernal, D., Greenberg, D.: Formulating and solving routing problems on quantum computers. *IEEE Trans. Quant. Eng.* **2**, 1–17 (2021)
38. Havlíček, V., et al.: Supervised learning with quantum-enhanced feature spaces. *Nature* **567**(7747), 209–212 (2019)
39. He, A., Nachman, B., de Jong, W.A., Bauer, C.W.: Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Phys. Rev. A* **102**(1), 012426 (2020)
40. Heim, B., et al.: Quantum programming languages. *Nature Rev. Phys.* **2**(12), 709–722 (2020)
41. Hinze, J.: Construction planning and scheduling, vol. 228 (2004)
42. Horn, D., Gottlieb, A.: Algorithm for data clustering in pattern recognition problems based on quantum mechanics. *Phys. Rev. Lett.* **88**(1), 018702 (2001)
43. Humar, J.: Dynamics of Structures (2012)
44. Jaschke, D., Montangero, S.: Is quantum computing green? An estimate for an energy-efficiency quantum advantage. *Quant. Sci. Technol.* **8**(2), 025001 (2023)
45. Jordan, S.P.: Fast quantum algorithm for numerical gradient estimation. *Phys. Rev. Lett.* **95**(5) (2005)
46. Karp, R.M.: Reducibility Among Combinatorial Problems (2010)
47. Keenan, N., Robertson, N.F., Murphy, T., Zhuk, S., Goold, J.: Evidence of Kardar-Parisi-Zhang scaling on a digital quantum simulator. *NPJ Quant. Inf.* **9**(1), 72 (2023)
48. Kerendis, I., Landman, J., Prakash, A.: Quantum algorithms for deep convolutional neural networks (2019). arXiv preprint [arXiv:1911.01117](https://arxiv.org/abs/1911.01117)
49. Kim, Y., et al.: Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**(7965), 500–505 (2023)
50. Kirk, D., et al.: NVIDIA CUDA software and GPU parallel computing architecture. In ISMM, vol. 7 (2007)
51. Kochenberger, G.A., Glover, F.: A unified framework for modeling and solving combinatorial optimization problems: a tutorial. In: Multiscale Optimization Methods and Applications, pp. 101–124 (2006)
52. Lamata, L.: Basic protocols in quantum reinforcement learning with superconducting circuits. *Sci. Rep.* **7**(1), 1609 (2017)
53. Lassoued, Y., Monteil, J., Gu, Y., Russo, G., Shorten, R., Mevissen, M.: A hidden Markov model for route and destination prediction. In: IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6 (2017)
54. Liu, J.-P., Kolden, H.Ø., Krovi, H.K., Loureiro, N.F., Trivisa, K., Childs, A.M.: Efficient quantum algorithm for dissipative nonlinear differential equations. *Proc. Natl. Acad. Sci.* **118**(35) (2021)
55. Liu, Y., Arunachalam, S., Temme, K.: A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* **17**(9), 1013–1017 (2021)
56. Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum principal component analysis. *Nat. Phys.* **10**(9), 631–633 (2014)
57. Madden, L., Simonetto, A.: Best approximate quantum compiling problems. *ACM Trans. Quant. Comput.* **3**(2), 1–29 (2022)

58. Mariella, N., Zhuk, S.: A doubly stochastic matrices-based approach to optimal qubit routing. *Quantum Inf. Process.* **22**(7), 264 (2023)
59. Mei, L., Wang, Q.: Structural optimization in civil engineering: a literature review. *Buildings* **11**(2), 66 (2021)
60. Nielsen, M.A., Chuang, I.L.: *Quantum Computation and Quantum Information* (2010)
61. O'Donncha, F., Sheehan, J., Lorenzi, F., Ploennigs, J.: Large language models and knowledge graphs to optimize commercial real estate portfolio. In: INFORMS An. Meeting (2023)
62. Outeiral, C., Strahm, M., Shi, J., Morris, G.M., Benjamin, S.C., Deane, C.M.: The prospects of quantum computing in computational molecular biology. *WIREs Comput. Mol. Sci.* **11**(1), e1481 (2021)
63. Paparo, G.D., Dunjko, V., Makmal, A., Martin-Delgado, M.A., Briegel, H.J.: Quantum speedup for active learning agents. *Phys. Rev. X* **4**(3), 031002 (2014)
64. Patel, D., Lin, S., Shah, D., Jayaraman, S., Ploennigs, J., Bhamidipati, A., Kalagnanam, J.: AI model factory: scaling AI for industry 4.0 applications. In: AAAI, vol. 37, pp. 16467–16469 (2023)
65. Pathak, N., Ba, A., Ploennigs, J., Roy, N.: Forecasting gas usage for big buildings using generalized additive models and deep learning. In: IEEE SMARTCOMP, pp. 203–210 (2018)
66. Peruzzo, A., et al.: A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**(1), 4213 (2014)
67. Postler, L., Heußen, S., Pogorelov, I., Rispler, M., Feldker, T., Meth, M., Marciniak, C.D., Stricker, R., Ringbauer, M., Blatt, R., et al.: Demonstration of fault-tolerant universal quantum gate operations. *Nature* **605**(7911) (2022)
68. Preskill, J.: Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018)
69. Preskill, J.: Quantum computing 40 years later. In: Feynman Lectures on Computation, pp. 193–244 (2023)
70. Schuld, M., Sinayskiy, I., Petruccione, F.: The quest for a quantum neural network. *Quant. Inf. Process.* **13**, 2567–2586 (2014)
71. Shaydulin, R., Alexeev, Y.: Evaluating quantum approximate optimization algorithm: a case study. In: International Green and Sustainable Computing Conference (IGSC), pp. 1–6 (2019)
72. Shor, P.W.: Algorithms for quantum computation: discrete logarithms and factoring. In: An. Symposium on Foundations of Computer Science, pp. 124–134 (1994)
73. Simpson, B.G., Zhu, M., Seki, A., Scott, M.: Challenges in GPU-accelerated nonlinear dynamic analysis for structural systems. *J. Struct. Eng.* **149**(3), 04022253 (2023)
74. Smarsly, K., Dragos, K., Wiggenbrock, J.: Machine learning techniques for structural health monitoring. In: 8th European Workshop on Structural Health Monitoring (2016)
75. Sun, H., Burton, H.V., Huang, H.: Machine learning applications for building structural design and performance assessment: state-of-the-art review. *J. Build. Eng.* **33**, 101816 (2021)
76. Toth, P., Vigo, D.: *Vehicle Routing: Problems, Methods, and Applications* (2014)
77. Treiber, M., Kesting, A.: Traffic flow dynamics. In: *Traffic Flow Dynamics: Data, Models and Simulation*. Springer Berlin Heidelberg, pp. 983–1000 (2013)
78. Van Den Berg, E., Minev, Z.K., Kandala, A., Temme, K.: Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors. *Nat. Phys.* 1–6 (2023)
79. Wang, Y., Kim, J.E., Suresh, K.: Opportunities and challenges of quantum computing for engineering optimization. *J. Comput. Inf. Sci. Eng.* **23**(6) (2023)
80. Wiseman, H., Milburn, G.: Interpretation of quantum jump and diffusion processes illustrated on the bloch sphere. *Phys. Rev. A* **47**(3), 1652 (1993)
81. Yarkoni, S., Neukart, F., Tagle, E.M.G., Magiera, N., Mehta, B., Hire, K., Narkhede, S., Hofmann, M.: Quantum shuttle: traffic navigation with quantum computing. In: ACM SIGSOFT International Workshop on Architectures and Paradigms for Engineering Quantum Software, pp. 22–30 (2020)
82. Zienkiewicz, O.C., Taylor, R.L., Zhu, J.Z.: *The Finite Element Method: Its Basis and Fundamentals* (2005)



A Contribution to Process-Oriented Graph Data Management for Structural Data Exchange

Marcel Heiß^(✉), Christian-Dominik Thiele, and Uwe Rüppel

Institute of Numerical Methods and Informatics in Civil Engineering, Technical University of Darmstadt, Darmstadt, Germany
heiss@iib.tu-darmstadt.de

Abstract. Digital transformation in the construction industry has led to the introduction and use of heterogeneous software systems. Unified interfaces are essential to enable their use. These require a high degree of standardisation and have little flexibility in mapping data, so interfaces are often a bottleneck for data. Parallel modelling results in asynchronous data processing and often redundant data that must be manually reconciled on an ongoing basis. This unconnected data in different software systems leads to the accumulation of heterogeneous, distributed, and inaccessible data sources and knowledge silos. The presented approach aims at developing a single source of truth in structural design based on a graph database using labelled property graphs (LPG). In the context of this work, the suitability of graphs for the holistic, digital representation of structural building design data is investigated by integrating structural design mindsets and processes. In addition to an overarching methodology showing the future integration of a graph database into existing structural design data exchange structures, a data structure is developed for mapping structural design data into a LPG. A prototypical implementation demonstrates that a graph-based digital representation of structural design data is suitable for data management and exchange of basic structural data, and can be extended to process-oriented data. The focus of graph databases on relationships between data corresponds to the high relevance of connections between components within the design. The developed approach integrates the previous way of working with partial models and positional planning, and also enables bi-directional data exchange between BIM and structural analysis throughout the process. This allows for holistic and interconnected data management without loss of context and data exchange on a unified database. This form of linked data also enables the use of data-driven machine learning approaches, particularly graph-based approaches such as graph neural networks.

Keywords: Graph database · Structural engineering · Standardisation

1 Introduction

Heterogeneous data from disparate software systems is a challenge for interoperable data management and collaboration in the Architectural Engineering and Construction (AEC) industry [1]. Building Information Modelling (BIM) and its associated digital twins

have increasingly established as a central method of data exchange as part of the digital transformation. However, it is becoming apparent that further expansion of the associated Industry Foundation Classes (IFC) standard is increasing its complexity and reducing its manageability [2]. As a result, its suitability for mapping and enabling data exchange in individual disciplines is limited. However, these require a high degree of standardisation and therefore have little flexibility in mapping data, so interfaces are often a bottleneck for data. Interfaces and data exchange formats also ensure that data is only exchanged at certain points in time. Individual and parallel modelling leads to asynchronous data processing, often resulting in redundant data that must be manually reconciled on an ongoing basis. This disconnected data across software systems leads to the accumulation of heterogeneous, distributed and inaccessible data sources as well as knowledge silos, both across and within domains. The Single Source of Truth (SSoT) concept addresses this problem by collecting and aggregating data or knowledge from different systems in a central location, breaking down data silos. The benefits are many, including improved data accuracy, data reliability and improved collaboration. However, implementing SSoT is often associated with challenges. Integrating multiple systems into a single source of truth is complex and time-consuming due to the large number of systems in use and the associated different data models, formats and standards. In structural engineering, the synergy between data exchange of BIM models of architecture, structural design and their analysis models is mainly seen to automate recurring, manual and time-consuming tasks [3]. In structural engineering, different software from one or more vendors is used to solve different problems. Depending on the construction project, this may include wind flow analysis, bracing to ensure the overall stability of a structure, determination of profile parameters, analysis of connections and seismic analysis. While it is often still possible to exchange data between software from one vendor in proprietary formats, the ability to exchange data with software from another vendor is limited. This is important when using software from different vendors within a company, or when working with another structural engineer on a project. In addition, the integration of the digital transformation of work processes in structural engineering must be considered, including the partial model-based way of working, which contrasts with the BIM-based way of working. While BIM models primarily include the architectural components of a building, in structural analysis, only partial models are often modelled and individual components are dimensioned and grouped in a position plan [3].

The approach presented is based on a specially developed *Labeled Property Graph* (LPG) model for the holistic representation of structural design data to create a universal data model in structural engineering. Furthermore, the developed concept includes the integration of structural design processes and mindsets. This enables holistic and networked data management without loss of context, and data exchange to share data between designers and structural engineers across multiple systems in a unified database. Firstly, several existing standards, approaches and methodologies are presented. The advantages and disadvantages of the existing standards and methods and their areas of application are then analysed and a rationale for the choice of LPG is given. An overarching methodology of this approach, which envisages the use of a LPG-CDE (*Common Data Environment*) in structural design, is presented and the representation of structural data in a LPG database is introduced. A data interface concept is then presented that shows the use of LPG in interaction with different application programming interfaces

(API) of example software based on a possible workflow. In an application with some example data the data interface between different software is tested by following the process workflow. In a following chapter the results are presented and discussed by comparing the use of LPG with existing data exchange formats. A conclusion summarises some of the main aspects and results of this work and presents possible further research.

2 Related Work

Data mapping and data exchange in the AEC industry is a key issue in digital transformation and a current research topic. Various formats and approaches have been developed to exchange data and information between software systems and stakeholders. As a global standard, the IFC format plays a central role in data exchange in the AEC industry. In this schema it is possible to describe semantic and geometric information for the building industry [1]. In addition, an official standard, the ifcOWL, serialises the IFC data schema in a graph-based manner using the Semantic Web concept and associated ontologies. As the ifcOWL is an exact translation into OWL (Ontology Web Language) and RDF serialisation of the IFC schema, the limitations and advantages of the IFC also apply to the ifcOWL [4, 5]. IfcOWL offers the possibility to link the IFC schema with other standards of other domains and to link building data with the help of applied technologies and to extend it more flexibly than the IFC schema itself. However, the translation does not simplify the objectified relationships of IFC, resulting in a rather verbose and overly complicated ontology for which several simplifications have been proposed [6]. Various approaches use SM and ontologies to extend building data representations for specific applications. For structural design, approaches such as [7, 8] process geometric and semantic information for automated structural analysis using an open graph-based description. In [9], an ontology-based approach is investigated to support holistic structural design. In addition to the RDF format, so-called labelled property graphs (LPG) are another way of representing data based on graphs, allowing attributes to be assigned directly to individual nodes and edges [10]. Compared to ifcOWL, LPG has a relatively compact data structure and BIM information in a general sense in many applications [11]. Previous research has focused on the use of LPG to map IFC data, while its use for data management and exchange in structural engineering is completely unexplored. Other research is already using graph representations of building structures as a basis for graph neural networks in structural design to support structural analysis [12, 13]. There is also a need for a standardised data exchange format in structural engineering. There are a number of proprietary and vendor-neutral exchange formats for exchanging different types of data in structural design, such as bar structures, surface structures or reinforcement data. The best-known vendor-neutral exchange format is the IFC *Structural Analysis View*, a so-called *Model View Definition* (MVD) of IFC. The development of the *Structural Analysis Format* (SAF) has provided the first open format for the exchange of structural analysis models only. Outside of these formats, structural data exchange is limited within a single vendor's system or is enabled by direct interfaces or various other formats in the AEC industry. While IFC is typically used in the form of horizontal interoperability to enable data exchange between BIM software

environments, so-called vertical interoperability refers to data exchange between BIM software environments and structural software environments. Data exchange between different structural software environments is also a form of horizontal interoperability, for example with SAF.

3 Analysis

A gap in research is horizontal interoperability between different structural software environments with a neutral data exchange format for holistic structural design data, which includes any type of structural data such as bar structures, surface structures and reinforcement data. Additional requirements for a neutral data exchange format include its use for vertical interoperability. Figure 1 schematically illustrates the horizontal and vertical data exchange that currently takes place in practice. A distinction is made between horizontal interoperability at the BIM design level, where different roles often use different software (SW) and enable data exchange with IFC and *Common Data Environments* (CDE), and horizontal interoperability at the domain-specific software level, for example with SAF. Another option is the IFC Structural Analysis View, which contains a description of the structural model, loads and load combinations, and can be used for vertical interoperability to derive a structural model and horizontal interoperability at the domain-specific software level. Collaboration and data exchange between design and structural analysis models is a current research topic and is investigated in several research papers such as [14]. Due to the dramatically different representations of the building structure, a transfer from a BIM model to the structural analysis profile centre-line model is not straightforward [14]. Automated pre-processing is therefore required to avoid much manual work, as investigated in [15]. Over the years of development, the complexity of the IFC data model has reached a critical point. With hundreds of classes, thousands of attributes and a dense network of relationships between classes [16], only a few experts fully understand the functionality and behaviour of the entire data model. Further extension would increase its complexity and reduce its manageability [2]. The SAF format has therefore proven to be an easy way to exchange data in structural engineering, alongside the IFC Structural Analysis View. It is an Excel-based file format based on a relational database model. The Excel format does not allow quick access and querying of relationships, nor is it easy to manage a large number of projects. In addition to exchange formats such as SAF and IFC, direct interfaces, which are often used for vertical interoperability, do not require an exchange format, but are always software dependent and therefore do not provide a means for neutral data exchange, and are therefore primarily an option for closed systems from a single vendor or in a multi-vendor collaboration. Besides data exchange in structural engineering, which is the focus of this paper, other domains use their own data exchange formats to supply various applications with data.

In order to represent the data in a more comprehensible and understandable way compared to IFC, graph-based research approaches using RDF and LPG seem to offer a promising solution. RDF can be better used in situations where the focus is on linking heterogeneous global data and machine inference, while LPG can be better used in situations where the focus is on data storage and retrieval [11]. Compared to ifcOWL,

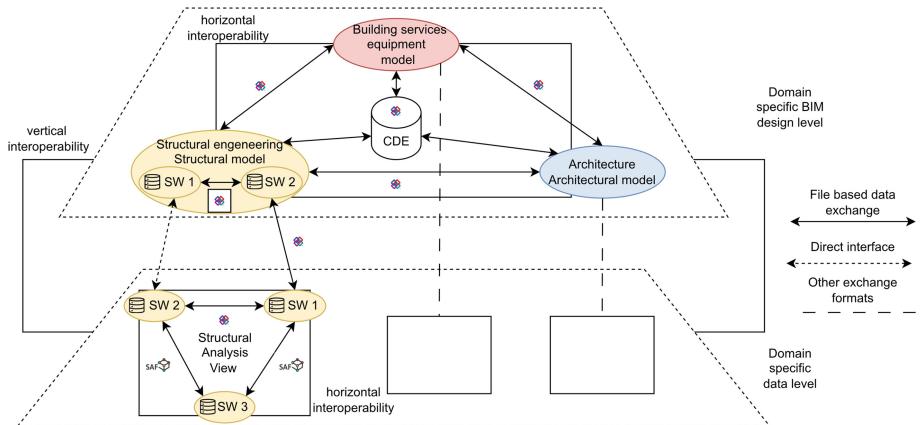


Fig. 1. Schematic representation of the horizontal and vertical data exchange that currently takes place in practice.

LPG has a relatively compact data structure allowing information such as an *IfcPerson* instance to be stored within a single node in LPG, whereas an RDF graph must use multiple nodes and relationships, as shown in [11]. LPG provides easy access to information using common database APIs and effective information retrieval using advanced query languages. Compared to relational and other NoSQL databases, the graph database is better at representing relationships. LPG databases often offer a high degree of flexibility in implementing a schema-free approach, where instance graphs can be populated with literally any data. However, this flexibility comes at a price, namely the potential risk of incompatibility with applications that rely on specific data structures. As a result, many graph databases implement a “metamodel” that clearly defines the type of nodes, their attributes, and the potential connections between nodes, mirroring the concept of a schema.

First, structural design is highly dependent on the entire building structure with its components and their connections. In order to manage structural design data, the relationships between components are central, hence a graph-based approach is adopted in this study. Secondly, the aim of this study is to enable analytical model access and querying, rather than linking data across domains for machine reasoning, so LPG is used as the basis for further work.

4 Methodology

The overarching methodology underlying this approach envisages the use of a LPG-CDE at the domain-specific software level in the area of structural engineering and is shown in Fig. 2. It is based on the idea that different data models reflect the different ways of thinking in different specialist roles. For this reason, a central, cross-domain data model is very complex and challenging. Instead, different domain-related data models can store domain-relevant data within one data model and form the data basis for different software systems within a domain. The LPG-CDE is intended to prioritise

horizontal interoperability with an understandable and comprehensible data structure at the domain-specific software level in structural engineering. In a further step, BIM-specific structural design software will also be linked to the LPG-CDE through vertical interoperability. In a final stage, the LPG-CDE data can also be made available to other domains via restricted access in the form of vertical interoperability. Data relevant to other professional roles can be retrieved to share data intersections. While in Stage 1 the LPG-CDE is considered as the sole data source and thus SSoT, in level 2 it is considered as a complementary data source to the BIM software data to access structural analysis data and BIM relevant data from the LPG-CDE to enrich the BIM model with additional data. The main objective is to improve the collaboration and data exchange between the designer using the BIM software and the structural engineer using the structural analysis software. The aim of this work is to investigate the suitability of a LPG in terms of its potential and limitations for such a level 1 and 2 CDE. Analogue to LPG, other data bases can be used for domains like architecture as central data storage at the domain-specific data level.

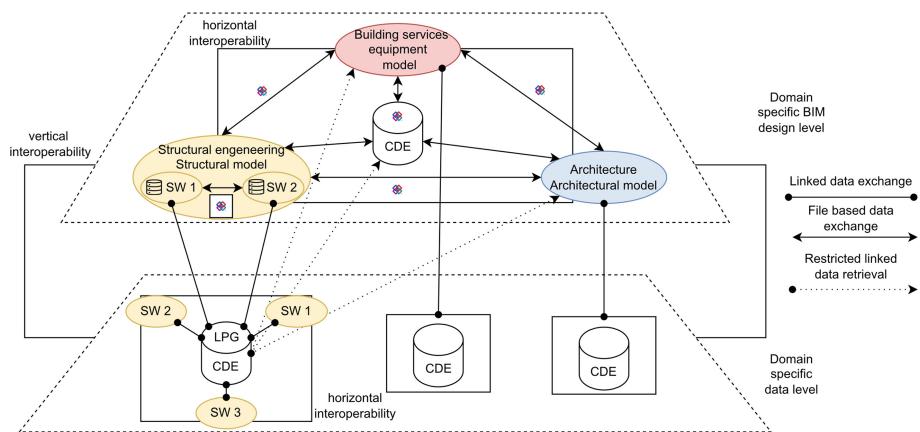


Fig. 2. Scheme of the horizontal and vertical data exchange with LPG-CDE.

4.1 Data Structure

This work is based on a specially developed LPG model for the holistic representation of structural analysis models and data. This enables holistic and networked data management without loss of context and data exchange on a unified database within the environment shown in Fig. 2. The presented approach pursues the paradigm shift from a product model-centric to a process model-centric information space [17] by integrating structural design processes and mindsets into the LPG model. In order to capture the structural design mindset, in contrast to most previous approaches, the connection of individual components via relations is at the centre of the LPG model, as shown in Fig. 3. In this way, the representation in the computer is most similar to the way a structural

engineer or designer thinks, and the result is an intuitive model with a high degree of comprehensibility.

Product model-centric information

The product model-centric information described below is shown in Figs. 3 and 4. The point, member, cross-section, material and support label elements form the basis of the model. While point labelled elements have x, y and z coordinates as properties and define the centreline model, member nodes represent different structural elements arranged between points. Directional relationships between points and members contain information about the direction of motion and member hinges, which are associated with member ends and limit the internal forces transferred from one member to another. Elements marked as points can be further defined by a possible relation called *Is_Support*. This relation points to a support element which defines the properties for the support. Member elements are further described by two relations called *Has_Material* and *Has_CrossSection*. While *Has_Material* points to a material element that defines the material and associated properties, *Has_CrossSection* points to an element called *CrossSection* that defines the section and associated properties. In structural software it is common to number all elements such as points, members and materials in ascending order. While member and point elements are unique to each project and therefore receive their number as a node attribute, member, section and material elements are often part of a database as types and are used in multiple projects. Therefore, the project specific number of these elements is given as a relationship attribute.

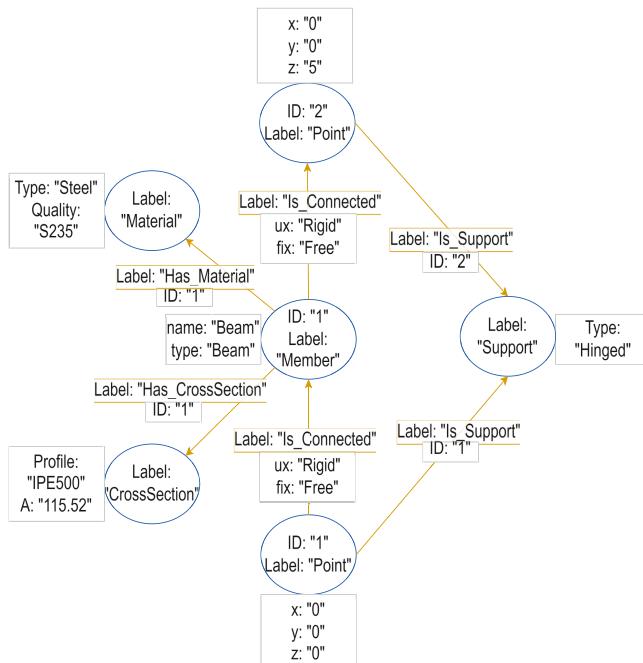


Fig. 3. Load-bearing structure in LPG.

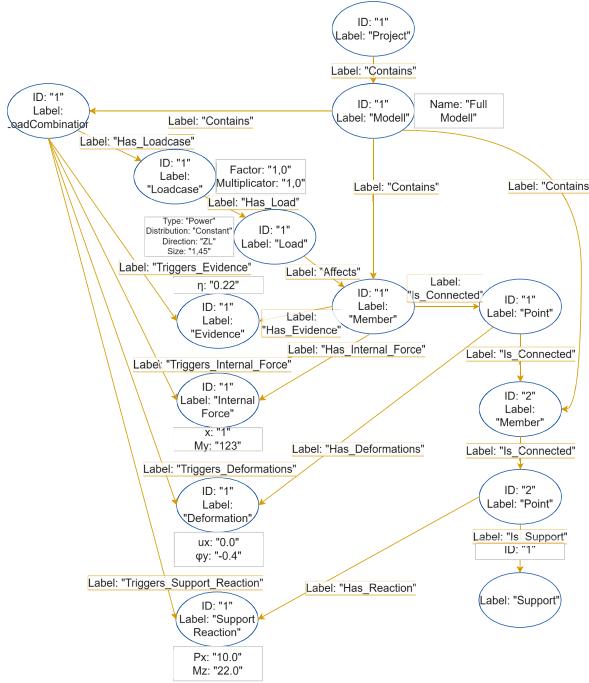


Fig. 4. Loads and analysis data in LPG.

Figure 4 shows the integration of additional elements associated with loads and resulting analysis data. This is done by adding a project node and a model node. The project node uniquely identifies the parent project by ID and name. The project node contains the model node, which allows inference to the complete model through relationships to all members and load combinations of the project. Load combination nodes are linked to load cases, which in turn are linked to the loads they contain. The load elements point to the member or point elements on which they act. The results of the calculation are divided into nodes labelled *evidence*, *deformation*, *internal force* and *support reaction*. Each of these node types is connected on one side to the member or point element to which it belongs and on the other side to the load combination from which it results. Nodes labelled *evidence* contain information such as the type of evidence and the stress ratio of the member to which they are linked. Nodes labelled *deformation* and *internal force* contain corresponding information for the member to which they are linked. In addition, nodes labelled as *deformations* can also be linked to nodes labelled as points to describe their deformations and rotations. Points with *Is_Support* relations are also linked to a node labelled as support reaction to provide information on support forces and moments.

Process model-centric information

The process model-centric information described below is shown in Fig. 5. Depending on the actual workflow, additional model nodes may be assigned to a project node. In this case, it is not necessary that new model nodes contain all the members of the project. This

allows members to be assigned to submodels, with member nodes remaining the same within different submodels. Another component of process model centric information is the assignment of different members to nodes called positions. These position nodes are associated with material and section nodes that are identical to the members shown in Fig. 3.

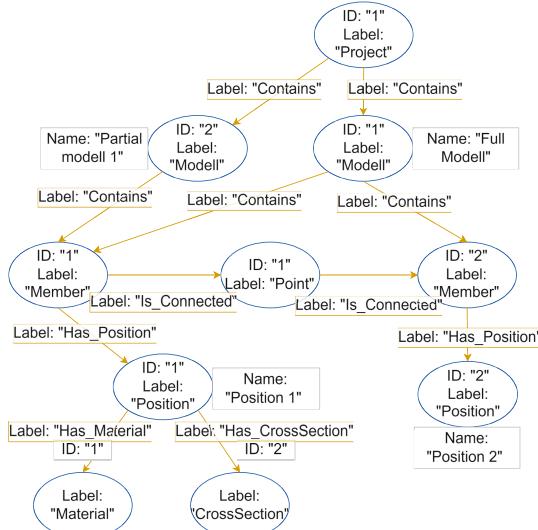


Fig. 5. Partial models and positions in LPG.

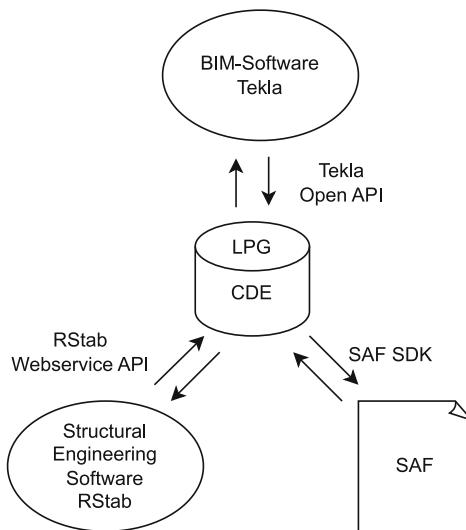


Fig. 6. Data interfaces.

5 Implementation

5.1 Data Interfaces

In order to evaluate LPG as a universal format for data exchange, three different ways of data exchange are investigated as shown in Fig. 6. Based on the structural design process, the structural BIM model within the authoring software is the starting point for data exchange. This is either derived from the architectural model or created from scratch by the structural engineer. The first step is to transfer the structural model into the LPG. As with the previous approaches to deriving analysis models, various correction steps are required to transfer the BIM model to the structural analysis profile centreline model, due to the dramatically different representations of the building structure within different BIM software. This paper does not deal with the exact derivation of structural models to analysis models, but focuses on the data exchange after successful derivation. Trimble's *Tekla Structures 2023* is chosen as the BIM authoring software, and the Tekla Open API is used to transfer data to a *Neo4J* database with the help of the query language cypher. After that the data of the model in *Neo4J* database is transferred to the *Dlubal RStab* structural software using RStab Webservice API. Various changes and additional data such as results within the structural software are transferred back to the *Neo4J* database using the same API. Using the Tekla Open API, these changes and data can be transferred back to the BIM software. In addition, the data interface includes creating and reading SAF files using the SAF SDK.

5.2 An Exemplary Application

The LPG database structure presented in 4.1 is illustrated in an example process flow using example data and the data interfaces presented in 5.1. The example data is a bar structure consisting of four columns and four beams that together form a three-dimensional frame structure as shown in Fig. 7 on the left. After modelling the example data as a structural model in *Tekla Structures*, the data is transferred in the LPG by creating a corresponding project node and a model node that represents the complete model and contains all the currently modelled components and the information needed to represent them. This includes points, sections and materials. This information allows the creation of a first analysis model in RStab, as shown on the right in Fig. 7.

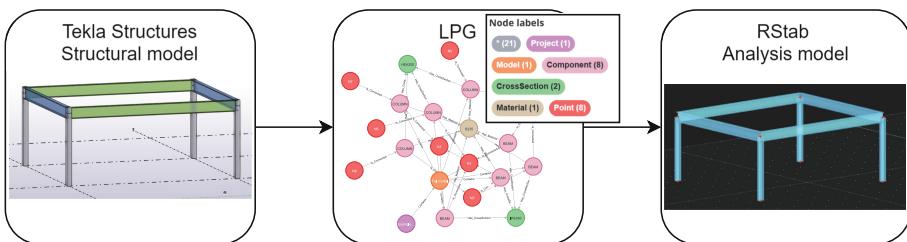


Fig. 7. Transferring structural model in LPG and analysis model.

At the analysis level, further semantic enrichment takes place by modelling information relevant to the analysis. This includes the support conditions of the four columns and the loading of the structure, including loads, load cases and load combinations. By manually assigning positions to each component by adding a position number to each component, components with similar characteristics and loads are grouped together. In the example data, the four columns are grouped into one position, while the beams are grouped into two further positions according to their different lengths and loads. The colours of the components on the left of Fig. 7 illustrate the division of the components into positions. The new information is then transferred to the LPG and at the same time a new submodel is created consisting of the three components highlighted in the left panel of Fig. 8. This results in the LPG shown in the centre of Fig. 8 with the process centric information according to the scheme shown in Fig. 5. Within the same step, an analysis submodel consisting of the three components is created by reading the component nodes and related information contained by the previously created submodel node.

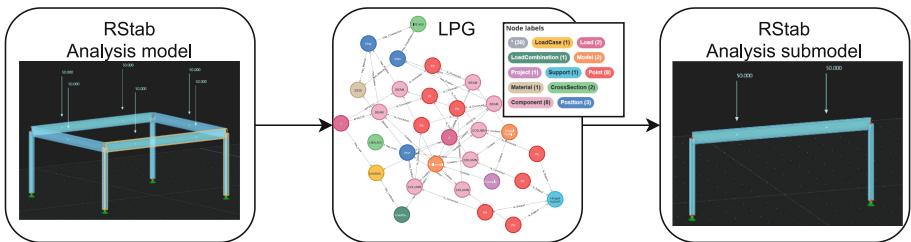


Fig. 8. Semantic enrichment and process-model-centric information modelling in analysis model.

The refined model can be used for further calculations and can also be transferred to the LPG as shown in Fig. 4. Any changes required as a result of the design can also be incorporated into the analysis submodel by further editing of members, materials, cross-sections or loads. To do this, the analysis submodel created is extended to include a new column and beam in the centre between the two existing columns with a new position and a new section. This is done by splitting the beam in two. The load on the beam is also reduced. With the refined model, shown on the left in Fig. 9, calculations can be carried out again and also transferred to the LPG according to the diagram in Fig. 4. Changes made in the submodel are also transferred to the LPG- and assigned to both the submodel and the overall model. The changes within the submodel result in the overall model shown in the centre of Fig. 9. A new beam is also created in the centre. The changes made within the submodel result in the overall model shown in the centre of Fig. 9. The changes made in the submodel and the model are always transferred to the LPG-CDE and can be transferred back to the structural model in a final step, as shown in Fig. 9, right.

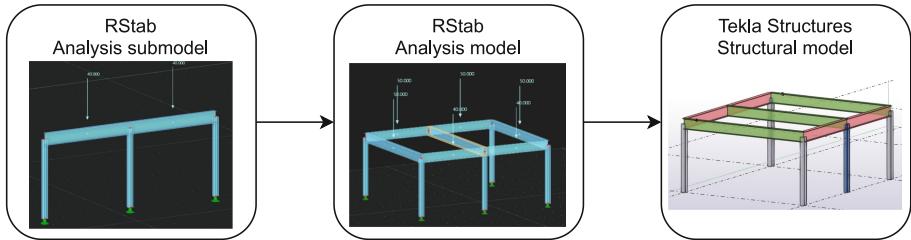


Fig. 9. Merging submodeling and transferring into structural model.

6 Results and Discussion

In the following, the use of LPG is compared with the previous neutral data exchange formats presented in Sect. 2 in terms of various features in order to discuss the advantages and disadvantages. Neutral data exchange formats currently in use are the IFC Structural Analysis View and the Structural Analysis Format (SAF). Like these, the LPG approach is a neutral, open data format that can be integrated by any software. Two important criteria are standardisation and flexibility. Unlike the two formats, the use of LPG is generally a schema-free approach that does not follow any standardisation. This allows for a high degree of flexibility, where graphs can be populated with literally any data, but also carries the potential risk of incompatibility with applications that rely on specific data structures. The metadata model introduced in Sect. 3.1 is intended to address this incompatibility by providing a basic structure, but also the flexibility to extend the model to include any data and create the necessary entities. The IFC Structural Analysis View has limited flexibility due to the addition of custom property sets and properties, but these are to be attached to predefined entities. The SAF format generally follows a standardised form that specifies entities, their properties, and relationships. Since the exchange is based on Excel, it is in principle also possible to extend the format with one's own entities, properties and relationships. Other criteria are human readability, editability, and the associated file format. As mentioned, the Excel file based SAF format allows for easy human readability as well as free and easy editing of the SAF format. The EXPRESS file based IFC Structural Analysis View is in generally human readable, but a basic understanding is required for readability and subsequent editability, although reading and editing is not intended. In contrast to the two file-based formats, LPG is database based. Depending on the LPG database used, data is stored in different ways. However, the graphical representation in the form of a diagram enables human readability. Visualisation allows users to visually explore the stored data, identify significant structures, and gain a better understanding of relationships. It is also often possible to interactively manipulate the stored data by modifying the diagram without in-depth knowledge of the underlying storage technology or query language (e.g. *Cypher*, *GraphQL*, *Gremlin* or *SPARQL*). At the data level, the use of LPG focuses on the relationships between data, which play a crucial role in structural design. In contrast the IFC Structural Analysis View only implicitly includes relationships between components. LPG also differs from SAF and IFC in the way it can be used and how data is transferred. While SAF and IFC allow

for file-based data exchange at a given point in time, the database-based LPG can form the basis of a web-based CDE that serves as a data exchange platform and general data storage, similar to the big open BIM idea. In addition, the use of a LPG-CDE has an impact on the workflow, particularly through the integration of process-model-centric information that is not part of SAF or IFC. LPG achieves the holistic mapping of structural design data in a universal format. This integrates the previous way of working with partial models and position planning, and also enables bi-directional data exchange between BIM and structural analysis throughout the process.

LPG is used in this work as a representative of graph databases. At this stage of implementation, it would also be possible and conceivable to use RDF. For technical reasons, there are differences in the implementation of LPG and RDF as graphs and also in the exact scope of application. An example of this is the subject predicate object triple on which RDF is based. This makes it impossible, for example, to arrange the same relationship between two relations. LPGs were originally designed for fast queriers, deep or variable length traversals and path queries. This is achieved by the dense key-value structure that can be modelled in a relational setting. Since the aim of this work is the direct storage of data in a graph database for different applications and thus the dynamic change of the graph database, fast data exchange and thus retrieval from the database is required, LPGs have been used. Depending on the further development and use of the graphs, the deep path queries for determining load paths in complex structures, for example, prove to be an added value. On the other hand, RDF is also suitable for the current application due to other properties. RDF graphs are very good for slowly and gradually changing datasets and can be easily extended with new data from different data sources thanks to their simple, atomised, decomposed, and shareable format. Their application to linking heterogeneous data sources also has the potential to be used in the same way as the LPG in the context of this work.

7 Conclusion and Outlook

The work presented shows that the graph-based digital representation of structural design data is suitable for data management and data exchange of basic structural data in this domain and can be further extended by integrating process-oriented data. It integrates the current way of working with partial models and positional planning, and also enables bi-directional data exchange between BIM and structural analysis throughout the process. The focus of graph databases on relationships between data corresponds to the high relevance of connections between components within a design. A high degree of flexibility allows graphs to be populated with literally any data, with the presented metadata model providing a schema for basic data to ensure compatibility with different applications. Graph visualizations allow users to visually explore the stored data, ensuring human readability. It is also often possible to interactively manipulate the stored data by modifying the diagram without in-depth knowledge of the underlying storage technology. The focus on data relationships enables analysis, semantic search or recommendation engines to recognise complex relationships.

The previous results need to be further confirmed by analysing the mapping of additional data such as surface structures and joint details. In this context, the use of RDF for this purpose is also a potential area of research. In general, however, the use of graph databases and thus the holistic mapping of data for structural design leads to further possible applications of data-driven machine learning approaches. The dissolution of heterogeneous data silos through graph-based approaches results in linked data, which in particular enables the use of using machine learning through graph neural networks.

References

1. Borrmann, A., Beetz, J., Koch, C., Liebich, T., Muhic, S.: Industry foundation classes: a standardized data model for the vendor-neutral exchange of digital building models. In: Borrmann, A., König, M., Koch, C., Beetz, J. (eds.) *Building Information Modeling*, pp. 81–126. Springer International Publishing, Cham (2018)
2. Laakso, M., Kiviniemi, A.: The IFC standard—A review of history, development, and standardization. *Electron. J. Inf. Technol. Constr.* **17**, 134–161 (2012)
3. Stracke, J., Kepplin, R.: Der BIM-Prozess in der Tragwerksplanung. *Beton und Stahlbetonbau* **115**, 324–331 (2020)
4. Beetz, J., van Leeuwen, J., de Vries, B.: IfcOWL: a case of transforming EXPRESS schemas into ontologies. *Artif. Intell. Eng. Des. Anal. Manuf.* **23**, 89–101 (2009)
5. Pauwels, P., Terkaj, W., Krijnen, T. & Beetz, J. : Coping with lists in the ifcOWL ontology. In: Proceedings of the 22nd EG-ICE Workshop, pp. 113–122. EG-ICE, Eindhoven, The Netherlands (2015)
6. Pauwels, P., Roxin, A.: SimpleBIM: From full ifcOWL graphs to simplified building graphs. In: Proceedings of the 11nd European Conference on Product and Process Modelling, pp. 11–18, ECPPM, Limassol, Cyprus (2016)
7. Huyeng, T.-J. et al. : Interlinking geometric and semantic information for an automated structural analysis of buildings using semantic web. In: ECPPM 2021—eWork and eBusiness in Architecture, Engineering and Construction, pp. 91–97, ECPPM 2021, London, Great Britain (2021)
8. Huyeng, T.-J. et al. : An approach to process geometric and semantic information as open graph-based description using a microservice architecture on the example of structural data. In: Proceedings of the 27th EG-ICE Workshop (2020)
9. Zhang, J., Li, H., Zhao, Y., Ren, G.: An ontology-based approach supporting holistic structural design with the consideration of safety, environmental impact and cost. *Adv. Eng. Softw.* **115**, 26–39 (2018)
10. Junghanns, M., Petermann, A., Teichmann, N., Gómez, K., Rahm, E.: Analyzing extended property graphs with Apache Flink. In: Proceedings of the 1st ACM SIGMOD Workshop on Network Data Analytics, pp. 1–8. ACM, San Francisco California (2016)
11. Zhu, J., Wu, P., Lei, X.: IFC-graph for facilitating building information access and query. *Autom. Constr.* **148**, 104778 (2023)
12. Song, L., Wang, C., Fan, J., Lu, H.: Elastic structural analysis based on graph neural network without labeled data. *Comput.-Aided Civ. Infrastruct. Eng.* **38**, 1307–1323 (2023)
13. Chang, K.-H., Cheng, C.-Y.: Learning to simulate and design for structural engineering. In: Proceedings of the 37th International Conference on Machine Learning (2020)
14. Sibenik, G., Kovacic, I.: Interpreted open data exchange between architectural design and structural analysis models. *ITcon* **26**, 39–57 (2021)
15. Lakusic, S.: Automated preprocessing of building models for structural analysis. *JCE* **74**, 211–226 (2022)

16. Amor, R., Jiang, Y., Chen, X.: BIM in 2007—are we there yet? In: Proceedings of the 24th W78 Conference, CIB W78—Information Technology for Construction, pp. 159–162. Maribor, Slovenia (2007)
17. Scherer, R.J., Schapke, S.-E.: A distributed multi-model-based management information system for simulation and decision-making on construction projects. *Adv. Eng. Inform.* **25**, 582–599 (2011)



Machine Learning and Internet of Things for Construction Air Pollution Monitoring and Prediction

Ruichuan Zhang¹(✉), Irene Paek², Samuel Park³, and Jenna Krall⁴

¹ School of Construction, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

ruichuanz@vt.edu

² Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

³ Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

⁴ Department of Global and Community Health, George Mason University, Fairfax, VA, USA

Abstract. The construction industry significantly contributes to air pollution by releasing substantial quantities of pollutants that adversely affect both urban and natural environments. Among these pollutants, particulate matter (PM) emerges as a crucial concern. Such air pollution has broader negative consequences, impacting the health of workers and nearby communities, leading to environmental compliance costs, and threatening sustainability. Regulatory and research efforts have been initiated to tackle construction air pollution. However, the absence of comprehensive analysis of construction sites, ambient environmental data, and PM pollution hampers effective pollution detection and control measures. To address this challenge, this paper introduces an intelligent system that leverages machine learning and Internet of Things (IoT) technologies for monitoring and predicting PM concentrations associated with construction sites. The proposed system comprises three key components: (1) an IoT network for real-time data collection on ambient environmental conditions, construction-related air pollutants, and construction activities, (2) an objective detection model for detecting construction workers, vehicles, and heavy machinery that could contribute to air pollution, and (3) a deep learning-based predictive model to capture the relationships between air pollution and ambient environmental conditions as well as detected objects from construction sites. Experimental validation on a construction site at the Virginia Tech Blacksburg campus demonstrates promising results for the proposed system.

Keywords: Construction air pollution · Machine learning · Internet of Things · Particulate matter

1 Introduction

The construction sector significantly contributes to air pollution by releasing substantial amounts of pollutants that negatively impact urban and natural environments [1]. Among these pollutants, particulate matter (PM) emerges as a crucial concern. PM comprises

various elements or forms such as dust, dirt, soot, and smoke. Some particles are visible, while others require sensors for detection—they fall into three size categories [2]: (1) PM₁₀ (diameters generally 10 µm or smaller), (2) PM_{2.5} (diameters between 2.5 and 1 µm), and (3) PM₁ (diameters generally 1 µm or less). Construction-related air pollution has broader negative consequences, impacting the health of workers and nearby communities, leading to environmental compliance costs, and threatening sustainability [3].

Numerous efforts have been undertaken to tackle construction air pollution. Regulatory bodies such as the Environmental Protection Agency (EPA) enforce crucial regulatory frameworks for managing air pollution, which apply to construction projects, such as the Clean Air Act (CAA) and the National Ambient Air Quality Standards for Particulate Matter (PM NAAQS). Recent updates by the EPA have tightened PM_{2.5} standards to a range of 9.0–10.0 µg/m³, aiming to safeguard public health [4]. However, the absence of comprehensive analysis of construction sites, ambient environmental data, and PM pollution hampers effective pollution detection and control measures. Current approaches rely heavily on lab tests, predetermined emission inventories, and/or simulations, lacking mechanisms to predict air pollution caused by construction activities [1]. Addressing this gap necessitates an integrated intelligent system capable of collecting, processing, and analyzing multimodal data related to construction sites and their environmental impact.

To address this need, this paper introduces a new intelligent system that leverages machine learning and Internet of Things (IoT) technologies for monitoring and predicting construction air pollution. The proposed system comprises three key components: (1) an IoT network for real-time data collection on ambient environmental conditions, construction-related air pollutants, and construction activities, (2) an objective detection model for detecting construction workers, vehicles, and heavy machinery that could contribute to air pollution, and (3) a deep learning-based predictive model to capture the relationships between air pollution and ambient environmental conditions as well as detected objects from construction sites. We further conducted experimental validation on a selected construction site at the Virginia Tech Blacksburg campus, which demonstrates promising results for the proposed system.

The subsequent sections of this paper are organized as follows: Sect. 2 reviews related research efforts in construction air pollution, and smart technologies (e.g., machine learning and IoT) for managing construction pollution; Sect. 3 details the proposed intelligent system for monitoring and predicting air pollution on construction sites; Sect. 4 presents preliminary experiments conducted to validate the proposed system and showcases experimental results; Sects. 5 and 6 highlight the contributions and conclude the paper.

2 Background

2.1 Construction and Air Pollution

Civil and construction projects are essential for urban expansion and revitalization. However, they create significant threats to urban environments in rapidly urbanizing regions and impose substantial challenges to sustainable development. They result in resource

wastage, environmental pollution, and considerable expenses in environmental compliance and health costs for construction workers and the public [1]. PM pollution is a primary concern associated with construction activities. The civil and construction industry is one of the significant sources of PM pollution—PM encompasses microscopic particles of solid or liquid matter suspended in the air, which can vary in size, forms, and sources. Classified into three size categories—(1) PM_{10} (diameters generally 10 μm or smaller), (2) $PM_{2.5}$ (diameters between 2.5 and 1 μm), and (3) PM_1 (diameters generally 1 μm or less)—construction-generated PM includes diverse forms such dust, dirt, soot, and smoke. The sources of PM in construction activities are primarily twofold: (1) PM-generating construction tasks (e.g., site preparation and earth-moving), and (2) diesel-powered vehicles (e.g., trucks) and heavy machinery (e.g., lifters) employed in these tasks [5–7]. For instance, site preparation and earth-moving activities cause substantial soil disturbance, leading to open areas prone to wind-driven dispersion of dust particles, manifesting as both $PM_{2.5}$ and PM_{10} ; additionally, the use of trucks in loading, transporting, and depositing excavated earth releases combustion-related particulates, comprising both PM_1 and $PM_{2.5}$ [6].

2.2 Smart Technologies for Construction Pollution Management

Existing research efforts for construction pollution management have centered on monitoring and estimating PM emissions from construction sites. For example, [3] linked increased concentrations of $PM_{2.5}$ and PM_{10} in urban areas with construction activities by employing clustering analysis of PM concentration data obtained near construction sites. Ahmed and Arocho [7] investigated PM emissions from different construction materials (e.g., timber and steel). Yoon et al. [8] applied emission factors to gauge $PM_{2.5}$ emissions at construction sites and juxtaposed these estimations with data obtained from sensors. However, current efforts predominantly rely on laboratory tests [5], predefined emission inventories and factors [8], and simulations [9], which lack precision in capturing real-time PM emissions. Second, these efforts primarily focus on siloed data, such as concentrations of specific PM sizes (e.g., $PM_{2.5}$) or specific forms (e.g., dust) [9], neglecting the integrated analysis of construction activities, ambient environmental data, and PM pollution—a crucial aspect for comprehensively understanding their correlation. Recent advancements in object detection have exhibited potential in accurately identifying elements at construction sites, including workers and vehicles, despite challenging conditions such as varied viewing angles and distances, lighting conditions, and occlusions in visual data [10]. Thus, there exists an untapped opportunity to integrate detected construction workers and vehicles to capture construction activities that potentially contribute to PM emissions. Third, existing efforts concentrate on data collection techniques and statistical analysis but lack the capability for pollution analysis and prediction. Progress in deep learning methodologies, such as long short-term memory [11] and transformer [12] models, offers prospects for achieving these capabilities by processing, integrating, and analyzing sequential data more effectively.

3 Proposed System

The proposed intelligent system for construction air pollution monitoring and prediction comprises three key components (as per Fig. 1): (1) an IoT network for real-time data collection on ambient environment data (e.g., temperature, humidity, and lighting), construction-related air pollutants (e.g., PM₁, PM_{2.5}, and PM₁₀), and construction site videos, (2) an objective detection model for detecting construction workers, vehicles, and heavy machinery that could contribute to air pollution, and (3) a deep learning-based predictive model to predict air pollutant concentration based on ambient environmental conditions as well as detected objects from construction sites.

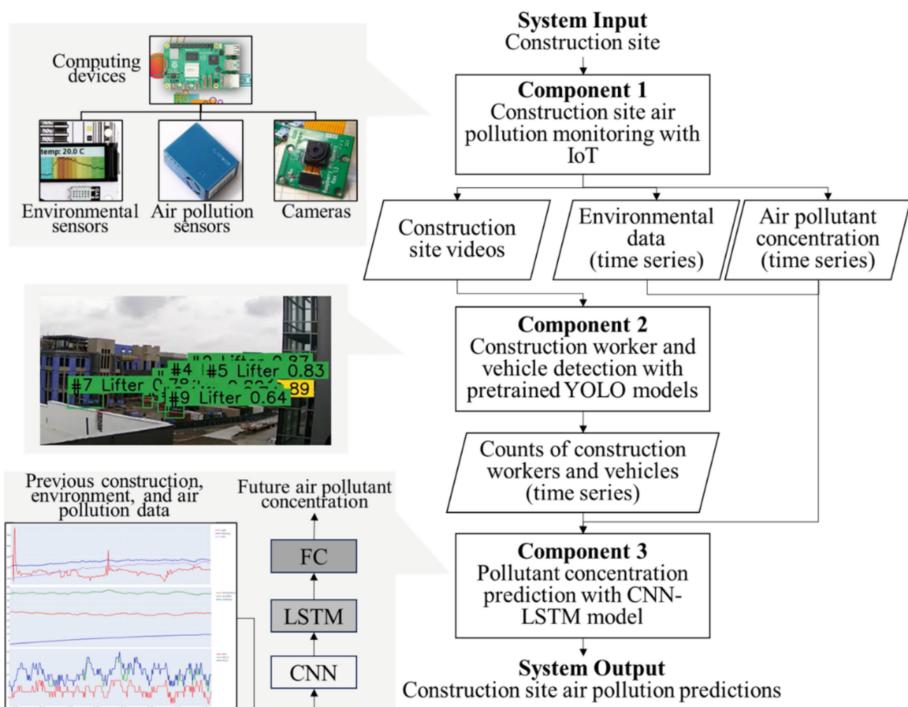


Fig. 1. Proposed intelligent system for construction air pollution monitoring and prediction.

3.1 Construction Site and Air Pollution Monitoring

The first component, IoT for construction site and air pollution monitoring, aims to diverse data from various sources related to air pollution at construction sites. This component encompasses four types of devices: (1) environmental sensors for collecting ambient environmental data, (2) environmental sensors for collecting PM concentrations,

(3) visual sensors for recording construction site videos, and (4) computing devices for managing and controlling data collection processes.

Ambient Environmental Data. The sensors capture multidimensional environmental conditions, including temperature, pressure, humidity, lighting conditions, and concentrations of volatile organic compounds (e.g., carbon monoxide, nitrogen dioxide, and ammonia). The BME 280 sensor records temperature, barometric pressure, and humidity, providing yearly stability ratings within specified margins for the proposed system. The analog gas sensor MICS 6814 was employed to record VOC concentrations.

PM Concentrations. The PMS 5003 sensor was utilized to measure PM concentrations, including PM₁, PM_{2.5}, and PM₁₀. The PMS 5003 particulate sensor is a laser scattering sensor that gives digital outputs of the concentration of suspended particles in the air. The output is mainly in the form of the number of each particle with different sizes per unit volume. The effective range is 0–500 µg/m³ and the resolution is 1 µg/m³, and thus suffice the proposed system. The outputs from the employed sensor were compared to that from ambient monitors operated by EPA through the AirNow website.

Construction Site Videos. Cameras compatible with Raspberry Pi (e.g., Pi Camera) were utilized for real-time video recording and transmission. Four cameras were set up at different angles, capturing frames of the construction site every three seconds.

3.2 Construction Worker, Vehicle, and Heavy Machinery Detection

The second component, a construction worker, vehicle, and heavy machinery detection model, aims to extract crucial information for predicting air pollution linked to construction activities.

Model Architecture. We leveraged pretrained computer vision model—You Only Look Once (YOLO) [13]—and finetuned the model using a domain-specific dataset (i.e., an annotated video and image dataset of construction sites). YOLO is optimized for swift and effective object detection in images or video frames, making it ideal for applications requiring real-time processing. Unlike conventional methods that involve applying a classifier across different image sections, YOLO divides the image into a grid, predicting bounding boxes and classifying objects directly from the entire image in a single pass through a convolutional neural network (CNN) model.

Model Finetuning. The model finetuning involves refining a YOLO model with the domain-specific dataset by substituting the last fully connected (FC) layer of the model with a new FC layer. This new layer corresponds to the number of output classes in our proposed system, such as construction workers, vehicles, and heavy machinery. During this process, the pretrained YOLO model’s previously learned weights in other layers, such as the CNN, can adapt to construction site images, while the last layer is specifically trained for construction site elements. This strategy, known as transfer learning, capitalizes on general-domain knowledge while utilizing limited annotated data, a common scenario in domain-specific applications [14].

3.3 Air Pollutant Concentration Prediction

The third component, a deep learning-based model for predicting air pollutant concentrations, aims to forecast future PM concentrations by considering ambient environmental conditions at construction sites, the detected count of construction workers, vehicles, and heavy machinery, and previous PM concentrations. This model comprises three key steps: data fusion, representation learning, and PM concentration estimation.

Data Fusion. The data fusion step integrates data from various sensors and aligns disparate time intervals. To achieve synchronization, we perform subsampling. For instance, video intervals occur every three seconds per frame, while ambient environmental data and PM concentrations are recorded every single second. The combined data form a multivariate time series containing ten variables: temperature, humidity, air pressure, light levels, concentrations of PM_1 , $\text{PM}_{2.5}$, and PM_{10} , and counts of construction workers, vehicles, and heavy machinery.

Representation Learning. Our model employs CNN and LSTM to learn significant temporal representations of the multivariate time series resulting from data fusion. Both CNN and LSTM are effective in sequential data analytic tasks such as prediction, classification, and anomaly detection. The CNN layer condenses the time stamps in the input time series for improved computational efficiency. The LSTM layer aggregates a single representation of the entire input time series, facilitating future pollutant concentration prediction.

PM Concentration Estimating. We use an FC to predict the air pollutant concentration level. This layer consumes the representations learned from the CNN-LSTM model, which captures the ambient environmental conditions of the construction site, the construction workers, vehicles, and heavy machinery, and PM concentrations in the past time stamps, and generates three real-number values representing the concentration of PM_1 , $\text{PM}_{2.5}$, and PM_{10} of the future time stamp.

4 Preliminary Experiments and Results

Data Collection and Preprocessing. We chose the Virginia Tech's Hitt Hall construction site in Blacksburg, VA, USA for implementing and validating the proposed system and collecting ambient environmental data, PM concentrations, and construction site videos. The data collection involved three separate sessions, each lasting between 30 and 90 min. For our construction worker, vehicle, and heavy machinery detection model (Component 2), we invited three domain experts to annotate 600 images extracted from videos. For our PM concentration prediction model (Component 3), we utilized a sliding window algorithm to segment the time series data into samples of 200 s each, resulting in a total of 2,491 data samples. To ensure consistency, all data underwent normalization using a min-max approach. We split the images and time series data samples into training, validation, and testing subsets in an 80:10:10 ratio respectively.

Experimental Design and Evaluation. To comprehensively assess the performance of the proposed system, we conducted three experiments to evaluate three specific aspects:

(1) the effectiveness of the construction worker, vehicle, and heavy machinery detection model (Component 2); (2) the accuracy of the PM concentration prediction model (Component 3); (3) the overall performance of the proposed system, focusing on the impact of different data modalities.

In the first experiment, we used the mean average precision (mAP) and F1 score evaluation, which are commonly used in object detection. mAP is defined as the average of areas under curve (AUC) of the precision-recall curves for all types/classes of objects. Precision (P) (Eq. 1) is the ratio of correctly detected objects to the total detected objects. Recall (R) (Eq. 2) is the ratio of correctly detected objects to all actual objects in the images. mAP@50 denotes the assessment conducted at an intersection over union threshold of 50%, which is the overlap of bounding boxes of the detected objects with respect to the actual objects considered as successful recognition when determining true positives (TP), false positives (FP), and false negatives (FN). F1 score is the harmonic average of P and R. Ideally, a well-performing model achieves a balance between precision and recall—such a model correctly identifies a significant portion of positive instances (high recall) while also minimizing the number of false positives (high precision) across all confidence levels for achieving an optimal F1 score.

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

In the second and third experiments, we used mean square error (MSE) (Eq. 3) and mean absolute error (MAE) (Eq. 4) to measure the discrepancy between the predicted PM concentration (c^p) and actual PM concentration (c^a) for evaluate the overall performance of the proposed system in the second and third experiments. Here, n is the total number of data samples. For this paper, we calculated normalized MSE and MAE, where c^p and c^a are normalized. Lower MSE and MAE indicate better model performance.

$$MSE = \frac{1}{n} \sum_i^n (c_i^p - c_i^a)^2 \quad (3)$$

$$MAE = \frac{1}{n} \sum_i^n |c_i^p - c_i^a| \quad (4)$$

Experimental Results and Analysis. Figure 2 shows the entire precision-recall curve and F1-confidence curve of the construction worker, vehicle, and heavy machinery detection model on the testing data in the first experiment. The model achieved an mAP@50 of 0.860 (based on the areas under the curve in Fig. 2a) and an optimal F1 of 0.810 at the confidence level of 40% (based on the curves in Fig. 2b). The model exhibits competitive performance based on the original pretrained YOLO model.

Table 1 shows the results of the comparative experiments between the proposed model and selected baseline models (linear regression and support vector regression) for air pollutant concentration prediction in the second experiment. The proposed CNN-LSTM model achieved an MSE of 0.007 and an MAE of 0.069, exceeding the selected baseline models, indicating strong performance.

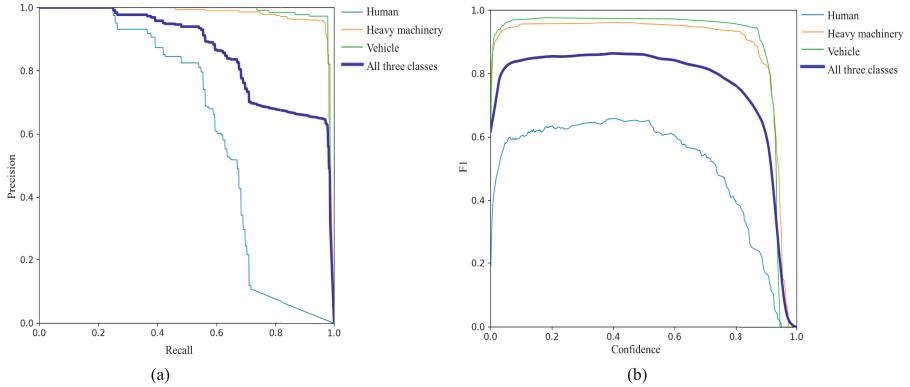


Fig. 2. Testing performance of the object detection model: **a** precision-recall curve; **b** F1-confidence curve.

Table 1. Performance of proposed and baseline model architectures in PM concentration prediction.

Model	MSE	MAE
Proposed model: CNN-LSTM	0.007	0.069
Baseline 1: Linear regression	0.099	0.247
Baseline 2: Support vector regression	0.011	0.084

Table 2 shows how various data modalities affect the predictive performance of the proposed system for air pollutant concentration in the third experiment. The removal of any of the three data modalities—ambient environmental data, past air pollutant concentrations, and detected objects—results in an increase in MSE and MAE, indicating the significant contribution of all data modalities to the performance of the proposed system. The results further indicate that including more data modalities and metadata (e.g., geographical locations of where the data were collected and the presence of construction workers, machinery, and vehicles) can potentially improve the performance of the proposed system. By incorporating these data, knowledge about the spatial distribution of PMs can be integrated into the predictive models.

5 Contributions

The proposed intelligent system contributes to the evolving body of knowledge at the intersection of IoT, machine learning, and construction pollution management. Its contributions unfold in three significant ways. First, the system utilizes IoT devices to accurately capture real-time, multimodal, and multidimensional conditions prevalent at construction sites. This capability encompasses the collection of ambient environmental data, PM concentrations, and construction site videos, all pertinent to construction-related air pollution. Second, the system underscores the potential of deep learning by

Table 2. Impact of data modalities in PM concentration prediction.

Data modality	MSE	MAE
All data modalities (Ambient environmental data + Past PM concentrations + Detected objects)	0.007	0.069
Ambient environmental data + Past PM concentrations	0.010	0.079
Past PM concentrations + Detected objects	0.012	0.085
Ambient environmental data + Detected objects	0.012	0.088

harnessing its ability to extract valuable representations from both acquired data and publicly available pretrained models. This involves a fusion of finetuning YOLO models for detecting workers, vehicles, and heavy machinery on construction sites and training CNN-LSTM models for analyzing time series of detected objects and environmental conditions. Third, the system not only monitors and analyzes PM concentration levels but also possesses the capacity for predictive forecasting. Such predictive abilities are crucial for subsequent actions like preemptive or remedial strategies in managing construction-related pollution.

6 Conclusion

This research introduces a new intelligent system utilizing machine learning and Internet of Things (IoT) technologies to monitor and predict construction-related air pollution. The system consists of three main components: (1) an IoT network gathers real-time data on environmental conditions, construction-related air pollutants, and construction activities; (2) an objective detection model identifies construction workers, vehicles, and heavy machinery linked to air pollution; and (3) a deep learning-based model for predicting future PM concentrations based on ambient environmental conditions at construction sites, detected objects, and past PM concentrations. The experiments demonstrated an F1 score of 0.810 and an mAP@50 of 0.860 for the detection model, and a normalized MSE of 0.007 and MAE of 0.069 for the PM concentration prediction model. The models exhibit competitive performance in predicting air pollution, with all data modalities contributing to the high performance.

Despite the promising results, there are areas for potential improvement in future work. First, integrating data from mobile sensors/devices (e.g., air pollution and GPS sensors on construction workers, vehicles, and machinery) could offer more comprehensive spatial insights into construction-generated air pollutant concentration distributions. Second, enhancing the deep learning model's capability to handle outliers or anomalies in the collected time series data could improve its accuracy in forecasting air pollution over larger time scales. Third, future research could explore the system's performance across various construction project types, site conditions, and project stages. Expanding the system's testing with more diverse construction site data could improve its overall applicability and generalizability.

Acknowledgement. The authors would like to thank 4-VA. This material is based on work supported by 4-VA, a collaborative partnership for advancing the Commonwealth of Virginia. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of 4-VA.

References

1. Cherian, D., Choi, J.H.: A review of research on particulate matter pollution in the construction industry. *J. Clean. Prod.* **254**, 120077 (2020)
2. U.S. Environmental Protection Agency.: Particulate Matter (PM) Basics (2023). <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>. Accessed 15 October
3. Azarmi, F., Kumar, P., Marsh, D., Fuller, G.: Assessment of the long-term impacts of PM 10 and PM 2.5 particles from construction works on surrounding areas. *Environ. Sci.: Process. Impacts* **18**(2), 208–221 (2016)
4. U.S. Environmental Protection Agency.: National Ambient Air Quality Standards (NAAQS) for PM (2023). <https://www.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqs-pm>. Accessed 15 Oct 2023
5. Azarmi, F., Kumar, P.: Ambient exposure to coarse and fine particle emissions from building demolition. *Atmos. Environ.* **137**, 62–79 (2016)
6. Luo, Q., et al.: Occupational health risk assessment based on dust exposure during earthwork construction. *J. Build. Eng.* **44**, 103186 (2021)
7. Ahmed, S., Arocho, I.: Emission of particulate matters during construction: a comparative study on a Cross Laminated Timber (CLT) and a steel building construction project. *J. Build. Eng.* **22**, 281–294 (2019)
8. Yoon, S., Won, D., Chi, S.: Empirical studies on emission factors for real-time particulate matter 2.5 monitoring at construction sites. *J. Clean. Prod.* **384**, 135546 (2023)
9. Tong, R., et al.: The construction dust-induced occupational health risk using Monte-Carlo simulation. *J. Clean. Prod.* **184**, 598–608 (2018)
10. Zhang, M., Shi, R., Yang, Z.: A critical review of vision-based occupational health and safety monitoring of construction site workers. *Saf. Sci.* **126**, 104658 (2020)
11. Lipton, Z.C., Kale, D.C., Elkan, C., Wetzel, R.: Learning to diagnose with LSTM recurrent neural networks (2015). arXiv preprint [arXiv:1511.03677](https://arxiv.org/abs/1511.03677)
12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, 30 (2017)
13. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016)
14. Shin, H.C., et al.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imag.* **35**(5), 1285–1298 (2016)



Exploring the Potential of BIM Models for Deriving Synthetic Training Data for Machine Learning Applications

Simon K. Hoeng^(✉), Friedrich Eder, Marc Schmailzl, and Mathias Obergrießer

Ostbayerische Technische Hochschule Regensburg, Regensburg, Germany
simon.hoeng@oth-regensburg.de

Abstract. To increase the efficiency and quality of design and construction tasks, the use of *Artificial Intelligence* (AI) and *Machine Learning* (ML) offers a way to automate both repetitive and complex tasks. Many of these ML models rely heavily on large amounts of suitable, machine-readable, and labeled training data. Therefore, a variety of conceivable use cases for ML in the Architecture, Engineering and Construction (AEC) industry are difficult to implement due to a lack of freely and directly usable training data. The process of manually structuring and labeling existing data is time-consuming and needs in some cases skilled personnel to ensure the quality of the labeled data. Due to these factors, approaches for utilizing artificially generated data, referred to as synthetic data, are becoming more prevalent. Since *Building Information Models* contain a large amount of information, deriving training data from these models presents an obvious route for generation of this data. There are many ML applications whose implementation is inhibited due to a lack of training data, for which model-based synthetic data offer a possible solution approach. The *Industry Foundation Classes* (IFC) standard provides a powerful exchange format for models independently of their authoring software. Parametric and generative approaches to model creation enable the generation of numerous different building models within a short period of time and with low effort. This paper presents a workflow for automated derivation of synthetic training data from rule-based or parametrically generated models combined with existing IFC datasets as a multimodal data repository. The method is validated by testing automated synthetically labeled image data for a plan detection task, which is carried out with the Object Detection Framework *YOLOv8*. The suggested workflow has the potential to enhance data accessibility, thereby contributing to the implementation of ML applications in the AEC industry.

Keywords: Artificial Intelligence (AI) · Machine Learning (ML) · Synthetic data · Building Information Modeling (BIM) · Parametric modeling · Industry Foundation Classes (IFC)

1 Introduction

Building Information Modelling (BIM) and BIM models are establishing their role as a central and comprehensive data repository for building related information. This data presents a valuable resource of knowledge, which can be leveraged to optimize architectural processes and decision making. Throughout the lifecycle of a building structure related data is stored in various models and formats. To improve data exchange throughout the Architecture, Engineering and Construction (AEC) industry, *Building Smart International* is establishing certain standards. One standard that increases interoperability for authoring software is the *Industry Foundation Classes* (IFC).¹ Even if IFC has still some limitations, it is emerging as an essential tool for interoperability. IFC is widely supported therefore allowing the archiving of building information in a consistent and accessible manner. Consequently, it connects different stages of a building's lifecycle and fosters collaboration.

Simultaneously *Artificial Intelligence* (AI) and *Machine Learning* (ML) are emerging as powerful tools, offering solutions to challenges across various domains. However, their successful implementation relies heavily on the availability of comprehensive and fitting datasets. Curating these datasets manually poses a time-consuming process, which may require domain specific knowledge.

Utilizing the information contained in BIM-Models seems like a possible approach, to improve availability of high quality labeled and unlabeled training data.

Particularly researchers or startups, have no access to a substantial quantity repository of BIM or IFC models. Limited access to such datasets can hinder the training and optimization of AI and ML algorithms, constraining their efficacy in architectural applications. Parametrically generated models combined with accessible BIM or IFC models may help to overcome this challenge for implementing ML applications.

2 State of the Art

The potential applications of data derived from BIM models are extensive, as the considerable number of publications exploring the utilization of synthetic data show.

Noichl et al. introduced a method for generating labeled synthetic point cloud data based on highly detailed models. This approach proved valuable for AI systems tasked with remodeling intricate heating, ventilation, and air conditioning equipment [1].

Hong et al. delved into the generation of synthetic image data using BIM models for the classification of building components. Employing *Generative Adversarial Networks* (GANs), they produced images closely mirroring real-world scenarios. The dataset's robustness was further enhanced by merging the generated synthetic image data through pixel splicing [2].

It's noteworthy that both studies utilized existing BIM models from real projects. In prior work Hoeng and Eder presented a method to generate "true" synthetic data by leveraging a parametric model for generation of randomized models for derivation of multimodal training data. The shown principle was validated by using derived 2D plans for object detection tasks for extraction of building objects in architectural plans [3].

¹ <https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/>.

In a comprehensive literature review on the automated reconstruction of As-is BIM models using AI tools, Schönfelder et al. emphasized the underutilization of information stored in existing drawings and text documents as potential sources for the reconstruction process. They advocate to address this gap in future research projects [4].

Alawadi and Yan examined the use of a parametric BIM to generate synthetic data for categorizing exterior building objects from real-world images. Their experiments showed that very large artificially generated datasets tend to outperform curated smaller datasets [5].

In a different application, Clever et al. used a parametric modelling approach to generate varying BIM models of train stations. These models were used to predict pedestrian flow density using deep learning techniques [6].

3 Workflow for Automated Derivation of Labeled Synthetic Data from BIM Models

In this work we further explore the method of deriving synthetic training data for ML applications from BIM-Models by proposing a workflow for automated derivation of synthetic labeled training data by pooling parametrically and randomized generated models with existing models in IFC-Format. The pool of models is used as a multimodal information source for derivation of labeled training data.

BIM models have a high density of information. Each single model object stores various semantic and geometric information. The model itself represents an implicit contextual description of structural rules and architectural concepts. By interpreting a building model, valuable information about a given building can be extracted in the form of plans, images, or semantic structures. Subsequent processing of these models and the derived information facilitates the creation of training data. This data can be employed by numerous ML applications, like shown in Sect. 2. The use of multimodal data sets from combining existing models with parametrically generated BIM models is intended to represent a flexible, adjustable, and scalable approach to cover a variety of use cases and therefore may help do boost data accessibility for implementation of ML applications.

The proposed workflow is shown in Fig. 1 and will be elaborated in the following subchapters.

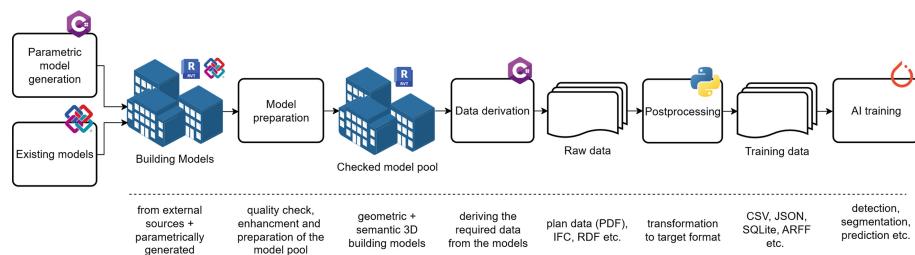


Fig. 1. Multisource model based synthetic data pipeline

3.1 Data Base: IFC and Parametric Models

IFC Models

Various sources offer freely available IFC models, such as the *buildingSMART Example files*² and the *OpenIFC Model Repository*³ provided by the university of Auckland. Numerous companies in the construction and planning sectors, as well as governmental planning authorities, utilize modern BIM methods. Typically, they store their models in proprietary formats, necessitating licensed software for access. However, all major authoring software supports IFC export. This not only provides a model exchange but also may serve as viable access point for batched processing of models, facilitating their transformation into the IFC format. Even if there are still some issues regarding information loss during export and import processes, IFC is a very promising format for interoperability and it is expected that these issues can be overcome by improving the mapping process [7]. Consequently, amassing extensive repositories of IFC models, especially for those with access to BIM data, appears feasible and holds great promise as a source of building related information stored in the form of interoperable models.

Parametric Models

Nevertheless, high quality models that match the intended use cases and contain the required information may still be difficult to come by due to copyrights concerns or simply by missing data sources. If there are no or not enough existing models to saturate the intended use case, specifically tailored parametric, generative design methods provide a way to synthetically create or expand the needed model basis for the data derivation.

Parametric and generative approaches for model creation, such as those presented by [8], enable the generation of numerous different building models within a short period of time. The use of rule-based generative models combined with randomized input enables the generation of large numbers of synthetic models in compliance with specified restrictions. *Autodesk Revit API*⁴ offers a well-documented and accessible way to automate modeling processes and workflows and poses one solution for implementation of such a model generation system.

3.2 Model Preparation and Checking

The synthetically generated structural models and the IFC models serve as the basis for subsequent process steps.

The IFC-Files are imported automatically and batchwise via *Revit API*, as *Autodesk Revit* is used as an intermediate software. During the import, *Revit* assigns internal *Revit* categories to the IFC classes. The mapping of these classes can be adjusted and finetuned via a mapping file that is used by *Revit* during the import. These internal categories will be used later to label the model objects.

The parametrically models are created directly in *Revit*; therefore, the categories are controlled by the API, and there is no further mapping needed.

² <https://github.com/buildingSMART/Sample-Test-Files>.

³ <https://openifcmodel.cs.auckland.ac.nz>.

⁴ <https://aps.autodesk.com/developer/overview/revit>.

To ensure the quality and usability of an IFC model the *Information Delivery Specifications*⁵ (IDS) standard provides a method for automated model validation and model checking. This ensures that the models contain the needed information (e.g. model objects with certain semantic specifications) and meet the quality requirements for the intended use case. Kremer and Beetz introduced an extension to the IDS format, enhancing the checking process by incorporating additional verification resources. This extension facilitates not only semantic checks but also includes geometric and topological assessments [9].

If the model does not pass the suitability test defined as part of the IDS checking, the model can either be discarded or (if feasible) modified so it becomes usable as well. Once the models are imported respectively created in Revit the models can be modified and enriched.

This process is applied to every model which results in a pool of fitting and accepted models (Fig. 2).

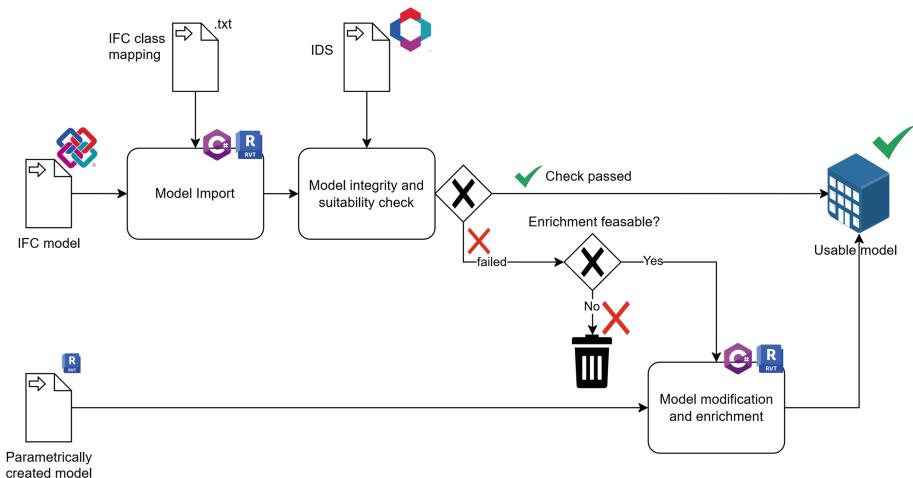


Fig. 2. Model preparation and checking

3.3 Data Derivation

The models are now ready for derivation of the raw data. For this study ceiling views are derived from the models and placed on a prepared plan sheet. The plan sheets are then exported in PDF format. The export of other formats like OBJ, STL or rendered images is also feasible.

Parallel to the authentic plan sheets, color labeled version of the sheets are exported as well. Using the building element category information stored in the model, different RGB colors are assigned to the different categories via a view template. The color values

⁵ <https://technical.buildingsmart.org/projects/information-delivery-specification-ids/>.

of the associated classes are stored in a machine-readable JSON file, which will be used in post-processing (see Sect. 3.4) for the correct mapping (assignment of the color values to a class). In this way, two related plans are created for each model.

3.4 Postprocessing

The postprocessing and derivation of data can be adjusted and automated for the respective use cases. In this study labeled object segmentation data in the *COCO-JSON* format is created for *Computer Vision* (CV) Tasks (Fig. 3). COCO-JSON is a common and standardized annotation format for CV tasks, facilitating interoperability by providing a structured representation of object detection and segmentation annotations in images.

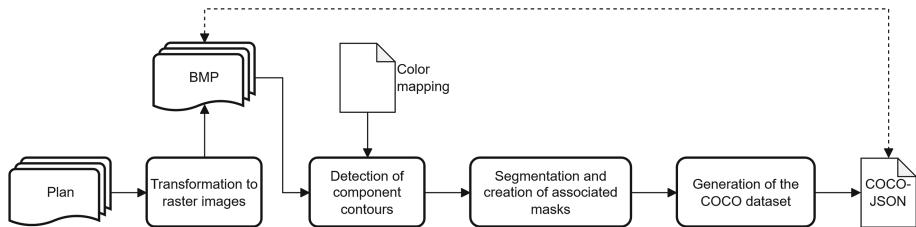


Fig. 3. Postprocessing to generate machine readable image training data

The basis for the automated processing of the plan data is the RGB labeling mask described in Sect. 3.3. The PDF plans are first converted into the Bitmap (BMP) graphic format. BMP is a lossless raster graphics format, which ensures that no compression artifacts occur that would cause a loss of quality in the downstream process. Semantic segmented masks are then generated according to the classification stored in the RGB mapping. Binary coded masks can then be derived for each classification mask, in which the contours of the respective components are clearly recognizable (Fig. 4). Using the OpenCV⁶ library, the contours of the components on the binary masks are extracted and saved as polygons in COCO format according to their class as well as the corresponding bounding boxes.

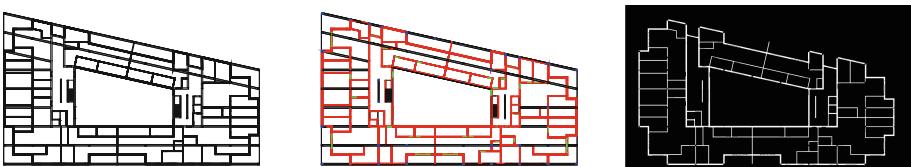


Fig. 4. Postprocessing masks

⁶ <http://www.opencv.org>.

4 Application and Validation

The described workflow is validated by using an application example. *YOLOv8*⁷ is a state-of-the-art object detection framework. It is easy to handle and fast to train, which makes it suitable for experiments and validation of usability of data. *YOLOv8* also supports instance segmentation, which presents an important technique for extracting information from a plan in a remodeling process. The segmentation boundary could then be the basis for creation of new model objects in a downstream process.

A dataset containing of 50 plans was created. Half of them were created parametrically the other half was derived from existing IFC models (Fig. 5). For this study the IFC models were gathered by combining models from student projects and freely available IFC files from open access online repository. As a result, the quality of the building model and the underlying authoring software varies broadly, as well as the IFC schemas and export mappings used.

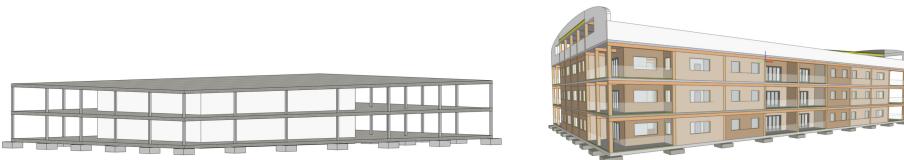


Fig. 5. Parametrically created model (left) and IFC model from a student project (right)

A self-developed Add-In for Autodesk Revit[®] was used to generate the parametrically controlled models via *Revit API*. Modeling was carried out according to programmatically implemented rules, whose limiting parameters can be defined by user input. These rules ensure formal technical correctness in the modeling process. Initially, a grid is generated for the placement of building components. The number of nodes and the spacing in the X and Y directions are randomly determined based on predefined step sizes. Using this grid, columns, walls, foundations, and slabs with openings are subsequently placed. Various types for each type of component (column, wall, foundation slab, isolated footing, slab) are available in the form of a component catalog. In addition to the localization (via the grid), the type of the corresponding component is randomly assigned. Furthermore, the number of gridlines, spacing between gridlines, dimensions, and position of the bracing wall core, as well as the number and height development of the floors, are randomly chosen. The resulting models resemble buildings constructed in a grid-based skeletal structure. Additional annotation such as dimensioning and item numbers of the components are added to the views.

The size of the object instances that are supposed to be detected are small compared to the size of the whole plans. Therefore, the plans were transformed to the JPEG image format and divided into tiles with a size of 2048 * 2048 pixels. The annotations were accordingly tiled as well and transformed to the *YOLOv8* annotations format. *YOLOv8* takes only the bounding boxes of the corresponding classes as an input, the segmentation

⁷ <https://github.com/ultralytics/ultralytics>.

information is not used. The tiling and annotation transformation was performed with the CV Dataset tool *Roboflow*,⁸ which is free to use for datasets that are made publicly available. The Dataset is now ready to be feed into *YOLOv8* as training data (Fig. 6).

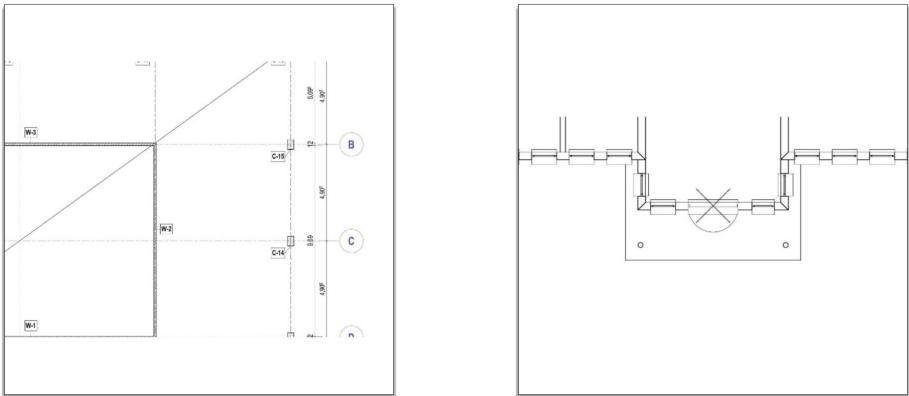


Fig. 6. Example Images from the dataset (left: derived from parametric model; right: derived from IFC)

The model was trained on two classes: Columns and walls. The pretrained *YOLOv8n-seg* (which is the smallest and therefore the fastest, but most imprecise versions of *YOLOv8*) model was used as a starting point. To further improve the performance the image size of the train input was reduced to 1024 * 1024 pixels. The mean Average Precision (mAP) is a common metric to evaluate the overall performance of object recognition models.

The model was trained for 200 epochs on a *NVIDIA T4 GPU* which took about one and a half hour to complete and resulted in a promising *mAP@0.5:0.95* of 0.42, with the classification loss continuously decreasing during training. This demonstrates that the data is readable hence usable for the ML framework to learn from the data provided via the proposed workflow.

For practical implementation, the inference, i.e. the performance of the trained model on real and unknown images, is an important factor. As can be seen in Fig. 7, the results show promise but are not yet suitable for practical use. The confidence for this inference run was set to 0.4, Meaning that only detections with a confidence score with at least 40% are shown. This decreases the risk of false positives but increases the risk of missed objects.

5 Conclusion and Future Work

The automated model based synthetic pipeline for generating synthetic labeled training data from IFC files combined with parametrically created models has proven to be feasible. The proposed workflow was validated by creating a semantically segmented plan dataset, which was used by a *YOLOv8* segmentation network for instance segmentation.

⁸ <https://roboflow.com/>.

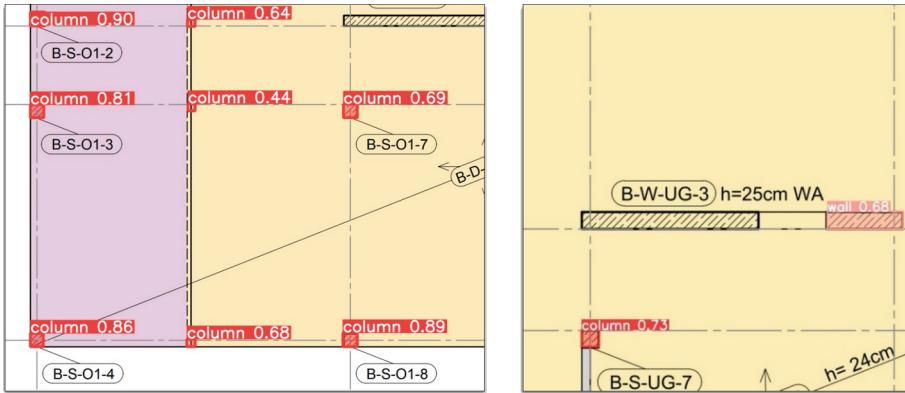


Fig. 7. Inference on realistic real-world plan data

The chosen segmentation type as polygons has the downside that only whole objects can be described therefore closed continuous building elements like surrounding walls are treated as one big wall object. For this work the plans were divided into smaller tiles and only the bounding boxes were used as training input, therefore these unrealistic holeless segmentation areas are not too significant. COCO also supports the so-called Run-length encoding (RLE). RLE is compression algorithm that can embed pixel wise segmentation information. This could solve this issue and will be tested in future work. Wei et al. investigated and compared the possibilities of detection and segmentation in a CV based plan analysis context and showed challenges and possible solutions for the further downstream remodeling process [10]. The generation of labels from of pixel-based images through RGB-mapping poses a potential source of error due to compression methods that generate compression artifacts or down sample the color spectrum. This was avoided by leveraging high resolution BMP image formats for creation of the labels. Another approach would be trying to extract the labeled information directly from the model or from the vector information stored in the PDF format.

The IFC-Models for this work were collected and checked manually, as the requirements for the tested use case were very low. Automatic scraping of the models from databases and implementation of automated model checking for complex requirements should improve the scalability of the proposed workflow.

To further enhance the inference of the training data, several options are feasible. Firstly, more input data could be gathered. The workflow is mostly automated so there is no significant additional expense when using a bigger model base. Furthermore, common data augmentation practices, such as alterations in color, blurring, and tilting, are recognized for their efficacy in improving object detection performance. Therefore, it can be reasonably assumed that the application of these techniques would likely enhance the quality of detection in our context as well.

With this work, we outline a scalable and flexible end to end approach to generate substantial amounts of specifically tailored labeled training data with minimal effort. Our findings show the urge for improved interoperability and standardization of model data. IFC provides a comprehensive data structure for building related information.

Providing these models on centralized and freely accessible platforms may help to overcome challenges in data acquisition on the road to leverage the potential of modern machine learning technologies for the AEC sector and beyond. The presented approach, a multimodal, automated workflow for generating synthetic training data from randomly controlled parametric BIM models combined with existing IFC models, offers a promising opportunity to expand the database for AI applications. The source code for the described workflow will be provided on request.

Acknowledgments. This work is supported by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag as part of the KIFa-Project. Our explicit thanks go to Dr. Michael Kraus for his support and input in the early stages of this research.

References

1. Noichl, F., et al.: “BIM-to-Scan” for Scan-to-BIM: generating realistic synthetic ground truth point clouds based on industrial 3D models. In: Presented at the 2021 European Conference on Computing in Construction July 26 (2021). <https://doi.org/10.35490/EC3.2021.166>
2. Hong, Y., et al.: Synthetic data generation using building information models. *Autom. Constr.* **130**, 103871 (2021). <https://doi.org/10.1016/j.autcon.2021.103871>
3. Hoeng, S., Eder, F.: Methode zur Generierung multimodaler synthetischer Daten aus parametrischen BIM-Modellen zur Nutzung in KI-Systemen. In: Tagungsband 34. Forum Bauinformatik, pp. 398–405. Ruhr-Universität Bochum (2023). <https://doi.org/10.13154/294-10130>. [In German]
4. Schönenfelder, P., et al.: Automating the retrospective generation of As-is BIM models using machine learning. *Autom. Constr.* **152**, 104937 (2023). <https://doi.org/10.1016/j.autcon.2023.104937>
5. Alawadhi, M., Yan, W.: Deep learning from parametrically generated virtual buildings for real-world object recognition. <http://arxiv.org/abs/2302.05283> (2023). <https://doi.org/10.48550/arXiv.2302.05283>
6. Clever, J., et al.: Deep learning approach for predicting pedestrian dynamics for transportation hubs in early design phases. Presented at the EG-ICE 2021 July 14 (2021)
7. Gerbino, S., et al.: On BIM interoperability via the IFC standard: an assessment from the structural engineering and design viewpoint. *Appl. Sci.* **11**, 23, 11430 (2021). <https://doi.org/10.3390/app112311430>
8. Ma, W., et al.: Generative design in Building Information Modelling (BIM): approaches and requirements. *Sensors*. **21**(16), 5439 (2021). <https://doi.org/10.3390/s21165439>
9. Kremer, N.C., Beetz, J.: Extending—information delivery specification—for linking distributed model checking services. In: Presented at the EC3 Conference 2023 (2023). <https://doi.org/10.35490/EC3.2023.266>
10. Wei, C., et al.: Interoperability between deep neural networks and 3D architectural modeling software: affordances of detection and segmentation. *Buildings* **13**(9), 2336 (2023). <https://doi.org/10.3390/buildings13092336>



Natural Language Communication with Sensor Data Through a LLM-Integrated Protocol: A Case Study

Fanglai Jia, Arianna Fonsati^(✉), and Kjartan Gudmundsson

KTH, Royal Institute of Technology, 100 44 Stockholm, Sweden
ariannaf@kth.se

Abstract. The ability to share data can facilitate cooperation and decision making throughout the entire life cycle of buildings, from initial stages of planning, through design and construction, up to the management of assets and towards end of life and recycling or reuse. Accordingly, the different actors of the Construction industry can share data and functionalities across software platforms through automated processes. Such processes involve stakeholders with heterogeneous backgrounds; for this reason, it is of value to make data available to people without expert knowledge of specific programs or computer systems.

This study is concerned with digital protocols for automated capturing of real-time sensor data for the assessment of building performance. The research explores a Large Language Model (LLM) driven protocol for sensor data acquisition to evaluate building performance, leveraging the existing database where building data is stored. Using pre-written prompts which instruct how to carry out tasks in a step-by-step manner, the LLM can utilize pre-defined functions built upon API calls for communication with sensor data, including data acquisition, post-processing, and interpretation. The LLM receives instructions by the User using natural language. Initial tests underscore the protocol feasibility, highlighting its potential utility to improve the data communication of existing digital twins for individuals without professional expertise in building management. A case study exemplifies how human instructions in natural language trigger the LLM to invoke a request for indoor climate sensor data in a building.

The developed tool was tested by professionals working in the Construction industry and the Educational Sector to provide feedback on its practical application. This evaluation helped to identify weaknesses for future developments and proved the flexibility and adaptability of the method.

As advancements in Artificial Intelligence (AI), particularly in LLMs, continue to surge, there is anticipation for further refinements. This includes cost reduction, enhanced stability, and the integration of more advanced functionalities such as advanced data analysis using machine learning coded by LLM.

Keywords: Digital Twin (DT) · Artificial Intelligence (AI) · Natural Processing Language (NPL) · Large Language Models (LLMs)

1 Introduction

Digital twin (DT) has emerged as a transformative concept in various industries, including the Architecture, Engineering and Construction (AEC) sector, offering different levels of digital representation of physical entities. Accordingly, digital twins can facilitate monitoring, understanding, and optimizing the functions of their physical counterparts, thereby enhancing the quality of life and well-being of humans [1]. A crucial aspect of DTs is the ability to capture a wide range of data. In this research, data capture includes the following indoor climate-related variables: temperature, relative humidity, and CO₂ concentration. Thus, data communication plays a pivotal role in the functionality of DTs, and Human-Computer Interaction (HCI) employing artificial intelligence (AI) is a desirable solution for a higher level of data communication [2].

In the past years, the evolution of AI has ushered in the era of Large Language Models (LLMs), which are pre-trained self-supervised systems that can be prompted or fine-tuned to a wide range of natural language tasks, without requiring a separate network model [3]. LLMs possess unparalleled capabilities in Natural Language Processing (NLP), which could be exemplified by the revolutionary GPT-3.5 released on November 28, 2022 by OpenAI. This NLP expertise by LLMs could bridge the gap between computer and humans in data communication, which means that a LLMs-integrated protocol might help different stakeholders in the building management sector understand and interact with building information such as sensor data. Compared to the traditional NLP, LLM does not require customized and complicated algorithms on semantics and syntactics and it necessitates less labeled training data than conventional machine learning (ML) approach [4].

Within this framework, researchers have delved into the possibility of applying NLP and LLMs to address specific challenges in the construction industry. For instance, in the field of project and construction management, Amer et al. employed a NLP-based approach to automatically correlate master schedule activities with look-ahead planning tasks [5]. Meanwhile, Wang et al. integrated case-based reasoning (CBR) and NLP to enhance knowledge retrieval and capture in the Building Information Modeling (BIM) environment [6].

Despite the potential benefits of employing LLMs in the construction sector, there is a notable absence of research on their integration with Digital Twins specifically geared towards establishing human-building interaction. This interaction might leverage sensor data acquisition and analysis for optimizing building performance and energy management. To address this research gap, this paper introduces a method designed to streamline HCI and facilitate decision-making in building energy management for different kind of stakeholders.

The developed tool focuses on digital protocols for automated real-time sensor data capture to monitor building performance. It illustrates how the Application Programming Interface (API) of a building control system can populate the elements of a DT with real-time sensor data. By utilizing the existing Building Management System (BMS) and its database capturing energy efficiency-related indicators, the study explores a LLM-driven protocol for sensor data acquisition. A case study demonstrates how human instruction in natural language, facilitated by a LLM, can request indoor climate sensor data within a building.

This tool underwent testing with researchers to gather feedback on its practical utility. The evaluation not only identified weaknesses but also outlined potential avenues for future development, highlighting the adaptability and flexibility of the proposed method.

2 Method

This research mainly concerns the integration of LLM and systems for data acquisition in order to improve data communication among different stakeholders in the building management sector. The aim is to develop a tool to enhance efficient communication and make building data easily available upon request in natural language, using a chat application. To develop such an application, the LLM needs to be prompted in a step-by-step way which generally includes identification, execution, and interpretation. Identification means to grasp the Users' intention in their instruction, which is a specific action like data acquisition on specific building objects. After having identified the task, the LLM triggers the execution of predefined functions and sends back the execution results to the User in natural language. By encapsulating all the technical details required in its prompt, the LLM is able to answer questions on building data for those who are not familiar with specific proprietary software or platform.

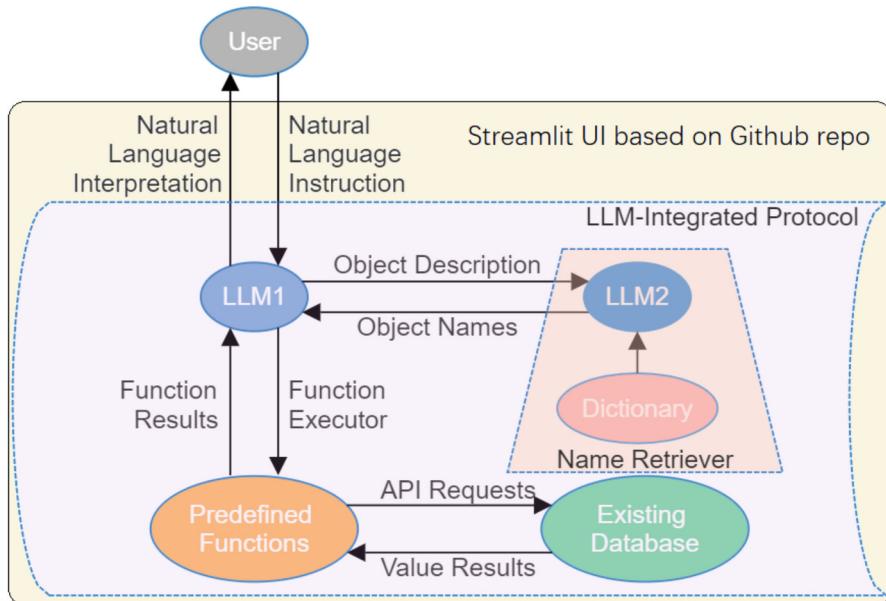


Fig. 1. The paradigm of the proposed LLM-integrated protocol

In this research, the case study of Undervisningshuset (U-building) was used to develop the application. The proposed protocol (Fig. 1) described below was imposed upon the existing Building Management System (BMS). The building management system of the U-building is provided by Nordomatic, a BMS service integrator based in

Europe. The cloud-based system for data retrieval and consultation is called “Styrportalen”. With its 767 sensors installed in the U-building, the BMS is capable of delivering both real-time and historical data via web interfaces or API endpoints using HTML5 [7]. Accordingly, the data retrieval from the BMS Cloud can be effectively pursued through both Web Scraping and API Calling. Compared to the potentially limited data availability determined by the API provider, the web scraping approach has expansive data accessibility and flexibility, allowing virtually any data rendered in a browser to be accessible through an automated and choreographed way. However, this approach faces the following challenges: (i) delay in response: notable delays might occur during the data extraction processes; (ii) potential legal issue: it may encounter legal challenges and activate anti-scraping mechanisms on websites; (iii) scalability concerns: each procedural step necessitates explicit hardcoding; (iv) intricate data management: for tabular data, the method mandates substantial time to sequentially load and parse each cell, considering that HTML elements require unfolding and meticulous searching; (v) stability issues: there is a tangible risk to stability if the scraping process becomes complicated, particularly when encountering not clickable elements due to overlapping entities on the webpage; (vi) external executable requirement: dependency on external executables, such as Chromedriver, is requisite.

Fortunately, the APIs defined by Nordomatic includes a comprehensive set of features. In general, API calling ensures: (i) rapid data acquisition: swift response owing to the adherence to well-defined connectivity protocols; (ii) systematic data requests: structurally defined requests facilitate scalability and coherence of the code; (iii) ease of implementation and stability: when implementing according to the providers’ instruction, it guarantees a stable data retrieval process; (iv) facility for preliminary testing: initial testing on platforms, such as Postman, is supported with convenient conversion into code using its code snippet feature.

Therefore, for efficient and reliable data collection, this study utilizes HTTP POST request methods directed at the API endpoints defined by Styrportalen to retrieve sensor data from its cloud database. The POST method enhances data transmission security by concealing request parameters within the request body, rather than exposing them in the URL, thus preventing sensitive information from being leaked out. Also, it exhibits versatility in handling diverse data types including text, JSON, XML, and binary, offering flexible format options in data communication. The specific URL for the API varies for different actions. Initially, Users must send a login request to the designated login URL, embedding the API Username, password, and a client ID - generated using an online GUID generator - within the request payload. The login session needs to be maintained for further data extraction. The session ID and security token, extracted from the returned JSON response from the login request, must be integrated into the payload and headers of subsequent requests. Additional necessary parameters, such as sensor names, should also be included.

This research used LLM models developed by OpenAI [8] in order to implement the LLM-integrated protocol proposed in * MERGEFORMAT Fig. 1. The LLM needs to have some sort of memory for conversation in order to perform complex tasks. However, the LLM API provided by OpenAI does not have inherent memory. For this reason, every time the API is called, all the relevant memory, including the prompt that guides the LLM,

needs to be attached to the message to provide context for the LLM. The prompt for the LLM in the system message section is well-structured in a step-by-step way, along with commonly used sentences like “take a deep breath and think step by step” to elicit LLM’s Chain of Thought (CoT) feature, which lowers the risk of producing random make-up answers, technically called “hallucination”. To facilitate the commissioning of the protocol by making the process transparent and explicit, the function-calling feature by LLM through fine-tune is not applied in this research, instead, this part is hard coded in the prompt. The complete version of the prompt for the LLM is available upon request.

Each API call requires the use of “tokens”, which are defined by OpenAI as “pieces of words” and are limited based on the language model. For example, the API gpt-3.5-turbo LLM is limited to 4K tokens, which correspond to about 3K words. To execute the task and conversation, the prompt takes up around 2K tokens, so there are around 1,5K words left, enabling about ten rounds of exchanges between human and LLM. Even though OpenAI released a newer version “gpt-3.5-turbo-1106” that has 4 times token limit, this version does not have the same intelligence and flexibility during the test, so it might only serve as the auxiliary LLM (LLM2) not the main LLM (LLM1).

This study proposes a protocol that works as shown in Fig. 1, where the information flows in a clockwise manner. The User sends the natural language instruction to the protocol as input, which is received by LLM1. LLM1 is employed as the building assistant that is responsible for the whole conversation, and it acts as prompted. However, LLM1 has no prior knowledge of the desired data. On the other hand, the Users only know what data they want to retrieve (e.g. indoor temperature or relative humidity of a specific room) without knowing the sensor names and express their wish implicitly in their instruction.

By its knowledge, LLM1 can only decompose the User’s instruction into intention (e.g. looking up current value or historic value) and sensor description (e.g. temperature or relative humidity of which room), but it does not know the exact sensor name. After that, the LLM2 comes into play as “Name Retriever”. The LLM2 would first load the dictionary, which is a predefined list of sensor names with a corresponding description. Based on this dictionary, the LLM2 filters out the sensor names that best fit the User’s demands and send them back to LLM1.

After having identified the action and the object, LLM1 has enough knowledge to send a function executor to executor (JSON string) to execute the predefined functions. The function utilizes API requests to gather data, and the results of the execution are sent back to LLM1 to decide whether the current loop is completed or not. The steps in the process described above are virtually invisible to the User except for the interpretation by LLM1. If LLM1 decides that the mission has been completed at least for that specific round, it would return its interpretation, which is a reply in the form of complete sentence that would incorporate the returned data value in natural language. If further post-processing is needed, the LLM1 would decide to re-enter the loop. In this protocol, integrated LLMs are utilized for specific purposes. LLM1 has the function to smooth the communication process in HCI by following the set agenda described in the instruction prompt. On the other hand, LLM2 filters out the intended object names from the dictionary.

By modularizing the data retrieval and post-processing tasks into functions, which ranges from data acquisition of current and historic values to trending data presentation, including spreadsheets exportation and graphical plot-generation, a mapping between functions and their respective names has been established. Subsequently, the LLM1, based on the mature “gpt3.5-turbo-0613” model, is equipped with a handle called JSON string that enables it to make an impact in the external Python environment by executing functions that are predefined in Python code. As prompted, when the LLM1 tried to satisfy User’s demand, it executes predefined functions by returning a JSON string to the parser function. The JSON string would be parsed and loaded as Python dictionary that consists of function name and arguments. Using the mapping’s get methods, it would elicit the execution of the function by its name and arguments. After the execution triggered by the executor JSON string, the returned value of the predefined function would be sent back to LLM1 as the execution results, so that the LLM1 would serve the User as intended.

The integration of the LLMs is also intended to broaden User scenarios to accommodate individuals unfamiliar with the specialized sensor nomenclature. Consequently, a ‘translator’, converting User’s instructions in natural language into specific sensor names, is realized by LLM2. However, the LLM2 itself does not have any knowledge about the U-building, the translator needs to refer to a csv list acting as a dictionary between sensor names and their descriptions. In this context, the csv agent, based on the newly released “gpt3.5-turbo-1106” model and developed under the open-source Langchain framework, is tailored for the job. By referencing the dictionary, it is able to identify and retrieve the specific sensor names that the User desires.

Therefore, the synergy of the two LLMs functions as follows: LLM1 receives the User’s natural language instruction, it discerns the intended actions, and then it retrieves object (sensor) names with the assistance of LLM2. LLM1 then produces a pertinent JSON string to initiate the desired functions. Subsequently, LLM1 either generates a JSON string for further post-processing or directly provides an interpretation to conclude the ongoing conversational exchange.

To preserve confidentiality, particularly concerning sensitive credentials like API usernames and passwords, a webpage application developed with Streamlit was conceived to serve as the frontend. The entirety of the execution loop in the backend is isolated within an independent threading that is free from state refreshing, employing queues to facilitate information exchange between the frontend and backend. All the pertinent files, including frontend and backend components, are securely stored in a private GitHub repository. This repository is linked to Streamlit Cloud, whose convenient deployment and maintenance enable the program to operate as a webapp easily accessible without any download or installation.

3 Results

Using the method described above, the U-Building (U-B) Assistant tool was developed, with the aim to establish a connection between building Users and the Building Monitoring System. Figure 2 shows an exemplar interaction using the tool. Using only natural language, the U-B Assistant interacts with the User as prompted by answering questions

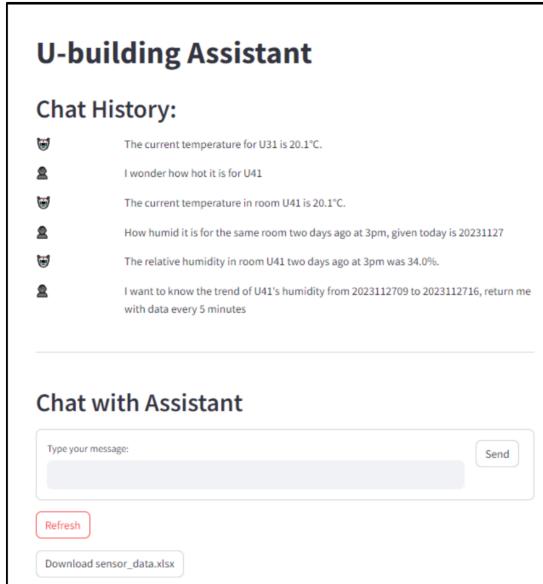


Fig. 2. User interface of the U-building assistant

or generating files based on the data retrieved from the BMS. The data retrieved could be validated by manually executing the predefined functions to preclude hallucination. The use of a LLM integrated protocol is visible when the User asks for the humidity trending value in a given time span of 5 min. This request required some simple arithmetic reasoning to get the right result, which could not have been achieved by using traditional NLP methods. With all the features shown above, it is also vital to host the program for Users in a proper manner. By using the free Streamlit cloud, the administrator could decide who has the access to the webapp and every User would use different threading so that they do not interfere with each other, which is vital for security and privacy issues.

The researchers involved in the study tested the U-Building assistant in order to evaluate its performance in five main aspects, mainly those concerning HCI. Despite the highlighted potential demonstrated by the overall high score, there are still several drawbacks as highlighted in Sect. 3.1.

For the primary purpose of testing the tool, the aesthetic design of the User interface was not elaborated upon. This includes features such as the layout and alignment of the buttons and sections, which might cause minor confusion when testers are unfamiliar with the web app. Due to the same reason, the way of organizing files generated by the assistant could sometimes cause confusion. The newly generated file would overwrite the existing file of the same type so that the User might miss the desired file, but this situation could be prevented if the User downloads the files timely. Owing to the session-state feature of Streamlit, the User has to refresh manually every time they expect an update in the chat message section. Another issue that might cause frustration is the high traffic time of OpenAI API service, which might take a long time for the API to respond when it is overwhelmed by requests. Even if the OpenAI API itself has no intrinsic

memory in theory, in practice, the feedback by the User suggests that there might be a warming up time for the LLM to get more familiar with the task.

Another limitation is due to the token limits of 4K, which are just sufficient for around ten cycles of tasks. This might be improved by implementing the mechanism to get rid of the unnecessary and irrelevant memory. However, the tester still needs to reboot the application if they expect the token limit is about to be surpassed, otherwise the LLM would return an error. The similar updated model, (1106) with larger context capacity of 16K tokens, is not capable of carrying out the task, because it has been fine-tuned to be conservative so that it tends to resort to “I don’t know” whenever it feels uncertain. Nevertheless, the currently employed model (0613) is not perfect; sometimes it would calculate wrong interval value for the trending data, so it needs the tester to correct it by reinforcement learning from human feedback (RLHF). GPT-4 API might outperform the current model, but it is unsuitable for this application due to cost and slower response. For the reasons mentioned above, the developed web app is quite dependent on the proprietary LLM, and it is vulnerable to any unexpected situation happening to the OpenAI API.

Therefore, the LLM itself should also be equipped with better knowledge. For instance, the LLM can only make interpretation rather than evaluation (i.e. if the value exceeds the acceptable range), for its lack of knowledge about the values meaning in the real world. Besides, this web app was developed as proof of concept; the sensor names included in the “dictionary” do not cover all 767 sensors. Accordingly, the dictionary might be expanded to the full list of sensors in practice.

3.1 Tool Assessment

The U-B Assistant showed good results in terms of Human-Computer Interaction. To assess its performance, researchers involved in the study tested the tool. After the testing, the participants were asked to fill out a questionnaire to provide feedback and evaluate the experience with the Assistant. To assess the tool, the following factors have been used for the evaluation: Perceived Usefulness (PU), Perceived Ease-of-Use (PEU), Human-Computer Interaction (HCI) and Communication. A five-point Likert scale was used to evaluate the selected parameters. The scale for the evaluation included numbers and verbal labels from 1: “Strongly disagree” to 5: “Strongly agree”.

Perceived usefulness and perceived ease-of-use were evaluated using the Technology Acceptance Model (TAM) developed by Davis (Davis, 1989). PU is defined as “the degree to which a person believes that using a particular system would enhance his or her job performance”, while PEU is “the degree to which a person believes that using a particular system would be free of effort” [9]. To evaluate the HCI and more specifically the “experience of interacting with a technology via its interface during use” [10] the Technology-based Experience of Need Satisfaction (TENS)-Interface was used. This model was developed by Peters et al. as an adaptation of the PENS [11] for non-game technologies. Therefore, a new measure was elaborated by the authors to evaluate the “Communication” with the tool, to assess the level of understanding and interpreting questions of the tool in natural language.

The overall perceived usefulness ratings were high (mean = 4.17, SD = 0.29) and the overall perceived ease-of-use ratings were even higher (mean = 4.50, SD = 0.35). The

evaluation of the HCI parameter using TENS-Interface was divided into Competence (mean = 4.00, SD = 0.81) and Autonomy (mean = 3.80, SD = 0.35). A significant difference was found in the evaluation of the overall autonomy, which got the lowest rating. The two statements that received the lowest scores are “The U-B assistant provides me with useful options and choices” and “I can get the U-B assistant to do the things I want it to”. This value means the participants were not fully satisfied with the tasks provided by the tool. However, the Communication with the tool was highly ranked (mean = 4.00, SD = 0.29), which means the tool showed a good level of understanding and interpreting natural language.

Furthermore, participants were asked for suggestions on further implementation of the tool. For this purpose, two open-ended questions were included in the survey, “From your point of view, what are the possible future developments of the U-B assistant?”, and “What tasks would you implement to increase the perceived usefulness of U-B assistant?”.

The responses provided detailed suggestions on future developments as follows: (i) faster and more stable response from the tool in the chat; (ii) improved User interface, including, for instance, visual placement of devices in the building; (iii) improvement of the plot function, since the tool can provide trends and show those in diagrams but this function is not clear when using the application; (iv) adding more data post-processing features, such as comparison between real-time and historic values or implementing predictions on future scenarios by changing external conditions.

Amongst the tasks to implement in order to increase the perceived usefulness of the U-B assistant, the participants included:

- Indoor comfort evaluation: the U-B assistant should be able to evaluate whether the indoor environment is suitable or not;
- Alert provider: the U-B assistant should be able to provide alerts if the data retrieved by Users is out of a predefined range;
- Energy performance gap evaluation: the U-B assistant should be able to use data captured for building performance simulation and to compare it with design data, in order to evaluate the actual building performance and estimate the eventual energy performance gap.

To conclude, the User-testing phase showed good results in terms of PU, PUE and Communication of the tool. However, it highlighted limitations in terms of HCI, especially in relation to the “Autonomy” parameter, since the tool got lower ratings when participants where asked if they can get the U-B assistant to do the things they wanted.

4 Conclusion and Further Developments

As one of the first Natural Language toolset for data consultation, the U-Building assistant has both good promise and a series of limitations. The target of the U-Building includes several stakeholders, such as building occupants, energy researchers/experts, real estate/building managers etc. On one side, the tool enables an easy access to information on building performance for any professional in the building sector. The tool also provides an easy-to-use interface to “chat” with the building. This function might

help Users' raise awareness towards building performance-related issues. On the other side, it is helpful for energy experts/managers to easily consult data captured through the BMS system and use such information for further elaboration and analyses.

For this reason, this protocol of communicating the sensor data powered by LLM would facilitate the management of the U-building, due to its natural language inputs and outputs and relatively easy-to-use feature. Based on the existing BMS, the protocol in this study could be easily populated with other buildings that share the same BMS provider. Accordingly, the presented method has potential of wide application for a better communication in building management due to its adaptability and scalability. Furthermore, the proposed integration of LLMs and BMS would enhance responsiveness to Users and simplify data collection, which is still one of the most critical phase of design and asset management, in buildings and infrastructures equipped with such a protocol. Although current LLMs have not achieved Artificial General Intelligence (AGI), its integration in BMS might get a step closer to the concept of intelligent buildings. Along the way to a wide application, the integration of LLM could catalyze a paradigm shift in Architecture, Construction and Engineering professions. The key challenge lies in effectively delegating routine tasks to LLMs, while ensuring humans retain control over major decisions.

At the same time, there is great room for improvement and further development. One of the limits is related to the BMS system; without proper planning, installation and monitoring of devices for data capture and related web-portal, the tool could not be implemented. This means the applicability of the tool is limited to specific and high standard buildings. Another limitation is the number of tasks of the tool, which is able to retrieve and communicate data using a natural language processing, but it is not capable of any kind of data analysis. As far as the User interface is concerned, the User's experience should be analysed more in detail, together with the way of organizing the file both in aesthetical and practical aspects to ensure clear data communication. As for the LLM itself, in the near future, proprietary LLMs such as ChatGPT, Claude3, Comate, and Gemini might be the only option for task completion of this complexity, but open-source LLM like LLaMA should also be assessed. For the prompt part, since it has been developed in a way that adds complexity gradually, it turned out to be quite long. For this reason, there is room for prompt compression [12] and modularization [4]. Furthermore, the knowledge of LLM should be improved. An automated mechanism might provide current time for the LLM as well as an expanded dictionary in which more sensors names would be included along with their normal range of values. Since the predefined functions are highly adaptable, they might be easily expanded to be able to alert the User about abnormal sensor value or evaluate the thermal comfort by calculating Predicted Mean Vote (PMV) or Predicted Percentage of Dissatisfied (PPD).

Aside from being employed to smooth data communication, it is of interest to dig deeper into LLM's potential in data mining and analytics to extract more information in building's data in the future research. While it is guided and overseen by human, the current GPT-4 is now able to receive the file and do the data mining applying modularized machine learning techniques by running the code on its code interpreter. This feature might be exploited to mine the sensor data as well as to extract patterns, detect anomalies, make predictions, and propose strategies for improvement, which are of paramount

importance in decision making when managing the building. In this context, it remains largely unexplored how to efficiently prompt the LLM in order to carry out data mining and its validation.

Furthermore, multi-modal feature of LLM might also be applied in building management. In this case, by analysing and learning from the pictures of common building damages and grasping the patterns explicitly, the LLM might identify these damages. Being equipped with validated domain knowledge in building physics that could be applied to discern the cause of common building damages, the LLM may analyse the causes and bring up possible solutions based on pictures. Under some circumstances where it is dangerous for human to step in, a humanoid robot such as Optimus by Tesla or “Alter3” by Tokyo University [12] could even leverage this trained LLM to act as human for building maintenance. What domain knowledge does it take for the LLM to achieve an acceptable accuracy and how would the HCI shift under this scenario are questions that matter and require further investigations. Therefore, wider application of LLMs to asset management would lead to higher efficiency and reduction, both in terms of resources (e.g. energy) consumption and costs.

Along the way, it is indispensable for professionals to scrutinize the results generated by AI, while also considering the data security and privacy issues.

To increase the tool capabilities, the authors plan to implement its functionalities together with machine-learning approaches. This would increase the usefulness of the tool for several stakeholders, fostering the data-driven approach towards energy management in the building industry.

References

1. El Saddik, A.: Digital twins: the convergence of multimedia technologies. *IEEE Multimedia* **25**(2), 87–92 (2018)
2. Wang, T., Li, Y., Deng, Y., Wang, C., Snoussi, H., Tao, F.: Digital twin for human-machine interaction with convolutional neural network. *Int. J. Comput. Integr. Manuf.* **34**(7–8), 888–897 (2021)
3. Sejnowski, T.: Large language models and the reverse turing test. *Neural Comput.* **35**, 309–342 (2022)
4. Zheng, J., Fischer, M.: BIM-GPT: a prompt-based virtual assistant framework for BIM information retrieval (2023). arXiv preprint [arXiv:2304.09333](https://arxiv.org/abs/2304.09333)
5. Amer, F., Jung, Y., Golparvar-Fard, M.: Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. *Autom. Construct.* **132** (2021)
6. Wang, H., Meng, X., Zhu, X.: Improving knowledge capture and retrieval in the BIM environment: combining case-based reasoning and natural language processing. *Autom. Construct.* **139** (2022)
7. Zhang, K.: Studying building behaviors by using the Building Management System of a new teaching building: a study case of a school building in Stockholm. KTH, Stockholm (2020)
8. OpenAI Models: <https://openai.com/product>. Last accessed 22 Jan 2024
9. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **13**(3), 319–340 (1989)
10. Peters, D., Calvo, R.A., Ryan, R.M.: Designing for motivation, engagement and wellbeing in digital experience. *Front. Psychol.* **9** (2018)

11. Rigby, S., Ryan, R.M.: *Glued to games: how video games draw us in and hold us spellbound.* Praeger, Santa Barbara, CA (2011)
12. Jiang, H., Wu, Q., Lin, C. Y., Yang, Y., & Qiu, L.: Llmlingua: Compressing prompts for accelerated inference of large language models (2023). arXiv preprint [arXiv:2310.05736](https://arxiv.org/abs/2310.05736)
13. Yoshida, T., Masumori, A., Ikegami, T.: From text to motion: grounding GPT-4 in a humanoid robot” Alter3” (2023). arXiv preprint [arXiv:2312.06571](https://arxiv.org/abs/2312.06571)



Integrating Machine Learning for Cyber Risk Analysis in Construction 4.0

Dongchi Yao^{1,2}(✉) and Borja García de Soto^{1,2}

¹ S.M.A.R.T. Construction Research Group, Division of Engineering, Experimental Research Building, New York University Abu Dhabi (NYUAD), Abu Dhabi, United Arab Emirates
dongchi.yao@nyu.edu

² Department of Civil and Urban Engineering, Tandon School of Engineering, New York University (NYU), Brooklyn, USA

Abstract. The construction sector is transitioning into a data-centric industry characterized by integrating sophisticated technologies such as digital twins, robotics, and cloud computing with traditional construction methods. This shift has resulted in construction firms generating and managing an unprecedented volume of data in a digital environment, thereby intensifying their vulnerability to cyber threats. Consequently, the imperative for cyber risk analysis, which encapsulates risk identification, estimation, and mitigation, has grown to protect data and prevent potential losses. While risk identification and estimation form the cornerstone for developing effective risk mitigation strategies, existing literature in the construction sector primarily focuses on methods that rely heavily on lengthy human involvement, which can result in subjective outcomes. To bridge this gap, this study explores Machine Learning (ML) techniques to refine and optimize these tasks, leveraging their inherent automation capabilities. Our exploration begins with an examination of the broader application of ML in general cyber risk analysis, identifying popular ML algorithms through a simplified bibliometric analysis. Following a standard ML system design approach, this study proposes four ML frameworks specifically designed for risk identification within the construction sector, ranging from multi-class classification methods to deep learning-driven generative models inspired by advancements in Natural Language Processing (NLP). For risk estimation, we also propose four distinct ML frameworks, each characterized by the specific format of the input threat-vulnerability pair pertinent to construction scenarios. The proposed frameworks serve as invaluable assets for construction stakeholders, enabling even those with limited cybersecurity expertise to enhance the cybersecurity robustness of their projects.

Keywords: Cybersecurity · Automation · Cyber-physical systems · Natural language processing · Machine learning

1 Introduction

The global construction industry is transitioning towards Construction 4.0, which aims to construct, operate, and maintain assets faster, cheaper, and with higher quality [1]. In Construction 4.0, the cyber-physical system is the core component, which involves

the interaction between cyberspace and physical space through data. Protecting data is critical in the system, consisting of three layers: physical, connecting, and digital. The physical layer captures on-site data using sensors and drones, while the connecting layer stores data using BIM and Common Data Environment (CDE), and the digital layer analyzes the data using cloud computing, big data, and virtual reality (VR) technologies [2, 3]. As these technologies perform vital functions in construction projects, it is essential to safeguard the data [4].

However, the construction industry is vulnerable to cyberattacks that can lead to economic losses, physical property damages, and even loss of human life, and the construction sector does not prioritize cybersecurity as a high business priority, as indicated by a recent survey by the Department for Digital, Culture, Media, and Sport [5]. Cybersecurity incidents in construction have repeatedly occurred globally in recent years. These incidents include Turner Construction falling victim to a spear-phishing scam [6], Jewson discovering an abnormal piece of code compromising personal data [7], Marous Brothers Construction not receiving payment due to maliciously changed routing numbers [8], Bird Construction being breached by ransomware [9], and Hoffmann Construction reporting unauthorized access to employee information [10].

Given the frequent cyber incident cases mentioned above, construction companies should be able to conduct systematic cyber risk analyses to identify and evaluate potential cyber risks. Although general cybersecurity solutions can tackle some issues, they fall short in mitigating the unique cyber risks arising from the specific characteristics of the construction industry. Furthermore, existing literature on cyber risk analysis within the construction industry is limited, and many proposed methods require substantial human involvement, leading to inefficiencies and time-consuming processes. Therefore, there is a need to develop an efficient technique tailored to the specific needs of the construction industry, one that can achieve automation capabilities while minimizing the need for extensive cybersecurity domain knowledge among construction practitioners.

ML techniques can be utilized to address this need. Given a database with sufficient risk-related records (e.g., [11]), ML can learn from the database, find regularities, and then make reliable predictions. It could also automatically complete risk decision-making and launch risk management measures. The data-driven results from ML are objective, adaptive, and accurate. Therefore, it is believed that ML is a good option to address the challenges in cyber risk analysis in construction [12]. This study proposes ML frameworks for cyber risk identification and estimation within the construction industry, which can enhance industry awareness and focus on cybersecurity by providing user-friendly, automated tools to help risk managers protect construction projects and improve cybersecurity practices across the industry.

2 Related Works

This section gives an introduction to cyber risk analysis within the construction industry, explores existing works on cybersecurity in construction, and conducts an ML investigation to identify popular algorithms for general cyber risk analysis.

2.1 Cyber Risk Analysis in Context

Cyber risks in construction refer to any financial, reputational, and safety loss caused directly by data loss or indirectly by data tampering in cyberspace. Cyber risks are committed by cyber threats exploiting vulnerabilities in an asset. According to different sources, cyber threats can come from the surrounding environment of a construction company or site (e.g., electromagnetic interference), malicious attackers (e.g., external opponents or retaliatory internal employees), or unintentional personnel (e.g., those with low-security awareness). Vulnerabilities can be classified into technological vulnerabilities (e.g., coding mistakes) and managerial vulnerabilities (e.g., disorders of data exchanges) [13]. Threats and vulnerabilities in construction are mostly attributed to strong interactions among participants, a large number of stakeholders, long supply chains, different phases of a project, and a lack of cybersecurity awareness and interoperability [14]. If not properly identified and addressed, cyber risks taking place in earlier phases can accumulate continuously or manifest into later phases [4]. Therefore, performing an efficient cyber risk analysis is crucial for ensuring the performance, productivity, and business of construction companies in all phases.

Cyber risk analysis refers to the process of evaluating the security attributes of a cyber system, such as confidentiality, integrity, and availability, and according to the results, corresponding risk management measures/solutions are selected and implemented [15]. Cyber risk analysis should be performed across the whole life cycle to assess different risks in different phases. The entire cycle of cyber risk analysis [13] comprises three main phases. First, the risk identification phase involves the identification of threats and vulnerabilities. Second, the risk estimation phase computes the degree of risk (risk score or level). Third, the risk management phase entails actions to mitigate risks that exceed a certain risk threshold. This study proposes suitable ML frameworks for the first two phases (i.e., risk identification and estimation). For risk management, it is typically more relevant to the specific company to devise particular risk strategies after the cyber risks have been identified and assessed.

2.2 Existing Works in Construction

Many studies, tools, and standards on this topic have been proposed, but mostly for the information technology industry, such as those by the National Institute of Standards and Technology (NIST) [16] and the International Organization for Standardization's Information Security Management Systems (series of 27000s codes) [17]. Nonetheless, cybersecurity in the construction industry poses unique challenges due to the dynamic nature of construction projects. Factors such as changes in teams, varying levels of cybersecurity knowledge among employees, frequent communication within the supply chain, exchange of digital information among stakeholders, and the overlapping of different projects make it challenging to address cybersecurity concerns using standard frameworks and methodologies. General cybersecurity frameworks designed for more static industries may not adequately address the specific needs of the construction industry. Instead, tailored and industry-specific approaches are necessary.

However, only a few studies have been developed specifically for the construction industry within the last few years. The purpose of these studies can be classified into

general discussions, review papers and specific solutions [18]. General discussions on cybersecurity topics in construction include works by Bello and Maurushat [19], EI-Sayegh et al. [20], García de Soto et al. [21], Mantha and García de Soto [4], Yao and García de Soto [22], etc. Review papers include works by Pärn and Edwards [23], Sonkor and García de Soto [24], Goh et al. [25] etc. Specific methods or solutions cover blockchain technology [23, 26], machine learning or deep learning algorithms [27, 28], threat modeling [29, 30], framework proposal [31], CVSS score [15], etc.

In 2023, Pargoo and Ilbeigi [18] conducted a scoping review, highlighting the limited research in this field. They found only 19 studies that proposed specific technical solutions, with just four of them related to ML. However, within the subset of studies focused on cyber risk analysis specifically, none utilized ML, and all relied on more static and manual methods that might lead to bias and subjectivity, such as the CVSS score [15] and threat modeling [29, 30]. The limited number of studies on cyber risk analysis in construction, along with the static and manual nature of existing methods, underscores the need to develop a more efficient technique tailored to meet the unique requirements of the construction industry. To bridge this gap, this study proposes ML frameworks for cyber risk identification and estimation tailored to the construction industry.

2.3 ML Investigation for General Cyber Risk Analysis

Although cyber risk analysis tasks in the construction industry differ from general tasks, Goh et al. [25] point out that looking at recent technology applications for general cyber security problems and then adapting them to the construction industry is beneficial. In this study, we identify popular ML algorithms by performing a simplified bibliometric analysis of the relevant literature (Keyword and Topic Analysis). This involves extracting and analyzing keywords, abstracts, titles, and other textual content from a set of scientific and scholarly publications to identify common themes, research topics, and emerging trends within a field. This simplified bibliometric analysis allows for the quick and efficient identification of popular ML algorithms. We utilized the WoS database to retrieve relevant papers using the following search criteria:

- **Searchable field:** “Topic”. When searching, we can decide on the searchable parts of the works. We chose the “Topic” part, which includes the title, abstract, and keywords of a paper.
- **Search keywords:** cybersecurity or “cyber security” or “cyber risk”. In this combination, the quotation marks mean that the system should return the exact match of the phrase, excluding its variants. The logic gate “or” means that papers containing any keywords can be returned.
- **Time:** 2018 to 2023.

To investigate popular ML algorithms, besides using the WoS search criteria above, we added specific ML algorithm keywords like “Support Vector Machine” or “SVM” using the “and” operator. We looked at 12 ML algorithms, and the number of publications returned by WoS is presented in Table 1. It is apparent that Neural Networks, Support Vector Machine, Random Forest, and Decision Tree are the top 4 algorithms commonly used in this field. Neural Networks stand out, as it is over 250% more prevalent than the second most popular algorithm, indicating its potential as an algorithm suitable for

various tasks and worth exploring further. These identified algorithms serve as references for cyber risk analysis tasks in the construction industry.

3 Steps of Proposing ML Frameworks

Following a standard ML system design approach [32], the process of proposing ML frameworks for a cyber risk analysis task (i.e., risk identification and risk estimation in this study) is divided into four steps: (1) Understand the task within the context of the construction industry; (3) Approach the task from an ML perspective; (4) Define the ML inputs; and (5) Define the ML outputs. Each of these steps will be discussed in the following sections.

3.1 Understand Each Task

For each task in the cyber risk analysis cycle, i.e., risk identification and risk estimation in this study, it is vital to deeply understand the following aspects of the task.

- **Interconnections with the construction industry:** Each task in the cyber risk analysis cycle has unique aspects within the construction context. This includes understanding the industry's specific characteristics, processes, and challenges, and how they impact the task. For example, risk identification may focus on technologies and tools like Building Information Modeling (BIM), project management software, and IoT devices.
- **Information needed from a construction project for analysis:** This may include data on physical infrastructure, network architecture, and communication patterns among stakeholders. For instance, risk estimation may require information about supply chain network topology, personnel socioeconomic levels, and connected devices.
- **Expected output and results of each task:** Each task in the cyber risk analysis cycle aims to achieve a specific outcome. For example, risk management seeks to reduce the impact of identified risks on a project. Understanding these outcomes and their relation to the cyber risk analysis cycle helps in selecting appropriate ML techniques and designing a tailored cybersecurity framework for the construction industry.

By gaining a comprehensive understanding of these elements, we can develop tailored and effective strategies that specifically address the unique challenges faced by the construction industry in terms of cybersecurity.

3.2 Approach Each Task from an ML Perspective

For each task, we approach it from an ML perspective, demonstrating how tasks can be reformulated into questions addressable by ML techniques. When approaching the task, we need to think about the following aspects.

- Comprehend the specific objectives of the cyber risk analysis task and determine the most suitable ML architecture to apply, such as classification for categorizing risk levels or regression for quantifying risk scores.

Table 1. Number of publications on ML applications for general cyber risk analysis

Keyword 1: cybersecurity or “cyber security” or “cyber risk”	Keyword 2: different algorithms	Linear regression	Logistic regression	Decision Tree	Support Vector Machine	Naïve Bayes	K-nearest Neighbors	K-Means	Random Forest	Principle Component Analysis	Gradient Boosting	AdaBoosting	Neural Networks
41	65	160	216	76	85	47	178	55	48	1	773		

- Explain the ML-based risk analysis approach in terms that are relatable and clear to construction industry professionals, ensuring that the approach and results are understandable to the target audience.
- Evaluate the advantages and limitations of potential ML frameworks for cyber risk analysis in the construction industry, considering factors such as interpretability, data requirements, and adaptability to emerging threats.

3.3 Define the Inputs for the ML Model

For each task, it is essential to identify and extract pertinent features of a construction project to serve as input variables for the ML model. In designing these features, we recommend considering the following aspects:

- When selecting features for cyber risk analysis in construction projects, tailor them to fit each task's unique scenario and objectives, ensuring relevance and adaptability.
- Choose features that can be easily obtained from readily available construction project data, facilitating timely and practical risk assessment.
- Avoid multicollinearity by selecting features that are not highly correlated with one another, enhancing the model's interpretability and accuracy.

By taking into account these aspects, we can determine the most appropriate and informative features that will contribute to the development of an effective ML model. This ensures that the ML models are better equipped to handle the challenges presented by each task and provide accurate and actionable insights for improved cybersecurity in the construction industry.

3.4 Define the Outputs for the ML Model

For each task, we have comprehensively explored the various types of outputs that the ML model can generate to effectively convey the desired outcomes. These outputs may include risk levels, risk scores, changes in risk over time, or a comprehensive list of identified risks. By doing so, we aim to offer risk managers a clear understanding of what to expect from the ML model and how these outputs can be utilized in their risk decision-making processes. The model outputs are designed to provide valuable insights and quantifiable measures that allow risk managers to evaluate the effectiveness of their cybersecurity strategies and make informed decisions to mitigate potential threats. This approach ensures that the ML models not only generate meaningful results but also contribute to a more robust and resilient construction industry in terms of cybersecurity.

4 ML Frameworks for Risk Identification and Estimation

Following the steps in Sect. “[Steps of proposing ML frameworks](#)”, this section proposes ML frameworks for risk identification and risk estimation, respectively.

4.1 ML Framework for Risk Identification

Understand cyber risk identification. Cyber risk identification for a construction project is the process of identifying potential cyber risks that could impact the project and cause harm to its assets or operations. It involves analyzing the project's characteristics, such as the project phase, the construction delivery method, the IT infrastructure, and the supply chain networks, to determine the possible vulnerabilities and threats that may exist. One unique aspect of cybersecurity tasks in the construction industry is that a construction project typically progresses through different phases from start to finish. A construction project consists of well-defined phases, such as planning, design, construction, maintenance, operation, and demolition. Each phase poses different cyber risks for unique reasons. For instance, in the design phase, the cyber risks may include the data breach of Building Information Modeling (BIM) and blueprint leakage; in the construction phase, the risk might involve a malicious insider sabotaging the on-site sensors [33]. It is important to identify potential cyber risks for each phase of a specific project to plan ahead and take proactive measures.

ML approach to risk identification. Identifying potential cyber risks is a crucial aspect of risk management in the construction industry. However, due to the complex nature of construction projects, manually performing this task can be challenging. Additionally, consulting with experts every time can be time-consuming and expensive. Therefore, ML models can play a significant role in assisting construction companies in identifying potential cyber risks for each phase of a construction project. By utilizing ML models, we can achieve a more efficient, more accurate and more automatic identification of cyber risks. The ML model can identify potential cyber risks by analyzing the characteristics or features of the current phase of the construction project. These features are extracted and sent to the ML model, which then generates outputs that indicate the categories of potential risks. One crucial aspect is determining the features to be extracted as inputs for the ML model and the corresponding outputs it should provide, given a specific phase of a construction project (hereinafter referred to as risk scenario).

Define the features as inputs. The risks of a scenario can be identified using various features that change throughout the project's lifecycle, and these features should be extracted from the risk scenario and sent as inputs to the ML model for analysis. For example, during the planning phase that has higher uncertainties, high-level features such as administrative structure, site layout, and budget can be employed. As the project progresses and uncertainties decrease, more specific features can be used, like IT personnel skill level, internet traffic volume, and the number of actual sensors and monitors for the operational phase. Additionally, images from monitoring cameras and text from working email communications can be incorporated as features in the ML model. To enhance the model's performance, a combination of inputs of various modalities can be utilized to increase input diversity, which necessitates the use of multiple model architectures. This study presents a list of features corresponding to different phases of a construction project, as depicted in Fig. 1. The inputs initially consist of general features but become more specific as the project advances, allowing for improved risk identification.

Define the outputs indicating risk. Risk identification can be classified as a classification problem or a generation problem. Based on this, we propose to use four kinds of ML frameworks, each of which corresponds to a different kind of output, shown in

Fig. 2. Risks can be identified based on the extracted features of a specific phase of a construction project.

Framework 1: multi-class classification through one model. We can train one model that performs multi-class classification, where the output is a K -dimensional vector $[y_1 \ y_2 \dots \ y_{K-1} \ y_K]^T$ that contains K number of risks pre-defined in a database, and the element y_k represents the probability of having the k -th risk in the database (Fig. 2a). A representative architecture is a Neural Network [34];

Framework 2: multi-class classification through multiple binary classifiers. We can also train multiple binary classifiers to perform the multi-class classification, where each classifier addresses one kind of risk. In this case, the output is 0 or 1 (0 represents not having this risk, and 1 represents having the risk) for each kind of risk. However, the classification needs to be performed K times to cover all risks in the database (shown in Fig. 2b) [35], which is not time-efficient;

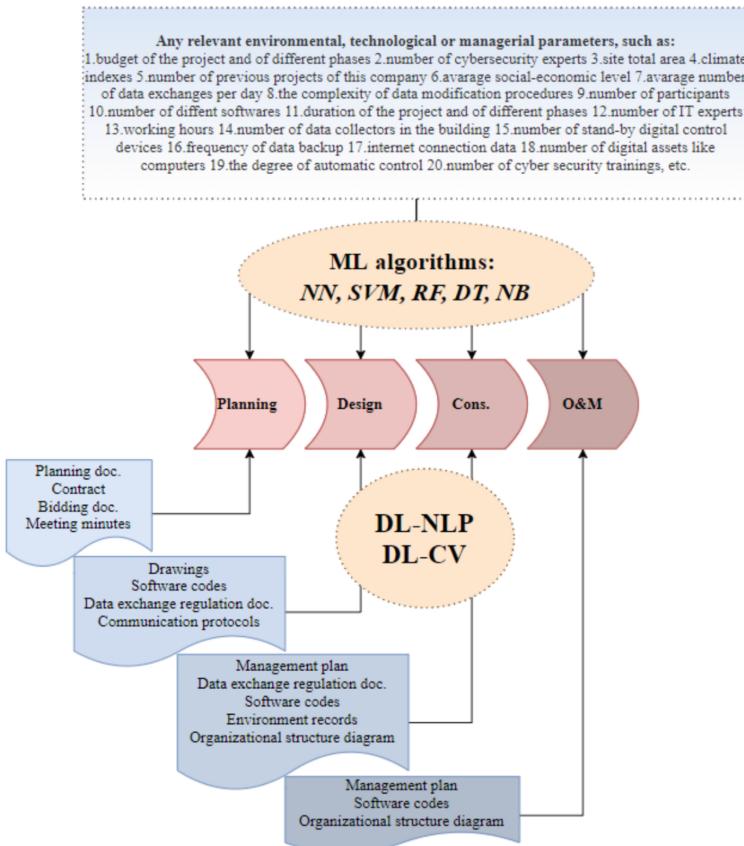


Fig. 1. Summary of ML model inputs in different phases of a construction project

Frameworks 1 and 2 require the manual design of features to be extracted from a given phase of a construction project. The selection of these features must be carefully considered to ensure optimal performance. Features can be high-level, such as the “number of personnel,” “IT expertise,” and “frequency of cybersecurity training” at the beginning of a project, or more specific, such as the “number of data exchange instances between the designer and owner” or “frequency of BIM software logins” during the design phase. The selection of features can be repeatedly experimented with to determine the optimal combination.

Framework 3: copying the risks from the most similar sample (risk scenario) in the database after computing sample similarity [36], which can be achieved using algorithms like K-Nearest Neighbors (KNN) (Fig. 2c). While this framework is easy to implement, it may not be completely suitable for a new risk scenario because the identified risks are simply copied from a previous, similar scenario.

Framework 4: automatically generating the risks to a list, which involves DL-based generative models used in Natural Language Processing (Fig. 2d). We feed in the model with a paragraph describing the project risk scenario, and the model analyzes the text and then generates a list of risks. For example, the trained generative model can automatically generate a list of numbers representing the risks, e.g., NO. 1_3_5_11_14_104. Representative generative models include LSTM [37] and GPT-2 [38].

Framework 4 is deemed the most promising due to several advantages. First, it eliminates the need for manual feature extraction from a project, which can be both time-consuming and challenging. Second, it only requires text describing the risk scenario as input, and the model can automatically extract needed features. Third, there are numerous text sources within a construction project, such as management regulation documents and cybersecurity commitment sections in contracts, that can serve as inputs. Furthermore, the model can even process code from construction-related software, like CoConstruct, Revit, and Autodesk BIM 360, by treating them as text, which can help identify software vulnerabilities and flaws. In a word, any pertinent materials, including source code, security plans, design specifications, management plans, and communication regulations, can be utilized as inputs for risk identification, and the risk identification results will be more precise.

4.2 ML Frameworks for Risk Estimation

Understand cyber risk estimation. Once the risks (threats and vulnerabilities) of a specific phase in a construction project (risk scenario) are identified, the holistic risk degree of the risk scenario can be determined and compared to a pre-established benchmark. If the calculated risk degree is higher than the benchmark, then risk measures have to be taken. Each pair of identified threats and their corresponding vulnerability constitutes a risk event with a certain risk degree. The interactions between various risk events are propagated to calculate the holistic risk degree of the risk scenario. The risk degree can be represented as a score between 1–100 or a level on a scale from 1–5, where a higher score represents a riskier degree. Accurate risk estimation requires an understanding of the risk propagation process, i.e., how the risk events interact with each other to impact the holistic risk degree of the risk scenario. To achieve this, an effective risk propagation

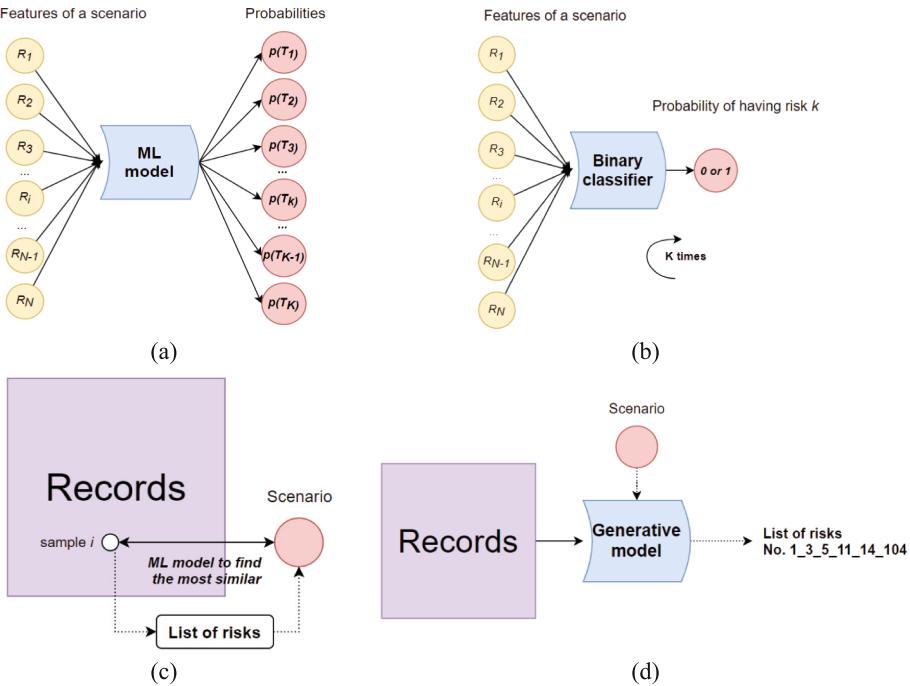


Fig. 2. Risk identification as **a** multi-class classification through one model (Framework 1), **b** multi-class classification through binary classifiers (Framework 2), **c** copying the list of risks from the most similar sample (Framework 3), and **d** risk identification through ML generative models (Framework 4)

function needs to be designed, taking into account all features extracted from the current project phase.

ML approach to cyber risk estimation. As previously mentioned, designing an effective risk propagation function is essential, and it requires a comprehensive understanding of the construction project, the entities involved, their communication patterns, and their interactions with each other. In addition to knowledge of the construction project itself, a solid understanding of cybersecurity is also necessary, which is typically challenging for construction practitioners. As a result, many risks are assessed only qualitatively, excessively relying on the experience and subjective judgments of experts. ML can aid in this process by learning the mapping functions for risk propagation from available historical data and objectively predicting the overall risk level of the project based on identified patterns. In ML terminology, predicting the risk level is a classification problem, while predicting the risk score is a regression problem [39].

Define the features as inputs. To effectively learn the risk propagation function, the appropriate inputs for the ML model must be determined. Based on the different forms of inputs, this study proposes four ML frameworks to tackle this problem. It should be reiterated that each threat-vulnerability pair leads to one type of risk event, and each risk event should have a risk degree as well. All of these risk events combined are propagated

to calculate the final holistic risk degree of the given phase of a construction project. Since there are different pathways to propagate the threat and vulnerabilities to form the holistic risk, different inputs are required for the ML model to accurately predict the risk degree. This study proposes 4 ML frameworks corresponding to 4 pathways.

Pathway 1: The 1st pathway is shown in Eqs. (1) and (2): firstly, compute the risk degree of each risk event r_i from each threat-vulnerability pair, and then use function f_1 to propagate and integrate all the risk degrees of r_1 to r_N to calculate the holistic risk degree of the current risk scenario R_T .

$$r_i = g(T_j, V_k) \quad (1)$$

$$R_T = f_1(r_1, r_2, r_3 \dots r_{N-1}, r_N) \quad (2)$$

Pathway 2: The 2nd pathway is shown in Eqs. (3) through (5): firstly, use function m to propagate and combine all the threat degrees (i.e., likelihoods of threats) of T_1 to T_J to form the total threat degree, and then use function h to propagate and combine the all the vulnerabilities degrees of V_1 to V_K to form the total vulnerability degree V , then use function f_2 to compute the holistic risk degree of the scenario from T and V .

$$T = m(T_1, T_2, T_3 \dots T_{J-1}, T_J) \quad (3)$$

$$V = h(V_1, V_2, V_3 \dots V_{K-1}, V_K) \quad (4)$$

$$R_T = f_2(T, V) \quad (5)$$

Pathway 3: The 3rd pathway is shown in Eq. (6); the holistic risk degree is directly computed using a more complex function f_3 . The inputs of this function are all the threat degrees and vulnerability degrees.

$$R_T = f_3(T_1, T_2, T_3 \dots T_{J-1}, T_J, V_1, V_2, V_3 \dots V_{K-1}, V_K) \quad (6)$$

where T_j is the degree of threat j ; V_k is the degree of vulnerability k ; r_i is the risk degree of risk event i ; R_T is the holistic risk degree of the current risk scenario; T is the holistic degree of all threats; V is the holistic degree of all vulnerabilities; N is the number of identified threat-vulnerability pairs; J is the number of identified threats; K is the number of identified vulnerabilities.

Pathways 1, 2, and 3 (Eqs. 1–6) require the computation of threat value (likelihood) and vulnerability value (severity) beforehand. However, in the field of risk analysis, determining the likelihood and severity is cumbersome, and the results are not always reliable because they involve the determination of uncertainty [40]. Therefore, if determining the likelihood and severity is difficult for a given scenario, the authors suggest adopting Pathway 4 (Eq. 7), which directly absorbs the most basic features and maps them to the output.

Pathway 4: All of the three pathways above compute the holistic risk degree from the identified threats and vulnerabilities, which are identified based on features of the given scenario. Therefore, it is reasonable to propose an additional pathway that computes the

holistic risk degree directly from the basic features identified, skipping the intermediate processes shown in Eq. (7). This is an effective way of conducting risk estimations for personnel unfamiliar with cybersecurity knowledge since little IT expertise is involved.

$$R_T = f_4(d_1, d_2, d_3 \dots d_d \dots d_{D-1}, d_D) \quad (7)$$

where d_d is the value of the feature d

In the equations above, the mapping functions $g, m, h, f_1, f_2, f_3, f_4$ can be learned by ML algorithms. Specifically, computing a risk level can be deemed as an ML classification problem, and computing a risk score as an ML regression problem. The general learning process is shown in Fig. 3. The inputs are the threat and vulnerability degrees (or basic features in Pathway 4), the computation model can be one ML model or a combination of several ML models, and the output is the holistic risk degree of the scenario.

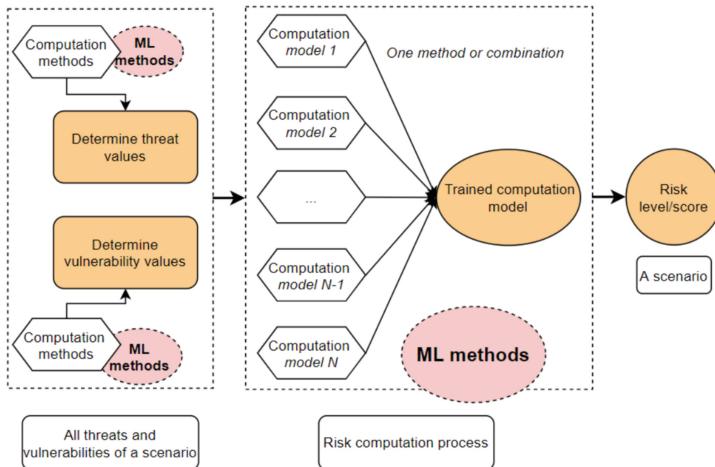


Fig. 3. The process of risk estimation using general ML methods

Define the outputs indicating risk degree. The output for this task is straightforward, as it involves only a risk level or score. The risk level should be represented on a scale from 1 to 5, while the risk score should range from 0 to 100. As previously mentioned, if the ML model predicts a risk level, it is classified as a classification problem. Conversely, if it predicts a risk score, it is considered a regression problem in ML terminology. The output then reflects the severity of the cyber risk for the current phase of the construction project. Regarding the selection of a specific ML algorithm, experimentation with the top four algorithms identified in Sect. “Related works” can be conducted to choose the one that leads to the highest accuracy. The authors suggest prioritizing NN-based architectures because they can learn more complex functions to model the intricate relationships among features compared to other ML algorithms.

5 Conclusions, Discussions and Future Works

This study proposes ML frameworks for cyber risk identification and risk estimation that are tailored to the construction industry. Through a simplified bibliometric analysis, it was found that the top four commonly used ML algorithms for general cyber risk analysis are Neural Networks, Support Vector Machine, Random Forest, and Decision Tree, with Neural Networks being the most prevalent. Following the standard ML system design approach, for risk identification, we proposed 4 distinct frameworks based on the various outputs a model can generate, highlighting the generative model as the optimal choice for this task. For risk estimation, we proposed 4 frameworks depending on the different input types a model can receive, indicating that the fourth framework is preferable due to its ease of implementation and comprehension.

Once implemented, the proposed ML frameworks can provide an important tool to support data-driven decision-making. These tools can enable an automatic cyber risk analysis process, significantly reducing the time and resources required for cyber risk identification and assessment. ML tools can provide a highly secure digital infrastructure for construction projects due to their accuracy and adaptability in identifying and assessing cyber risks. Enhanced security not only protects sensitive data but also significantly reduces the likelihood of interruptions caused by cyber threats, creating a more reliable and resilient environment for stakeholders. Also, integrating ML tools into construction projects enhances the collection and analysis of data, which is typically done inefficiently in the industry.

This study contributes to the growing body of interdisciplinary research in construction, cybersecurity, and machine learning. The proposed ML frameworks provide an automated, adaptable, and user-friendly solution to construction practitioners for cyber risk analysis. Future work will include applying these frameworks to actual construction projects to evaluate their effectiveness and efficiency. The long-term goal of this study is to develop an intelligent system for web and mobile applications that can integrate various ML modules to automatically identify, estimate, and formulate risk mitigation strategies for construction projects.

Acknowledgments. This work was supported by the Center for Cyber Security at NYUAD (CCS-AD), funded by Tamkeen under the NYUAD Research Institute Award G1104 and in collaboration with the NYUAD Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001. This work benefited from the High Performance Computing (HPC) and Research Computing Services, part of the New York University Abu Dhabi (NYUAD) Center for Research Computing.

References

1. Kline, R., Turk, Ž.: Construction 4.0—digital transformation of one of the oldest industries. *Econ. Bus. Rev.* **21**(3), 393–496 (2022)
2. Zhao, J., Feng, H., Chen, Q., García de Soto, B.: Developing a conceptual framework for the application of digital twin technologies to revamp building operation and maintenance processes. *J. Build. Eng.* **49**, 104028 (2021)

3. Vandegrift, R. A.: *Construction 4.0: An Innovation Platform for the Built Environment*. Routledge, New York (2020)
4. Mantha, B. R. K., & García de Soto, B.: Cyber security challenges and vulnerability assessment in the construction industry. In: Creative Construction Conference 2019, pp. 29–37. Budapest University of Technology and Economics (2019)
5. Emma, J.: Cyber Security Breaches Survey 2020. Department for Digital, Culture, Media & Sport 4.1, London (2020)
6. Phishing Attacks in the Construction Industry. Infosec. <https://resources.infosecinstitute.com/topic/phishing-attacks-construction-industry/>, Accessed 15 March 2021
7. Kunert, P.: Shut the Front Door: Jewson 'Fesses Up to Data Breach. The Register. https://www.theregister.com/2017/11/14/jewson_suffers_data_breach/, Accessed 15 March 2021
8. Sawyer, T., Rubenstone, J.: Construction Cybercrime is on the Rise. Expertise, Solution, and People For Success. <https://www.cbc.ca/news/politics/ransomware-bird-construction-military-1.5434308/>, Accessed 23 April 2021
9. Tunney, C.: Ransomware attack on construction company raises questions about federal contracts. CBC. <https://www.cbc.ca/news/politics/ransomware-bird-construction-military-1.5434308/>, Accessed 15 March 2021
10. Korman, R.: Hoffman Construction Reports Hack of Self-Insured Health Plan Data. Engineering News-Record. <https://www.enr.com/articles/51232-hoffman-construction-reports-hack-of-self-insured-health-plan-data/>, Accessed 15 March 2021
11. Yao, D., García de Soto., B.: A corpus database for cybersecurity topic modeling in the construction industry. In: Proceedings of the 40th International Symposium on Automation and Robotics in Construction, IAARC Publications, Chennai, India (2023)
12. Onwubiko, C., Ouazzane, K.: Challenges towards building an effective cyber security operations centre. arXiv preprint arXiv (2022)
13. Zhao, G.: *Information Security Management and Risk Assessment*. Tsinghua University Press, Beijing (2020)
14. García de Soto, B., et al.: Construction cybersecurity and critical infrastructure protection: significance, overlaps, and proposed action plan. vol. 0, no. May, pp. 1–20 (2020)
15. Mantha, B. R. K., García de Soto, B.: Assessment of the cybersecurity vulnerability of construction networks. Engineering, Construction and Architectural Management, vol. 28, no. 10, pp. 3078–3105 (2020)
16. Cybersecurity, C. I.: Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1. NIST (2018)
17. ISO/IEC 27000:2018 Information technology—Security techniques—Information security management systems—Overview and vocabulary. ISO (International Organization for Standardization). https://standards.iso.org/ittf/PubliclyAvailableStandards/c073906_ISO_IEC_27000_2018_E/, Accessed 11 Oct. 2021
18. Salami Pargoo, N., Ilbeigi, M.: A scoping review for cybersecurity in the construction industry. J. Manag. Eng. **39**(2) (2023)
19. Bello, A., Maurushat, A.: Technical and Behavioural Training and Awareness Solutions for Mitigating Ransomware Attacks. Springer International Publishing (2020)
20. El-Sayegh, S., Romdhane, L., Manjikian, S.: A critical review of 3D printing in construction: benefits, challenges, and risks. Archiv. Civil Mech. Eng. **20**(2) (2020)
21. García de Soto, B., et al.: Implications of construction 4.0 to the workforce and organizational structures. Int. J. Constr. Manag. **22**(2), 205–217 (2022)
22. Yao, D., García de Soto, B.: A preliminary SWOT evaluation for the applications of ML to Cyber Risk Analysis in the construction industry. In: IOP Conference Series: Materials Science and Engineering, vol. 1218, no. 1, p. 012017 (2022)

23. Parn, E. A., Edwards, D.: Cyber threats confronting the digital built environment: common data environment vulnerabilities and block chain deterrence. *Eng. Constr. Arch. Manag.* **26**(2) (2019)
24. Sonkor, M. S., García de Soto, B.: Is your construction site secure? A view from the cybersecurity perspective. In: Proceedings of the 38th International Symposium on Automation and Robotics in Construction, pp. 864–871. ISARC, Dubai, United Arab Emirates (2021)
25. Goh, G.D., Sing, S.L., Yeong, W.Y.: A review on machine learning in 3D printing: applications, potential, and challenges. *Artif. Intell. Rev.* **54**(1), 63–94 (2021)
26. Shemov, G., García de Soto, B., Alkhzaimi H.: Blockchain applied to the construction supply chain: a case study with threat model. *Front. Eng. Manag.* **7**(4), 564–577 (2020)
27. Pan, Z., Hariri, S., Pacheco, J.: Context aware intrusion detection for building automation systems. *Comput. Secur.* **85**, 181–201 (2019)
28. Sheikh, A., et al.: Cyber attack and fault identification of HVAC system in building management systems. In: 9th International Conference on Power and Energy Systems (ICPES), pp. 1–6. IEEE, Perth, Australia (2019)
29. Mantha, B., García de Soto, B., Karri, R.: Cyber security threat modeling in the AEC industry: An example for the commissioning of the built environment. *Sustain. Cities Soc.* **66**(December 2020), 102682 (2021)
30. Mohamed Shibly, M. U. R., García de Soto, B.: Threat modeling in construction: an example of a 3D concrete printing system. In: Proceedings of the 37th International Symposium on Automation and Robotics in Construction. IAARC Publications, Kitakyushu, Japan (2020)
31. Turk, Ž., et al.: A systemic framework for addressing cybersecurity in construction. *Autom. Constr.* **133**, 103988 (2021)
32. Rebala, G., Ravi, A., Churiwala, S.: An Introduction to Machine Learning. Springer International Publishing (2019)
33. Kurtz, S.: Cybersecurity vulnerabilities in the construction industry. Total IT Information Technology. <https://totalit.com/cybersecurity-vulnerabilities-in-the-construction-industry/>, Accessed 19 March 2021
34. Abiodun, O.I., et al.: State-of-the-art in artificial neural network applications: a survey. *Heliyon* **4**(11), e00938 (2018)
35. Grandini, M., Bagli, E., Visani, G.: Metrics for multi-class classification: an overview (2020). arXiv preprint [arXiv:2008.05756](https://arxiv.org/abs/2008.05756)
36. Ontañón, S.: An overview of distance and similarity functions for structured data. *Artif. Intell. Rev.* **53**, 5309–5351, 123AD (2020)
37. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
38. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. <https://github.com/codelucas/newspaper/>, Accessed 13 July 2022
39. Leo, M., Sharma, S., Maddulety, K.: Machine learning in banking risk management: a literature review. *Risks* **7**(1), 29 (2019)
40. Aven, T., Genserek, R.: How to define and interpret a probability in a risk and safety setting. *Saf. Sci.* **51**(1), 223–231 (2013)



Automated Hierarchical Object Clustering for Multi-Image Documentation: A Comprehensive Image Analysis Pipeline

Michael Disser^(✉), Maximilian Gehring, and Uwe Rüppel

Institute of Numerical Methods and Informatics in Civil Engineering, Technical University of Darmstadt, 64287 Darmstadt, Germany
disser@iib.tu-darmstadt.de

Abstract. This paper presents an automated comprehensive image analysis pipeline designed for the spatially separated detection and hierarchical clustering of objects within multiple images taken for different kinds of documentation. While the original application derives from German crime scene documentation, the presented concept describes the application for various fields and use cases containing hierarchical structures of objects. In civil engineering and construction documentation, the as-built documentation, and the automated evaluation of images of construction diaries or issue documentation are useful use cases, improving the way data from multiple images in construction documentation is analyzed and interpreted. The presented approach closes the gap in existing solutions relying on additional information like BIM-models or specific photogrammetry-usages to gather (building) object relations. The parts of the pipeline are Single Image Analysis, Image Comparison, and Evaluation of Image Comparison. The Single Image Analysis includes the methods Metadata Extraction, computer vision methods such as object detection and -segmentation, a QR code reader as well as an image-rating of probability of being an overview image. The Image Comparison is executed for all image pairs and allows for the comparison of different types of information between two images and calculates the affinity-score (based on probability) for matching information for every single one of the following single comparison-methods: Metadata Comparison, Feature Matching, Object Segmentation Comparison (also combined with Feature Matching) and QR Code Comparison. The Image Comparison is done by comparing all image pairs with all available methods or more intelligently by using smarter resource-saving approaches described in the paper. The Evaluation of Image Comparison sorts the detected objects in the images hierarchically based on the previous steps. Additionally, the images get sorted room-wise into clusters. As a proof of concept, a demonstrative application with the use case of crime scene documentation is introduced, and the results are discussed.

Keywords: Computer Vision · Clustering Analysis · Automated Image Analysis

1 Introduction

The phrase “a picture is worth a thousand words” applies to documenting a current state in a way that it can be interpreted by anyone in a matter of seconds. Visual materials, such as pictures and videos, are one of the most common and vivid mediums to not only capture important milestones in your personal life, but also the current status or progress of a project in a professional setting. In the case of the construction industry, images are used to document a certain state e.g., in the sense of a construction diary, documenting issues as well as binding evidence in a legal dispute. Images in the construction field are usually created room by room and if necessary, sorted manually room-wise for documentation purposes for easy assignment to information that might be needed in the future.

A similar approach is also used by another profession, the crime scene documentation. The crime scenes are meticulously recorded in order to document all evidence and objects at the time of the crime scene documentation, supported by camera footage. In contrast to construction site documentation, the procedure for crime scene photography must be carried out in a very specific way. The following procedure is prescribed for German crime scene documentation by [1, p. 230]: The crime scene is recorded with cameras in a spatially structured manner in order to visually capture rooms and objects. Initially, overlapping overview shots are taken that depict large areas of the room. This is followed by increasingly detailed images of the objects in the room. Important traces that are relevant for the crime scene work are marked with tracers that are turned towards the camera for each shot. These photos are usually taken with DSLR (digital single-lens reflex) cameras.

A solution that automates the time-consuming manual room and object-wise sorting of the images is to be developed to reduce the amount of work involved in image documentation. In addition to the spatial sorting of the images, it should be possible to make statements about the content and the relationship between the images and the objects depicted. A concept for an automated *Image Analysis Pipeline* was designed and implemented demonstratively for this purpose. Each picture taken is analyzed by itself using metadata information and computer vision methods in the *Single Image Analysis*. Afterwards, in the *Image Comparison*-component image pairs are compared based on the *Single Image Analysis* as well as various other (mainly computer vision) methods resulting in probabilities (shown as an affinity-score) for the images in the image pair to be related. The following step, the *Evaluation of Image Comparison*, consists of the (room-wise) clustering of the images based on the previously made comparisons. Further hierarchical conclusions are then drawn for the depicted objects. While the use case of the demonstrator, crime scene documentation, is quite specific, the concept and the demonstrator can also be used to generate semantic information and relations about constructions sites and other fields inside and outside the construction industry.

2 Methodology

To clarify the methodological approach of this paper, the following section briefly describes the approached steps that were taken, based on Johannesson et al.’s Design Science Research Methodology [2]. At first, the main problem of gathering information

from a collection of images documenting a certain kind of hierarchical information was identified and partial issues were derived. This was supported by the use case of crime scene documentation, guided by crime scene investigators, and afterwards expanded to the broader use case of general image documentation. The requirement analysis in chapter 5 was created based on these issues. Based on the requirement analysis, relevant technical basics and the state of the art were researched. The results of the research can be found in chapters 3 and 4. Possible forms of data input such as the image-data from smartphone and DSLR-camera images, were analyzed. Based on the previous steps, a universally valid concept was developed, that would fulfill the requests from the requirement analysis. While general methods and procedures are described in the concept, it is still independent of specific software-implementation. This implementation is described in chapter 6, where specific feature detection algorithms, object segmentation algorithms with specific pre-trained models and clustering algorithms are applied and described.

3 Technical Basics for the Image Analysis Pipeline

The following section describes the research for the necessary subjects used in the components of the *Image Analysis Pipeline*. The research of each subject consists of a short general description, various methods that are possible to be used in the subject as well as a brief description of the specific method used in the implementation.

The term metadata describes data that stores information on other data and is used in various fields such as file system metadata, internet metadata, document metadata and picture metadata [3, pp. 179–181]. In digital cameras, e.g., in smartphones and DSLR-cameras, metadata describes information about the file, the parameters of the image and camera settings. Although other standards such as the IPTC Photo Metadata Standard [4] are used [5, p. 82], the Exchangeable Image File Format (EXIF) is the most common metadata scheme for digital photography [6, p. 95] and is used by data formats like JPEG and TIFF [5, p. 84]. Camera model, resolution, creation date, aperture and focal length as basic metadata information [5, pp. 83–84; 7] can be used for the *Metadata Extraction* of the pipeline. A complete list of the EXIF-values and specifications for a variety of different image data formats, e.g., JPEG, can be found in [7]. Multiple libraries support the metadata extraction in multiple programming languages (Pillow [8] for Python, exif-js [9] for JavaScript, and metadata-extractor [10] for Java).

Feature detection is a component of computer vision and image processing where important features of an image like points, lines and edges are detected [11, p. 419]. In broad terms, the features can be identified by using different kinds of kernels that calculate the gradient of brightness or color for each pixel by comparing the current pixel value with its neighbors' [12]. Some detectors additionally combine the results of the kernels with so-called descriptors creating a numerical representation of the content around the feature points, usually described in so-called bins, allowing for a certain room for errors in contrast and deformations when trying to find the same features in different images [13, pp. 341–344]. Canny Edge Detector, Harris Corner Detector, SIFT (Scale-Invariant Feature Transform) [13, pp. 336–340], SURF (Speeded-Up Robust Features) and ORB (Oriented FAST and Rotated BRIEF) [11, p. 443] are commonly used examples of feature detection algorithms; each using different kinds of kernels and algorithms. A

detailed description of the functionality of detectable features, descriptors and mentioned feature detectors can be found in [11, Ch. 7] and [13, Ch. 13]. The functionality of the SIFT algorithm used in the pipeline was introduced by Lowe in [14], also described in [15].

Feature matching can be achieved by using the described descriptors of feature detection methods from two images and comparing them to find matches. While simple matchers like the “Brute-Force Matcher” are simply comparing features of one image to all features on the other image, returning the best matching point, other matchers like the FLANN (Fast Library for Approximate Nearest Neighbors) based matcher use optimizations with algorithms optimized for fast nearest neighbor search [16]. RANSAC (Random sample consensus) can be used to find errors in matching points (outliers) and exclude them from the final fit [13, p. 411]. The results from RANSAC can determine the homography matrix based on the matching features. The perspective transformation needed to transform the view of one image to the other can be described this way [17].

The methods of identification of objects in images and the corresponding probability can be separated into three main categories: The image classification labels the complete image with a suggestion of the detected object and its probability. Object detection describes the identification of classes of objects in an image and presenting it with a bounding box, a label, and a probability. The results for detected images are presented in a pixel-accurate manner in the semantic object segmentation [11, pp. 345–346, 387]. Viola–Jones object detection and Histogram of oriented gradients (HOG) are examples of non-learning vision algorithms for recognition that use a feature-based approach for specific detections [18, pp. 34–35]. Modern object recognition mainly utilizes convolutional neural network (CNN) methods [19, p. 326] which can be achieved by training a model based on labeled images with datasets such as the Microsoft-COCO-dataset. It includes 91 different common object types in 2.5 million labeled instances in 328k images [20] and is publicly available on their website [21]. Broadly speaking, object recognition models use CNNs that combine various convolution- and pooling layers to identify and extract features and their combination to classify regions of the image, which is described in detail in [11, Ch. 6], [19, Ch. 9] and [18, Ch. 2]. Two-stage methods in deep object detection like R-CNN, SPPnet, Fast(er) RCNN estimate candidate object proposals and afterwards classify these proposals resulting in a higher detection accuracy but slower speeds. One-stage methods e.g., YOLO, SSD, and Overfeat extract and classify the object proposals at once, resulting in higher speed yet lower accuracy [18, p. 21]. In practical use Object detection and segmentation methods are often implemented by using deep learning frameworks like TensorFlow or PyTorch. While the supported models are listed in the associated websites (TensorFlow [22], PyTorch [23]), pre-trained models designed for research exploration can be found on the Pytorch Hub [24] and the TensorFlow Hub [25]. For reading out QR codes in images, the QR codes must first be detected and then be decoded in order to read out the text or the URL it is representing. This can be done by various libraries e.g., for Python with pyzbar [26] based on zbar [27] and the OpenCV class QRCodeDetector [28].

Cluster analysis or (data) clustering is utilized in various fields of data mining and machine learning to combine or sort all kinds of related information, find similarities and group them in so-called clusters [29, p. 501]. In [30, Ch. 3–4], [31, Ch. 7] and [29,

Ch. 3, 14], various data clustering methods and their functionality are described in detail [31, Ch. 7]. Differentiates clustering algorithms between hierarchical and partitional clustering. Partitional clusters have non-overlapping subsets and hierarchical clustering creates subclusters where the nested clusters can be organized as a tree. Additionally, the types of clusters in “well-separated”, “prototype-based”, “graph-based”, “density-based”, and “shared-property” are separated. K-Means is a prototype-based, partitional clustering technique that allows a separation by a user-given cluster-number [31, p. 533]. In the graph-based clustering, nodes and edges are used to describe objects and their relations and the spectral clustering method can be used for clustering [29, p. 544]. Commonly used libraries are the Python based library scikit-learn [32] as well as the stats-package [33] for hierarchical clustering and igraph-package [34] for graph-based clustering in R.

4 Related Work

In the following section, different papers and publications are described to show research with similar objectives, i.e., the identification of (hierarchical) dependencies of objects based on images. In [35] an in-depth review of used image-based analysis techniques, main research areas, as well as application areas in construction, are analyzed. The study examined over 9000 journal papers, where 100 were selected to be suitable for the study. A mind map of the different types of image analysis techniques is shown, where 87 methods are declared as detection/classification, while quantification, object tracking, and edge detection are only represented in 25 papers. For each of the various applications, the technology used and the recognized objects, security aspects or structures are listed in table form. The main use cases for image-based analysis in construction are identified as construction safety, progress monitoring, and damage assessment. The “Imaged-based verification of as-built documentation of operational buildings” [36] publication includes a workflow where photogrammetric image processing methods are used to automate as-built verification for indoor and outdoor information of buildings. The image-based survey includes image processing (a combination of feature point selection and stitching, structure and camera pose calculations, as well as camera calibration) and dimension extraction. The results of the automated measurement and a separate manual measurement are then used to verify the dimensions of an as-built BIM model. In [37], as-built video data and 4D-BIM-models are used to recognize the actual state of the construction progress. For each frame of the video, the pose is calculated by pose estimation based on the starting point and the objects get detected and combined with the resulting objects in the BIM-model. Relevant tasks are identified and their states are displayed [38]. Introduces a hierarchical and contextual model to gather information from aerial images based on the groupings of displayed objects like cars, roofs, and roads. By detecting the displayed objects of the image and sorting them into groups that belong to a scene, a three-level contextual hierarchy is created. The usage of a minimax entropy framework allows statistical constraints and relationships between displayed objects to get automatically added. The context of the scene is then interpreted by evaluating the constraints by the “Clustering via Cooperative and Competitive Constraints” algorithm based on the Swendsen-Wang cluster sampling, described in [39],

while different bottom-up detectors are used for different types of objects. This results in a hierarchical sorting and clustering of the objects in the image scenes as well as a validation and completion of objects that might have been missed in the detection. In [40], images from a collection are sorted hierarchically, based on a representative image describing some of the clusters. Images not matching an existing cluster automatically create a new cluster, done by an agglomerative hierarchical clustering algorithm. This is made possible by combining Feature Extraction, building a similarity matrix, cluster representation, and the identification and creation of missing clusters.

5 Concept of the Image Analysis Pipeline

In the introduction chapter, the research topic and the procedure on how the concept of the *Image Analysis Pipeline* tries to solve the problem is briefly explained. In the current introduction, the requirements of the pipeline are defined. The pipeline should be able to gather information plainly based on the images of the documentation, that are created for the specific use case of the user, e.g., documentation of construction progress monitoring or crime scene documentation. While the gathered information can be linked to information based on semantic and/or geometric information such as BIM-models, as shown in [37], the application should only function with the image documentation. This ensures that the pipeline can be used independently from heterogeneous structured or unavailable data, which is a requirement given by the crime scene documentation use case. The results of the pipeline should include a room-wise documentation and hierarchical relationship of the objects displayed in the images. Additionally, the images themselves should be sorted by context, e.g., clustered room-wise. The methods that are used in the different elements of the pipeline should be configurable dependent on the specific use case of the implementation of the concept, e.g., the QR marker detection of the *Single Image Analysis* is only convenient if QR codes are part of the images. The QR markers can be placed to identify specific objects. In the context of the crime scene documentation, they can be placed on tracers next to evidence. Additionally, certain comparison methods are more practical to a certain use case, therefore should be weighted more than other methods. This results in the requirement of a modular pipeline, for which the used methods of each step of the pipeline can be selected. The parameters of the respective method should be adjusted to the use case and the weights of the different methods should be adjustable. Based on the use case, the different methods should also be outsourced as microservices in this modular pipeline with defined connection points, allowing for easy scalability. The results of the analysis should be saved or stored in a database if they are used in a microservice architecture and not given directly to the next component of the pipeline. If saved as a file, JSON or XML-schema, is a common standard. This allows the results to be restored if further images are included to the scene. In order to fulfill these requirements, the *Image Analysis Pipeline*, consisting of three parts, is introduced in the following sections and an overview can be seen in Fig. 1.

5.1 Single Image Analysis

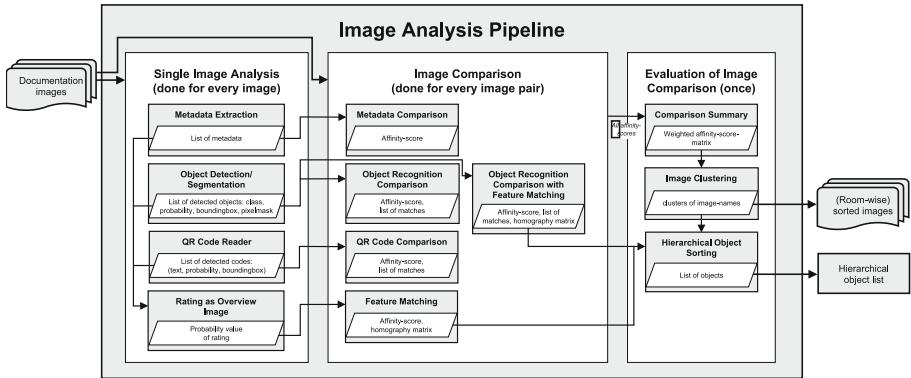


Fig. 1. Components of the Image Analysis Pipeline.

The *Single Image Analysis* enables statements to be made about individual recorded images. First, the *Metadata Extraction* reads out the metadata by using the described methods from chapter 3. The most significant metadata that can be extracted on all common images taken by modern cameras are filename, date time, camera model name, orientation, and resolution (X and Y). Additionally, camera parameters like aperture, exposure, focal length, focus distance, and geo-location are useful for differentiating very different scenes, such as indoor and outdoor photos.

The *Object Detection/Segmentation*-algorithms and models described in chapter 3 generate labels for the various objects in the image. Depending on the use case or the implementation, pre-trained models might be suitable. For the use case construction site documentation, pre-trained datasets like the MOCS (Moving Objects in Construction Sites) [41] could be used to identify construction objects. Furthermore, the ADE20K dataset of MIT [42] can be integrated, where in addition to 18850 images of building parts, separated into 37 different object types, various other beneficial objects are labeled [43]. For road construction or urban planning, the Cityscapes Dataset [44] could be utilized. Taken a step further, the “Joint 2D-3D-Semantic Data for Indoor Scene Understanding” [45]-dataset can be used to detect 2D indoor building information as well as 3D information from point clouds. If a construction company wants to automatically generate a labeled dataset for training their own model, the approach of [46] could be used, where images are automatically labeled based on the representing objects of the BIM-model. Existing object detection algorithms for a specific use case in construction can be researched from [35] and applied.

For the detection of QR codes, the methods described in chapter 3 can be used to read out the represented text or the linked URL. QR codes in the real world can be used as localization as well as semantic labeling of objects, even if they are not detected by the object recognition methods. Since the QR codes in the image might be comparably small and not detectable by most of the detectors, an image manipulation is expedient; further described in chapter 6.

The probability of the image being an overview image is assessed based on the previously described methods. Here, an affinity-score is calculated based on the number of recognized elements, QR codes, and timestamp-information (like the timestamps to filter for the first images of a series). The categorization of the overview images enables statements about the hierarchical structure of the image scheme for the subsequent steps. In the case of a large number of images, this can be used to avoid having to compare all images with each other by giving preference to certain images.

5.2 Image Comparison

In the *Image Comparison*, two images are compared with each other by using different methods, and affinity-scores (in percentage) are calculated to assess whether the two images are related. This is done for every compared image pair. Hereby, either all images can be compared with all images in a brute force manner, or new images can first be compared with the determined overview images. Another approach is to use some methods of the *Image Comparison* components first and use the more computing-intensive comparisons like the Feature Matching only if the other methods are proposing a relationship between the images.

The read-out metadata from the *Metadata Extraction* is compared, and conclusions can be drawn about the affiliation of the images in the *Metadata Comparison*. For example, if the two images were taken with the same camera in a series, it is very likely that they were at least taken in the same room. If the camera also saves the recording geo-information, the distance between the images can be determined. Based on the similarity of the metadata of the two images, an affinity score is calculated.

Feature Matching makes the contents of the two images comparable even if the objects on the image are not recognized. The researched feature detectors and matchers from chapter 3 can be applied. If a certain number of matching features are found and one image is a part of the other, based on the homography matrix, the image section representing the other image can be calculated and displayed. Based on the number of matched features and on the fact that a homography matrix could be calculated, an affinity-score is assigned.

The *Object Recognition Comparison* can be used to compare the detected object tags with each other. The detection probability specified in the object recognition method is also considered. In a similar procedure, the recognized QR code texts can be compared with each other and their matches transfer to an affinity-score. A large number of identical objects indicates that the images belong together, represented in a determined affinity-score. In a combined approach, the bounding boxes or the segmentation of the matching objects can be cut out and compared using feature matching to determine whether they are the same instance and not only the same class.

In addition, the consideration of further image parameters such as the color distribution via the RGB values in the histogram as well as a photogrammetric reconstruction like in [36] are possible, but were not further researched.

5.3 Evaluation of the Image Comparison

The affinity-scores of all image comparisons and included object relations are evaluated in the third component, the *Evaluation of Image Comparisons*. The first task of the component is the *Comparison Summary*. Here, each comparison method from the *Image Comparison* gets a weighting for its impact on the relation-assumption. The sum of the weightings should result in 1, so the totally calculated affinity-score can go from 0% to 100%. For each image pair, the resulting affinity-score (AS_i) of the methods are then multiplied by their weightings (w_i), resulting in a total affinity-score (AS_{total}) for the image pair as shown in the following equation:

$$AS_{total} = \sum_{i=1}^{\text{no.ofmethods}} AS_i * w_i \quad (1)$$

If all images are to be compared with each other in order to cluster the images room-wise, all total affinity-scores of the image pairs must be determined and evaluated. To do this, a symmetrical $n \times n$ matrix is formed, where n is the number of total images. So, for each image pair the total affinity-score is inserted in a matrix e.g., the affinity-score of the image pair of image 1 and 5 is inserted in the matrix in position (1/5) as well as (5/1).

The clustering algorithms introduced in chapter 3 can be used to bundle the images based on the matrix. When using a hierarchical clustering, the first level is the number of rooms, and further levels are parts of the certain room. When using k-means, the number of rooms could be given by the user or estimated with the silhouette coefficient. This results in different clusters of the images that represent the rooms.

Based on that, the images of each room and their depicted instances can be hierarchically arranged based on the *Object Recognition Comparison* combined with the *Feature Matching*. Based on the relative positioning of the bounding boxes/segmentations and based on the recognized object class certain relationships can be recognized. For example, if a smaller bounding box is placed in or on top of another bounding box, a child: parent relationship is assumed.

6 The Weegee Demonstrator

The technical description and the experiences implementing the demonstrator in the use case crime scene documentation based on the research and the concept are described in this chapter. In the construction context, it could also be used as an as-built documentation of objects and building structures based on image documentation. The demonstrator is named after “Weegee”, the pseudonym of a well-known New York crime scene photographer of the 20th century. The complete demonstrator is written in Python code since the required methods are already available in various Python libraries. OpenCV functionality is used for the computer vision methods. All results of the methods are stored in a JSON schema.

6.1 Single Image Analysis

In the *Single Image Analysis*, the metadata is read out with the help of the Pillow library [8]. Filename, date time, camera model name, orientation, and resolution (X and Y) are read out for the comparison. The implementation of the object segmentation utilizes the Mask R-CNN algorithm. A pre-trained model of the COCO dataset [21] is used. To recognize further object types in the crime scene sufficient image material and its annotations are needed. This can technically occur by training a model with images based on evidence, which is usually already categorized by the required paperwork of the crime scene documentation. The object tags, the probability, the segmentation, and the bounding box are saved for each recognized object of an image.

QR codes can be included in the tracers on the crime scene for easier detection of evidence. Various methods were tested for the recognition and readout of the QR codes. The existing solutions were unable to deliver sufficiently satisfactory results as the QR codes on the marker signs in the images are partially too small or difficult to recognize. The QR code has to occupy a defined percentage of the image in order to be recognized by some of the QR code decoders. The *QR Code Reader* was implemented with the pyzbar library after researching and testing various solutions. To improve the insufficient recognition of the QR codes on the images due to poor exposure and lack of contrast, various filter configurations including grayscaling, improving contrast and illumination were tested.

In the *Rating as Overview Image* the time stamps of the images are compared after evaluating all the images. A higher rating is assigned to the first images in a temporal image series and is combined with the rating of the amounts of objects in the image and the QR codes on the marker tags. This rating can be extended by evaluating the focus distance and the other metadata mentioned above.

6.2 Image Comparison

In the *Image Comparison* the brute force method was used, so that all images are compared with all images. Although more time-consuming, this method delivers more significant results compared to the method favoring some image pairs over others due to the previously defined assumptions. The metadata information, e.g., similar timestamp and same camera, is compared in the *Metadata Comparison*, and certain thresholds are determined to result in a certain affinity-score for the relation of the images.

In the *Feature Matching* component, the feature detection occurs via SIFT detector. A FLANN-based matcher compares the feature points and their description and searches for matches; displayed in Fig. 2. A high number of matches results in a higher probability of image dependency and therefore in a higher affinity-score based on reached previously set thresholds. An included termination condition finishes the feature matching if a certain number of matches are detected. The homography matrix is calculated based on the comparison of the positions of the respective points using the RANSAC algorithm. The matrix describes the geometric dependency of the two images and can be used to derive the image location of one image in the second image, displayed in Fig. 2. If detected, a higher affinity-score of the image affiliation is assigned. A high number of undesired matches are detected due to the matching algorithm easily detecting the QR

marker based on its sharp distinctive lines and corners, regardless of the semantic content of the QR code. To avoid the problem, the image gets preprocessed and the previously detected QR codes are covered up by white boxes. Different thresholds for the used algorithms were tested to gain the most significant results.

In the *QR Code Comparison* the read-out information of the QR code is compared and an affinity-score is given set on the number of matches. The equivalent procedure is followed in the plain *Object Recognition Comparison*, where the confidence of the object detection results is integrated. The results get saved in the data scheme after the execution of the methods. The next method combines *Object Recognition Comparison* and *Feature Matching*. The bounding boxes of each matched object-category in the *Object Recognition Comparison* are extracted from both images. Subsequently these extracted bounding boxes are compared by the *Feature Matching* Algorithm. To assure the recognition of the specific individual instance, the bounding boxes are utilized to include the context of the surroundings of the instance, hindering the misrecognition as another similar instance.

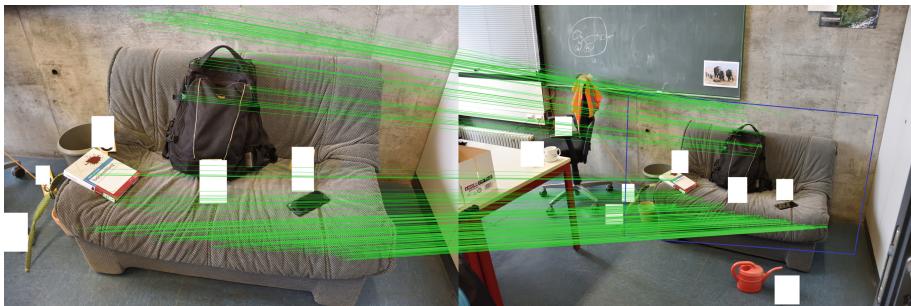


Fig. 2. Visualization of the Feature Matching. Green dots and lines display the matched features. The blue quadrangle displays the localization of the image section. QR codes are painted over to prevent false feature matches.

6.3 Evaluation of the Image Comparison

The *Evaluation of Image Comparison* utilizes the Python NumPy-package for the matrix calculations. The k-means algorithm, supported by the silhouette coefficient, analyses the matrix of all probabilities created in *Comparison Summary* and clusters the images in the *Image Clustering*. The bounding boxes of instances for each room-image cluster detected in the combined approach of *Object Recognition Comparison* and *Feature Matching* algorithm are compared. The relations of the instances described in the concept are carried out based on their position. As a result, a hierarchical object list is put out. In a case study 50 images were taken in four rooms for the proof of concept. In the images different kinds of rooms with COCO dataset detectable objects were taken, partially with and without tracers with QR codes. The measurable results like consumed time, accuracy of the methods and used weightings are strongly dependant on the used methods and assumptions, making them statistically not significant. Therefore, non-measurable results as well as problems are presented in chapter 7.

7 Discussion and Outlook

The introduced pipeline functions plainly on the input of images, making it generally applicable in various fields, where image documentation is used, e.g., through introduction of multiple strategies and procedures for different use cases in civil engineering and construction for the pipeline. While there are some publications that have the aim to reconstruct the objects of the construction site, they rely on additional sources of input information like BIM-models and pose estimation [37] or rely on point clouds or photogrammetry algorithms [36]. Approaches like [38] are only detecting a small amount of specific objects, do not consider comparisons over multiple images and do not allow flexibility of hierarchical relations. Other solutions like [40] have a similar approach and allow hierarchical relations and clusters, but don't have a matching use case or are not taking multiple objects into consideration.

The structure of the pipeline is conceptualized modularly, which allows adaptability and expandability for the implementation for the users specific use case. Various alternatives for the methods of the pipeline as well as their integration are being introduced by the research and the concept. By applying the images of the case study to the demonstrator, the functionality and the advantages and problems of the concept, were shown: The *Metadata Comparison* and the *Rating as Overview Image* is based on assumptions. The significance of the results is dependent of the execution documentation procedure. The weightings of the methods were introduced in the concept to prevent single methods such as the *Metadata Comparison* to falsify the results. The weightings are hardcoded or manually set by the user subjectively based on assumptions in the current state of the development of the pipeline. The quality of the results may vary due to large references and experience not being available yet. The recognition and readability of the QR codes is still not as reliable as needed. This leads to limitations especially in the analysis of the overview images where the QR markers only fill out a small amount of the image. The *Object Detection/Segmentation* can only detect classes the model is trained for. Hence, the non-identified objects will not be sorted into the *Hierarchical Object List*. The user either has to use or train a target-oriented model that detects the desired objects or be aware that unidentified objects not being depicted and included. Non-identified objects will not influence the image clustering since it still can be achieved through other comparison methods such as *Feature Matching*. The thresholds for feature detection and matching, rating the number of matched features and the termination condition for the affinity-score, are difficult to determine. The combined approach of the *Object Recognition Comparison* and the *Feature Matching*, is supposed to identify differing instances instead of only objects. During the testing limitations were discovered. Multiple instances of the same object, e.g., same models of chairs in an office, were not recognized as different instances depending on the surrounding or positioning of the objects. Further testing of the functionality of the cluster analysis for the room-wise separation needs to be performed with a larger set of data to gain statements on the accuracy of the method. The silhouette coefficient was not able to clearly identify the amount of rooms needed for the k-means algorithm depicted in the images of the case study. The relations between two objects could only be specified in the *Hierarchical Object List* if both were recognized. If not recognized in the same image, the objects were only organized in the room without hierarchical dependencies.

The further development of the presented pipeline, a multitude of representative case studies for the implemented use cases are needed so that valid weightings and set thresholds for the affinity-values can be determined more objectively. The approach can be used to merge the gathered information with data, which is created during construction planning, e.g., planning data sources such as BIM-models or contract documents where the different planned building components are described. This allows the dimensions of the current as-built state to be automatically compared to the as-planned state by including photogrammetric algorithms. An example of this work is described in [37]. If a large enough amount of manual documentation is available, e.g., hierarchical lists of objects due to the extraction of various BIM-models that are also documented by images, an automated approach of the estimation of weightings and thresholds can be performed by creating a neural network. Hereby, the *Single Image Analysis* data could be used as input, the *Image Comparison* methods, and the clustering as hidden layer nodes and the manually created *Hierarchical Object Lists* from previous documentations as training data for the output. The training could result in an automated determination of weightings which prevents errors made on false assumptions. A lot of available manual documentation would be needed, which is not available to every use case. This can be automated by applying [37], also creating labelled images for the object detection simultaneously.

8 Conclusion

The requirements for sorting and detecting objects in a multi-image documentation were defined, and the *Image Analysis Pipeline* was developed to meet these requirements. A demonstrator was implemented to prove the concept's functionality. The combination of different methods used in the pipeline is resulting in the desired outcome. With extensive testing and identification of correct affinity-scores and weightings for each method, the pipeline's accuracy is increased. Without testing, the accuracy fluctuates based on estimated values, because of human error. Instructions and recommendations have been provided for implementing the pipeline in various civil engineering and construction use cases. Integrating BIM models or other semantic information is currently a new idea that requires further specification before implementation but has great potential to automate documentation in various fields.

References

1. Grassberger, M., Schmid, H.: Todesermittlung: Befundaufnahme und Spurensicherung; mit 27 Tabellen, 2. Wiss. Verl.-Ges, Stuttgart (2014)
2. Johannesson, P., Perjons, E.: A method framework for design science research. In: An introduction to design science, pp. 77–93. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-78132-3_4
3. Daniel, L., Daniel, L.: Digital forensics for legal professionals: understanding digital evidence from the warrant to the courtroom. Syngress/Elsevier, Waltham, MA (2012)
4. IPTC Standard: IPTC. [Online]. Available: <https://iptc.org/standards/photo-metadata/iptc-standard/>. Accessed 16 Jan 2024

5. Hoffmann-Walbeck, T., et al.: Standards in der Medienproduktion. In: X.media.press. Springer Berlin Heidelberg, Berlin, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-15043-2>
6. Pomerantz, J.: Metadata. In: The MIT Press essential knowledge series. The MIT Press, Cambridge, Massachusetts; London, England (2015)
7. ExifTool Tag Names. [Online]. Available: <https://exiftool.org/TagNames/index.html>. Accessed 16 Jan 2024
8. Clark (Alex), J.A.: pillow: Python Imaging Library (Fork). [Online]. Available: <https://python-pillow.org>. Accessed 16 Jan 2024
9. exif-js/exif-js: exif-js. Jan. 15, 2024. [Online]. Available: <https://github.com/exif-js/exif-js>. Accessed 16 Jan 2024
10. Noakes, D.: drewnoakes/metadata-extractor. Jan. 15, 2024. [Online]. Available: <https://github.com/drewnoakes/metadata-extractor>. Accessed 16 Jan 2024
11. Szeliski, R.: Computer vision: algorithms and applications, 2nd edn. In: Texts in computer science. Springer, Cham (2022). [Online]. Available: can be downloaded for free: <https://szeliski.org/Book/download.php>
12. OpenCV: Understanding Features. [Online]. Available: https://docs.opencv.org/3.4/df/d54/tutorial_py_features_meaning.html. Accessed 16 Jan 2024
13. Prince, S.J.D.: Computer vision: models, learning, and inference. Cambridge university Press, New York, 2012. [Online]. Available: can be downloaded for free: <http://www.computervisionmodels.com/>
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
15. OpenCV: Introduction to SIFT (Scale-Invariant Feature Transform). [Online]. Available: https://docs.opencv.org/3.4/da/df5/tutorial_py_sift_intro.html. Accessed 16 Jan 2024
16. OpenCV: Feature Matching. [Online]. Available: https://docs.opencv.org/4.x/dc/dc3/tutorial_py_matcher.html. Accessed 8 Jun 2022
17. OpenCV: Feature Matching + Homography to find Objects. [Online]. Available: https://docs.opencv.org/3.4/d1/de0/tutorial_py_feature_homography.html. Accessed 16 Jan 2024
18. Jiang, X., Hadid, A., Pang, Y., Granger, E., Feng, X. (eds.): Deep learning in object detection and recognition. Springer, Singapore (2019)
19. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. In: Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts (2016). [Online]. Available: <https://www.deeplearningbook.org/>
20. Lin et al., T.-Y.: Microsoft COCO: Common Objects in Context. arXiv, Feb. 20, 2015. [Online]. Available: <http://arxiv.org/abs/1405.0312>. Accessed 23 Jan 2024
21. COCO - Common Objects in Context. [Online]. Available: <https://cocodataset.org/#home>. Accessed 25 Jun 2023
22. models/official at master tensorflow/models: GitHub. [Online]. Available: <https://github.com/tensorflow/models/tree/master/official>. Accessed 23 Jan 2024
23. Models and pre-trained weights — Torchvision 0.16 documentation. [Online]. Available: <https://pytorch.org/vision/stable/models.html>. Accessed 23 Jan 2024
24. PyTorch Hub: PyTorch. [Online]. Available: <https://pytorch.org/hub/>. Accessed 23 Jan 2024
25. TensorFlow Hub: TensorFlow. [Online]. Available: <https://www.tensorflow.org/hub>. Accessed 23 Jan 2024
26. Hudson, L.: pyzbar: Read one-dimensional barcodes and QR codes from Python 2 and 3. [Online]. Available: <https://github.com/NaturalHistoryMuseum/pyzbar/>. Accessed 27 Jan 2024
27. ZBar bar code reader - About. [Online]. Available: <https://zbar.sourceforge.net/about.html>. Accessed 18 Feb 2024

28. OpenCV: cv::QRCodeDetector Class Reference. [Online]. Available: https://docs.opencv.org/3.4/de/dc3/classev_1_1QRCodeDetector.html. Accessed 27 Jan 2024
29. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning. In: Springer series in statistics. Springer, New York, NY (2009). <https://doi.org/10.1007/978-0-387-84858-7>
30. Scitovski, R., Sabo, K., Martínez-Álvarez, F., Ungar, Š.: Cluster analysis and applications. Springer Nature, Cham, Switzerland (2021). <https://doi.org/10.1007/978-3-030-74552-3>
31. Tan, P.-N., Steinbach, M., Karpatne, A., Kumar, V.: Introduction to data mining, 2nd edn, Global edition. Pearson, New York, NY (2020). [Online]. Available: <https://www-users.cse.umn.edu/~kumar001/dmbook/index.php>
32. scikit-learn: machine learning in Python — scikit-learn 1.4.0 documentation. [Online]. Available: <https://scikit-learn.org/stable/index.html>. Accessed 23 Jan 2024
33. stats package - RDocumentation. [Online]. Available: <https://www.rdocumentation.org/packages/stats/versions/3.6.2>. Accessed 23 Jan 2024
34. igraph R manual pages. [Online]. Available: <https://igraph.org/r/doc/>. Accessed 23 Jan 2024
35. Mostafa, K., Hegazy, T.: Review of image-based analysis and applications in construction. Autom. Constr. **122**, 103516 (2021). <https://doi.org/10.1016/j.autcon.2020.103516>
36. Klein, L., Li, N., Becerik-Gerber, B.: Imaged-based verification of as-built documentation of operational buildings. Autom. Constr. **21**, 161–171 (2012). <https://doi.org/10.1016/j.autcon.2011.05.023>
37. Kropp, C., Koch, C., König, M.: Interior construction state recognition with 4D BIM registered image sequences. Autom. Constr. **86**, 11–32 (2018). <https://doi.org/10.1016/j.autcon.2017.10.027>
38. Porway, J., Wang, Q., Zhu, S.C.: A hierarchical and contextual model for aerial image parsing. Int. J. Comput. Vis. **88**(2), 254–283 (2010). <https://doi.org/10.1007/s11263-009-0306-1>
39. Barbu, A., Zhu, S.-C.: Swendsen-Wang algorithm. In: Ikeuchi, K. (ed.) Computer vision, pp. 1227–1229. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-63416-2_721
40. Pandey, S., Khanna, P.: A hierarchical clustering approach for image datasets. In: 2014 9th International Conference on Industrial and Information Systems (ICIIS), pp. 1–6. IEEE, Gwalior, India, Dec. 2014. <https://doi.org/10.1109/ICIINFS.2014.7036504>
41. Xuehui, A., Li, Z., Zuguang, L., Chengzhi, W., Pengfei, L., Zhiwei, L.: Dataset and benchmark for detecting moving objects in construction sites. Autom. Constr. **122**, 103482 (2021). <https://doi.org/10.1016/j.autcon.2020.103482>
42. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ADE20K dataset. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5122–5130. IEEE, Honolulu, HI, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.7544>
43. ADE20K dataset. [Online]. Available: <https://groups.csail.mit.edu/vision/datasets/ADE20K/>. Accessed 25 Jan 2024
44. Cordts, M., et al.: The cityscapes dataset for semantic urban scene understanding. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3213–3223, Jun. 2016. <https://doi.org/10.1109/CVPR.2016.350>
45. Armeni, I., Sax, S., Zamir, A.R., Savarese, S.: Joint 2D-3D-semantic data for indoor scene understanding. arXiv, Apr. 05, 2017. [Online]. Available: <http://arxiv.org/abs/1702.01105>. Accessed 25 Jan 2024
46. Braun, A., Borrmann, A.: Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning. Autom. Constr. **106**, 102879 (2019). <https://doi.org/10.1016/j.autcon.2019.102879>



Image Segmentation for Enhanced Visualization and Accessibility of Historical Municipal Development Plans

Phillip Schönfelder¹(✉), Husan Duski², Jonas Maibaum¹, and Markus König¹

¹ Department of Civil and Environmental Engineering, Ruhr University, Bochum, Germany
phillip.schoenfelder@rub.de

² Department of Computer Science, Ruhr University, Bochum, Germany

Abstract. A significant portion of publicly available municipal development plans is currently presented solely as raster graphics, derived from scanned paper development plans. This practice presents a challenge in terms of both accessibility and visualization of crucial urban planning information. For planners and citizens alike, it is imperative that both new and old development plans be made available within a unified XML-based information system. However, as of today, only new plans are machine-readable in that sense. This limitation hinders statistical analysis, impedes visualization, and restricts accessibility. In response to this issue, this study focuses on improving the representation of and accessibility of these plans, with the goal of enhancing their utility for professionals and the public. The presented approach involves the integration of deep learning-based methods, such as the pre-trained Segment Anything Model (SAM), which eliminates the need for extensive training and allows for direct application in digitization processes. For fine-grained information extraction, a specialized YOLOv8 image segmentation model is trained. This paper explores potential applications of this technology within urban planning contexts and offers insight into potential solutions for the segmentation of development plans. For example, the application of this deep learning-based approach not only enhances the visual representation of development plans but also facilitates their linkage with textual specifications, making them more informative and user-friendly. This research contributes the broader discourse on leveraging artificial intelligence to improve the utility of historical planning documents and engineering drawings, ultimately fostering more informed decision-making and enhanced public engagement.

Keywords: Document Analysis · Development Plans · Deep Learning · Image Segmentation · Text Extraction · Visualization

1 Introduction

Urban planning stands at the intersection of historical legacy and contemporary innovation, shaping the foundations of our cities. In Germany, a notable challenge persists in the presentation of municipal development plans, where a significant share of these documents exists only in the form of raster graphics, artifacts of scanned paper development

plans. This practice, while preserving historical records, causes limitations to the accessibility and visualization of vital urban planning information, hindering the effective utilization of these plans for professionals and the public alike. While new development plans benefit from machine readability within a unified XML-based information system, the existing raster-graphic plans remain to be converted into machine-readable data [1, 2].

This research paper addresses the stated issue by applying computer vision techniques to the historical documents. Our approach pivots on the integration of deep learning-based methods, offering a solution to the challenges posed by raster-based development plans. First, the full-size plan is cropped to a Region of Interest (ROI) by means of the Segment Anything Model (SAM). Second, to achieve fine-grained information extraction, we train a specialized YOLOv8 image segmentation model, subdividing the development plan into individual parcels. This synergy of advanced computer vision technologies holds the promise not only to enhance the visual representation of development plans but also to enable their seamless integration with textual specifications. Therefore, in a final step, the text information given on the plan is extracted using a method presented in a prior study. The result is a more informative and user-friendly representation of historical and contemporary urban planning documents, partially overcoming the limitations imposed by traditional raster graphics. The enriched representation can be queried for certain information and offers some interaction with the development plan.

The paper is organized as follows: Sect. 2 provides the basic facts about the structure and contents of development plans, and reflects the state-of-the-art in document analysis methods, focusing on document types similar to development plans. Section 3 documents the proposed analysis method, which is demonstrated in Sect. 4. The results are discussed in Sect. 5, deducing implications for this research subject, and suggesting possible future studies. The paper concludes with final remarks in Sect. 6.

2 Background

2.1 Development Plan Digitization

Spatial planning systems differ from country to country in terms of involved actors, regulations, and institutional frameworks. However, while there are country-specific regulations for the development of building areas, the planning systems are similar in certain aspects. For instance, most include some sort of municipal plans, showing designated areas for certain types of land use. For instance, in Germany, a two-stage local planning system includes the preparatory land-use plan as a first stage, outlining the intended future use of land in accordance with the needs of the municipality. The second stage, binding land-use plans (*B-Plans*), i.e., development plans, while based on preparatory land-use plans, provide the legally prescribed specifications for district development. In contrast to land-use plans, these development plans focus on smaller parts of a municipality and give precise instructions on how and where development is allowed within properties [3]. Other examples are the *Local Area Plan* in Canada, *Zoning Regulations* in the USA and the *Plan local d'urbanisme* in France, which can be seen as comparable regulations for building areas and structures, even if there are differences in certain aspects [4, 5]. The remainder of this paper focuses specifically on B-Plan style

development plans (referred to as *development plans* in the following); however, it is expected that core features of the methodology can be applied to international equivalent documents as well.

In general, the development plan consists of two basic components: a plan drawing and a textual part. The drawing provides a graphical representation of the spatial arrangement of areas, clearly outlining the specific purpose of each area (Fig. 1). It specifies the type and extent of area usage, the areas of property that can be built upon, the architectural shape of the buildings, and the planning of local traffic areas, while visualizing these regulations through various graphic elements. These elements include different types of lines, shapes, hatching, colors, and specific symbols, combined with text and numbers such as abbreviations and numerical values, playing a central role in the illustration. The textual part of the plan augments the graphical representation, offering detailed instructions that are not easily conveyed through visual means. Furthermore, the development plan must define the boundaries that represent its ROI. This is crucial, as the regulations only apply to the areas within the defined area.

The trend towards the digitization of development plans has been a major focus recently. To digitally represent the spatial data in these plans, XML-based formats have been in existence for years and serve as both a semantic data model and an object-oriented data exchange medium for spatial planning [1]. These XML-based formats not only standardize the digital representation of plans but also enhance administrative processes in urban planning in terms of cost-effectiveness and quality. A key benefit of this format is the ease of data reusability in future projects. Consequently, numerous local authorities now also offer development plans in these formats on their websites. However, most of the plans are still only available in the form of scanned PDF documents, in which only a georeferenced outline of the area covered by the development plan is available. Nevertheless, for optimal utilization of the advantages presented by XML-based development plans [1], complete vectorial capture of all plans is required.

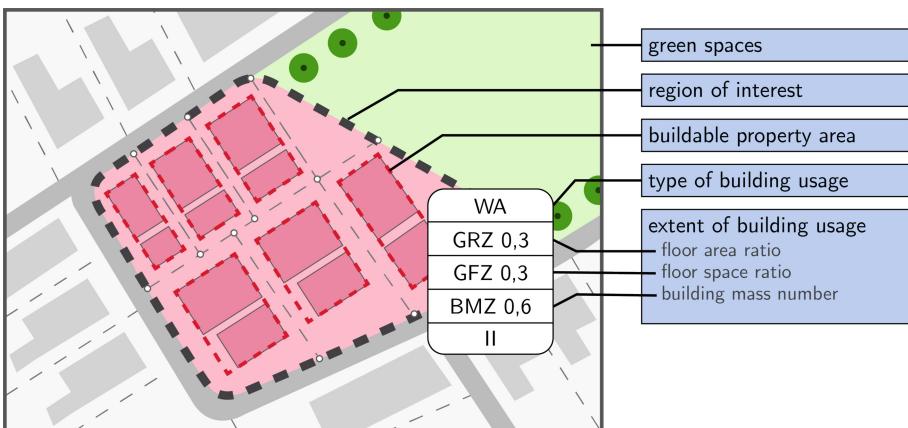


Fig. 1. Schematic view of an example development plan drawing.

2.2 Related Works

The retrospective digitization of existing planning documents is often addressed in the relevant literature, with a strong focus on construction documents. For instance, various studies are concerned with extracting geometry from floor plans [6–8] or construction drawings [9]. Especially in recent years, these tasks are typically handled with deep learning-based computer vision techniques, which have proven useful in this regard. Since the underlying extraction tasks like object detection and image segmentation are applicable to arbitrary geometries, it is reasonable to employ similar techniques to development plans as well.

Naturally, development plans cover a much larger physical area compared to single-building plans. A natural source of such large-scale images is satellite imagery, which has been subject to many computer vision approaches. For example, satellite images can be used to detect agricultural parcel outlines [10] and reconstruct building extents as polygons [11, 12]. Although these images exhibit features that regular development plans do not have, e.g., illumination changes, noise, colors, the ratio of the objects to be detected with respect to the whole image is quite similar to the case of parcels in development plans. While the mentioned studies report satisfactory results, it is worth exploring whether development plans pose unique challenges.

Considering the existing literature, it becomes evident that development plans have not been processed with deep learning-based methods, while input data with similar features has been successfully used for information extraction, in particular by means of object detection and image segmentation. Motivated by the lack of an automated digitization method of development plans, this study addresses the identified research gap.

3 Methodology

3.1 Processing Pipeline Overview

The proposed development plan processing pipeline consists of three main modules, i.e., ROI cropping, geometry extraction, and text extraction, as shown in Fig. 2.

A development plan raster graphic is a rather complex document. Along with the needed information, it also contains a drawing header, additional text descriptions, a legend, and surrounding land portions which remain unspecified by the document at hand. Therefore, the first step of the processing pipeline is to limit the following segmentation efforts to the region of the plan that actually contains defined zoning, i.e., the ROI. This pre-processing has multiple advantages, e.g., it reduces the need for computational resources and is necessary to avoid false positive zoning recognition outside this region.

Once the ROI is identified, multiple aspects about it can be analyzed in parallel, e.g., parcel outlines, the road network, types of land use, and textual details. Within the scope of this study, parcel geometries and text elements are extracted. Regarding the parcel outlines, a pixel segmentation mask is created for each parcel, making them separable from each other and quantifiable in size if the drawing scale is provided. Independently from any parcel geometry, text elements are located and converted to machine-readable strings.

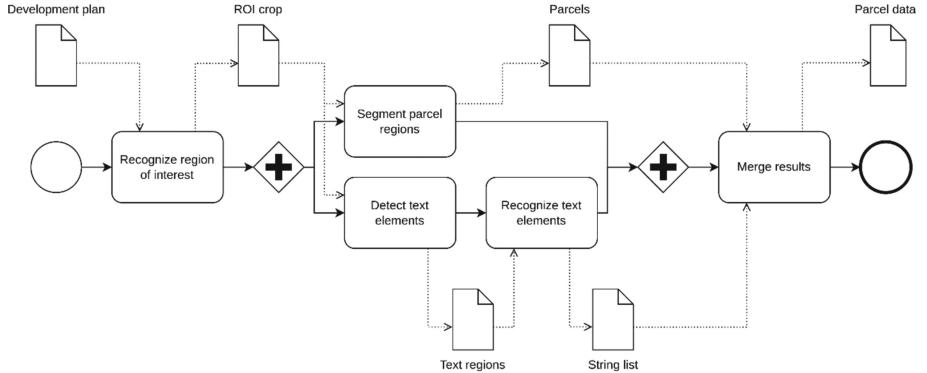


Fig. 2. Schema of the proposed processing pipeline.

In a final step, the results are merged, creating links between parcels and text elements enclosed by them. The individual substeps of the pipeline are described in the following sections.

3.2 Region of Interest Identification

The method used to extract the ROI consists of two stages: first, object detection for coarse localizing, and second, instance segmentation for refinement (Fig. 3). In the first stage, a YOLOv8 [13] model is used that is trained specifically for the task of detecting the ROI in development plans by means of an axis-aligned bounding box. Following this preliminary localization, the process advances to a second stage where the identified region is further refined to the arbitrarily shaped ROI.

This second stage uses the pre-trained SAM [14] to finalize the ROI segmentation. SAM emerges as a key protagonist in our method, obviating the need for extensive model training and facilitating direct applicability in the digitization process. It utilizes a foundational model that obtains its strength from the comprehensive SA-1B dataset, which consists of more than 11 million images and one billion segmentation masks, rendering it useful for many different image types.

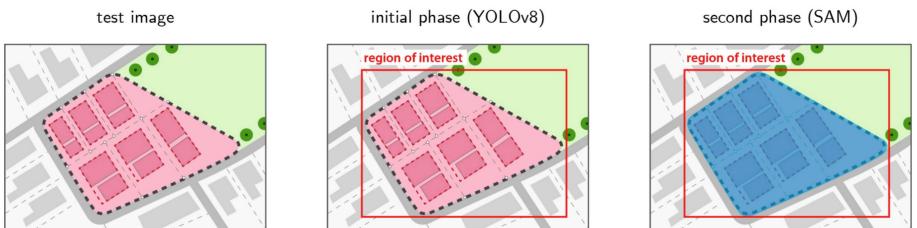


Fig. 3. Two-stage ROI extraction.

3.3 Parcel Segmentation

To segment the development plan into individual parcels and to identify them, another YOLOv8 model is trained on a custom labeled dataset. This model is offered in architectures of five different sizes, with varying network complexity and parameter count. The impact of model size on the datasets is evaluated by utilizing both the smallest (YOLOv8n-seg) and largest (YOLOv8x-seg) pre-trained models. Three different experiments are conducted for the training datasets, each with patch sizes of 512×512 and 1024×1024 pixels, which differ mainly in the number of epochs and the model sizes. The other hyperparameters are set to default values.

Since the model was originally developed for small image patches, a direct inference to the entire image is not feasible. Therefore, inspired by existing tiling methods [15, 16], high-resolution images are divided into overlapping patches corresponding to the training size used by the model. The model processes each patch individually, using the trained knowledge to segment parcels and generating a list of recognized parcels for each patch, which are then converted to coordinate-based polygons, each representing a detected object. Since detection is patch relative, not image relative, polygon coordinates are realigned to the original image context. The final step, which is critical for combining individual parcels from different patches into a complete image, is the merging of overlapping polygons created by patch overlaps.

3.4 Textual Information Extraction

The text elements are extracted based on a two-step method, including text detection and recognition, with models pre-trained as documented in [17]. In the first step, a YOLOv8 model [13] pre-trained on floor plan text detection is applied to the development plan ROI. The model is expected to generalize well since the two document types share key characteristics in terms of text appearance, e.g., the text is scattered throughout the plan, appears in small snippets, and may be partially occluded by other lines. The result of this processing step is a list of bounding boxes indicating the text positions. In the second step, the detected text snippets are individually fed to the Optical Character Recognition (OCR) model PARSeq [18], which was pre-trained on images of a scene text task. Finally, this results in a list of strings with corresponding locations on the image.

Based on the text locations and the identified parcel segmentation masks, each parcel is associated with the encapsulated text. This poses a straightforward merging of the two information types, which makes certain types of querying possible, e.g., searching for parcels with a certain number on the whole development plan.

4 Demonstration

The following sections document how the deep learning models are trained and show the results of applying the trained models to test images. While displaying the overall applicability of the method, we also highlight issues in the segmentation results. Furthermore, the pre-trained text recognition pipeline is put to a test, and the extracted information is reported by means of a web application.

4.1 Data Acquisition

Data acquisition is a crucial stage in this research and focuses on the reduction of inconsistencies by sourcing the data exclusively from the development plans of a single local authority, initially in PDF format. These plans are converted to JPG, which, as a raster image format, makes them align with the image processing pipeline. Since the data is initially unprocessed and lacks annotation, data preparation involves labeling the images for model training. To ensure reliability, considerable attention is given to this stage due to the potential influence of human error on data quality. However, not all of them are suitable for training the models, as development plans can be inconsistent even within a single municipality. Therefore, the suitability of a development plan as an example is assessed by visual inspection, focusing on image quality and consistency.

From the selected 119 plans, this study entails generating two identical datasets from the raw image collection, each undergoing distinct pre-processing steps to meet specific model requirements. The images of the first dataset are split into patches of two sizes, specifically 512×512 and 1024×1024 , to create uniform subsets. These subsets are used to train the YOLOv8 model for parcel segmentation. Conversely, the second dataset is used for ROI detection training and only requires a minimal image processing step of scaling down the image (Fig. 4).

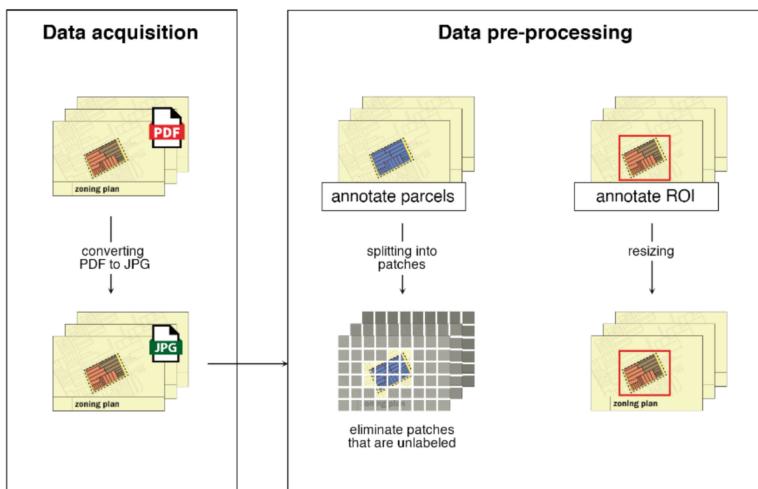


Fig. 4. Workflow of dataset preparation.

4.2 Deep Learning Model Training and Results

Regarding the ROI identification step, the results indicate that the YOLOv8 model is able to identify relevant areas in the test images with considerable accuracy (Fig. 5b). Moreover, the use of SAM allows for a highly accurate delineation of the labeled areas. However, in some cases, it is observed that the accuracy of segmentation relies on the

YOLOv8's bounding box quality and may be impaired if the inferred bounding box is too small (Fig. 5d). Thus, to ensure precise segmentation by the SAM model, the bounding box must accurately localize the relevant area and ensure that it is completely enclosed.

Furthermore, the results of the models from the experiments with patches of 512×512 and 1024×1024 pixels show the ability of the models to segment parcels in development plan images. Figure 6 features a selected input patch from the test dataset, alongside its ground truth annotation, and the corresponding model prediction. True labels are indicated by blue (Fig. 6b) to accurately designate labeled parcels, while model predictions present random colors (Fig. 6c) to illustrate the parcel instances segmented by the model.

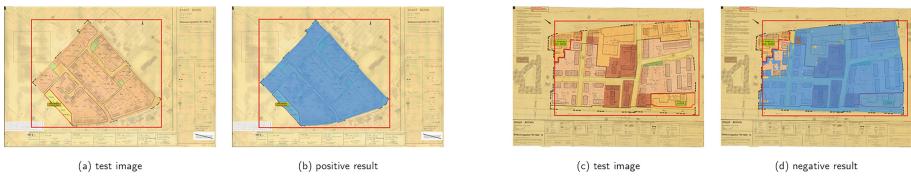


Fig. 5. Example results of the ROI segmentation step. b) showcases a positive segmentation result, for the development plan in a), while the plan in c) is not correctly recognized as seen in d).

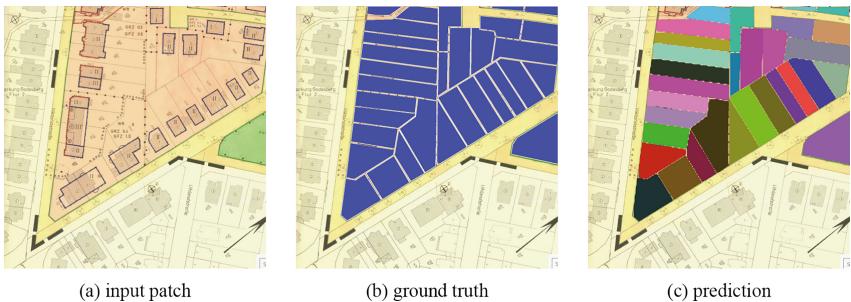


Fig. 6. Example results of the parcel segmentation step. Predictions are given in random colors to account for instance segmentation masks.



Fig. 7. Results after merging the individually processed plan tiles.

The results indicate that the models are generally capable of recognizing and segmenting the structure of parcels. Nevertheless, the entire image should be analyzed, not just individual sections of the parcels, which requires merging the segmentation results.

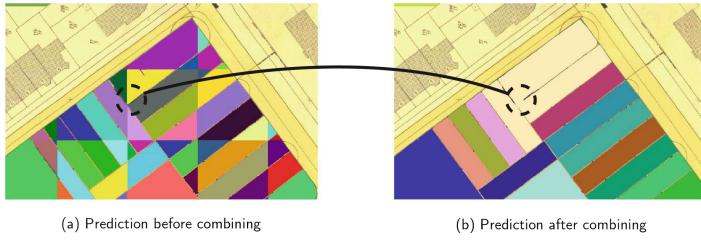


Fig. 8. Merging parcels divided by tiling step.

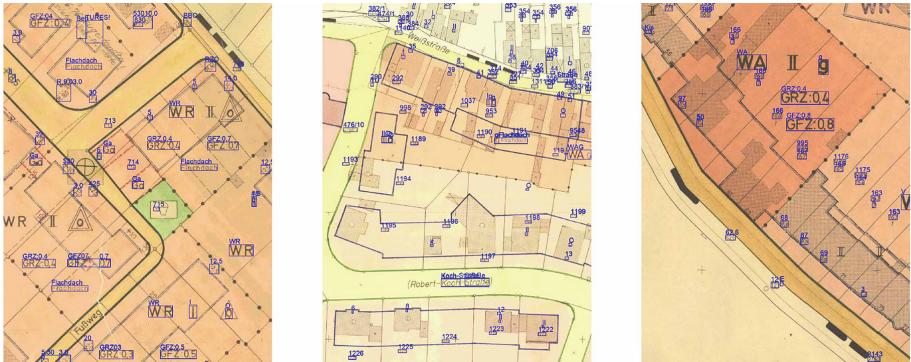


Fig. 9. Example results of the OCR pipeline.

As shown in the merged segmentation in Fig. 7, the inference with the tiling strategy is able to recognize parcels in smaller 512×512 patches quite well. However, the segmentation accuracy suffers in larger 1024×1024 patches.

One problem with the tiling strategy is the lack of clear guidelines for merging detected sub-parcels into a complete parcel (Fig. 8). Relying on overlap can lead to errors, such as the incorrect merging of neighboring parcels, which reduces the accuracy of the parcel segmentation process.

In Fig. 9, the YOLOv7 model accurately located the majority of text elements in the test images, as evidenced by the blue bounding boxes. The labels positioned above these boxes represent the texts that the PARSeq model extracted, thereby confirming the ability of the model to accurately extract text, even when hardly legible.

4.3 Web-Based Visualization

The demo web application aims to provide an intuitive user interface for the visualization of development plans, with features to upload these plans and automatically detect parcels using selected models from each task. Figure 10 shows that moving the cursor over detected parcels highlights them in blue, and clicking on them reveals their area in square meters to the right. While this is admittedly a rather limited interface, any extension to the proposed method, e.g., the extraction of further information, reflects in additional features in the interface, enriching the accessibility of the data.

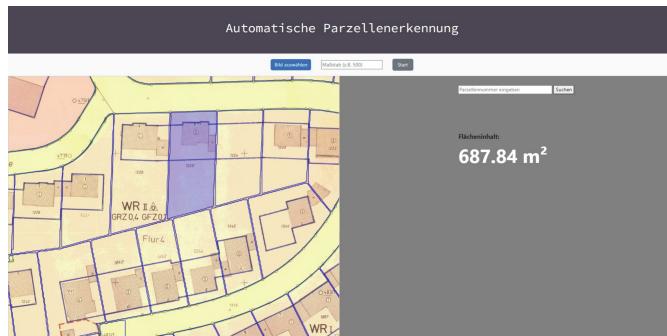


Fig. 10. View of the web application for results visualization and querying.

5 Discussion

The study outlines the successful development of an image processing pipeline using deep learning for automatic parcel recognition in development plans. The process involves several subtasks, from data acquisition to final visualization in a web application. In this way, the task was effectively broken down into manageable subtasks and the potential of the developed solution was proven.

Several implications arise from the study's results. First, it is noted that this proof of concept underscores the general applicability of deep learning-based tools for aiding the digitization in the context of urban planning documentation. Second, it becomes evident that methods previously developed for analyzing civil engineering drawings or architectural floor plans, such as the text recognition pipeline, generalize well to other 2D plan documents like development plans. Third, exploring the use of SAM for document analysis allows for the conclusion that, while the pre-trained model has proven useful for the task of ROI detection, it cannot provide satisfactory parcel segmentation results. Finally, this paper is not merely a technical exploration; it is a contribution to the broader discourse on the role of deep learning in improving the utility of historical planning documents and engineering drawings. Unlocking the potential of advanced computer vision can foster more informed decision-making processes and amplify public engagement in the context of urban development.

Although the study has shown promising results, there are still challenges that need to be addressed for an effective deployment. First, the inference method used for parcel segmentation lacks definitive criteria for merging detected parcels, potentially leading to inaccuracies. Therefore, future works should address more fine-grained rules for consistency and plausibility checks. Also, a human-in-the-loop approach could prove advantageous since, after all, development plans are legal documents which should be handled thoroughly. Second, incomplete area coverage using the YOLOv8 model in ROI detection could limit the segmentation capabilities of the SAM model, emphasizing the need for plausibility checking in this regard as well. Third, one challenge with text extraction is that the model identifies all text, instead of extracting only specific parcel numbers, resulting in false positive matches. This could be alleviated by querying regular expressions to filter the recognition results.

Possible future research includes extending the current focus on parcels to further information in the development plans. For instance, soil parameters could be extracted from as-built plans, or administrative information given in the drawing header might be used to categorize plans automatically. This would promote the comprehensive digitization and automation of planning processes.

Since the proposed method relies on tedious data annotation, one future research endeavor is particularly suggested here: Since a portion of development plans already exists in vectorized, XML-based formats, it is possible to overlay this information onto their pixel-based equivalents, creating labelled segmentation data automatically. This could unlock large amounts of training data, potentially leading to high-precision parcel segmentation results. In turn, these results could also be used for a quantitative analysis based on legally binding development plans, validating the method for the use in real scenarios.

6 Conclusions

Motivated by the potential benefits of digitizing existing planning documents, this study develops a method for information extraction from development plans with deep learning-based computer vision techniques. This fills the research gap of an automated processing pipeline for this specific document type. Applying YOLOv8 and SAM for ROI cropping and parcel segmentation proved effective in general, but also revealed certain limitations that should be addressed in future studies. For instance, rule-based post-processing is required for a meaningful final output, and manual data annotation remains an issue. A promising research endeavor is to use development plans that already have existing digitized versions, alleviating the need for manual annotation. All in all, this study contributes to the ongoing efforts to tackle digitization challenges with deep learning-based techniques, offering potential solutions in particular for the case of municipal development plans.

Acknowledgements. This scientific work was prepared as part of the research project “ProGeo – Implementation of product responsibility through circularity in geosynthetics”. It is funded by the Federal Ministry of Education and Research (BMBF) [FKZ 033R336] as part of the funding guideline “Resource-efficient circular economies – plastics recycling technologies (KuRT)” in the framework program “Research for Sustainable Development – FONA3”. The responsibility for the content of this publication lies with the authors.

References

1. Horenczuk, J.: Xplanung: Bedeutung, Organisation, Mehrwerte, Hemmnisse und Herausforderungen. In: Flächennutzungsmonitoring XIII: Flächenpolitik - Konzepte - Analysen – Tools, IÖR Schriften, vol. 79, pp. 131–325. Rhombos-Verlag, Berlin, Germany (2021). <https://doi.org/10.26084/13dfns-p029>
2. Hersperger, A.M., Thurnheer-Wittenwiler, C., Tobias, S., Folvig, S., Fertner, C.: Digitalization in land-use planning: effects of digital plan data on efficiency, transparency and innovation. Eur. Plan. Stud. **30**(12), 2537–2553 (2022). <https://doi.org/10.1080/09654313.2021.2016640>

3. Pahl-Weber, E., Henckel, D. (eds.): The planning system and planning terms in Germany: a glossary. In: Studies in spatial development, vol. 7. Verlag der ARL, Hannover (2008)
4. Krawchenko, T., Schumann, A.: Land use planning systems in OECD countries. In: The Palgrave encyclopedia of urban and regional futures. Palgrave Macmillan, Cham (2021). https://doi.org/10.1007/978-3-030-51812-7_216-1
5. Pecenik, P.: Spatial planning in France from the renaissance to the present day. Selected aspects. *Space Form* **45**, 193–212 (2021). <https://doi.org/10.21005/pif.2021.45.C-07>
6. Pizarro, P.N., Hitschfeld, N., Sipiran, I., Saavedra, J.M.: Automatic floor plan analysis and recognition. *Autom. Constr.* **140**, 104348 (2022). <https://doi.org/10.1016/j.autcon.2022.104348>
7. Gimenez, L., Hippolyte, J.-L., Robert, S., Suard, F., Zreik, K.: Review: Reconstruction of 3D building information models from 2D scanned plans. *J. Build. Eng.* **2**, 24–35 (2015). <https://doi.org/10.1016/j.jobe.2015.04.002>
8. Dodge, S., Xu, J., Stenger, B.: Parsing floor plan images. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), pp. 358–361, IEEE, Nagoya, Japan (2017). <https://doi.org/10.23919/MVA.2017.7986875>
9. Zhao, Y., Deng, X., Lai, H.: Reconstructing BIM from 2D structural drawings for existing buildings. *Autom. Constr.* **128**, 103750 (2021). <https://doi.org/10.1016/j.autcon.2021.103750>
10. García-Pedrero, A., Lillo-Saavedra, M., Rodríguez-Esparragón, D., Gonzalo-Martín, C.: Deep learning for automatic outlining agricultural parcels: exploiting the land parcel identification system. *IEEE Access* **7**, 158223–158236 (2019). <https://doi.org/10.1109/ACCESS.2019.2950371>
11. Li, Z., Wegner, J.D., Lucchi, A.: Topological map extraction from overhead images. In: 2019 IEEE/CVF International conference on computer vision (ICCV), pp. 1715–1724. IEEE, Seoul, South Korea (2019). <https://doi.org/10.1109/ICCV.2019.00018>
12. Zhao, W., Persello, C., Stein, A.: Building outline delineation: from aerial images to polygons with an improved end-to-end learning framework. *ISPRS J. Photogramm. Remote. Sens.* **175**, 119–131 (2021). <https://doi.org/10.1016/j.isprsjprs.2021.02.014>
13. Jocher, G., Chaurasia, A., Qiu, J.: YOLOv8. <https://github.com/ultralytics/ultralytics/>. Last accessed 15 Dec 2023
14. Kirillov, A., et al.: Segment anything. arxiv 2304.02643 (2023). <https://doi.org/10.48550/arXiv.2304.02643>
15. Akyon, F.C., Altinuc, S.O., Temizel, A.: Slicing aided hyper inference and fine-tuning for small object detection. In: 2022 IEEE International conference on image processing (ICIP), pp. 966–970. IEEE, Bordeaux, France (2022). <https://doi.org/10.1109/ICIP46576.2022.9897990>
16. Ünel, F.Ö., Özkalayci, B.O., Çığla, C.: The power of tiling for small object detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 582–591. IEEE, Long Beach, CA, USA (2019). <https://doi.org/10.1109/CVPRW.2019.00084>
17. Schönenfelder, P., Stebel, F., Andreou, N., König, M.: Deep learning-based text detection and recognition on architectural floor plans. *Autom. Constr.* **157**, 105156 (2024). <https://doi.org/10.1016/j.autcon.2023.105156>
18. Bautista, D., Atienza, R.: Scene text recognition with permuted autoregressive sequence models. In: 2022 European Conference on Computer Vision (ECCV), pp. 178–196. Springer, Tel Aviv, Israel (2022). https://doi.org/10.1007/978-3-031-19815-1_11



Optimization of BIM Collaboration Format Data Analysis Through Advanced Classification and Information Extraction

Benedikt Kandler^(✉), Marcel Heiß, and Uwe Rüppel

Institute of Numerical Methods and Informatics in Civil Engineering, Technical University of Darmstadt, Darmstadt, Germany
kandler@iib.tu-darmstadt.de

Abstract. In BIM projects, BIM Collaboration Format (BCF) files are a central resource for interdisciplinary coordination and communication between different project participants. During a project, complex information is collected in the BCF files, but this information is often insufficiently analysed. An optimized, cross-project evaluation of the different BCF files can be achieved based on fundamental classifications of the BCF files and the information contained within them. However, the existing classification options in the BCF format, especially for complex construction projects, as well as the actual classification process in the construction process, often prove to be insufficient. This is not only due to the additional effort required for classification but also to the difficulty of accessing specific information, such as building components described in the BCF. The direct linking of further information resources is difficult in the BCF format, which makes an automated classification of the BCF files even more difficult. This paper presents an innovative classification scheme that takes into account the structure of BCF files and extends it with various classification parameters. The primary goal is to extract the information from BCF files and convert it into a structured data format. This enables efficient storage in a database and provides the basis for later analysis. The classification of BCF files will be partially automated by using information extraction methods that utilize natural language processing (NLP) to extract essential text segments from the unstructured information in the BCF files. This methodology will be used to simplify the future evaluation of the BCF information and to classify the different BCF topics in terms of their technical characteristics. In the first phase of the evaluation, the focus is primarily on semantic information, with the potential integration of other data and information sources already in prospect.

Keywords: BCF · LLM · Knowledge Management

1 Introduction

In BIM projects, the BIM Collaboration Format (BCF) serves as an essential tool for communication and issue documentation, enabling stakeholders to exchange project-specific comments and address issues during the construction process. Although effective

for coordination and incident recording, the broader potential of BCF for post-project analysis often remains untapped, leading to the underutilization of valuable information. The challenge is that this information, though recorded, is in an unstructured format, limiting its analysis and reuse. To effectively use this information for future projects, it is crucial to transform this unstructured data into a structured format, to make the information more accessible and therefore enable efficient analysis creating a foundational knowledge base.

This paper aims to develop a strategy for extracting and analyzing knowledge from BCF files laying the groundwork for subsequent analysis and classification tasks. The process will utilize BCF XML files, though is adaptable for BCF data in general. The classification will target data like titles and comments within BCF files, aligning with predefined classification categories, to ensure consistency across multiple BCF files. This classification itself will be conducted by utilizing existing Large Language Model (LLM) tools like ChatGPT, reducing the necessity for model training for basic Natural Language Processing (NLP) tasks.

Additionally, we aim to mine further information from the BCF files using various algorithms, enhancing the information basis for further analysis. Crucially, the extracted information, along with the classifications and derived insights, will be formatted into a simple, structured data format. This enables seamless integration into various analysis tools, allows for reintegration into the original BCF files, or facilitates storage in databases. This structured approach will significantly improve data organization, retrieval, and utility, enhancing the potential for comprehensive analysis and application in future projects. Crucially, the system for storing metadata and categorizing BCF files and issues is designed to be dynamic, evolving to suit different construction project types and conditions. The open structure of the data format enables storage of additional generated information when data generation algorithms and classification categories are updated and extended in future works. Additionally, it allows to adapt to different conditions in construction projects as well as the diversity of information in BCF files.

Identifying potential challenges is crucial in this context. These encompass the need for flexible classification categories, the efficiency of information extraction algorithms, and the integration of structured data into diverse analytical tools. Being aware of these issues highlights the complexities involved in developing a comprehensive and adaptable classification system in the ever-evolving field of construction project management.

2 BCF – State of Affairs

This section offers a detailed overview of the current technologies in BCF usage and analysis, focusing on data storage, information extraction methods, and general text analysis within BCF files.

2.1 Current State of BCF

BCF's primary application in construction projects predominantly revolves around communication and information exchange in issue management. A BCF file can contain

multiple issues or ‘topics’ as referred to in the BCF documentation. A topic is a specific problem or point of discussion related to the construction project, which can be caused by planning errors during the design phase or execution mistakes in the construction phase. While the root level of a BCF file contains multiple files defining its structure, the focus here is on the ‘topic’ folder for each issue. The ‘topic’ folder can include multiple files like markup.bcf (storing topic information including GUID, title, and stakeholder comments), viewpoint.bcfv (containing IFCGuid of related model components), and snapshot files [1]. The markup file is particularly crucial as it encompasses the core information about the ‘topic’. Additionally, the viewpoint file links the ‘topic’ to specific model components, highlighting the selected elements within the BCF.

Beyond the documentation of issues during the construction phase, BCF can be extended for documentation and asset management in the operation and maintenance phase, though its primary focus remains in the planning stage, which highlights the different conditions under which BCF can be used [2]. The type of information within BCF files is dictated by the specific use case. Therefore, the information can vary from design choice annotations to construction process issues and handover documentation [3]. For example, when BCF is utilized for renovation work, no model is available; however, BCF still is a viable tool for exchanging information and images of the construction site. The different use cases highlight the variety of data that can be stored within a BCF file, which can be used to generate new knowledge.

2.2 BCF – Information Extraction

While recognizing the value of BCF file analysis is common in industry reports and academic literature, detailed methodologies for such analysis are notably sparse. Currently, there are two primary technical approaches for accessing BCF file information: via BCF XML files or through the BCF API on a server.

Accessing BCF files in the XML format typically involves using a dedicated application or script or doing it manually. This often necessitates local storage of BCF files to ensure accessibility for scripts or applications. However, this method requires downloading the entire BCF file to access specific sections like comments, which can result in longer processing time while querying the BCF data. The BCF API, on the other hand, facilitates direct server retrieval through HTTP requests. Despite its convenience, to access all of the BCF files directly, the API’s limited query capabilities, particularly for in-depth data like comments, often necessitate retrieving the entire BCF file, leading to increased data traffic. To bypass this, enhancements such as new routes in the BCF API have been proposed, specifically for direct access to sections like comments [4]. According to the BCF API documentation, in version 3, routes directly to the comments, for example, were implemented [5].

An alternative approach proposed to improve data access is to extract data from the BCF File and store it in a relational-based data storage. This could be either directly in a relational database reflecting the BCF Schema structure or in data storage based on an ontology, such as BCF OWL, which mirrors the BCF schema into an ontology. The advantage of this method is having data storage where individual information entities, like comments, pictures, and labels, can be directly queried. Additionally, functionalities

such as interpreting information with an ontology reasoner and linking the BCF information to building data via other ontologies like CTO (Construction Task Ontology), BOT (Building Topology Ontology), or IFC OWL could be further implemented and enrich the contextual understanding of BCF information. The exploration of using a relational database like PostgreSQL for this purpose hasn't been thoroughly investigated yet [6].

2.3 Classification by Natural Language Processing

An integral task in this research is the classification of text information within BCF files, which are mostly unstructured. To interpret and structure this data effectively, NLP is essential. In NLP, text categorization into labels and categories involves understanding and interpreting the context and nuances of language data. Initially, rule-based methods that rely on manually crafted linguistic rules, were intended for classification in this paper.

Named Entity Recognition (NER), a specific form of text classification plays a crucial role in identifying and categorizing key information such as names, locations, and dates. Traditional NER methods range from rule-based systems, leveraging linguistic expertise, to supervised learning models that require annotated training data. However, NER systems face certain limitations: they typically have a limited context understanding, are often language and domain-specific, need substantial training data, and their classification categories are usually pre-defined, making them less dynamic. In addition, for optimal performance, NER systems require customization for specific fields like construction. Although there are advancements in research suggesting an NER for building construction, a universally recognized system in this sector is yet to be established [7].

Advancements in artificial intelligence, in special in the field of LLMs have led to a shift to the integration of LLMs such as GPT for NER tasks. LLMs are pre-trained on extensive text data, giving them a deep understanding of language contexts. However, their text-generation nature poses challenges in tasks like NER, particularly in overgeneralization and lack of granularity in sequence labeling. GPT-NER addresses this by transforming NER into a text-generation task, utilizing special tokens in text sequences to mark entities. This approach adapts NER to LLM capabilities, enhancing precision and reliability through a self-verification strategy, where the model is asked to verify the correctness of its identification. GPT-NER shows comparable performance to traditional supervised models, particularly beneficial in low-resource environments with limited labeled data. This adaptability is crucial for real-world NER applications, where improving demonstration quality and prompt structures in in-context learning proves more effective than merely increasing training data volume [8].

Despite the advancements in NER, certain drawbacks, like the requirement for pre-defined classification categories and their possible values, persist. However, another promising approach involves using LLMs for classification tasks. The CARP (Clue and Reasoning Prompting) method exemplifies this, employing a progressive reasoning strategy. Initially, it prompts LLMs to identify superficial clues—keywords, tones, semantic relations. These clues lead to a diagnostic reasoning process for decision-making. This method showcases LLM's effectiveness in text classification, particularly with limited datasets, by leveraging their extensive pre-training for keyword recognition

and analysis. Different setups and engines for LLMs were explored, including Zero-Shot, Few-Shot, and Fine-Tuning methods, and their capabilities were demonstrated. Coming to the results that the LLM gets better classification with the Few-Shot demonstration or Fine-Tuning approach. In the Few-Shot approach, a minimal dataset is used within prompts. A k-nearest neighbors model supplements this to select appropriate examples, addressing token limitations inherent in LLM prompts. Fine-tuning the LLM, while resource-intensive, yields the best results by customizing the model more closely to specific requirements. Compared to traditional NER, LLM-based classification, such as CARP, offers greater flexibility and is capable of handling limited data scenarios. This research will follow the approach of LLM-based text classification, overcoming language-specific barriers of NER and facilitating analysis with constrained datasets. Notably, the Zero-Shot approach, which operates without example data, offers a valuable starting point for classification tasks, especially in situations where data availability is constrained [9].

Compared to traditional NER, this approach offers advantages in predefining classification options, leading to system flexibility despite potential inaccuracies in results. Considering data availability and analysis tools, this research opts for LLM-based text classification. This method not only overcomes language barriers inherent in NER but also allows analysis based on limited data.

3 Methodology

The core idea of this paper is to extract the implicit knowledge embedded in BCF communication protocols, which are generated by engineering specialists and planners during projects. A methodology is presented to accumulate and process this partly unstructured information and convert it into a format suitable for further analysis. While traditional NER systems might seem ideal for this task, their application is not feasible in this context due to the lack of available data and a trained model specific to this domain.

In contrast, the capabilities and accessibility of LLMs like ChatGPT offer a more practical solution. LLMs operate on pre-trained data, providing a flexible classification ability not limited to specific training rules but rather reliant on the LLM's broader understanding of language and context. This flexibility, along with their widespread recognition and usage, positions LLMs as a competitive choice for this research.

This research employs a hybrid classification strategy, focusing on LLMs for information extraction using a categorization system. The methodology involves crafting precise prompts to effectively direct the LLMs, a crucial factor for accurate classification. In this process, we initially extract data from BCF topics. Depending on the nature of this data, it undergoes either direct processing or is fed into the LLMs for further classification. Key text elements are then integrated into these prompts, ensuring they are tailored for effective LLM analysis. Additionally, to enhance the classification's context, IFC element types are extracted from corresponding IFC files. The comprehensive process, from data extraction to the generation of new insights, is depicted in Fig. 1 and will be detailed further in this chapter.

It's important to note that this research is conceptual. While test data from BIM-Collab [10] is mentioned to illustrate the practical application of LLMs in classifying

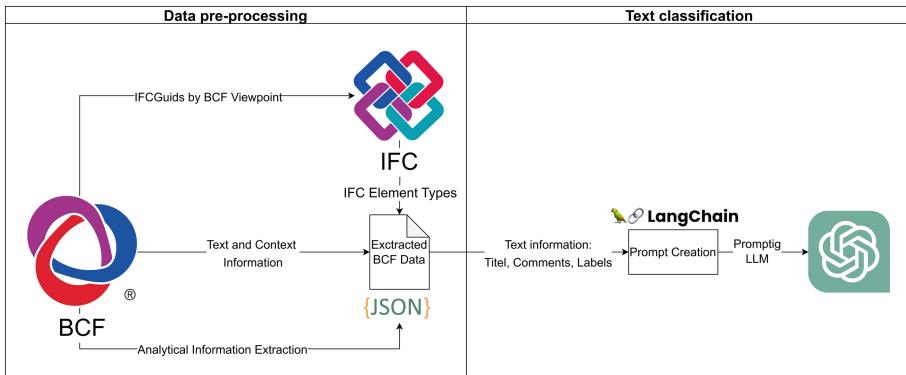


Fig. 1. Information extraction process

information from BCF files, this research does not include actual test data but rather proposes a conceptual framework for future implementation.

3.1 Data Preprocessing

To create a prompt, it's essential to first identify the relevant attributes in a BCF schema. These attributes can then be extracted from the BCF files. When developing the information extraction process, it's important to differentiate between the attributes that are useful for prompt construction and those that will be needed for later evaluations. This includes not only direct textual information but also metadata and contextual details that can aid in comprehensive analysis. For prompt creation, text information from BCF files such as titles and accumulated comments over the project's course is essential. It is important to note that a BCFZIP file contains several individual issues, each ideally focusing on a specific 'topic'. An issue is divided into multiple files containing various information, with central descriptive information in the Markup file. The following attributes have been identified as relevant:

- Title: Provides a specific description of the topics.
- Comments: Documents the problem-solving process and stakeholder communication.
- Labels: Offers additional metadata describing the BCF issue.

While the Markup file also contains attributes like 'TopicType', 'Priority', and 'Author', these are deemed less critical for prompt construction due to their limited semantic content. Nevertheless, they hold value for broader context understanding. For example, 'TopicType' categorizes the issue but lacks direct semantic detail. On the other hand, author information, while potentially useful, was unavailable in our datasets. Additionally, timestamps within BCF and comments are insightful for understanding the timeline of stakeholder interactions and issue resolution, although they are not directly leveraged in prompt creation.

The Visualization or Viewpoint file is another central component that contains linked model elements identified during BCF creation. These elements are stored with their IFCGuid and include attributes describing their visibility and selection status, making

selected model elements particularly relevant to the issue at hand. Extracting the selected IFCGuids from the IFC model gives us insight into the types of model elements involved, which contributes to prompt construction.

Besides the Markup and Visualization files, project-specific information can be sourced from either the root level of the BCF file or directly from the IFC file, adding another layer of contextual detail for our analysis.

3.2 Prompt Building

After extracting information from the BCF file, the next step is its further classification. Unlike NER, which classifies text into predefined classes and values based on rules, using LLMs involves a text generation task for classification. The prompt for generating this text, which subsequently represents the classification task of the input, must be carefully crafted. Since LLMs enable a more open-ended classification not limited to a specific set of possibilities, it's crucial to specify the framework and criteria for the LLM to classify the text. Additionally, the LLM's output must be defined in a way that can be processed by programs.

The process of classifying information extracted from BCF files differs from traditional NER systems. Unlike NER, classification through LLMs involves generating a response based on a carefully constructed prompt. This prompt not only guides the LLM in text generation but also serves in this case as the basis for classification. Therefore, it is essential to define the framework and criteria for the LLM's classification task and the required output structure clearly to enable the processing by subsequent programs.

A comprehensive prompt includes four critical components: the definition of the task, the criteria for classification, structured text information from the BCF, and instructions for the output format. When previously classified information from BCFs is available, these examples can be used in a method called Few-Shot Learning or In-Context Learning. This approach leverages the LLM's pre-trained knowledge to align the responses with the given examples, without necessitating specific training for each task. When utilizing this method, it's important to consider the token limit by selecting only relevant examples while maintaining a consistent format. These requirements can be met with automated prompt generation and the use of prompt templates.

The prompt typically starts with an introduction of the task, informing the LLM about the nature of the construction issue to be classified. The format for the expected output, such as a dictionary in JSON format, is specified upfront to ensure the usability of the processed information. The prompt also outlines the categories for classification, each accompanied by a brief definition to aid the LLM's understanding. Currently, the LLM is tasked with classifying information based on predefined categories as detailed in Table 1. While providing value ranges for each category can make the LLM's output more predictable, it also limits its capability to those predefined values.

Afterward, the LLM is provided with the text information to be classified, along with details about the type of information it represents. If already classified information is available, both the classified data and the results can be presented here to the LLM as examples, indicating that they are for reference.

To enhance accuracy, a self-verification method may be incorporated. Here, the LLM revises its classifications, confirming their appropriateness with simple affirmative or

Table 1. Classification categories

Classification category	Description
Initial Identified Issue	The first identification or description of the issue
Issue History	A chronological sequence of events related to the issue
Final Issue	The current or unresolved aspect of the issue
Additional Needed Resources	Any additional resources, planning, or analysis required to address the issue
Actions Taken	Specific actions taken in response to the issue
Final Resolutions	Proposed or implemented solutions to resolve the issue

negative responses. This step is particularly valuable given LLMs' tendency to sometimes overgeneralize or create spurious entities. The prompt design is modular, allowing for incremental enhancements or additional instructions to optimize results while ensuring core components are included.

Finally, prompts are executed through the ChatGPT web application or via an API. Tools like LangChain facilitate the creation of flexible prompt templates. Using an API also provides control over the LLM's engine and parameters, such as 'temperature,' which influences response variability, and 'frequency_penalty' and 'presence_penalty,' which discourage repetition and encourage the discussion of new topics.

4 Implementation

In this chapter, we detail the implementation process, starting with the extraction and preprocessing of data from BCF files and later on classifying it with LLMs, using BIM-Collab's test data based on BCF Version 2.0. This version might differ from the latest schema of Version 3.0, as reviewed in the official BCF-XML Github repository [1]. The focus here is on extracting insightful information from these files, including the analysis of IFCGuids and their relevance to various project components and topics.

4.1 Data Preprocessing

The process begins with the extraction of data from BCF files. Since the structure is XML-based, it's possible to search for specific nodes, their subnodes, and attributes, and iterate through the text information. For text information like comments or labels, iteration over the corresponding nodes is required. Attributes like comments, titles, and 'TopicStatus' are directly accessible. However, the aim is to derive information beyond what is immediately visible.

To provide the LLM with a better understanding of the component types affected by a particular 'topic', all component IFCGuids contained in the Viewpoints were extracted. These extracted IFCGuids must be matched with the corresponding IFC files. This will allow us to locate specific IFC element instances and derive their IFC element types, which can then be attached to our information structure. In addition, we determined the

```
{"GUID": "1c948e4d-18fc-4b7c-88f1-4a74c7d3e8e9",
 "Title": "Intersections of column",
 "Comments": ["Issue related to issue #11",
    "Ventilation duct clashes with structural columns. Please move the ducts."],
 "Labels": ["Architecture"],
 "Topic_Status": "Active",
 "Topic_Type": "Clash",
 "Ifc_Elements": ["32Z0kEuDcJGuqA2F891ld7", "3w1MgpXIT6U8u4qlxKu2sF"],
 "Ifc_Element_Types": ["IfcColumn", "IfcFlowSegment"],
 "Issue_Timespan": "719 days, 6:09:01")}
```

Fig. 2. Extracted data dictionary

period between the creation of the BCF and the last comment. While this metric does not directly contribute to prompt generation, it offers valuable insights for subsequent analyses. Finally, the details of individual issues are stored in a dictionary for easy reference as displayed in Fig. 2. As observed, some comments mention related BCF topics, which could be utilized to identify connections between them. However, this approach is not further explored in this context.

In addition to directly extracting data, a secondary layer of information is derived by analyzing the appearance of IFCGuids across various issues. This method aids in identifying elements that frequently occur, offering insights into potentially critical components. The compiled information includes the IFCGuids, their associated model element types, and the corresponding BCF titles and GUIDs, all of which illustrate the involvement of IFC element instances in different ‘topics’. Figure 3 showcases examples of these recurring IFC element instances. The interpretation and analysis of this aggregated data offer a promising avenue for future research.

```
{"IFC_Element_GUID": "2GkxjylDH6seFyg9HVc3Nt",
 "BCF_GUIDs": [
    "0fa51a7a-5e82-420b-bf33-da94aa82a93", "63979943-df22-446b-944c-e4fff92402ea",
    "7031d024-7c99-47d5-a379-359d73017852", "ebb1a8bf-6d1d-4aad-a875-61ad3cc40d30"],
 "BCF_Titles": [
    "Hollow core slab and sewer are clashing.",
    "Ventilation duct too close to wall",
    "Sewer and hollowcore slab intersections",
    "Ventilation and prefab concrete"]}
```

Fig. 3. Multiple referenced IFCGuids

The project and site information stored in the IFC file and project.bcfp files at the root of the BCF file could offer additional context for the LLM, their extraction isn’t feasible due to the insufficient availability of these data.

4.2 Prompting

The text from each BCF ‘topic’ undergoes classification through an LLM. This is done by constructing a prompt that guides the LLM in its classification task. The prompt needs to be well-structured to address all classification categories, output format requirements,

and any provided few-shot examples. Creating the prompt involves several key components. While a basic approach using f-strings with placeholders is possible, larger prompts require more complexity. To help with this, we use the LangChain [11] library, which provides a robust framework and diverse methods for developing LLM-based applications. The library offers features like PromptTemplates, Few-Shot Templates, and OutputParsers to facilitate prompt construction.

During the process of building a prompt, the first step is to define the actual task which needs to be performed. This involves creating a string without any placeholders that describe the task to be included in a Prompt Template. The description of the task can be a simple introduction sentence as shown in Fig. 4. To obtain better results, various task prompts can be created and tested.

Task: Given a construction issue, classify it based on the text provided into the given categories and consider their definitions.

Fig. 4. Task prompt

The subsequent phase is defining the output format for the LLM. To ensure the data is returned in a structured format a JSON parser is implemented. To establish a precise output format, a class was designed to specify the various attributes and their respective data types of the output format. This class is then passed to the OutputParser as the format structure, which is then integrated into the prompt. This output format also implicitly defines the classification categories that the LLM needs to focus on. An example of the prompt component, showcasing the standard JSON schema utilized, is illustrated in Fig. 5.

Here is the output schema:

```
```json
{
 "properties": {
 "Initial_Identified_Issue": {"title": "Initial Identified Issue", "description": "The first identification or description of the issue.", "type": "string"},
 "Issue_History": {"title": "Issue History", "description": "Chronological sequence of events related to the issue.", "type": "string"},
 "Final_Issue": {"title": "Final Issue", "description": "The current or unresolved aspect of the issue.", "type": "string"},
 "Additional_Needed_Resources": {"title": "Additional Needed Resources", "description": "Any additional resources, planning, or analysis required to address the issue.", "type": "string"},
 "Actions_Taken": {"title": "Actions Taken", "description": "Specific actions taken in response to the issue.", "type": "string"},
 "Final_Resolutions": {"title": "Final Resolutions", "description": "Proposed or implemented solutions to resolve the issue.", "type": "string"}
 }
}
```

```

Fig. 5. Classification schema

Next, the actual information for classification is introduced into the prompt. Utilizing a PromptTemplate with placeholders, this template is filled with data from a selected BCF ‘topic’. For instance, details from the ‘topic’ “Intersections check: column inside slab” are converted into text form and integrated into the prompt, as demonstrated in Fig. 6.

Given Information:

Title: Intersections check: column inside slab

Comments: Will adjust the slabs in next model refresh: Hybrid BIM office_BWK_v3.ifc, The columns should run through. Slab connects to the column. Will model this soon., The columns are intersecting with the slab on the first floor.

Labels: Architecture

Linked Model Elements: IfcColumn, IfcSlab

Fig. 6. BCF data prompt

To aid the LLM's In-Context Learning, few-shot examples are incorporated. Due to data constraints, we limited this to one example per prompt, utilizing LangChain's Few-Shot template class. While it's feasible to include multiple examples, the token limitations of LLMs necessitate careful selection. LangChain also offers embedding models that could assist in this process, though this approach was not extensively explored in our study. The culmination of this process, integrating the few-shot prompt template with examples, is depicted in Fig. 7.

Provided example information:

Title: Intersections of slab

Comments: This will cause some problems, need to have a meeting with the architect to discuss a different type of Air inlet., Air outlets MEP are clashing with slab.,

Label: Architecture

Linked model elements: IfcFlowSegment, IfcFlowFitting, IfcFlowTerminal

Example issue classification: Initial_Identified_Issue: Intersections of slab,

Issue_History: None,

Final_Issue: Air inlet and outlet MEP clashing with slab.

Additional_Needed_Resources: Meeting with architect, redesign of air inlets and outlets,

Actions_Taken: Identification of the issue and plan to meet with architect,

Final_Resolutions: None

Fig. 7. Few-shot example prompt

Once the prompt is crafted, it is forwarded to an LLM, like ChatGPT. The prompt can be submitted through the web application or an API interface. The LLM processes the provided prompt and generates output classification information in the predefined JSON file structure. The returned classification is displayed in Fig. 8.

```
{"Initial_Identified_Issue": "Intersections check: column inside slab",
 "Issue_History": "Columns intersecting with slab on the first floor identified",
 "Final_Issue": "Need to adjust slabs for column integration",
 "Additional_Needed_Resources": "Model refresh in Hybrid BIM office_BWK_v3.ifc",
 "Actions_Taken": "Plan to model the integration of slab and column",
 "Final_Resolutions": ""}
```

Fig. 8. Classified BCF text information

The LLM also produces additional semantic information beyond the basic format, which needs to be post-processed for implementation in further processes. However,

a few pieces of information could be classified and returned in the required output format. The information extracted by the LLM, along with data obtained through various algorithms, is then systematically integrated into a structured format, paving the way for in-depth analysis. An overview of the comprehensive data, encompassing both the LLM outputs and algorithmically derived insights, is presented in Table 2.

Table 2. Extracted information

| Schema Attribute | Data Acquisition Method |
|-----------------------------|-------------------------|
| GUID | BCF information |
| Labels | BCF information |
| Topic_Status | BCF information |
| Topic_Type | BCF information |
| Specialists | BCF information |
| Overlapping_IFCGuids | BCF Data extraction |
| IFC_Element_Types | IFC-BCF Data extraction |
| Resolution_Time | BCF Data extraction |
| Initial_Identified_Issue | LLM classification |
| Final_Issue | LLM classification |
| Additional_Needed_Resources | LLM classification |
| Actions_Taken | LLM classification |
| Final_Resolutions | LLM classification |

5 Results and Discussion

This research primarily focuses on automating the extraction and preparation of data from BCF files for advanced analysis. The methodologies presented serve as a conceptual foundation for more detailed evaluations.

When evaluating the classification results, it's observed that the LLM accurately extracted the title for the 'Initial_Identified_Issue'. This extraction logically aligns with the typical structure of BCF files as the title is defined when creating the BCF topic, often the first comment is related to the title, potentially offering further insights.

For 'Issue_History,' the continuing issue of columns intersecting with the slab was identified. Including 'TopicStatus' in the LLM's analysis might have offered crucial context to determine whether the issue had been resolved or was still ongoing. Additionally, it was observed that the comments are presented in reverse chronological order. Providing a detailed description of the conversation's progression, complete with information about the authors and the sequence of comments, could be immensely beneficial for the LLM's analysis.

In the ‘Additional_Needed_Resources’ category, while the need for remodelling was identified, the LLM did not deduce the implication of additional working hours. Moreover, the mention of a specific model name lacked contextual relevance, making this information redundant. Refining the prompt or using predefined classification values could lead to more accurate interpretations in such scenarios.

The ‘Actions_Taken’ category correctly reflected remodelling plans, but it was unclear if these actions were realized. The incorporation of ‘TopicStatus’ could have provided clarity on the current status of the issue.

A notable gap in the LLM’s interpretative capabilities was evident in the ‘Final_Resolution’ category, where it failed to recognize “The Columns should run through” as a viable solution.

The use of LangChain for structuring prompts is advantageous, particularly when it comes to the potential automated evaluation of larger datasets using predefined templates. However, these prompts need to be continually refined for optimal performance. There is still a lot of scope for exploration of LangChain’s resources, especially in the Few-shot examples aspect. To identify the examples that best fit the dataset’s characteristics, it will be crucial to create and carefully select these examples with human input.

To conduct a more thorough evaluation of the LLM’s performance in different contexts, it is necessary to have access to BCF files that contain complex information and pre-classified datasets. Such enriched data sources would not only facilitate a deeper analysis but also improve the results. The initial implementation of LLMs for unstructured text classification shows promise, with classifications that are consistent with the desired format. However, to achieve greater precision in these classifications, it is essential to further refine the prompt construction and adjust the LLM’s parameters.

Moreover, the inclusion of additional, detailed data, such as project or site-specific information and insights into the stakeholders involved in issue communication, would significantly advance this research. This extended data scope could provide the LLM with a richer context, leading to more nuanced classifications. Understanding the roles and backgrounds of individuals involved in these communications and integrating this information into the text data could add critical depth to the classification process, yielding a more layered and insightful analysis.

6 Future Work

The next step in the research process should focus on expanding the range of data extraction and refining the classification strategy. In both areas, it is crucial to collaborate with planners and project managers to identify key indicators with practical relevance. This involves implementing both simple algorithms and advanced heuristics to uncover implicit knowledge from the available data.

Developing a robust application capable of consistently processing data, extracting crucial information, and performing accurate classification is a pivotal goal. Central to this development is the optimization of the LLM used in the process. This includes fine-tuning the model and adjusting its parameters to bolster the relevance of the classified data. An integral part of this refinement is the incorporation of a linked data approach, which considers not only textual information but also component details, project insights, and actor perspectives, thus enriching the overall analysis.

Additionally, the enhancement of the classification system by incorporating value ranges into the categories is anticipated to significantly improve the quality and consistency of the output. A crucial element of this enhancement involves introducing a hierarchical structure within the input information format. This, coupled with establishing relationships between different data concepts, will provide additional context to the LLM. In this regard, implementing and extending the current BCF OWL could be a strategic move, leading to a more detailed representation and interconnectedness of data elements, in line with the project's overarching goals.

Adopting the LinkedData approach and integrating rule-based systems or ontologies that contain specialized expert knowledge holds the promise of yielding significant advancements in future research. Further synergies could be explored with emerging methodologies in the field of LLMs, such as Retrieval-Augmented Generation (RAG). The potential to integrate a wide range of information sources, including internal and public databases and various ontologies, using these advanced methods, is vast. Should the development of the application continue, looking into different data storage options and incorporating additional data sources through APIs would prove beneficial. This would not only enhance the richness of the dataset but also extend the application's utility and scope.

References

1. buildingSMART: BCF-XML GitHub. https://github.com/buildingSMART/BCF-XML/tree/release_3_0. Last accessed 20 Jan 2024
2. buildingSMART: Use Case Management. <https://ucm.buildingsmart.org/use-case-details/2518/en#:~:text=proper%20operation%20of%20the%20facility,%20last%20accessed%202023/12/04>. Last accessed 10 Jan 2024
3. buildingSMART: BIM Collaboration Format (BCF). <https://technical.buildingsmart.org/standards/bcf/>. Last accessed 4 Dec 2023
4. Schulz, O., Oraskari, J.T., Beetz, J.: Lessons learned from designing and using bcfOWL. In: 11th Linked Data in Architecture and Construction Workshop, LDAC 2023, Matera, Italy (2023)
5. buildingSMART, BCF-API: GitHub. <https://github.com/buildingSMART/BCF-API>. Last accessed 20 Jan 2024
6. Schulz, O., Oraskari, J., Beetz, J.: bcfOWL: A BIM collaboration ontology. In: Proceedings of the 9th Linked Data in Architecture and Construction Workshop (LDAC2021), Luxembourg (2021)
7. Jeon, K., Lee, G., Yang, S., Jeong, H.D.: Named entity recognition of building construction defect information from text with linguistic noise. Autom. Constr. **143**, 104543 (2022)
8. Wang, S., Sun, X., Wang, G.: GPT-NER: Named entity recognition via large language models (2023)
9. Sun, X., Li, X., Wang, G.: Text classification via large language models (2023)
10. BIMcollab Help Center: https://helpcenter.bimcollab.com/portal/en/kb/articles/example-projects-and-templates#Example_project. Last accessed 15 Jan 2024
11. LangChain: https://python.langchain.com/docs/get_started/introduction. Last accessed 23 Jan 2024



Optimizing Truck Allocation in Open-Pit Mining Using a Deep Reinforcement Learning Policy

Mohsen Hatami^(✉) and Ian Flood

M.E. Rinker, Sr. School of Construction Management, University of Florida, Gainesville,
FL 32611-5703, USA
{mohsen.hatami,flood}@ufl.edu

Abstract. The paper is concerned with the development, optimization and assessment of a deep neural network based policy for dynamic truck-to-process allocation in an open-pit mining system. The study addresses the challenge of the need for real-time control in a highly uncertain environment. The policy was developed using a reinforcement learning (RL) method working in tandem with a discrete-event simulation model of the mining operations. The simulation model was based on the logistical and performance data collected from the Sungun mine in Iran. Key performance indicators, such as truck utilization, waiting time, and production rate, serve as the metrics for evaluating the model's effectiveness. The performance of the RL-based policy was optimized in a series of sensitivity analyses ranging key modeling parameters such as the structure of the neural network, and RL algorithm reward period. The performance of the optimized RL-based policy is compared to that of a common rule-of-thumb policy over a series of lengthy validation runs. The findings clearly show that the RL-based policy delivers substantial improvements over the rule-of-thumb approach, increasing the production rate by an average of 17% per cycle. In addition to technical performance, the paper also explores practical considerations essential for the RL-based policy's successful deployment. These include its scalability, as well as the need for high-quality data and computational resources. A comparative critique of existing optimization methodologies further emphasizes the broader applicability of the RL-based policy approach in managing complex, dynamic optimization challenges. Importantly, the techniques and insights gained are not limited to mining; they have a potentially wide field of application in the construction industry as well. In summary, this research marks a significant step forward in the field of real-time control for both the mining and construction sectors.

Keywords: Artificial Intelligence · Decision Policy · Deep Artificial Neural Network · Discrete-Event Simulation · Open-Pit Mining · Reinforcement Learning · Truck Allocation Optimization

1 Introduction

Open pit mining, a primary technique for extracting minerals and metals from the earth's surface, is employed when resources are near the surface, offering a more feasible and cost-effective alternative to underground mining. Open pit mining operations involve

multiple stages, including overburden removal, drilling, blasting, excavation, hauling, and processing of extracted resources. Trucks play a critical role in these operations, impacting overall productivity and profitability. Optimizing truck allocation is crucial for maximizing open pit mine performance, necessitating a balance between truck utilization and road network maintenance.

This research integrates various strategies, including optimization techniques, simulation methodologies, and the application of reinforcement learning, to enhance the operational efficiency of open pit mines. The primary goal of optimization in this context is to increase output and efficiency while reducing both costs and environmental impact. Common optimization approaches involve the use of linear and integer programming along with various metaheuristic strategies. Nonetheless, these conventional methods might not adequately address the changing aspects of mining operations. For a more detailed analysis and modeling of such complex processes, Discrete Event Simulation (DES) proves essential. This forms the groundwork for the creation and assessment of Deep Reinforcement Learning (DRL) algorithms. Reinforcement Learning (RL), a branch of machine learning, emphasizes the learning process of agents through their interaction with their surroundings. This aspect makes RL particularly suitable for addressing intricate and dynamic optimization challenges, such as the allocation of trucks in open pit mining scenarios.

Truck allocation in open pit mines has traditionally been addressed using methods like dispatching rules, linear programming, and metaheuristic algorithms. However, these approaches may not adequately handle the uncertainties and variations in mining operations. Recent advancements in machine learning and artificial intelligence have opened new avenues for developing intelligent control and optimization algorithms for open pit mining operations. RL algorithms, for example, can learn optimal policies through interaction with the environment, making them suitable for dynamic and complex optimization problems like truck allocation in open pit mines.

RL enables agents to learn optimal actions through interactions with their environment to maximize cumulative rewards. DRL, combining deep learning techniques with RL algorithms, has been successful in various domains, including robotics and autonomous vehicles. In the context of mining and earthmoving operations, RL has been applied for tasks such as mine planning, scheduling, and truck dispatching. These studies demonstrate RL's potential in improving efficiency and effectiveness by learning from historical data and adapting to dynamic environmental changes.

The integration of RL techniques in open pit mining operations has shown promising results in improving efficiency and reducing operational costs. Studies have applied RL to truck dispatching and open pit block scheduling, demonstrating the benefits of RL-based approaches over traditional methods. However, there is still room for further research, particularly in the development and evaluation of DRL algorithms for truck allocation tasks in open pit mines.

1.1 Background and Literature Review

Open-pit mining is a vital process for mineral extraction, involving various stages from overburden removal to resource processing (Hustrulid and Kuchta, 2006). Efficient truck allocation in these operations is crucial for maximizing productivity and profitability

(Asad and Topal, 2012). Traditional methods, like dispatching rules and mathematical programming, are often limited in capturing the dynamic nature of mining operations and struggle with uncertainties and variations in the environment (Topal and Ramazan, 2010; Ramazan and Dagdelen, 2004).

The emergence of Reinforcement Learning (RL) and Deep Reinforcement Learning (DRL) has provided a novel approach to address these complex decision-making scenarios (Sutton and Barto, 2018; Mnih et al., 2015). DRL, which combines deep learning with RL, is particularly adept at handling high-dimensional data (Silver et al., 2016). Similar to advancements in open-pit mining, the construction industry is exploring deep learning AI as an innovative alternative to DES for complex project planning, demonstrating the versatility of these technologies across different fields (Hatami et al., 2022). The application of RL in mining, including tasks like mine planning and truck dispatching, has demonstrated significant improvements in operational efficiency and adaptability (Hatami et al., 2022; Shitole et al., 2019).

Integrating RL into open-pit mining operations has shown promise in enhancing truck allocation efficiency and reducing operational costs. Studies applying RL to truck dispatching and open-pit block scheduling highlight the advantages of RL-based approaches over traditional methods (Chang et al., 2015). However, there remains a gap in the development and evaluation of DRL algorithms specifically for truck allocation in open-pit mines.

This literature review examined the evolution of optimization methods in open-pit mining, with a focus on the transition from traditional approaches to advanced RL and DRL techniques. It underscores the need for more research in this area, particularly in refining and evaluating DRL models for specific tasks like truck allocation.

2 Theoretical Framework

2.1 Formulating Truck Allocation as a Reinforcement Learning Problem

Decision-Making Agents. In the realm of open-pit mining, decision-making processes are essential, particularly in areas like truck allocation and equipment scheduling. The crux of these decisions lies in the use of agents that analyze the state of the mining system to optimize operations. These agents are of two types: search-based and experience-based. Search-based agents delve into solutions systematically but are computationally intensive, thus not ideal for real-time applications. In contrast, experience-based agents rely on past experiences for quicker decision-making but might not offer the most optimized solutions. This study delves into experience-based approaches, specifically focusing on a rule-of-thumb method and a Deep Artificial Neural Network (DANN). While the DANN is experience-based, it uniquely incorporates search techniques from reinforcement learning to develop more effective solutions. The primary goal is to compare the effectiveness of search and experience-based approaches in the long-term control of open-pit mining systems, with a particular focus on truck allocation. The DANN, representing a complex end of the experience-based spectrum, is explored in detail for its potential in enhancing operational efficiency through reinforcement learning, offering a nuanced approach that blends experience with strategic search methodologies (Fig. 1).

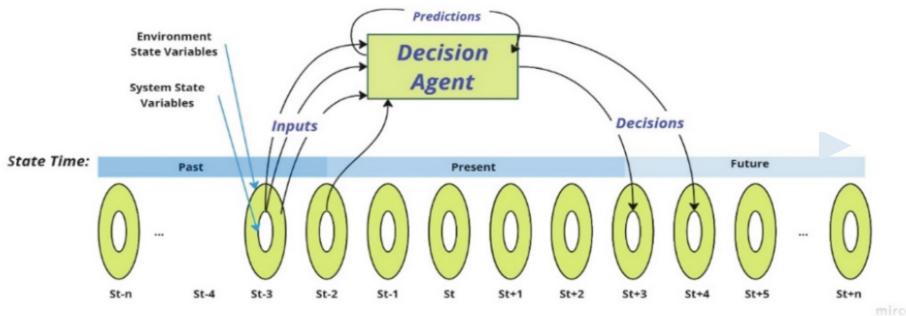


Fig. 1. Dynamic system control by a decision agent (adapted from Flood and Flood (2022)).

DRL Model Development Strategies. Developing a Deep Reinforcement Learning (DRL) model for truck allocation in open-pit mining involves unique strategies, particularly since direct observation or a priori solutions are often not feasible. This challenge necessitates adopting a hindsight strategy, where the agent explores decision paths in a simulated environment, learning effectively through trial and error. This approach for DRL model development is twofold. Firstly, direct adjustments to the model's structure or weights are made, echoing evolutionary methods, and secondly, modifications to the model's output are evaluated for their impact on decision-making, aligning with Reinforcement Learning (RL) methods. The latter, implemented in this study, is instrumental for iteratively refining truck-excavator allocation decisions in a simulation, leading to progressively informed and optimized decision-making. This RL approach balances exploration of diverse allocation strategies and exploitation of gathered knowledge for optimal decision-making. Such a strategy paves the way for advanced, intelligent decision-making in complex systems like open-pit mining. Simulation models serve as crucial platforms for testing various decision policies, providing a cost-effective, risk-free environment to evaluate and optimize their impact on operational efficiency. This simulated experimentation is key to developing and validating the decision policy, ultimately shaping efficient and productive mining operations (Fig. 2).



Fig. 2. Historic track of a real system to be modeled, followed by simulated alternative future tracks of the system (adapted from Flood and Flood (2022)).

3 Methodology for DRL Optimization Truck Allocation in Open-Pit Mining

The methodology employs a Reinforcement Learning (RL) approach, wherein an agent learns to make decisions through trial and error interactions within a simulated environment. RL optimizes decision-making by receiving feedback in the form of rewards or penalties, thereby refining the agent's policy towards achieving the highest cumulative reward. The RL model used in this study is based on a deep neural network structure, designed to process complex and dynamic operational data from open-pit mining operations.

The neural network within the RL framework is constructed to approximate value functions or policies, allowing it to handle high-dimensional data pertinent to mining operations. The reward period, a critical parameter in RL, is carefully calibrated to ensure that the learning agent can effectively balance short-term actions with long-term operational goals, such as maximizing truck utilization and minimizing waiting times.

The study also incorporates a discrete-event simulation (DES) model, developed using the Arena software. The model accurately mirrors the Sungun mine's operational intricacies, including vehicle logistics and loading activities, based on data and details provided by Azadi et al. (2013). The DES model serves as the testbed for the RL agent, providing a realistic and dynamic environment for policy development and evaluation.

Data collection is a crucial step in this methodology, involving the gathering of logistical and performance data from the Sungun mine. This data includes information about truck movements, loading times, route conditions, and production rates. The collected data is meticulously preprocessed and integrated into the DES model, ensuring that the RL agent is trained and evaluated on realistic and relevant operational scenarios.

3.1 Operation Simulator

The operation simulator, inspired by the work on construction manufacturing by Flood and Flood (2022), captures the complexities of truck allocation in open-pit mining at Sungun mine. It involves 40 trucks, nine excavators, various haulage routes, and accounts for variable travel and loading times, as well as truck breakdowns. The simulator, running on a weekly cycle, aims to optimize operational efficiency and costs. The operation simulator uses several variables and parameters as depicted in Tables 1, 2, 3, 4, and 5 (Mohtasham et al., 2017). Their respective probability distributions are also provided (Fig. 3).

3.2 Control Policy Types

The following are the various policies used in this study:

- **Random policy:** A benchmark policy using a uniformly distributed random variate for excavation point selection.
- **Rule-of-thumb policy:** Prioritizes excavation points based on factors like material type and location.
- **DANN policy:** Utilizes RL for optimizing truck allocation based on the system state.

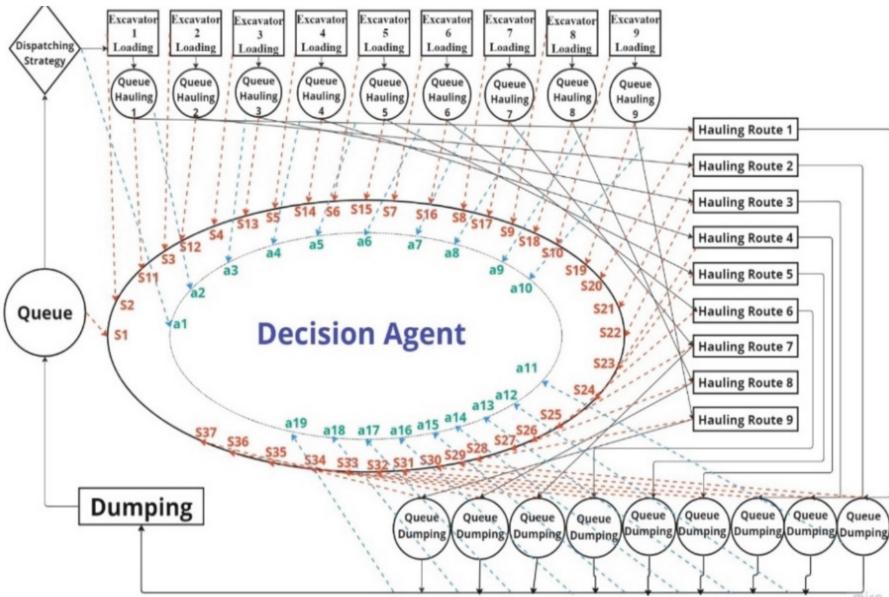


Fig. 3. Schematic model of the open-pit mining system.

Table 1. Operating characteristics of different loading points in target shift. (Data sourced from Mohtasham et al. (2017))

| Loading points (Excavators) | Bench level (m) | Type of material | Loading device model (Type of Excavator) | Maximum production rate (ton/shift) | Average production rate (ton/shift) | Average grade (%) |
|-----------------------------|-----------------|------------------|--|-------------------------------------|-------------------------------------|-------------------|
| 1 | 1912.5 | Ore | Komatsu-600A | 8650 | 7600 | 1.12 |
| 2 | 1950 | Ore | CAT-988B | 9570 | 8500 | 0.74 |
| 3 | 1962.5 | Ore | Komatsu-600A | 8650 | 7600 | 0.92 |
| 4 | 2100 | Ore | Komatsu PC-800 | 8250 | 7300 | 0.39 |
| 5 | 1937.5 | Ore | Komatsu-600A | 8650 | 7600 | 0.34 |
| 6 | 2237.5 | Ore | Komatsu PC-800 | 8650 | 7600 | 0.85 |
| 7 | 2262.5 | Ore | CAT-988B | 9570 | 8500 | 1.03 |
| 8 | 2287.5 | Ore | Komatsu-600A | 8650 | 7600 | 0.74 |
| 9 | 2312.5 | Ore | Komatsu PC-800 | 8250 | 7300 | 0.57 |

Table 2. Loading time distributions of trucks in loading points (s); (Triangular distribution)

| Loading points | Minimum loading time of 30 tons trucks (s) | Mode of loading time of 30 tons trucks (s) | Maximum loading time of 30 tons trucks (s) |
|----------------|--|--|--|
| 1 | 108 | 122 | 131 |
| 2 | 117 | 135 | 151 |
| 3 | 108 | 122 | 131 |
| 4 | 95 | 109 | 123 |
| 5 | 108 | 122 | 131 |
| 6 | 95 | 109 | 123 |
| 7 | 117 | 135 | 151 |
| 8 | 108 | 122 | 131 |
| 9 | 95 | 109 | 123 |

Table 3. Hauling time distributions of trucks in loading points (s); (Triangular distribution)

| Loading points | Minimum loading time of 30 tons trucks (s) | Mode of loading time of 30 tons trucks (s) | Maximum loading time of 30 tons trucks (s) |
|----------------|--|--|--|
| 1 | 385 | 409 | 431 |
| 2 | 340 | 362 | 385 |
| 3 | 355 | 376 | 393 |
| 4 | 351 | 373 | 388 |
| 5 | 364 | 371 | 379 |
| 6 | 390 | 396 | 399 |
| 7 | 389 | 397 | 409 |
| 8 | 341 | 373 | 391 |
| 9 | 359 | 371 | 382 |

These policies are analyzed through both RL and Arena simulation models to evaluate the DANN policy's effectiveness. The comparison aims to demonstrate the DANN policy's potential in enhancing truck allocation efficiency in mining (Fig. 4).

DRL Model Development Strategies. The development of the DRL model in open-pit mining involves unique strategies due to the lack of direct observational data. A key method used is the hindsight strategy, where the agent explores various decision paths in a simulated environment, learning through trial and error. The model adopts two approaches: evolutionary adjustments to the model's structure and weights, and supervised feedback adjustments to the model's output. This RL method allows the DRL model to iteratively improve truck-excavator allocation decisions in a simulated mining environment, striking a balance between exploration and exploitation. This strategy is

Table 4. Returning time distributions of trucks in loading points (s); (Triangular distribution)

| Loading points | Minimum loading time of 30 tons trucks (s) | Mode of loading time of 30 tons trucks (s) | Maximum loading time of 30 tons trucks (s) |
|----------------|--|--|--|
| 1 | 325 | 379 | 403 |
| 2 | 297 | 315 | 336 |
| 3 | 310 | 339 | 367 |
| 4 | 313 | 344 | 371 |
| 5 | 324 | 348 | 369 |
| 6 | 340 | 355 | 372 |
| 7 | 338 | 357 | 366 |
| 8 | 299 | 309 | 323 |
| 9 | 339 | 357 | 363 |

Table 5. Dumping time distributions of trucks in dump points in target shift (s); (Triangular distribution)

| Dump points | Minimum dumping time of 30 tons trucks (s) | Mode of dumping time of 30 tons trucks (s) | Maximum dumping time of 30 tons trucks (s) |
|-------------|--|--|--|
| Dump 1950 | 62 | 68 | 75 |

crucial for the DRL model to effectively learn and optimize truck allocation, promising enhanced operational efficiency and productivity in open-pit mining (Fig. 5).

4 Model Optimization and Evaluation

4.1 Sensitivity Analyses for Optimizing the RL-Based Policy

To optimize the RL-based policy for truck allocation, extensive sensitivity analyses were conducted. These analyses primarily focused on adjusting key modeling parameters to evaluate their impact on the overall performance of the system. Variables such as truck load capacity, excavator loading speed, and route travel times were varied systematically in the simulation model. By altering these parameters by $\pm 10\%$ and $\pm 20\%$ from their baseline values, the robustness and adaptability of the RL-based policy were assessed under varying operational conditions. The sensitivity analyses provided critical insights into the model's behavior in response to changes in operational parameters, thereby guiding the optimization of the RL policy for better performance in real-world scenarios (Fig. 6).

Key Modeling Parameters Adjusted. Key modeling parameters in the simulation included truck travel times between excavation and dumping sites, excavator cycle times for loading trucks, truck loading and unloading times, and queuing systems at the sites.

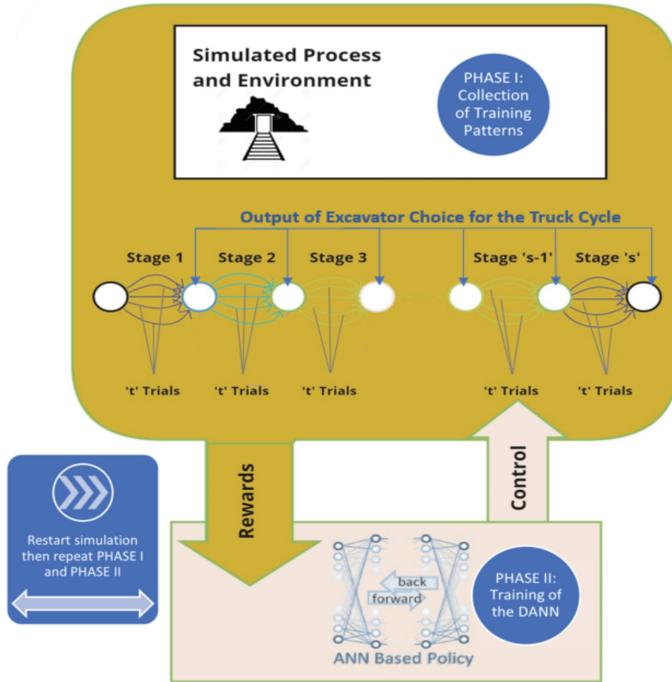


Fig. 4. Two phase reinforcement learning DANN development cycle (adapted from Flood and Flood (2022)).

| Inputs | Description |
|-------------------|--|
| Truck State | Current Process (PL, PH, PD, PR), Remaining Time, Associated Excavator |
| Excavator State | Number of Trucks Loading/Waiting |
| Process Durations | Duration Parameters (min, mode, max) for each process (PL, PH, PD, PR) |
| System Parameters | Number of Trucks, Excavators, Routes |

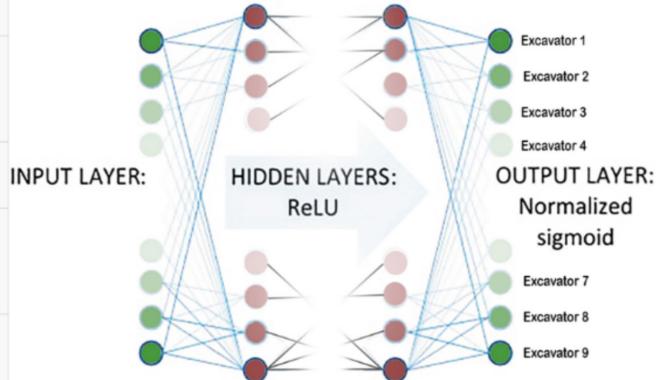


Fig. 5. Structure of the DANN.

The RL algorithm's learning rate, discount factor, and reward structure were also meticulously fine-tuned to ensure effective learning and decision-making by the RL agent. This fine-tuning was crucial in aligning the RL policy with the operational objectives of maximizing productivity while minimizing resource idle time.

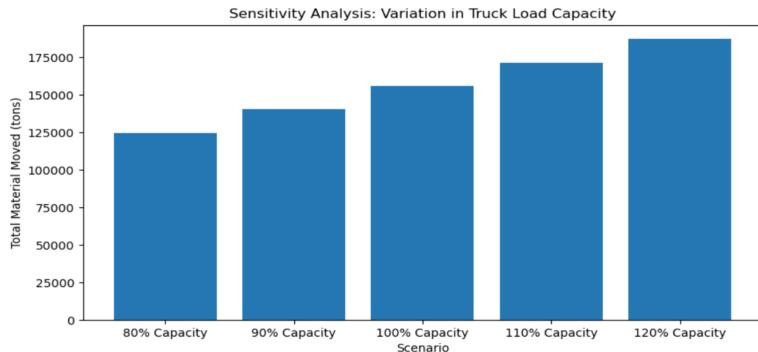


Fig. 6. Sensitivity Analysis by Variation of Truck Load Capacity.

Metrics for Evaluating Model Effectiveness. The effectiveness of the RL-based model was evaluated using several key performance indicators (KPIs), including truck utilization, average truck waiting time, and production rate. These metrics were vital in quantifying the improvements achieved by the RL-based policy over traditional methods. Truck utilization was measured by the proportion of time trucks were actively involved in operations, while waiting time was assessed by the average duration trucks spent idle. The production rate was evaluated based on the total material moved per unit of time. This comprehensive approach to performance evaluation ensured a holistic assessment of the RL-based policy's effectiveness in optimizing truck allocation in open-pit mining operations (Table 6).

Table 6. Detailed time analysis for each excavator during the simulation run, encompassing loading, hauling, dumping, returning, and the calculated total time taken (in seconds).

| Simulation Run | Excavator | Loading Time (sec) | Hauling Time (sec) | Dumping Time (sec) | Return Time (sec) | Total Time (sec) |
|----------------|-------------|--------------------|--------------------|--------------------|-------------------|------------------|
| 1 | Excavator 1 | 122 | 409 | 68 | 379 | 978 |
| 1 | Excavator 2 | 135 | 362 | 68 | 315 | 880 |
| 1 | Excavator 3 | 122 | 376 | 68 | 339 | 905 |
| 1 | Excavator 4 | 109 | 373 | 68 | 344 | 894 |
| 1 | Excavator 5 | 122 | 371 | 68 | 348 | 909 |
| 1 | Excavator 6 | 109 | 396 | 68 | 355 | 928 |
| 1 | Excavator 7 | 135 | 397 | 68 | 357 | 957 |
| 1 | Excavator 8 | 122 | 373 | 68 | 309 | 872 |
| 1 | Excavator 9 | 109 | 371 | 68 | 357 | 905 |
| 2 | Excavator 1 | 108 | 385 | 62 | 325 | 880 |
| ... | ... | ... | ... | ... | ... | ... |

4.2 Evaluation of the Optimized DANN Model

In this section, the effectiveness of the Deep Artificial Neural Network (DANN) strategy integrated into the Deep Reinforcement Learning (DRL) model for open-pit mining, like Sungun mine, is assessed. The DANN model excels in optimizing truck allocation, balancing efficiency and productivity. Its superior performance is evident when compared with traditional heuristic methods, showcasing significant advancements in mining operations. Key metrics such as Learning Rate, Mini-Batch Size, Reward Length, Number of Hidden Layers, and Neurons per Layer are used to benchmark the DANN model against a Random Policy. The DANN model consistently demonstrates superior performance across these metrics. Additionally, the ‘Rolling Arithmetic Mean Difference’ metric is employed to gauge the continuous improvement of the DANN model. Figures 7 and 8 provide a comparative analysis of the DANN model’s performance over 6000 truck cycles, illustrating substantial improvements over the random policy model.

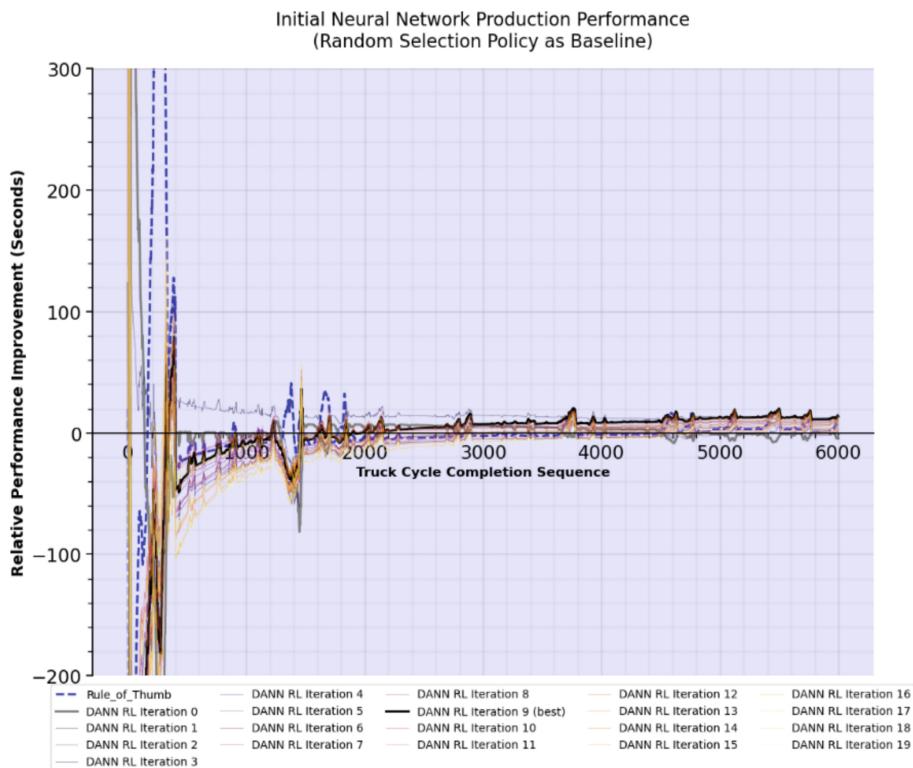


Fig. 7. Performance of the DANN Production in the Initial DRL Model over 6000 Truck Cycles.

5 Results and Discussion

5.1 Comparison of RL-Based Policy to Rule-Of-Thumb Policy

The results from the implementation of the RL-based policy were compared to a traditional rule-of-thumb policy commonly used in open-pit mining operations. The RL-based policy demonstrated a marked improvement in various operational aspects. Notably, the RL policy resulted in an average increase of 17% in the production rate per cycle compared to the rule-of-thumb approach (Fig. 8). This substantial increase highlights the efficacy of the RL-based policy in adapting and responding to the dynamic mining environment more effectively than traditional methods.

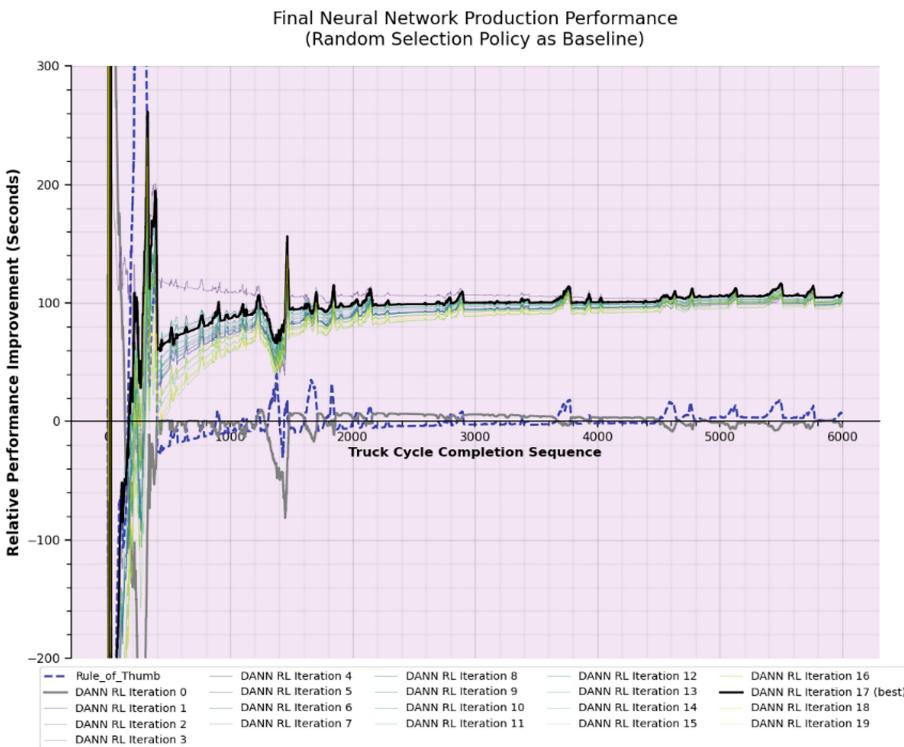


Fig. 8. Performance of the DANN Production in the Final DRL Model over 6000 Truck Cycles.

Improvements in Truck Allocation Efficiency and Production Rate. The RL-based policy's optimization led to significant improvements in truck allocation efficiency. The policy reduced average waiting times and increased truck utilization, ensuring that resources were employed more productively. These improvements are evident in the reduced cycle times and increased total material moved, as indicated by the simulation results (Table 6). The enhanced truck utilization directly contributed to the observed increase in the overall production rate, underlining the potential of the RL approach in optimizing operational efficiency.

Scalability and Data Requirements. The scalability of the RL-based policy is a critical factor for its application in diverse operational scales and environments. The policy's adaptability to different mining operation sizes and its performance under varying conditions indicate its scalability. However, the need for high-quality data and substantial computational resources is a consideration for its deployment. The accuracy and effectiveness of the RL-based policy are heavily dependent on the quality and comprehensiveness of the input data, necessitating reliable data collection and processing systems.

Evaluating the RL-Based Strategy Against Conventional Optimization Techniques. In comparison to traditional optimization methods like linear programming and heuristic strategies, the RL-based approach presents a more flexible and responsive solution. Conventional techniques typically struggle to manage the uncertainties and intricacies associated with open-pit mining activities. On the other hand, the RL-based policy is designed to perpetually learn and adjust to evolving operational scenarios, facilitating more informed and effective decision-making processes. This ability to adapt, combined with its proficiency in navigating complicated decision environments, clearly distinguishes the RL-based approach as an advanced alternative for optimizing truck allocation in open-pit mining contexts.

6 Broader Implications and Future Work

6.1 Applicability Beyond Mining

The success of the RL-based policy in optimizing truck allocation in open-pit mining operations opens the door for its application in other sectors. Particularly, the construction industry, with its similarities to mining in terms of logistical challenges and the need for efficient resource allocation, stands to benefit significantly from this approach. The RL-based policy can potentially optimize equipment allocation, scheduling, and logistical workflows in construction projects, thereby enhancing productivity and reducing operational costs.

Impact on Real-Time Control in Various Industries. The insights gained from this research extend to the broader field of real-time control in various industries. The ability of the RL-based policy to continuously learn and adapt to dynamic environments makes it an invaluable tool for industries where operational conditions are constantly changing. Industries such as manufacturing, transportation, and logistics could implement RL-based systems for real-time decision-making, leading to improvements in efficiency, responsiveness, and overall system performance.

Future Research and Development. Looking forward, several areas warrant further research and development in RL-based optimization methods. One key area is the exploration of hybrid models that combine RL with other machine learning or optimization techniques to tackle more complex and multi-faceted decision-making scenarios. Another area of interest is the enhancement of RL algorithms to better handle uncertainty and variability in real-world environments, which could significantly increase their applicability and robustness.

Moreover, future research could focus on the scalability of RL-based systems, especially in terms of their applicability to large-scale operations and their integration with

existing IT infrastructures. Finally, ongoing advancements in computational power and data processing technologies will likely open new avenues for the development of more sophisticated and capable RL-based optimization systems.

7 Conclusion

This research presents a significant advancement in the field of open-pit mining through the development and implementation of a Deep Reinforcement Learning (DRL)-based policy for optimizing truck allocation. The key findings from this study demonstrate that the DRL-based policy significantly outperforms traditional rule-of-thumb methods, particularly in terms of production rate, truck utilization, and waiting time reduction. Specifically, the RL-based policy achieved an average of 17% increase in production rate per cycle, showcasing its efficacy in real-time operational decision-making.

The shift towards Deep Reinforcement Learning (DRL) marks a significant departure from traditional optimization methods, catering to the pressing need for solutions that are both dynamic and adaptable in the face of the uncertainty and complexity prevalent in operational environments. This study demonstrates the effectiveness of coupling DRL with discrete-event simulation to create a comprehensive framework that accurately reflects and enhances the processes of real-world mining operations. Such an approach not only boosts efficiency but also sheds light on the complex interplay inherent in open-pit mining activities.

Moreover, the impact of this research is not confined to the mining sector alone. The techniques and insights derived from this work have broad applicability across a range of sectors that benefit from real-time control and optimization, including but not limited to construction, logistics, and manufacturing. The inherent capacity of DRL to continuously assimilate and adjust to new information renders it an invaluable asset across numerous fields where both efficiency and flexibility are of paramount importance.

In summary, this research contributes significantly to the field of real-time control and optimization, offering a novel and effective approach to managing complex operational challenges. The success of the DRL-based policy in open-pit mining serves as a testament to the potential of advanced machine learning techniques in revolutionizing traditional operational methodologies across various sectors.

References

- Asad, M.W.A., Topal, E.: An approach to improve truck-shovel productivity in open-pit mines. *Min. Technol.* **121**(3), 171–179 (2012)
- Azadi, N., Monjezi, M., Ataaeiipour, M.: Application of Arena software to optimize Sungon copper mine transport fleet. Presented at the International Mining Conference, Tehran, Olympic Hotel (2013, February)
- Chang, Y., Ren, H., Wang, S.: Modelling and optimizing an open-pit truck scheduling problem. *Discrete Dyn. Nat. Soci.* Article ID 745378 (2015). [Online] Available: <https://doi.org/10.1155/2015/745378>
- Flood, I.: A neural network-based production control system for construction. *J. Constr. Eng. Manag.* **115**(1), 75–94 (1989)

- Flood, I., Flood, P.: Intelligent Control of Construction Manufacturing Processes Using Deep Reinforcement Learning. In: Proceedings of the 12th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2022), pp. 112–122. Lisbon, Portugal (2022)
- Hatami, M., Franz, B., Paneru, S., Flood, I.: Using deep learning artificial intelligence to improve foresight method in the optimization of planning and scheduling of construction processes. *Comput. Civ. Eng.* **2021**, 1171–1178 (2022)
- Hatami, M., Paneru, S., Flood, I.: Applicability of artificial intelligence (AI) methods to construction manufacturing: a literature review. *Constr. Res. Congr.* **2022**, 1298–1306 (2022)
- Hustrulid, W., Kuchta, M.: Open pit mine planning and design, 3rd edn. CRC Press, Boca Raton (2006)
- Mnih, V., et al.: Human-level control through deep reinforcement learning. *Nature* **518**(7540), 529–533 (2015). <https://doi.org/10.1038/nature14236>
- Mohtasham, M., Mirzaei Nasirabad, H., Mahmoodi Markid, A.: Development of a goal programming model for optimization of truck allocation in open pit mines. *J. Min. Environ.* **8**(3), 359–371 (2017). <https://doi.org/10.22044/jme.2017.859>
- Ramazan, S., Dagdelen, K.: Production scheduling in open pit mining: a review. *Min. Technol.* **113**(3), 127–138 (2004)
- Shafiee, S., Nehring, M., Topal, E.: An econometric model of mine production performance: a case study of the Sungun copper mine. *Resour. Policy* **36**(3), 211–219 (2011)
- Shitole, V., Louis, J., Tadepalli, P.: Optimizing earth moving operations via reinforcement learning. In: Proceedings of the 2019 Winter Simulation Conference (WSC 2019). Oregon State University, Corvallis, OR, USA (2019)
- Silver, D., et al.: Mastering the game of Go with deep neural networks and tree search. *Nature* **529**(7587), 484–489 (2016)
- Sutton, R.S., Barto, A.G.: Reinforcement learning: an introduction. MIT Press, Cambridge (2018)
- Topal, E., Ramazan, S.: A new MIP model for mine equipment scheduling by minimizing maintenance cost. *Eur. J. Oper. Res.* **207**(2), 1065–1071 (2010)



An Image-Based Approach for Construction Site Monitoring and Documentation Using Machine Learning

Matthias W. Glunz, Florian Steinbach, Lina Wedekind,
Martina Mellenthin Filardo[✉], and Jürgen Melzner

Chair for Construction Engineering and Management, Bauhaus-Universität Weimar, Weimar,
Germany

martina.mellenthin.filardo@uni-weimar.de

Abstract. Given current shortages of skilled labour in the construction industry, this paper presents a study on the feasibility and application of an image-based, automated approach for construction site monitoring and documentation using machine learning methods. The study concentrates on object detection based on images of a specific construction site, taken multiple times a day periodically over the course of a year, that have been evaluated using the YOLOv8 technology, thus enabling progress monitoring for selected elements. Training and validation data have been created from annotated images for the object detection, which was accompanied by an evaluation of the chosen hardware and the observation viewpoint for future reference in the data acquisition. Further, a ground truth for the construction progress was generated manually to allow comparison with the results achieved by the machine learning approach. This study demonstrated, that the expected results were achieved without the need for writing a single line of code, which is meaningful given the aforementioned labour shortages in the construction industry and highlights the fast-paced nature of the machine learning field.

Keywords: Construction Site Progress Monitoring · Machine Learning · YOLOv8 · Object Detection

1 Introduction

Image-based monitoring of construction sites using machine learning technology can be a feasible approach to assess different stages of construction. If implemented thoroughly, it can address labour shortages by making the process more transparent and therefore more robust as well as supporting decision making and facilitating documentation. To ensure this application of machine learning, two main factors must be examined: Trust in the results and the threshold level for implementation. With these goals in mind, an experiment was undertaken to confirm results from a Convolutional Neural Network (CNN). For this, a data set (described in 2.3) consisting of recurring pictures of a construction site over the period of a year was used to train a CNN (described in 2.2) as

well as 3). To enable an assessment of the object detection results achieved by the CNN, a so-called ground truth of the construction process was manually generated. Through the comparison of the CNN results of the object recognition with the human-generated construction process, the construction of specific parts (columns) on the construction site illustrates the CNN results and where they deviate from the human-generated ones.

The other aspect of a broad application of machine learning in the field of construction monitoring is addressed by the tools applied to this experiment: The criteria for choosing the CNN as well as all adjacent tools was a strict ‘no coding’ policy, given that traditional construction personnel has no programming experience.

2 Background

In the Background section, the key background information for Object Detection is explained (2.1), and the technologies chosen in the experiment are outlined (2.2). Further, the used data is described in Sect. 2.3.

2.1 Object Detection

Before the detection of the objects can begin, the object itself needs to be defined. In this experiment, the use case of target-actual comparison was chosen to demonstrate the efficiency of image-based monitoring using machine learning technology to analyse the state of construction. To exemplify the process analysis, the columns in their different construction stages were selected. The process was classified into four states: ‘connecting reinforcement installed’, ‘reinforcement cage installed’, ‘formwork elements installed’, and ‘completion of the column’, as described in Table 1.

For image-based monitoring, a special kind of artificial neural network (ANN) was used. ANNs are computer-based systems inspired by biological nervous systems, consisting of many connected artificial neurons. The basic idea behind ANNs is that they receive information, process the given information, and forward the processed information to other neurons [1]. While in ANNs, the neurons are all connected to each other, the neurons in CNNs are partly connected to the other layers in the system. This local connection enables the CNN to capture spatial information more effectively and hierarchically learn features in images [1]. In contrast to conventional ANNs, the neurons in the layers of a CNN are organized three-dimensionally. These dimensions encompass the spatial dimensions of height and width, as well as depth. This promotes the effectiveness of pattern recognition in images. In this context, depth does not refer to the total number of layers in the network but rather to the third dimension of an activation volume. Each element in the depth of the activation volume represents a specific feature map or filter specialized in a particular feature pattern [2].

Usually, the basic structure of a CNN comprises three layers. In the input layer, the data to be processed is input. In the hidden layers, also called intermediate layers, decisions are made, influencing the final result. The output layer provides the ultimate output of the network [3].

There are different learning paradigms for neural networks, including supervised and unsupervised learning. In unsupervised learning, no labels are used, and the network

attempts to recognize patterns and structures in the input data by minimizing or maximizing a cost function. In supervised learning, the network is provided with pre-labelled input data along with corresponding goals or labels. Based on this, the network tries to learn a function that maps the input data to the correct targets. For supervised learning, a sufficiently large training dataset must be provided to the network. The size depends on how complex and variable the objects to be recognized are. In this experiment, the supervised learning paradigm was utilised [3].

For annotation, simple bounding boxes were used to define the position and extent of the objects to be detected. In addition to the bounding boxes, there are more detailed annotation options, such as polylines, which were excluded due to their time intensity. The bounding boxes were placed as accurately as possible, with little margin, to enable the detection to be as precise as possible. All annotated images come from the same pool, which was later used as the basis for training the YOLO model.

2.2 Chosen Technology

The CNN system used in this experiment is YOLO (You Only Look Once) version 8 (YOLOv8) [4]. It was selected because of its capabilities to detect objects in images or videos in real-time. To implement and conduct object detection, the manufacturer's platform was utilized, where various network sizes, such as 'nano (N)', 'medium (M)', and 'small (S)', were provided. Larger networks were not tested in this experiment. These network sizes indicate the number of hidden layers present in a CNN. The number of hidden layers can vary and depends on the complexity of the task [3]. To train the YOLOv8 model, the Google Colab platform was utilized (described in 3.2).

For the annotations, CVAT (Computer Vision Annotation Tool) was used to annotate and label data for the subsequent supervised learning step [5]. CVAT offers the advantage of being freely available and supports a variety of annotation types, including bounding boxes for object localization. Due to the absence of specific files (yaml ending) in the import from the annotation tool CVAT to the object recognition platform YOLO, information about the data's location, the number of classes, and their labels were missing (configuration data). As a result, this information had to be added manually to ensure that the model could be trained correctly.

2.3 Used Data

The object under consideration is a construction site of a multistorey building in a German city. For the following analysis, the focus was placed on the basement and the construction of reinforced concrete components (more precisely on the column construction).

The data used in this experiment comprises a storage capacity of approximately 75.7 GB, which corresponds to around 35,000 images. The storage volume includes images in the order of 3–3.1 MB per daytime image and around 1 MB for night images. The construction site was subdivided into two areas both for the ground truth generation and the YOLO results, as a division in the actual construction processes between these two areas (the back area represents area 1, the front area 2) can be recognised over time.

The choice of analysing the column production process is based on the visibility of most of the components from the camera position. Compared to the walls, for example, the step of the formwork elements can be identified more clearly. In addition, the analysed area was also limited to area 1 to avoid any distorted results. The images were captured and stored at 15-min intervals over the course of a year. Through appropriate data management, the required storage size can be further reduced accordingly (e.g., elimination of night images). The images document the progress of the construction site and were taken from a position next to the excavation pit. Figure 1 shows an example of the analysed construction site.



Fig. 1. Example image of the area under consideration, already including bounding boxes.

3 Implementation

In the implementation step, the states to be analysed are defined at the beginning. Comparative data, which is subsequently used to categorise the results, is recorded and documented using an excel file (3.1). In the next step, the existing data is integrated into the machine learning platform. The model was trained with the labelled training set accordingly in order to be deployed on the other images (3.2). Finally, the results generated by YOLOv8 are presented in 3.3.

3.1 Generating a Ground Truth

To analyse construction processes on a specific construction site, a shared understanding of the undergone (sub-) processes is required: ‘A process is understood to be the totality of interacting processes within a system’ [6].

Part of the control and optimisation of processes is the analysis and documentation of used resources (material, energy, and information) as well as the monitoring of progress within the process. The aim is to optimise the provision and use of available resources and to complete the process or component on time [6]. Machine learning can be a supporting tool in the control and optimisation of these processes.

For the documentation and analysis of column production, the following activities and states were identified with the help of existing images and the technological knowledge about the production of reinforced concrete work, as shown in Table 1. These were used to define the sub-processes relevant to the production of columns. Due to the time intervals between individual images (15 min), the activities (A1-A4) were not considered, and the focus was primarily placed on the construction stages (S1-S4). The choice of construction stages can also be explained by the fact that a hard-output-oriented description [7] can be used for the subsequent analysis with the support of machine learning.

Table 1. Process steps and construction stages used for the recognized object ‘column’.

| No. | Description |
|-----|---|
| A1 | Installing the connecting reinforcement |
| S1 | Connecting reinforcement installed |
| A2 | Production of the reinforcement cage and connection to the corresponding connecting reinforcement |
| S2 | Reinforcement cage installed |
| A3 | Formwork elements are installed around the reinforcement cages |
| S3 | Formwork elements installed |
| A4 | Concreting, formwork removal and curing of the column |
| S4 | Completion of the column |

A = activity, S = state of construction

Ground truth data in the form of a detailed schedule of the construction processes and sub-processes was used to categorise the results achieved by the machine learning approach. The ground truth data, consisting of a detailed, daily time schedule of the identified construction works (columns) within the considered area, was created manually. Figure 2 shows the change in columns over time. In the illustration, the emphasis is on the number of columns in the corresponding construction stages.

3.2 Machine Learning Application

The YOLOv8 model was trained on the Google Colab platform, given that functions from a powerful graphics processing unit (GPU) were necessary for the recognition. The required GPU memory, usually around 12 GB, depends on the complexity of the training data and the desired model accuracy. By using the chosen platform, the training

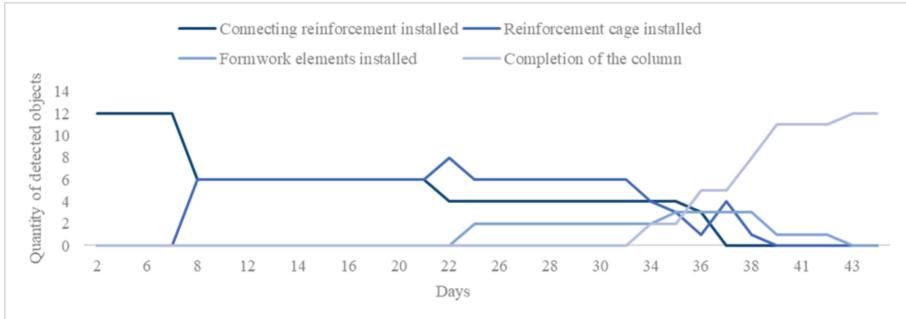


Fig. 2. States of construction for the recognized object ‘column’.

process was performed in a cloud infrastructure that offers scalability and resource efficiency. The model size chosen for this study was designated as ‘S’ (small) and a careful testing and optimisation process followed to find a balance between performance and resource consumption. Iterative training and testing procedures were carried out, with performance evaluation performed on a selected set of images.

In the context of the ‘S’ size model, clear trends emerged in object recognition. Excessive identification of connecting reinforcements was observed, while the detection of closed columns was found to be insufficient. The ‘reinforcement cage installed’ (S2) and ‘formwork elements installed’ (S3) classes showed comparable recognition quality, with an average maximum probability of 91.6% for object recognition.

The evaluation of the different models showed that the model of size ‘S’ outperformed the others and showed a commendable balance between accuracy and resource utilisation. To further improve the performance of the model, various tuning options were explored. In particular, larger installed models were prone to overfitting, which affected their ability to generalise features and apply them to unknown data. To reduce overfitting, the complexity of the ANN had to be reduced [1].

Another key setting parameter was confidence, which indicates the reliability of the model in recognising objects. Empirical tests and comparative analyses showed that a confidence level of 12.5% results in optimal recognition performance for number plates [8]. The Intersection over Union (IoU) setting, which indicates the correspondence between the recognised bounding box and the manually defined bounding box, was left at 25% in this study and enabled a robust evaluation of the object recognition [9].

3.3 YOLOv8 Results

The analysed domain parallels human analysis, with images retained in their unaltered state. Consequently, no image processing, such as adjustments to brightness or contrast, was executed. The relevance of natural illumination guided image selection, precluding the inclusion of images with pronounced contrast or shadowed regions encroaching upon the analysis area, given that no construction work takes place in the dark. Recognition quality was deemed susceptible to shadows or nocturnal images, although such factors did not result in complete failure or non-detection of objects. To maintain standardisation, images captured within a specific time window (04:30 a.m. to 07:00 p.m., as per

metadata) were chosen. Furthermore, image subsections were not provided, eliminating predefined boundaries for object location in subsequent analyses, thereby mitigating potential distortions. Notably, explicit definitions are warranted for elements like formwork at the construction site periphery, detailing the extent to which these should factor into the evaluation.

Aligning with human evaluation standards, identical classes were employed. Emphasis was placed on achieving a hard-output-oriented description and measurability to ensure clarity and unambiguous results. Figure 3 shows the results of the object recognition within the employed platform illustratively.

While objects were theoretically recognized, variations in recognition reliability were discerned among different classes. The expectation for classes such as ‘connection reinforcement installed’ (S2) and ‘completion of the column’ (S4) to exhibit maximum column count was not uniformly met. Conversely, classes like ‘formwork elements installed’ (S3) and ‘completion of the column’ (S4) demonstrated heightened certainty in correct recognition.



Fig. 3. Output from the Ultralytics-platform with results.

In summary, the image evaluation process was semi-automated, requiring methodological decisions. Despite encountering limitations and challenges, a foundational object recognition capability was attained. Recognition reliability, however, exhibited variability across distinct classes. The outcomes were notably influenced by perspective and material utilisation. The insights derived from this analysis serve as a groundwork for subsequent investigations and the refinement of future evaluation methodologies.

4 Comparison of YOLOv8 Results with Ground Truth

The results from the comparison between the ground truth data and the ‘generated’ results using machine learning show that all four classes/construction stages (S1-S4) were not fully recognised (see Fig. 4). As part of the study, 299 objects were correctly

recognised with machine learning. A total of 91 objects were recognised incorrectly or not at all.

A further 87 objects were under-recognised in the experiment in relation to the comparison data. This indicates that machine learning has potential for improvement.

These results suggest that the support provided by machine learning is not yet optimal and that corresponding potential for improvement needs to be tested to find the precise necessary adjustments.

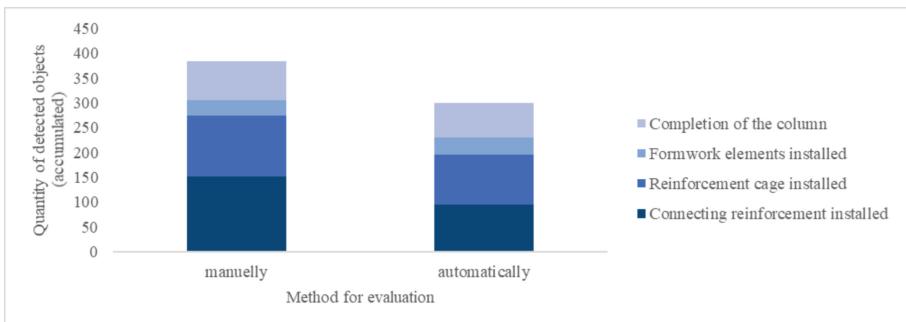


Fig. 4. Difference in the number of recognized classes.

The ‘formwork elements installed’ (S3) and ‘completion of the column’ (S4) classes are nonetheless used to analyse and document construction progress. It was found that more columns were recognised (13 in total) than were planned according to the original design (ground truth data: 12 columns). One possible cause can be attributed to e.g., inconsistent planning data. In this case, more columns were produced on the construction site than the original planning had anticipated. This particular column was therefore not included in the manual analysis. Nevertheless, a clear tendency can be recognised from the available data, as the graphs of the detected columns are close together or overlap. It would therefore be conceivable to report on this indicator. Despite the described differences, clear trends could be recognised through this experiment. This can be argued by the overlapping or slight deviations (or something similar).

The highest agreement between human evaluation and machine learning was found in the ‘formwork elements installed’ (S3) class (see Fig. 5). Only three misclassifications were recognised in this class over the entire observation period. In addition, all formwork elements used (also in different types) were recognised, which can be seen as a positive result. Figure 5, which shows both the manually (ground truth) and automatically (YOLOv8 results) recognised formwork objects, shows that no formwork could be identified at day 25, even though it is clear from the ground truth, that formworks had already been built on that date. One possible cause could be, i.e., that formwork elements in the rear area could be concealed by occlusion from existing reinforcement cages. As a result, the columns and formwork in the background would not be recognised. Figure 6 shows the same graph for columns.

The ‘connecting reinforcement installed’ (S1) class showed the least agreement between human evaluation and machine learning. Particularly at the beginning of the



Fig. 5. Recognized formworks per day.

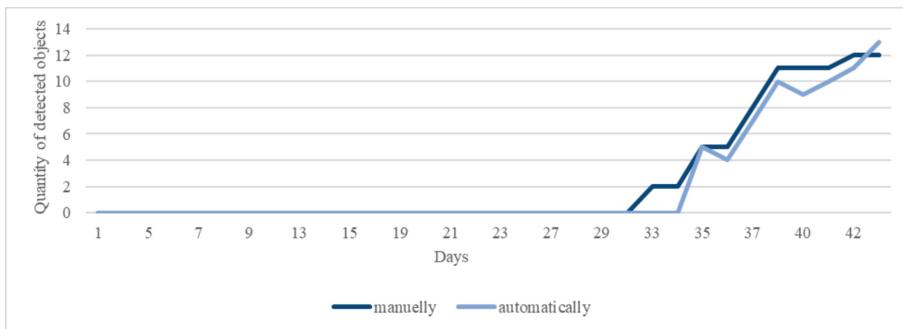


Fig. 6. Completed columns per day.

evaluation, significant discrepancies to the actual value were found, with a total of nine unrecognised connection reinforcements. The achieved results can be deemed unsuited for the aimed use case of construction progress monitoring. As shown in Fig. 7, the only common result between ground truth (manually graph) and the YOLOv8 results (automatically graph) is the recognition on day 37, from which on there are no more connecting reinforcements on the construction site. Recognising the connecting reinforcement understandably poses a particular challenge, as they only protrude a few centimetres from the floor slab and can easily be concealed by other objects. In addition, their colouring is very similar to the underlying floor slab. The increase in recognised connecting reinforcements towards the end of the evaluation could not be logically explained and could be due to incorrect recognition by other objects, such as reinforcements for walls.

5 Conclusion and Outlook

The experiment investigated the extent to which machine learning, in this particular case object detection results from the YOLOv8 technology, can be employed for construction site progress monitoring, mainly through detection of individual construction stages of defined components.

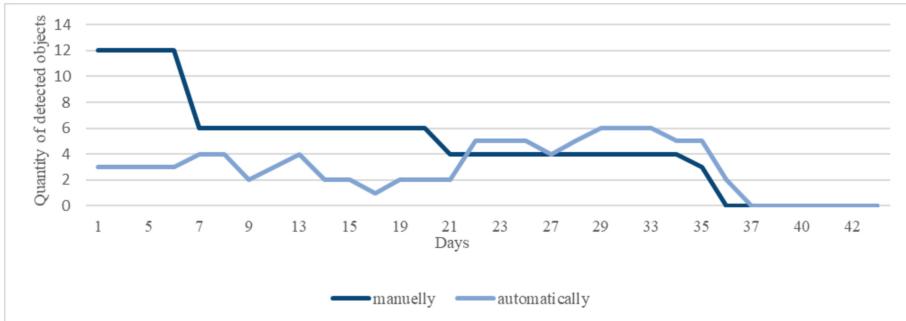


Fig. 7. Number of recognized connection reinforcements per day.

In general, it was established that object detection on the chosen data set can be achieved and the respective construction stages could be identified using the example of the column construction process. Differences were identified between the detection of individual construction stages (in YOLOv8: classes). As considered in more detail in Sect. 4, the formwork class showed the greatest match with the comparison data (ground truth data). On the other hand, there were differences between the ground truth data and the results from YOLOv8 in the construction stages ‘connecting reinforcement’ and ‘reinforcement’.

These findings demonstrate the potential of object detection with the support of machine learning for construction site progress monitoring. Moreover, it illustrates not only that, but also how it can be achieved without any coding, thus dismantling a presumed threshold implementation barrier.

For the future practical application and implementation in daily tools of construction project management or construction supervision, it is necessary to automatically analyse a minimum number of images. The increased use will result in corresponding time savings as the number of analysed images increases. With a small number of images, manual evaluation is more efficient, as the initial training and definition of the classes is correspondingly more time-consuming (‘initial investment’). On a side note, the conducted experiment showed that recognition with machine learning was possible after around 30 images, validating the general assumption, that the actual manual recognition is more time-consuming than the automated, machine learning approach. An additional programming-based implementation, e.g., Python-based, can be used in a subsequent experiment to further investigate the extent to which this addition influences time saving.

This study rests upon the utilization of image material generated for the specific objectives set out for this study, but without previous knowledge of machine learning specifics. The application of machine learning within this study is confined to a discrete subset of images, employed both for the training of the model and subsequent evaluation processes. To augment result accuracy in future analyses, the used dataset should be extended. This extension should encompass diverse construction site viewpoints to prevent some of the encountered limitations (e.g., occlusions). The incorporation of different construction sites could address the generalizability of the developed models, thereby enhancing their efficacy in practical applications as well as their robustness.

Future studies should prioritise an examination of the transferability of the developed models to projects beyond the scope of the training dataset. Scrutiny of the extent to which the trained network can be seamlessly applied to distinct construction sites without requiring specialized retraining is essential. The objective of this evaluation should be authenticating the applicability of the developed models in varied contexts and ascertaining their general validity for construction site progress monitoring. Another focus of future investigations should be the systematic analysis of viewpoint perspectives and its consequential effects on the accuracy of the object recognition.

Acknowledgement. This research was developed in collaboration with the Chair of Construction Chemistry and Polymer Materials at our University, whom we would like to thank in particular for the data acquisition. We would also like to acknowledge the collaboration of Anton Streit and Dominik Reimann.

References

1. Frick, D., Gadatsch, A., Kaufmann, J., et al. (eds.): Data science: Konzepte, Erfahrungen, Fallstudien und Praxis. Springer Vieweg, Wiesbaden, Heidelberg (2021)
2. Liu, W., Shao, Y., Zhai, S., et al.: Computer vision-based tracking of workers in construction sites based on MDNet. IEICE Trans. Inf. Syst. **106**, 653–661 (2023). <https://doi.org/10.1587/transinf.2022DLP0045>
3. Mockenhaupt, A.: Maschinelles Lernen. In: Mockenhaupt, A. (ed.) Digitalisierung und Künstliche Intelligenz in der Produktion, pp. 133–163. Springer Fachmedien Wiesbaden, Wiesbaden (2021)
4. Ultralytics Inc.: YOLOv8 Repository. <https://github.com/ultralytics/ultralytics> (2023). Accessed 15 Dec 2023
5. CVAT.ai Corporation: CVAT Repository. <https://github.com/opencv/cvat> (2023). Accessed 15 Dec 2023
6. Gabler Wirtschaftslexikon, 19. Auflage. Springer Gabler, Wiesbaden, Heidelberg (2019)
7. Erlei, M., Leschke, M., Sauerland, D.: Neue Institutionenökonomik, 2nd edn. Schäffer-Poeschel Verlag, s.l. (2007)
8. Laroca, R., Severo, E., Zanlorensi, L.A. et al.: A robust real-time automatic license plate recognition based on the YOLO Detector. In: 2018 International Joint Conference on Neural Networks (IJCNN): 2018 proceedings, pp. 1–10. IEEE, Piscataway, NJ, USA (2018)
9. Terven, J., Cordova-Esparza, D.: A comprehensive review of YOLO: from YOLOv1 and beyond (2023)



Enhancing Information Extraction from German Regulatory Documents Using Deep Learning

Sherief Ali and Markus König^(✉)

Ruhr-Universität Bochum, Bochum, Germany

Sherief.ali@ruhr-uni-bochum.de, koenig@inf.bi.rub.de

Abstract. The construction industry aims to fully digitize the processes that are time-consuming, error-prone, and labor-intensive. One of the processes that needs digitization, is verifying the regulations and standards to ensure the validation of the construction work. The common obstacle that most of the studies address is that the regulations are not structured for machine interpretation and that sparks the exploration to develop tools and methodologies to parse the regulations effectively, enabling accurate information retrieval. This paper proposes a deep learning-based approach, which relies on the sequence labeling of key text in regulatory documents, to empower information extraction. Our proposed method commences with comprehensive text preprocessing and deep understanding to ensure effective and accurate extraction of information. Additionally, an entity list is outlined to provide a guideline for the prominent text. We employed Named Entity Recognition tagging to label the key text in German regulatory documents from the Research Society for Roads and Traffic regulatory documents. Subsequently, we leverage the pre-trained word embeddings to transform the textual dataset into numerical representations, which could be understandable by a neural network. The numerical representations work as a foundation for our model, which is based on a Bidirectional Long Short-Term Memory network. The labeled data is trained on the network, which provides an evaluation of the extracted data.

Keywords: Bidirectional Long Short-Term Memory (Bi-LSTM) · Natural Language Processing (NLP) · Named Entity Recognition (NER) · Deep learning · Smart Standards

1 Introduction

Building Regulations Consent (BRC) is a crucial aspect in the construction sector, where the relevant authorities ensure that the buildings are constructed in compliance with safety, environmental, and social aspects. The process of obtaining BRC is normally done manually, which involves a comprehensive review of regulations and standards. Therefore, this process is time-consuming, labor-intensive, and error-prone due to human interpretation.

The construction sector aims to fully digitize the validation of regulations and standards according to the construction projects. While, the regulations are written in natural language, which enables only humans, to understand and extract information from them [1]. As the language of the regulations is only interpretable through human intervention, Eastman [2] stated that the regulations or rules need to be translated into machine-processable format. To achieve this, Natural Language Processing (NLP) techniques are used to provide text comprehension. NLP is a subfield of Artificial Intelligence (AI) that involves the development of algorithms and language models, to interpret and understand text written in natural language [3]. NLP facilitates the interpretation of information in regulations through various techniques, including text classification, similarity analysis, feature extraction, and information extraction [1]. One of the techniques used in information extraction is Named Entity Recognition (NER), which categorizes the entities in the text document, to provide semantic identification of these entities such as a person, place, or organization [4].

In this paper, we propose a methodology to interpret and extract information from German regulatory documents that relies on creating an entity list that provides a guideline to the fundamental keywords in the regulations. Based on the guideline, the crucial keywords are labeled using NER tagging, to create a dataset of different entities, where each entity has a list of words. Accordingly, the textual dataset is transformed into indices by creating two dictionaries to form an efficient numerical representation, to be suitable as an input to the neural network. The dataset is employed to train a model that is based on Bidirectional Long Short Term Memory (Bi-LSTM).

2 Related Studies

There are different approaches to interpreting and extracting information from regulations. The traditional way is the rule-based approach, which Eastman et al. [2] stated that the process involves four steps, namely (1) rule interpretation and logic structuring; (2) building model preparation; (3) rules checking, and reporting the checking results. The process of rule formulation is a manual way of data processing that has proved to achieve high-quality translations and interpretation of regulations [5]. While the objective of the deep learning-based approach is to automate the process of information extraction, the most common obstacle is the size of the dataset and the coherence of the data extracted from the regulatory documents, as these documents are not meant to be machine-interpretable. There are different methodologies to apply deep learning to interpret regulations, while Fuchs and Amor [1] stated that Transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) have a promising impact on automating information extraction. Another deep learning model that showed high performance, is the LSTM model, which can learn long-term dependencies in sentences.

In the following Sects. 2.1 and 2.2, several research studies are discussed concerning the deep-learning approach.

2.1 Bidirectional Encoder Representations from Transformers

A new language representation model, which is Bidirectional Encoder Representations from Transformers (BERT), was introduced by Devlin et al. [6]. BERT is a pre-trained language model that is trained on two unsupervised tasks, which are the Masked Language Model (MLM) [7] and Next Sentence Prediction (NSP) [6]. The masked language model works by masking a random portion of tokens in a sentence and then asking the model to predict the masked token. While next sentence prediction aims to predict the next logical sentence that should follow an input sentence or sequence, by jointly pretraining text-pair representations. Therefore, NSP teaches BERT to learn long-term dependencies across sentences. As BERT has shown powerful performance on eleven natural language processing tasks, it can be fine-tuned with one output layer for application to other NLP tasks such as question answering and language inference.

Based on that, there are diverse studies that have conducted trials on BERT with NLP tasks, particularly after Fuchs and Amor [1] claimed that transformation-based models are promising and show good performance in information extraction. For instance, Schönfelder and König [8] proposed a NER-based model that trained German building code documents on the pre-trained German corpus BERT. They used NER tags to label text in the building code as they aimed to train the network based on supervised learning. The study has outlined performance values of 95.7% precision and 95.2% recall. The authors stated that the study still needs further improvement, as the current model demonstrates optimal performance values with the German corpus employed.

2.2 Bidirectional Long Short-Term Memory

Bidirectional LSTM is a type of recurrent neural network (RNN) architecture that incorporates both forward and backward layers. These layers allow the network to effectively utilize and retain information from both past and future inputs. BiLSTM can capture efficiently sequential dependencies within a sequence of words or phrases in both directions. This makes the network a powerful tool for tasks involving understanding and processing sequential data. Hochreiter and Schmidhuber [9] introduced the gradient-based method of “Long Short-Term Memory”, which can learn, process and classify sequential data. Several studies tested LSTM networks on NLP tasks and achieved promising results. For instance, Zhong et al. [10] proposed a hybrid deep learning approach to extract construction procedural constraints in a Chinese code. The approach is to use bidirectional LSTM (Bi-LSTM) and Conditional Random Field (CRF) to train labeled entities using NER. The authors selected 14 types of national standards defined in the Code for Acceptance of Construction Quality in China. The study outlined that they achieved an average precision for construction constraints classification of 73.7%. The authors have stated that, in future work, the approach’s performance needs to be refined by increasing the size of the dataset.

Another framework proposed by Moon et al. [11], aims to extract information from construction specifications by developing an NER model based on Bi-LSTM. The authors collected a dataset that consists of 56 construction specifications, in particular 4659 labeled sentences. They achieved promising performance values of 0.919 precision and 0.914 recall. The authors also stated that the model needs improvement as the

size of the dataset plays a crucial role in the performance of the model. There is also another approach conducted by Song et al. [12], to extract predicate-argument structure from building design rule sentences. In this study, they not only used NLP techniques to identify the semantic rule elements of building design requirements but also used a Bi-LSTM model with a conditional random field layer to extract the logical elements in the design rule sentences. The study has conducted several trials with variable dataset sizes, while the highest precision achieved is 80% in a dataset of 350 sentences.

3 Methodology

3.1 Data Preparation

The regulatory documents selected for analysis are from the Research Society for Roads and Traffic (FGSV) [13]. FGSV is responsible for formulating the technical regulations for the entire road and traffic system in Germany, and they have free and paid versions of the regulations. We selected five regulatory documents from FGSV as part of our case study, with the observation that these regulations are written in German. In our selection process, we thoroughly analyzed the regulations, to determine which regulations would be rich in different information types such as unstructured text and tables. The average number of pages for the selected regulations is 100. The information is listed in the selected regulations in the form of unstructured text, figures, tables, and text inside figures.

3.2 Named Entity Recognition

In the realm of data organization and classification, the introduction of new tags can significantly enhance the precision and clarity of information. In this context, We created a list of entities that play a crucial role in structuring and categorizing data. The entities, namely Pfact (Primary factor), Sfact (Secondary factor), Cfact (Comparative factor), none, and target, form a comprehensive list designed to express diverse facets of information as shown in Table 1. INCEPTION web-based annotation platform is utilized for tagging words based on the specified entities [14]. The platform provides the user the possibility to upload a document in plain format, enabling them to create tags corresponding to the case study. Following the completion of the tagging process, users have the option to export the annotated document in multiple formats.

3.3 Word to Index

The regulations are written in natural language, which can't be analyzed by the machine. To allow the machine to read and interpret the text, these texts have to be converted into numerical values. Therefore, individual words are converted into indices to optimize the training and interpretation of the annotated data. This indexing approach streamlines the training procedure, enhancing the model's ability to process and understand the input data more efficiently replacing the need to encode words directly. Each word is associated with a specific index, and this mapping is utilized during the preparation of text data

Table 1. Entities list

| Entity | Description | Example |
|--------|------------------------------|---------------------------------|
| Pfact | Structure type | Infrastructure |
| Sfact | More data about the material | Location |
| Cfact | Comparative | Maximum or minimum |
| None | Nonmeaningful words | The (Die, Der, Das, eine, etc.) |
| Target | Definitive values | Thickness 2.5 m |

for machine learning models. By converting words to their corresponding indices, the text is transformed into a format suitable for input into neural networks. Hence, the sentences extracted from the regulations undergo tokenization. Each word is assigned a distinct index to form a dictionary, enabling the retrieval of the original words based on their respective indices. This process facilitates the interpretation of model predictions or analyses post-training.

3.4 Model Architecture

After conducting an extensive literature review, the selected model for training annotated words in regulatory documents is the Bi-LSTM, demonstrating promising outcomes as discussed in Sect. 2.2. The architecture of Bi-LSTM includes forward and backward LSTM layers whose outputs are combined using the concatenating function (σ) as illustrated in Fig. 1 [15]. LSTMs make use of memory cells and gates, including input, forget, and output gates, to selectively store, adjust, and recall information across time. This characteristic renders LSTMs particularly effective for tasks that require the learning of patterns over extended sequences. The network could store information from the future and past by using forward and backward propagation, which allows the capture of the dependencies and patterns in both directions. This feature enhances the ability to understand the context over other neural networks.

4 Case Study

4.1 Preprocessing

The FGSV regulations are structured in PDF format, but to facilitate the annotation, the documents must be formatted in TXT, since INCEPTION only accepts TXT format. A thorough reading and comprehension of the regulations are required to identify sentences relevant to the case study. Our primary emphasis is on extracting sentences that hold quantitative information. To transform the regulations from PDF to TXT format, we initiated the process of manually extracting the sentences related to points of interest from five FGSV regulatory documents. Following the conversion to TXT format, we begin examining the sentences individually, as our objective is to handle each sentence separately. Subsequently, we initiate the preprocessing of the sentences by modifying

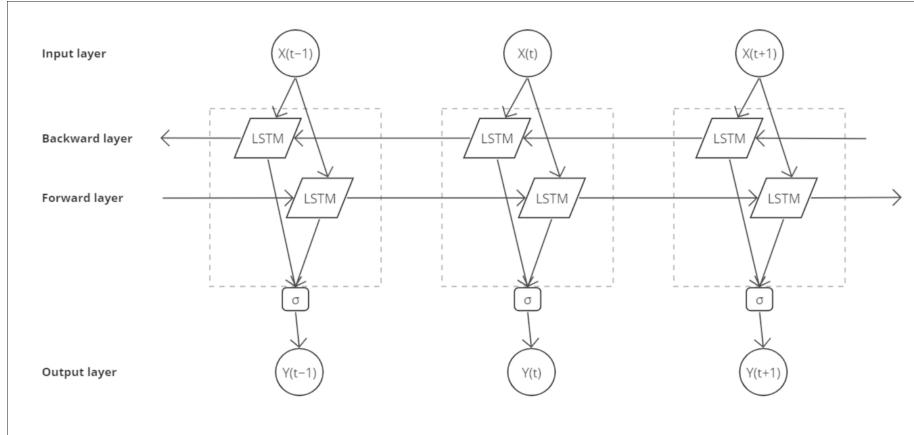


Fig. 1. Architecture of Bi-LSTM

the placement of commas or periods, as the presence of a period signifies the end of a sentence in INCEpTION. Accordingly, The extracted data from the regulations are cleansed and prepared for the next step, which involves annotation.

4.2 Data Labeling

The INCEpTION platform is utilized to annotate the quantitative sentences extracted from the regulations. The sentences are annotated by assigning NER tags from the list of defined entities provided in Sect. 3.2. We sought to achieve a balanced dataset of tagged entities, considering that certain entities are frequently encountered in the regulations while others are rare. Our goal is to find the appropriate balance in the number of tagged entities across the five categories explained in Sect. 3.2. As a result, we labeled instances with allocations of 200 words for Pfac, 250 words for Sfact, 70 words for Cfact, 450 words for None, and 880 words for Target as illustrated in Fig. 2.

4.3 Numerical Representation

For the model to grasp and interpret the content of annotated entities, which are written in natural language, a transformation into numerical values is imperative. This conversion involves the assignment of indices to each word, a process known as word-to-index elaborated upon in Sect. 3.3. The purpose behind this indexing methodology is to optimize the training phase, allowing the model to efficiently learn and understand the annotated data. By translating individual words into numerical indices, the machine gains the ability to navigate and process the information with enhanced accuracy and effectiveness. Following the annotation process outlined in the preceding section, the annotated words must be input into the network for training. However, it's important to note that the network is only capable of processing numerical data. As depicted in Fig. 3, this sentence serves as an example extracted from the regulations. Each word is assigned a unique number to form a dictionary of words and corresponding indices. This dictionary is then used to train the network.

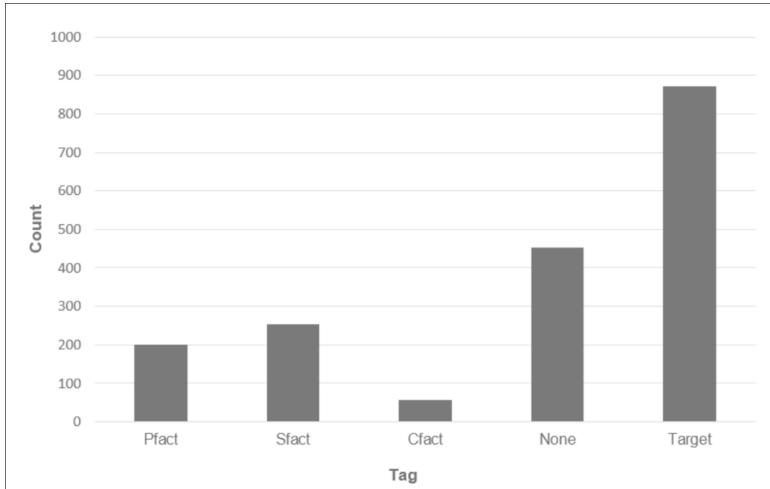


Fig. 2. Count of annotated words for each entity using NER tags

```
{"Bei": 1, 'Bauweisen': 2, 'mit': 3, 'vollgebundenem': 4, 'Oberbau': 5, 'soll': 6, 'bei': 7, 'Böden': 8, 'der': 9, 'Frostempfindlichkeitsklasse': 10, 'F3': 11, 'kritischen': 12, 'Wasserhältnissen': 13, 'auch': 14, 'F2': 15, 'eine': 16, 'Bodenverfestigung': 17, 'des': 18, 'Untergrundes': 19, 'bzw': 20, 'Unterbau': 21, 'Mindestdicke': 22, 'von': 23, '15': 24, 'cm': 25, 'vorgesehen': 26, 'werden': 27}
```

Fig. 3. Example for converting words into indices

4.4 Model Training

The Bi-LSTM model uses labeled sentences as an input and outputs the predicted labels of the words through training of the model. The split of the dataset is as follows 70% training, 20% validation, and 10% test. This split allows the model to learn from diverse examples during training while validating its performance on a separate set to detect overfitting. Furthermore, we monitored the maximum length of the sentence, because this ensures the model can effectively capture information from both short and lengthy sentences, enhancing its overall understanding. Hence, the selected maximum sentence length is set at 80 words, with any exceeding words subject to truncation. Regarding the dropout, it has been configured to a value of 0.1, a determination reached through several trials, to find the right balance between model regularization and maintaining the capacity to learn intricate patterns. Due to the small size of the dataset, the batch size has been configured to 16. This choice enables the model to update weights more frequently and improve generalization capabilities. The model undergoes training over the course of 120 epochs. To gradually decrease the learning rate during training for better convergence, we employed a learning rate scheduler, to reduce the learning rate by a factor of 0.95 at every tenth epoch with an initial learning rate of 0.01. The model is trained multiple times with different variations in the hyperparameters, to find the appropriate adjustment and prevent overfitting or underfitting of the model.

4.5 Results and Discussion

The model provides results by presenting true annotated and predicted entity labels, as a reflection of the training outcomes. Table 2 exemplifies a sentence prediction based on the entity list. Notably, some words are predicted incorrectly, attributed to the dataset's size constraints. In a broader observation of label predictions, random sentences were analyzed, revealing that a majority of sentences have two words predicted inaccurately. As stated earlier sentences should not encompass all entities, and there are instances where an entity may be repeated within the same sentence, thus there are entities that are not shown in this example.

Table 2. Example of a sentence's entity prediction

| Index | Word | True | Predicted |
|-------|-----------------------------|--------|-----------|
| 4 | Oberbau | Pfact | Sfact |
| 9 | Frostempfindlichkeitsklasse | Sfact | Sfact |
| 21 | Mindestdicke | Cfact | Cfact |
| 17 | des | None | 0 |
| 23 | 25 cm | Target | Target |

For an assessment of the model's performance, Table 3 provides a breakdown of the numerical classification performance for each entity. During the training process, the dataset is split multiple times randomly, to ensure that the model doesn't overfit on a particular order or size of training data.

The table reveals that the model exhibits exceptionally high precision in predicting Sfact, Cfact, and Target, achieving 92%, 91%, and 93%. In contrast, Pfact and None demonstrate lower precision. The performance values demonstrate that the predicted labels of Pfact and None, are misclassified due to the variability of words in these two categories. As these categories hold plenty of words with different meanings, thus the network could not categorize these labels correctly in all the sentences. Notably, the model achieves outstanding results when tested on different sentences, which indicates that the model is capable of accurately predicting the entities.

Table 3. Classification average performance scores

| Entity | Precision | Recall | F1 score |
|--------|-----------|--------|----------|
| Pfact | 0.88 | 0.87 | 0.87 |
| Sfact | 0.92 | 0.92 | 0.92 |
| Cfact | 0.91 | 0.9 | 0.9 |
| None | 0.80 | 0.78 | 0.79 |
| Target | 0.93 | 0.92 | 0.92 |

5 Conclusion

Building regulations consent requires human intervention nowadays, which is a time-consuming and error-prone process. In this paper, we employed a Bi-LSTM model to predict entities from a German regulatory document from the Research Society for Roads and Traffic. We used the NER tagging from NLP and created an entity list that is composed of five tags, which are Pfact, Sfact, Cfact, None, and Target. The Bi-LSTM model is trained on 200 words for Pfact, 250 words for Sfact, 70 words for Cfact, 450 words for None, and 880 words for Target. The training results of the model have demonstrated promising results of achieving an average validation accuracy of 92%. Following an in-depth analysis of the data, some tags are predicted with high precision such as Sfact, Cfact, and Target, while other tags cannot achieve more than 90%. The study demonstrated that the Bi-LSTM model can achieve high-precision results, however, the dataset size controls the level of precision that can be achieved.

Hence, there is a need to consider future enhancements. Primarily, our focus is on expanding the dataset size to enhance the model's capacity for training and assess its tag prediction capabilities more effectively. Moreover, in future research studies, we aim to establish relationships between entities within the same sentence. This endeavor aims to automatically extract meaning from sentences, potentially offering a methodology that can be relied upon without human intervention.

References

1. Fuchs, S., Amor, R.: Natural language processing for building code interpretation: a systematic literature review. CIB W78 (2021)
2. Eastman, C., Lee, J.-M., Jeong, Y.-S., Lee, J.-K.: Automatic rule-based checking of building designs. Autom. Constr. **18**(8), 1011–1033 (2009)
3. Chowdhary, K.: natural language processing. In: Fundamentals of artificial intelligence, pp. 603–649. Springer, New Delhi (2020)
4. Jehangir, B., Radhakrishnan, S., Agarwal, R.: A survey on named entity recognition—Datasets, tools, and methodologies (2023)
5. Fuchs, S.: Natural language processing for building code interpretation: systematic literature review report (2021)
6. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding (2019)

7. Taylor, W.L.: Cloze procedure: a new tool for measuring readability. **30**(4) (1953)
8. Schönfelder, P., König, M.: Deep learning-based entity recognition in construction regulatory documents. In 38th International Symposium on Automation and Robotics in Construction (ISARC 2021) (2021)
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory (1997)
10. Zhong, B., Xing, X., Luo, H., Zhou, Q., Li, H., Rose, T., Fang, W.: Deep learning-based extraction of construction procedural constraints from construction regulations. *Adv. Eng. Infor.* (2020)
11. Moon, S., Lee, G., Chi, S., Oh, H.: Automated construction specification review with named entity recognition using natural language processing. *J. Constr. Eng. Manage.* **147**(v)
12. Song, J., Lee, J.-K., Choi, J., Kim, I.: Deep learning-based extraction of predicate-argument structure (PAS) in building design rule sentences. *J. Comput. Des. Eng.* (2020)
13. F.V. GmbH: Forschungsgesellschaft für Straßen und Verkehrswesen (2023)
14. Klie, J.-C., Bugert, M., Eckart de Castilho, B.B.R., Gurevych, I.: INCEPTION platform: machine-assisted and knowledge-oriented interactive annotation. In Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics (COLING 2018), Santa Fe, New Mexico, USA (2018)
15. Cui, Z., R. K. Z. P., Wang, Y.: Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction (2018)



Evaluating the Capabilities of Machine Learning Surrogate Modeling to Support High-Resolution Building Performance Simulation

Elin Markarian¹, Seif Qiblawi¹, Shivram Krishnan², Anagha Divakaran²,
Albert Thomas², and Elie Azar¹(✉)

¹ Carleton University, Ottawa, ON, Canada
elie.azar@carleton.ca

² IIT Bombay, Powai, Mumbai, Maharashtra, India

Abstract. Many countries are attempting to adopt energy retrofit measures for their existing building stock to achieve low or net zero energy goals. In this regard, Building Performance Simulation (BPS) tools are often used to test and inform different retrofit measures. However, these tools can be time-consuming and computationally intensive, which limits their application in complex building retrofit and optimization applications. Machine learning (ML) algorithms used as surrogates to BPS models can supplement their capabilities and overcome some of their computational limitations. Recent studies have successfully trained and validated ML-based surrogate BPS models to predict building performance for different applications at low computational costs. However, most studies are limited to predicting total building energy consumption over large time windows (e.g., yearly), which is inadequate for applications that require more granular prediction windows (e.g., daily), such as net zero energy building design. The primary objective of this study is to develop and explore the capabilities of ML-based BPS surrogate models to support applications requiring high-resolution energy predictions. This is achieved by (i) developing a BPS model of an archetype office building in Canada, (ii) generating a dataset for ML surrogate model training, validation, and testing, (iii) applying and contrasting the capabilities of different ML algorithms at predicting energy loads using different prediction windows (e.g., annually, monthly, daily, and hourly). The results demonstrate that hourly, daily, and monthly ML models achieved competitive predictive accuracies, confirmed by adjusted R^2 values exceeding 0.9 and CV(RMSE) values within industry standards. In contrast, the annual ML models yielded lower adjusted R^2 values, which could be attributed to smaller sample sizes and a lack of granular weather information (e.g., hourly air temperature) provided as inputs to the models. The findings confirm the ability of ML-based surrogate models to mimic BPS models' performance for applications requiring high-resolution energy predictions.

Keywords: Surrogate Modeling · Machine Learning · Building Performance Simulation · Gradient Boosting · Energy Consumption · Office Building

1 Introduction

Buildings are major energy consumers, and are being responsible for approximately 30–40% of total energy consumption and greenhouse gas (GHG) emissions [1, 2]. These figures are expected to rise even more with rapid urbanization, making inefficient energy use in buildings a challenge [3]. Physics-based building performance simulation (BPS) tools, such as EnergyPlus and TRNSYS, are often used to help tackle this challenge. Using detailed building and weather information, BPS tools provide insightful predictions and information about building energy performance, and therefore facilitate the retrofitting process and energy-efficient design of buildings. However, these tools often have lengthy simulation runtimes, particularly when numerous variables and designs are to be simulated, which is inevitable when performing sensitivity/uncertainty analyses or design optimizations. Machine learning (ML) algorithms, when used as surrogate or meta models, can help overcome the limitations of physics-based models. Surrogate models are trained on a dataset consisting of simulation inputs and outputs, with the goal of emulating the behavior of a physics-based model. Once trained, these models provide reliable building performance predictions at a low computational cost [4].

Numerous studies have confirmed the capabilities of ML-based surrogate models in predicting building performance, covering different aspects and combinations of surrogate model development (e.g., number of features and targets) and applications. For instance, Ali et al. [5] trained several ML algorithms using a dataset of 10,000 annual EnergyPlus runs and indicated that extreme gradient boosting (XGB) can attain a coefficient of determination (R^2) of 0.99. In another study, Papadopoulos et al. [4] trained three ML ensemble models on a dataset including eight geometrical features as well as annual heating and cooling load intensities as targets, concluding that gradient boosting random tree (GBRT) algorithm has a better performance than its counterparts. Meanwhile, some other studies used surrogates to analyze various retrofit measures. Ali et al. [6] generated a dataset of one million buildings using four building archetypes to assess retrofit measures on an urban scale. The dataset was then used to train ten ML models, of which XGB reached the highest level of accuracy. Similarly, Saad et al. [7] attempted to enhance the performance of surrogate models for building retrofit purposes by focusing on different feature engineering and selection methods (e.g., backward-stepwise feature selection (BSFS), recursive feature elimination (RFE), and elastic net embedded regularization). The developed surrogates were able to achieve mean absolute percentage error (MAPE) values of 0.2–1.8% and reduce the computational time significantly when handling a dataset with a million samples. Shen and Pan [8] trained a light gradient boosting (LGBM) surrogate on a dataset created by DesignBuilder to optimize building energy performance, thermal comfort, and carbon emissions, concluding that their optimal solution can result in approximately 13% energy saving.

While the above studies predicted energy consumption on an annual basis, few studies attempted to use more granular time steps (e.g., hourly energy predictions). Li et al. [9] developed time-series surrogate models using hourly data to predict the heating and cooling loads of an office building. Four types of input features (i.e., static, dynamic, lagged, and autoregressive) were used to train the models. Initially, the models were solely trained on static (e.g., external wall U-value) and dynamic (e.g. outdoor air temperature) features. Lagged features (i.e., the past and current values of the input

variables, as well as autoregressive features (e.g., the energy use in the previous hour), were used as input features in the subsequent phases of their study. They found that a shorter time window can result in an R^2 of 0.99 and coefficient variation of root mean square error (CV(RMSE)) of 0.06 for both their targets. Westermann et al. [10] developed a deep temporal convolutional neural network using 569 weather files with an hourly resolution and building design parameters (e.g., wall insulation conductivity and heating and cooling setpoints) as model inputs to predict annual heating, annual cooling and hourly heating demand. The developed model could predict the heating demand with an error of less than 3% for locations that were not used during model training.

In summary, most existing literature using ML-based surrogate modeling and BPS data relies on annual energy performance data and prediction windows, overlooking the challenges, and potential, of applying such an approach on shorter time intervals. Although a few studies started addressing this issue [9, 10], to the best of the authors' knowledge, none provide a systematic assessment of the models with different time intervals. It is not clear what temporal granularity is feasible or needed to provide accurate predictions, and what are the challenges involved.

This study presents a systematic approach to investigate and compare the predictive performance of ML-based surrogate models developed using BPS data across different time granularities (hourly, daily, monthly, and annually). While the methodology is general, it is demonstrated through a case of an archetype office building in Ottawa, Canada. Two promising ML Algorithms are considered in this study, namely extreme and light gradient boosting. The contributions of this study help bolster building research efforts that require high-resolution energy predictions, such as parametric and net-zero energy design and retrofitting.

2 Methodology

The proposed methodology involves five main phases, as illustrated in Fig. 1, and detailed in the following sub-sections. Initially, a BPS model of an archetype office building in Ottawa, Canada is developed using EnergyPlus (Phase I). Subsequently, a dataset is created by running the BPS with different combinations of input variables while outputting hourly energy predictions (Phase II). Next, subsets are created by aggregating the hourly energy predictions into daily, monthly, and annually energy predictions (Phase III). Surrogate ML models are then trained and validated (Phase IV) using the data subsets generated in the previous step. Finally, the models are tested on data unseen during model training, using diverse accuracy metrics to reflect on the ML algorithm's abilities to predict energy consumption using different time prediction windows (Phase V).

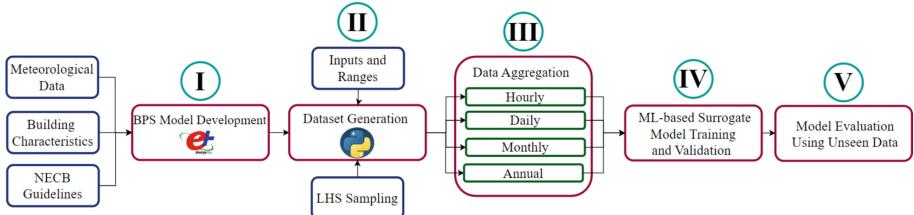


Fig. 1. Proposed methodology flowchart.

2.1 BPS Model Development

A BPS model, as illustrated in Fig. 2, is developed to simulate the performance of an archetype three-story office building in Ottawa, Canada based on the National Energy Code of Canada for Buildings (NECB) [11]. Each floor comprises five zones: one core and four perimeter zones. The building characteristics are presented in Table 1. The model's accuracy is validated by comparing the annual energy use intensity (EUI) with that of a similar building reported in [12]. The EUI of the simulated building is 105.3 kWh/m² which is approximately 5% different from that of [12], and is therefore considered adequate for this study.

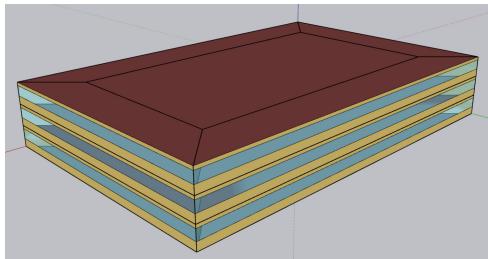


Fig. 2. BPS model of archetype building.

2.2 Dataset Generation

This phase consists of conducting a parametric variation of the BPS model developed earlier to create a dataset that includes BPS inputs (i.e., features) and outputs (i.e., targets) for use in the subsequent phases. A total of eleven BPS inputs are varied, following a uniform distribution, while monitoring two BPS outputs, namely electricity and natural gas consumption. Table 2 lists the stated variables, their ranges (i.e., upper and lower bounds), and the sources used to derive the inputs' chosen ranges.

Given the large number of features and their ranges, it is almost impossible to perform simulations for all possible scenarios. Latin Hypercube Sampling (LHS) is therefore adopted to draw samples from the specified ranges. LHS explores the sample space more effectively than traditional random sampling methods. It ensures that samples

Table 1. Archetype building specifications.

| Parameter | Value | Parameter (cnt'd) | Value (cnt'd) |
|-------------------------------------|-------|---|---|
| Total floor area (m ²) | 4982 | Lighting power density (W/m ²) | 8.8 |
| Floor area per person | 19.92 | Plug load density (W/m ²) | 10.27 |
| Solar heat gain coefficient (SHGC) | 0.31 | Boiler efficiency | 0.834 |
| Window U-value (W/m ² K) | 1.9 | Chiller COP | 4.51 |
| Window to wall ratio (WWR) (%) | 47.68 | Infiltration rate (m ³ /s-m ²) | 0.00025 |
| Wall U-value (W/m ² K) | 0.231 | HVAC system | Baseboards connective water heaters and VAV with Reheat |
| Roof U-value (W/m ² K) | 0.154 | Heating coil fuel | Natural gas |
| Occupant heat gain (W/person) | 130 | Chiller type | Scroll electric |

are distributed over the entire range of each feature, generating suitable representative samples [6]. In this study, the LHS functionality from Design of Experiments for Python (PyDOE) library is imported to generate 100 samples, which are then used to perform EnergyPlus simulations. To facilitate the simulation process, Eppy, which is a scripting language for EnergyPlus written in Python, is used. The simulations are performed on an hourly basis, producing 8760 datapoints for each run.

In the next step, the weather data, input features, and the outputs (electricity and natural gas consumption) are stored and processed. Weather data and day type (Table 3) are also added to the dataset as important determining factors in energy models. The day type feature distinguishes weekdays and weekends and is assigned a value of 0 for weekends and 1 for weekdays.

2.3 Data Aggregation

This phase consists of using the hourly dataset generated in the previous step to create subsets at the daily, monthly, and annual basis. Such subsets are later used to train and compare surrogate models. Throughout this process, building features remain constant across runs and require no further processing. The targets are summed, and weather features are either averaged or summed for daily and monthly datasets. Notably, weather features are excluded from the annual dataset, as they remain constant across all runs. In summary, the entire process results in four datasets: hourly (876,000 data points), daily (36,501 data points), monthly (1200 data points), and annual (100 data points).

Table 2. Input features and target variables used to generate the BPS dataset.

| Features/Targets | Range | Reference | Features/Targets | Range | Reference |
|--|-----------|-----------|---|----------------|-----------|
| X1-Wall U-value
(W/m ² K) | 0.1–1.41 | [13, 14] | X8-Cooling Setpoints
(time:6–21) (°C) | 22–28 | |
| X2-Roof U-value
(W/m ² K) | 0.1–1.89 | [13, 14] | X9-Cooling COP | 2.97–5.33 | [13] |
| X3-Window
U-value
(W/m ² K) | 0.79–6.42 | [13, 15] | X10-Heating efficiency | 0.72–0.95 | [13] |
| X4-SHGC | 0–1 | | X11-Infiltration
(m ³ /s.m ² exterior
area) | 0.000125–0.008 | [15] |
| X5- Heating
Setpoints
(time:21–6) (°C) | 16–20 | | Y1-Electricity
(kWh) | – | |
| X6-Heating
Setpoints
(time:7–21) (°C) | 18–24 | | Y2-Natural gas
(kWh) | – | |
| X7-Cooling
Setpoints
(time:21–6) (°C) | 25–35 | | | | |

Table 3. Weather features incorporated in the dataset; the numbering of variables is continued from Table 2.

| Weather features |
|---|
| X12-Outdoor air temperature (°C) |
| X13-Relative humidity (%) |
| X14-Wind speed (m/s) |
| X15-Diffuse Radiation (W/m ²) |
| X16-Direct Radiation (W/m ²) |
| X17-Day type |

2.4 ML-Based Surrogate Model Training and Validation

A wide variety of ML models have been utilized in the literature for building performance prediction purposes. Among these models, extreme gradient boosting (XGB) and light gradient boosting (LGBM) have been widely used in building context and have shown promising performance [4, 8]. XGB is a gradient boosting algorithm which transforms weak learners into strong learners [16]. It also has a built-in regularization term which

prevents over-fitting and controls complexity, therefore contributing to better prediction outcomes. LGBM, developed by Microsoft, is another gradient boosting algorithm known for its lower memory usage, high accuracy and fast training speed [8].

In this study, the multi-output regression method from the scikit-learn library is employed to simultaneously predict the two target variables (i.e., electricity and natural gas consumption). The XGB and LGBM ML algorithms are imported from xgboost and lightgbm libraries. In order to ensure the robustness of the selected ML models, a 10-fold cross-validation is performed on the training set. This additional analysis is only conducted to gain insights into the models' generalizability capabilities. The final model is trained on the 80% of each dataset, and the remaining 20% is reserved for testing (later used in Sect. 2.5). The default hyperparameters of the model are presented in Table 4.

Table 4. Default hyperparameters of the ML models.

| Algorithm | Parameter | Value |
|-----------|---|----------|
| XGB | Learning rate: the shrinkage factor controlling the contribution of each tree | 0.3 |
| | Max_depth: maximum depth of a tree | 6 |
| | N_estimators: the number of trees to be built | 100 |
| LGBM | Learning rate | 0.1 |
| | Max_depth | No limit |
| | N_estimators | 100 |

2.5 Model Evaluation Using Unseen Data

To ensure the reliability and accuracy of the trained models, the testing set (20% of the dataset) is used for evaluation. Error metrics including adjusted coefficient of determination (R^2), mean absolute error (MAE) and coefficient of variation of root mean square error (CV(RMSE)) are calculated using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N y_i - \bar{y}_i} \quad (1)$$

$$Adjusted R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - k - 1} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (3)$$

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N} \right)} \quad (4)$$

$$CV(RMSE) = \frac{RMSE}{\frac{1}{N} \sum_{i=1}^n y_i} \quad (5)$$

where y_i , \hat{y}_i , and \bar{y}_i are the actual, predicted, and mean energy values, respectively, N is the number of samples, and k is the number of independent variables. It should be noted that R^2 and root mean square error (RMSE) are calculated but not presented in the results. MAE can be considered as an absolute error, while adjusted R^2 and $CV(RMSE)$ are relative errors. Relative errors are preferred in this study as the datasets have different time intervals. Given the significant variations in electricity and natural gas consumption across these time scales, relative errors are deemed to provide a more meaningful interpretation of the model performance.

Finally, computational time is also an important factor in the performance of ML models. The computational time in this study includes both training and testing time. The model is run on an Intel Core i7-12700H, 2.30 GHz/14 Cores workstation.

3 Results and Discussion

The performance metrics of the ML algorithms (XGB and LGBM) applied to the hourly, daily, monthly, and annually datasets are presented in Table 5. Both models (i.e., XGB and LGBM) achieved adjusted R^2 values higher than 0.9 when trained on hourly, daily and monthly datasets. The results indicate that a high percentage of variance in the target variables are explained by the models' features. In contrast, the annual energy models show lower adjusted R^2 values, particularly for electricity use. This could be attributed to the absence of weather data fed to the ML models, unlike the models with more granular predictions windows (Refer to Sect. 2.2). In other words, without time-dependent weather features (e.g., air temperature and humidity), less of the variance in the predicted energy consumption could be explained. Another contributing factor is the low sample size; the annual dataset had only 100 datapoints, of which 20 are used to test the model (shown in the graph). The Adjusted R^2 metric considers the sample size and the number of predictors and penalizes when there are a large number of predictors as shown in Eq. 2. With only 20 samples and 11 predictors, the denominator is not large enough, resulting in relatively low adjusted R^2 values.

In terms of MAE, large variations can be observed for different time scales, which was expected given the significant difference in magnitude of electricity and natural gas consumption over different time periods. For instance, annual energy values being larger than monthly or hourly values resulted in larger MAE. This finding indicates the importance of relying on relative—rather than absolute—error metrics, especially to compare the results of models across different time prediction windows.

The CV(RMSE) results are also presented in Table 5. According to Standard ASHRAE Guideline 14 [17], thresholds or acceptable values of CV(RMSE) for hourly, daily, and monthly energy predictions are 30, 22.5, and 15%, respectively; there are no guidelines for annual energy predictions. As shown in the table, all model errors were below the stated thresholds, confirming that the average distance between the predicted values and a regression fit line for each model is acceptable.

Table 5. Performance metrics of the trained models for different time prediction windows.

| Time window | Model | Adjusted R ² | | MAE | | CV(RMSE) | | Train/test time (Sec) |
|-------------|-------|-------------------------|------|-------------------|----------|-------------|------|-----------------------|
| | | Electricity | NG | Electricity (kWh) | NG (kWh) | Electricity | NG | |
| Hourly | XGB | 0.94 | 0.96 | 7.68 | 8.69 | 0.20 | 0.24 | 1294 |
| | LGBM | 0.90 | 0.96 | 9.77 | 8.75 | 0.26 | 0.25 | 173.58 |
| Daily | XGB | 0.99 | 0.98 | 33.84 | 132.72 | 0.04 | 0.16 | 65.79 |
| | LGBM | 0.99 | 0.98 | 34.44 | 128.84 | 0.04 | 0.15 | 29.77 |
| Monthly | XGB | 0.97 | 0.98 | 775.08 | 3785.30 | 0.03 | 0.15 | 7.92 |
| | LGBM | 0.97 | 0.98 | 778.52 | 3402.07 | 0.03 | 0.12 | 17.40 |
| Annual | XGB | 0.55 | 0.72 | 8278.06 | 46868.25 | 0.02 | 0.13 | 4.91 |
| | LGBM | 0.35 | 0.81 | 9958.19 | 39926.14 | 0.02 | 0.10 | 1.35 |

To further confirm the goodness-of-fit of the models, the actual and predicted values of the XGB models are presented in Fig. 3. Overall, the points are scattered around the diagonal lines, indicating a positive relationship between the predicted and actual values. In the case of hourly values, large differences are observed in some instances, which was expected as hourly energy values are typically more difficult to predict than aggregate values (e.g., monthly or yearly). It is interesting, however, to observe that the error reduced for high energy values (e.g., > 100 kWh). This could be attributed to hours with high cooling loads (e.g., during summer), which are easier to predict given the dependence of the load on weather information that is provided to the model (e.g., air temperature and humidity). Annual results also showed a good alignment along the diagonal, despite a few instances with large errors. As discussed earlier, this could be attributed to the lack of weather-related features in the ML models and/or the small size of the annual energy dataset. Daily and monthly results demonstrated high correlations between predicted and actual energy values, leveraging weather data as features in the models while also benefiting from good sample sizes.

Finally, the train/test time for each trained model are also shown in Table 5. As expected, the computing time is proportional to the granularity of the models. For instance, the train/test time of the XGB hourly model is more than 250 times that of the XGB annual model. Also, overall LGBM is significantly faster than XGB while still achieving competitive results, which necessitates further investigation. It is worth noting that training time was not a critical factor in this paper but could become one for more complex models that require thorough hyperparameter tuning and optimization. Nonetheless, a strength of surrogate models is that once they are trained and tested, they can make predictions in fractions of a second.

4 Conclusion

ML models are increasingly employed to support BPS tools and applications, providing researchers with access to reliable energy predictions at minimal computational costs. This is particularly important when it comes to uncertainty/sensitivity analyses or optimization problems where numerous runs are required (e.g., designing low and net zero energy buildings). Despite the growing interest and applications in the literature, most studies have trained ML models using annual datasets, which limits their applicability where more granular information (i.e., hourly information) is required. This study proposed a systematic approach to evaluate the capabilities of ML-based surrogate models for various time resolutions. To do so, a BPS model of an archetype office building in Ottawa, Canada was developed and used to generate a dataset with 100 hourly runs, supported by an LHS sampling scheme. The dataset was then processed and used to train and test XGB and LGBM ML models at hourly, daily, monthly, and annual resolutions, comparing their performance using standard accuracy metrics.

The findings indicated high levels of predictive accuracy for the monthly, daily, and hourly models, confirmed by adjusted R^2 exceeding 0.9 and CV(RMSE) values within ASHRAE 14's acceptable thresholds. In contrast, the annual models yielded lower adjusted R^2 ones, which could be attributed to smaller sample sizes and a lack of granular weather information (e.g., hourly air temperature) provided as inputs to the models. Overall, given the same number of runs, hourly, daily and monthly models outperform the annual models and therefore are well suited for application where higher time resolutions are required. To enhance the proposed models' generalizability, future research should expand the scope of analysis to cover varying weather conditions (i.e., different climate zones and future weather scenarios) and building types (e.g., residential and retail). Additionally, hyperparameter optimization and dimension reduction techniques could be employed to achieve more accurate and robust predictions capabilities.

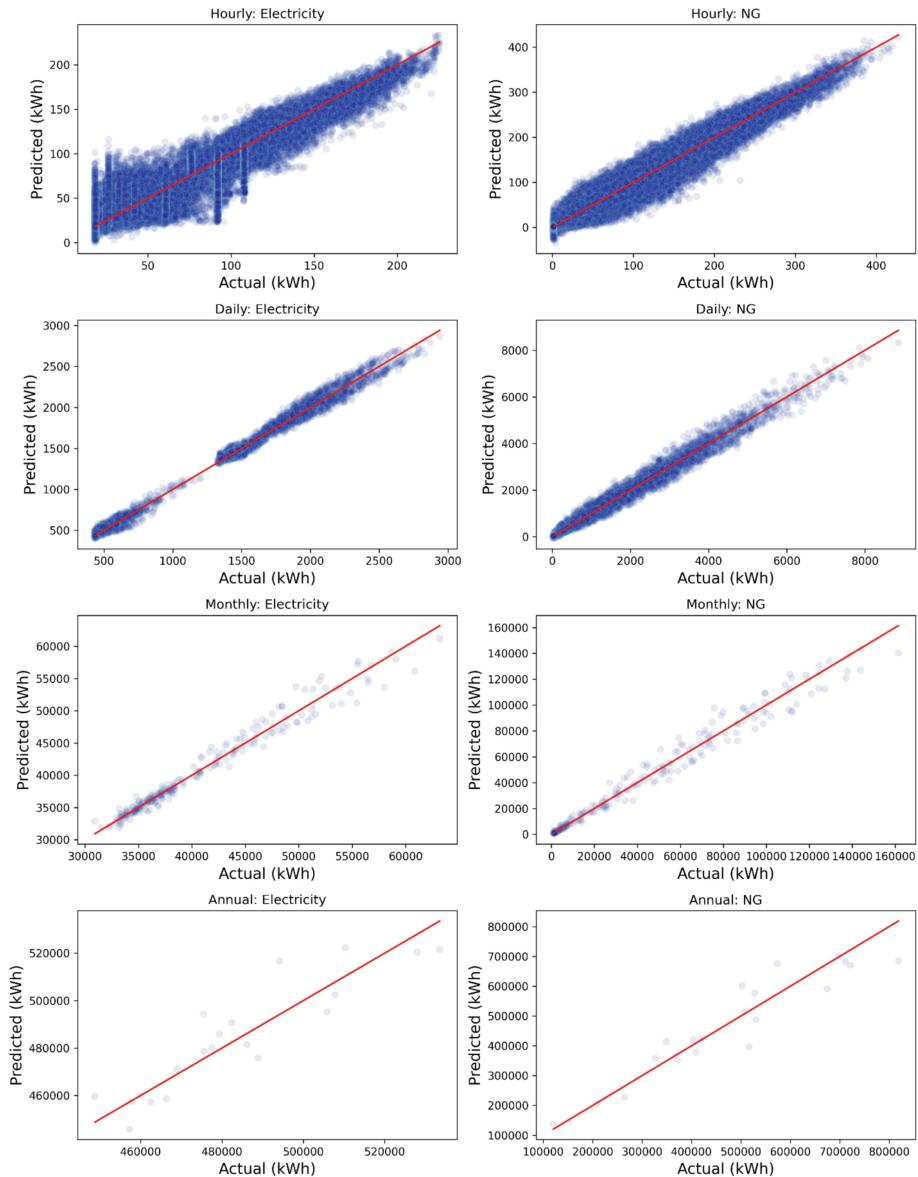


Fig. 3. Predicted vs. actual electricity and natural gas (NG) values of the XGB models for different time resolutions.

Acknowledgements. The authors acknowledge the funding support from IC-IMPACTS (India-Canada Centre for Innovative Multidisciplinary Partnerships to Accelerate Community Transformation and Sustainability) and DST (Department of Science and Technology), Government of India, grant number DST/IC/IC-IMPACTS/2022/P-7, as well as the Canada Research Chairs Program.

References

1. Chen, J., Gao, X., Hu, Y., Zeng, Z., Liu, Y.: A meta-model-based optimization approach for fast and reliable calibration of building energy models. *Energy* **188**, 116046 (2019). <https://doi.org/10.1016/j.energy.2019.116046>
2. Thrampoulidis, E., Hug, G., Orehoung, K.: Approximating optimal building retrofit solutions for large-scale retrofit analysis. *Appl. Energy* **333**, 120566 (2023). <https://doi.org/10.1016/j.apenergy.2022.120566>
3. Kazmi, H., Fu, C., Miller, C.: Ten questions concerning data-driven modelling and forecasting of operational energy demand at building and urban scale. *Build. Environ.* **239**, 110407 (2023). <https://doi.org/10.1016/j.buildenv.2023.110407>
4. Papadopoulos, S., Azar, E., Woon, W.-L., Kontokosta, C.E.: Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. *J. Build. Perform. Simul.* **11**, 322–332 (2018). <https://doi.org/10.1080/19401493.2017.1354919>
5. Ali, A., Jayaraman, R., Mayyas, A., Alaifan, B., Azar, E.: Machine learning as a surrogate to building performance simulation: predicting energy consumption under different operational settings. *Energy Build.* **286**, 112940 (2023). <https://doi.org/10.1016/j.enbuild.2023.112940>
6. Ali, U., et al.: Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. *Energy Build.* **303**, 113768 (2024). <https://doi.org/10.1016/j.enbuild.2023.113768>
7. Saad, M.M., Menon, R.P., Eicker, U.: Supporting decision making for building decarbonization: developing surrogate models for multi-criteria building retrofitting analysis. *Energies* **16**, (2023). <https://doi.org/10.3390/en16166030>
8. Shen, Y., Pan, Y.: BIM-supported automatic energy performance analysis for green building design using explainable machine learning and multi-objective optimization. *Appl. Energy* **333**, 120575 (2023). <https://doi.org/10.1016/j.apenergy.2022.120575>
9. Li, G., Tian, W., Zhang, H., Fu, X.: A novel method of creating machine learning-based time series meta-models for building energy analysis. *Energy Build.* **281**, 112752 (2023). <https://doi.org/10.1016/j.enbuild.2022.112752>
10. Westermann, P., Welzel, M., Evins, R.: Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones. *Appl. Energy* **278**, 115563 (2020). <https://doi.org/10.1016/j.apenergy.2020.115563>
11. Canada, N.R.C.: National Energy Code of Canada for Buildings 2017, <https://nrc.canada.ca/en/certifications-evaluations-standards/codes-canada/codes-canada-publications/national-energy-code-canada-buildings-2017>, last accessed 2023/12/18
12. Kharvari, F., Azimi, S., O'Brien, W.: A comprehensive simulation-based assessment of office building performance adaptability to teleworking scenarios in different Canadian climate zones. *Build. Simul.* **15**, 995–1014 (2022). <https://doi.org/10.1007/s12273-021-0864-x>
13. Love, C., Bozoian, S.: Defining energy efficiency measures and their impacts (2022)
14. Morrison, H.: Building envelope thermal bridging guide. <https://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/power-smart/builders-developers/building-envelope-thermal-bridging-guide-v1-6.pdf>, last accessed 2023/12/18
15. Chidiac, S.E., Catania, E.J.C., Morofsky, E., Foo, S.: A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings. *Energy Build.* **43**, 614–620 (2011). <https://doi.org/10.1016/j.enbuild.2010.11.002>
16. Chen, Y., Guo, M., Chen, Z., Chen, Z., Ji, Y.: Physical energy and data-driven models in building energy prediction: a review. *Energy Rep.* **8**, 2656–2671 (2022). <https://doi.org/10.1016/j.egyr.2022.01.162>
17. ASHRAE: Measurement of energy and demand savings. Guideline 14–2002. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA (2002)



Data-Driven Predimensioning: Applying Graph Neural Networks to Reinforced Concrete Design

Nils Schäfer^(✉), Jan-Friedrich Köhle, and Joaquín Díaz

THM University of Applied Sciences, Gießen, Germany
nilsschaefer12@gmail.com

Abstract. Predimensioning in structural engineering has traditionally been an experience-driven process that relies heavily on the skills of the engineer and general formulas. While machine learning methods, in particular Graph Neural Networks (GNNs), have been used for optimization in various engineering domains, their application to the predimensioning of reinforced concrete (RC) structures has been notably absent. This study fills this gap by developing a synthetic dataset specifically focused on the predimensioning of RC columns and slabs. The dataset is rigorously simulated using Finite Element Method (FEM) software, with reinforcement calculations that comply with deflection and concrete crack width limits closely aligned with Eurocode standards.

The generated data serve as an ideal basis for training, validation, and testing of GNNs. The primary goal is to automate the selection of optimal cross-sections for these RC elements. In doing so, this study pioneers the application of GNNs to the specific challenges of predimensioning in RC structures, thereby introducing a data-driven approach to this critical aspect of structural design.

Keywords: machine learning · artificial intelligence · graph neural networks

1 Introduction

Predimensioning is utilized to preliminarily determine component cross-sections and material quantities. The objective is to reduce the design space of the cross-sections to an economically and technically feasible range while meeting the structural requirements for load-bearing capacity and serviceability. This serves as a starting point for further iterative optimization and detailed calculation of the structural elements. Traditionally, predimensioning structures has relied heavily on personal knowledge and experience, limited to proven design methods and rarely incorporating larger data sets. General estimation formulas also result in imprecise dimensioning, requiring increased iterative optimization later in the project and consuming unnecessary time and resources. During the design phase, traditional predimensioning, long simulation times for optimization algorithms, and frequent design changes create a contrast, which often results in suboptimal designs and limited exploration of the design scope.

While Machine Learning (ML) and Graph Neural Networks (GNN) have been applied in various engineering disciplines, their potential in predimensioning RC structures remains largely unexplored. This paper investigates the use of GNNs for simulating

RC structures, aiming to pioneer data-driven predimensioning in this field. The study involves training a simulation GNN on synthetic FEM data, integrating Eurocode standards for deflection, crack width, and stability. The generation of the dataset is handled by a programming interface followed by a Graph representation of the structural models. Additionally the paper compares two GNN architectures and evaluates their effectiveness in simulating building structures. Integrating the trained GNN into existing optimization processes allows data-driven pre-dimensioning, enhancing the (early-stage) design of structures.

The goal is to introduce data-driven predimensioning into the design process of RC structures, allowing for more creative and optimized solutions while incorporating structural engineering early on.

2 Related Work

The field of structural engineering is increasingly incorporating ML techniques for the dimensioning and design of building structures. Several studies have demonstrated the effectiveness of various ML approaches in this domain.

Ampanavos [1] developed a Convolutional Neural Network (CNN) based ML system for the automatic generation of 2D steel building floor plans, aimed at assisting architects during the initial design phase. This iterative system progressively refines the floor plan, though it faces challenges in accurately scaling up for larger structures.

Pizarro and Massone [2] introduced a deep-learning neural network trained on Chilean residential building plans for predicting the dimensions of RC walls. This model effectively forecasts wall thickness and length, aiding in preliminary floor plan design. Subsequently, Pizarro et al. [3] enhanced this method using a CNN and two regression models to predict wall dimensions and displacements more accurately, thereby facilitating faster floor plan decision-making.

Chaillou [4] employed Generative Adversarial Networks (GAN) [5] for architectural floor plan creation, introducing ArchiGAN, a statistical design generation tool that allows for user intervention in the design process. Liao et al. [6] utilized GAN for the structural design of shear walls, noting a significant speed advantage without quality compromise. Similarly, Fei et al. [7] developed a GAN-based method for designing RC-shear wall structures, automating the schematic design process with high precision and efficiency. Lou et al. [8] proposed an optimization method for shear wall design in tall buildings using Support Vector Machines (SVM), reducing analysis time and material usage. Zhao et al. [9–11] applied GNN and deep learning to optimize the design of various structural elements, including beam arrangements, shear walls, and beams and slabs within shear wall structures. Li et al. [12] described an automated process for collision-free reinforcement optimization using GNN and Genetic Algorithms (GA), noting a significant efficiency increase.

Fisch et al. and Kraus [13, 14] showed that various ML algorithms, including Physics-Informed Neural Networks (PINN), are effective in the (pre)dimensioning and design of steel structures, with PINNs performing comparably or even better to established methods.

Chang and Cheng [15] developed a novel approach in structural engineering by conceptualizing structural models as graphs and utilizing two GNNs to determine the optimal

cross-sections for steel beams and columns. Their methodology involved representing the beams and columns as nodes in a graph, with their connections symbolized as undirected edges. In their suggested Optimization workflow, the first GNN, NeuralSizer, was tasked with suggesting suitable cross-sections for these structural components, guided by feedback from a second GNN, NeuralSim. Neural Sim's primary function was to simulate the horizontal movements of the building structure during seismic events.

One of the key benefits of this ML architecture was the accelerated and accurate estimation process of NeuralSim compared to traditional simulation methods. Chang and Cheng's approach demonstrated impressive scalability in three dimensions, addressing a common limitation observed in similar methods. Additionally, their results were similar to those obtained through conventional structural dimensioning techniques.

Despite these advancements, there remains a significant gap in the application of ML and GNNs within the field of structural engineering, especially concerning the predimensioning of RC structures. While there have been successful applications of these technologies in steel structures and the optimization of RC elements and floor plans, their use in the predimensioning phase for RC has not been explored. This research seeks to fill this gap by developing a GNN capable of being incorporated into existing optimization workflows, facilitating a data-driven and automated approach to the predimensioning of RC structures.

3 Methodology

The methodology of this study is structured into two segments: Data Generation and the GNN Application/Training.

3.1 Data Generation

In this research, we utilized a dataset originally derived from Chang & Cheng's work [15] on 4,000 buildings. This dataset was subsequently modified to better align with the specific requirements of RC design. Modifications included adjusting story heights to 3.5 m, varying slab span widths from 5.0 to 7.5 m, applying a concrete strength class of C30/37 XC1, setting slab thicknesses between 18 and 34 cm, and square column cross-sections between 30 and 60 cm. This dataset presents a base story replicated up to 10 stories and encompasses various loads and load combinations.

The loads on the structures include the self-weight as well as permanent and variable loads for the roof (permanent load (G) = 2 kN/m², variable load (Q) = 1kN/m²) and floors (G = 1.2 kN/m², Q = 5 kN/m²). Load combinations were formed for the ultimate limit state (ULS) and the serviceability limit state (SLS). Analog to Eurocode 2, the calculation of the RC slabs in the GZG was conducted under quasi-permanent load combination with criterion l/250. The column calculation was conducted considering stability, crack width limitation, and load transfer (Fig. 1).

The structural models consist of RC columns and slabs. The columns were created based on the endpoints of the scaled vertical lines of the data set. Ground floor columns were fixed, and all columns were assigned a random cross-section. The randomness minimizes the included bias and enables the GNN to effectively learn the influences of

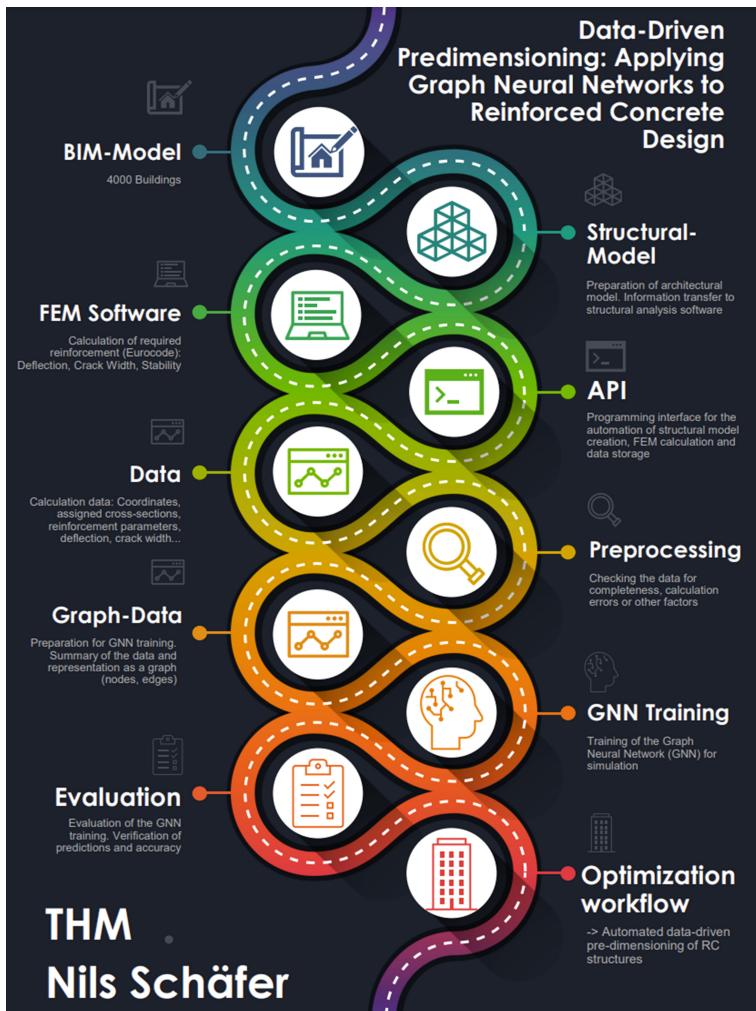


Fig. 1. Methodology overview

the individual parameters based on the diverse data. The buckling length coefficients required for stability were set to 0.7 or 1.0. The RC slabs were created based on the ordered point series of horizontal edge lines of a story, to which a deflection profile was assigned in addition to the slab thicknesses. This deflection profile was based on the shortest distance between the supports and the l/250 criterion (EC2-1-1). The indexing of the individual structural elements was saved for later tracking.

FEM-Simulation. The Simulations were carried out with Autodesk Robot Structural Analysis (RSA) and the simulation results served as labels for GNN training. The automated generation of the simulation results was conducted using the RSA API (Application Programming Interface) written in (Iron-) Python 2.7. The applied calculation uses a simplified, linear-elastic calculation algorithm with equivalent stiffness,

in which the change in stiffness as a result of cracking is considered. The calculated deflection is the product of elastic deflection and stiffness coefficient. This method leads to conservative deformations compared to the continuous stiffness update/verification of the program.

Calculation results for RC slabs include the required reinforcement in the X and Y directions, upper and lower layers respectively, as well as the deflection in cm. Calculation results of the RC columns include the required reinforcement concerning the width and height of the column as well as the percentage of reinforcement. In addition, calculation errors were saved as well as their identification (RC slab/column—finite element). The building identification as well as the calculated shortest distance, the assigned deflection profile, the number of slabs and columns, the design forces/momenta, and the calculation/simulation times were also saved.

Preparation and analysis. Post-simulation, data was saved in CSV files for GNN training, including necessary features and label information. The preparation involved examining the data, particularly for calculation errors, and organizing the results. Python libraries *pandas*, *numpy*, *matplotlib*, *seaborn* were used for data arrangement, formatting, and visualization. Classification labels for the slabs were determined based on the calculated deflection and the EC deflection criterion 1/250. For the columns, the percentage of reinforcement of 4% was used as the threshold for binary classification. Regression labels for RC slabs were determined by averaging the values of all finite elements in each slab, while the regression labels of the columns remained unchanged. Both regression labels are shown below:

Slabs: [Bottom layer X (cm^2/m), Bottom layer Y (cm^2/m), Top layer X (cm^2/m), Top layer Y (cm^2/m)]

Columns: [along the width (cm^2/m), along the height (cm^2/m), reinforcement ratio (%)]

3.2 Graph Neural Networks

Graph Creation. The conversion of the simulation into graph data was executed using Python 3 and the NetworkX library. In this process, each building was modeled as a directed graph, comprising nodes and edges, through the utilization of Python's class objects. The RC columns and slabs were represented as graph nodes and the connections of the structural elements as graph edges. Node Information includes coordinates, cross-sections, structure type (column/slab), load, story, calculation errors, deflection criterion, and reinforcement results either as numerical values or one-hot-encoding. In addition, a pseudo-ground node was created, with the feature vector values set to 0. The connections between the slabs, columns, and the pseudo floor node were represented with directed edges. The directed edges mimic gravitational load transfer and therefore include physical-related information. Graph data was saved as lists of dictionaries for efficient representation and access. All relevant information was represented within the specific feature vector of each node.

GNN training. GNN Training was conducted using Google Colab, leveraging Nvidia T4 GPUs and CUDA for GPU-accelerated computation, Pytorch, and PyTorch Geometric extension specifically designed for GNNs. During data pre-processing, the necessary

information was manipulated, modified, and represented as node-specific feature vectors. Features that originally had values greater than 1, such as the coordinates, were min-max scaled. The scaling minimized the disproportionate influence of larger feature values on the final GNN result. Following the scaling process, the data was transformed into PyTorch tensors. This included the creation of an edge tensor, which represented structural connections, and a node tensor for structural elements, both of which were formed using the existing feature data. *The scikit-learn library* was used to split the data set into training, validation, and test parts. The ratio of 80/10/10 (%) as well as a fixed random state of 20, which ensures the reproducibility of the data segmentation, was used. In addition, the order of the data in the training part was randomized to improve the robustness of the model.

The regression labels for the RC slabs and columns were merged and individually scaled for each index to create the final regression label incorporating the calculated required reinforcement values:

[(Slab Reinforcement-)Bottom layer X (cm^2/m), Bottom layer Y (cm^2/m), Top layer X (cm^2/m), Top layer Y (cm^2/m),(Column Reinforcement-) along the width (cm^2/m), along the height (cm^2/m), reinforcement ratio (%)].

For the classification label, the deflection criterion (slabs, 1/250) and the reinforcement ratio (columns, < 4%) were used.

Two different network architectures were investigated: Graph Convolutional Network (GCN) [16] and Graph Attention Network (GAT) [17], each consisting of three layers. The layers transform the feature vector of each node into sizes 128, 64, and 1 (binary classification) or 7 (regression). After the first and second layers, the ReLu activation function is used. A one-time dropout was applied, which set random inputs to 0 to prevent overfitting. The implementation of the Adam optimization was performed with learning rates in the range of 0.01 to 0.00001, a weight decay of 0 to 5e-4, a batch size of 1, and an epoch number between 5–10. These parameters were systematically varied using the grid search method to minimize the risk of remaining in a local minimum and maximize the likelihood of achieving the global minimum of the loss function. The selected loss functions were L1-loss (mean-absolute error) for regression and BCE-loss (binary-cross-entropy) for binary classification. The losses were consistently computed throughout the training process, utilizing both the training (80%) and validation data set (10%) for ongoing model evaluation. Upon completing the training, the test data set (10%) was used to assess training performance.

4 Results and Discussion

4.1 Results

Performance of GNN Architectures. Both GCN and GAT architectures were evaluated for their effectiveness in modeling RC structures. The GAT architecture marginally outperformed GCN in terms of classification accuracy (96.798%) and regression loss (~1.37% average deviation). These results align closely with those of NeuralSim referenced in [15], indicating the robustness of GNN in structural analysis applications (Tables 1 and 2).

Table 1. GNN Trainingsresults RC Design

| Model | L1 Loss x 1e-3 | Classification accuracy |
|-------|----------------|-------------------------|
| GCN | 13.74 | 96.19 |
| GAT | 13.73 | 96.798 |

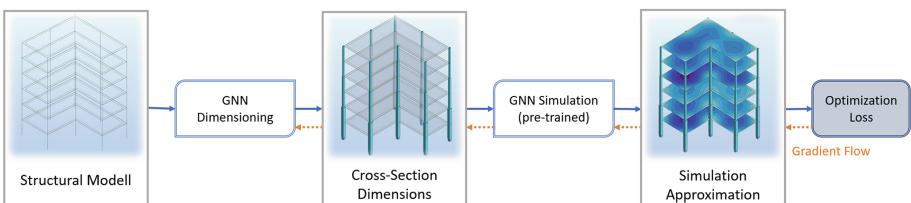
Table 2. GNN Trainingsresults Reference Literature [15]

| Model | L1 Loss x 1e-3 | Classification accuracy |
|------------|----------------|-------------------------|
| GCN | 1.601 | 89.22 |
| GAT | 1.087 | 93.35 |
| Neural Sim | 0.757 | 95.64 |

Classification Accuracy. The binary classification accuracy of the GNN, particularly the GAT architecture, was noteworthy. With a classification accuracy of 96.798%, the model was highly effective in correctly categorizing structural nodes based on predefined criteria. However, the error rate of 3.202% and varying levels of specificity and sensitivity across different structural elements highlight areas for potential improvement.

Deviation Analysis. The GNN output showed variability in average deviation based on structural elements, with columns presenting smaller deviations compared to slabs. The training loss consistently decreased over epochs and batches, validating the effectiveness of the model training. L1 test loss for most data fell within 0.00–0.02, with some outliers noted, which requires further investigation.

Simulation Time Efficiency. Our research highlights the significant efficiency advantage of GNN over FEM simulations in estimating RC simulation results. While FEM simulations averaged between 2 to 25 min, GNN simulations required only 5–10 ms. This represents a speed increase of approximately 12,000 to 24,000 times, suggesting that GNN can be effectively integrated into iterative optimization processes for real-time feedback and optimization of structural cross-sections, as suggested in the referenced literature [15] and depicted in Fig. 2.

**Fig. 2.** Data-Driven Predimensioning of RC Structures (inspired by [15])

4.2 Discussion

Implications of High-Speed GNN Simulations. The drastic reduction in simulation time using GNNs opens new avenues for integrating AI in structural design optimization. The ability to rapidly iterate and optimize structural elements could significantly enhance design efficiency.

Comparative Performance of GNN Architectures. The close performance metrics between GCN and GAT architectures underscore the potential of both in structural analysis. The slight edge of GAT in accuracy and regression loss indicates its suitability for more precise applications.

Addressing Deviations and Outliers. The observed deviations and outliers in GNN output, particularly in slab reinforcement, underscore the need for model refinement and validation against real-world data. Ensuring the reliability of predictions across all structural elements remains a critical challenge.

Enhancing Classification Accuracy. While the GNN demonstrates high classification accuracy, the presence of both false positives and negatives, especially in slab classifications, indicates the need for further model tuning. Understanding the underlying causes of these errors will be vital for improving the model's predictive accuracy.

Future Research Directions. Future research should focus on expanding the data sets, both real and synthetic, to encompass a wider range of structural scenarios. Incorporating more diverse data would likely improve the model's accuracy and robustness, making it more reliable for practical applications. Further, exploring the integration of GNN with other optimization methods could yield even more efficient design processes.

Limitations and Considerations. It is important to acknowledge the limitations inherent in our dataset and methodology, particularly the focus on specific structural elements and the exclusion of factors like creep and shrinkage in RC. These limitations suggest caution in generalizing our findings and highlight the need for continued research and validation.

5 Conclusion

In this paper, the use of a GNN for the simulation of RC structures was investigated. The aim was to explore the possibilities of data-driven pre-dimensioning by training a GNN for the simulation of RC structures. The research results demonstrate the potential of the GNN approach. The simulation GNN trained on a synthetic FEM dataset shows a classification accuracy of $\sim 96.8\%$ and a low average deviation of $\sim 1.3\%$. This means that compliance with the Eurocode limits for deflection, crack width limitation, and stability for the RC columns and slabs was predicted 97% correctly and the theoretically required reinforcement was predicted with an average deviation of $\sim 1.3\%$.

The time advantage of the simulation GNN compared to the FEM simulation is particularly remarkable. The simulation GNN is up to 24,000 times faster and can predict reinforcement data and compliance with Eurocode limits in 5 ms. The integration of the pre-trained simulation GNN into the existing, adapted optimization process (Fig. 2) enables automated pre-dimensioning of RC structures.

The limitations can be categorized into three areas and should be addressed in future studies. I. The quality of data-based pre-dimensioning is largely dependent on the underlying data set. The creation and publication of data sets that include the entire complexity of structures should be prioritized. FEM simulations can be used for the required amount of data and real projects can be used to reduce the distortion of the data set. II. The possibility of using ML methods in the design phase of structures, which includes the dimensioning of the components, but also the materials, connections, and localization of the structural elements as well as the constructability, should be investigated. III. The optimization process should include the consideration of various optimization objectives (e.g., economic efficiency, sustainability).

The use of automated, data-driven pre-dimensioning enables efficient exploration of the large design scope in the design phase. The complexity of modern buildings combined with the constant striving for efficiency and sustainability requires more powerful analysis and design tools. ML and GNN can help to automate iterative processes and make the design process more efficient. The use of automated, data-driven pre-dimensioning enables efficient exploration of the large design scope in the design phase. The efficiency and speed enable the evaluation of many different approaches, which are not or only to a limited extent comparable with traditional methods due to a lack of time and resources. This paper intends to contribute to the development of data-driven pre-dimensioning and to encourage the integration of digital methods, in particular artificial intelligence (AI), into everyday structural engineering.

References

1. Ampanavos, S., Nourbakhsh, M., Cheng, C.-Y.: Structural design recommendations in the early design phase using machine learning. In: Gerber, D., Pantazis, E., Bogosian, B., Nahmad, A., Miltiadis, C. (eds.) Computer-Aided Architectural Design. Design Imperatives: The Future is Now. 19th International Conference, CAAD Futures 2021, Los Angeles, CA, USA, July 16–18, 2021, Selected Papers. Communications in Computer and Information Science, pp. 190–202. Springer Singapore; Imprint Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-1280-1_12
2. Pizarro, P.N., Massone, L.M.: Structural design of reinforced concrete buildings based on deep neural networks. Eng. Struct. **241**, 112377 (2021). <https://doi.org/10.1016/j.engstruct.2021.112377>
3. Pizarro, P.N., Massone, L.M., Rojas, F.R., Ruiz, R.O.: Use of convolutional networks in the conceptual structural design of shear wall buildings layout. Eng. Struct. **239**, 112311 (2021). <https://doi.org/10.1016/j.engstruct.2021.112311>
4. Chaillou, S.: ArchiGAN: Artificial Intelligence x Architecture. In: Yuan, P.F., Xie, M., Leach, N., Yao, J., Wang, X. (eds.) Architectural Intelligence. Selected Papers from the 1st International Conference on Computational Design and Robotic Fabrication (CDRF 2019). Springer eBook Collection, pp. 117–127. Springer Singapore; Imprint Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-6568-7_8
5. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**, 139–144 (2020). <https://doi.org/10.1145/3422622>
6. Liao, W., Lu, X., Huang, Y., Zheng, Z., Lin, Y.: Automated structural design of shear wall residential buildings using generative adversarial networks. Autom. Constr. **132**, 103931 (2021). <https://doi.org/10.1016/j.autcon.2021.103931>

7. Fei, Y., et al.: Integrated schematic design method for shear wall structures: a practical application of generative adversarial networks. *Buildings* **12**, 1295 (2022). <https://doi.org/10.3390/buildings12091295>
8. Lou, H., Gao, B., Jin, F., Wan, Y., Wang, Y.: Shear wall layout optimization strategy for high-rise buildings based on conceptual design and data-driven tabu search. *Comput. Struct.* **250**, 106546 (2021). <https://doi.org/10.1016/j.compstruc.2021.106546>
9. Zhao, P., Liao, W., Xue, H., Lu, X.: Intelligent design method for beam and slab of shear wall structure based on deep learning. *J. Build. Eng.* **57**, 104838 (2022). <https://doi.org/10.1016/j.jobe.2022.104838>
10. Zhao, P., Liao, W., Huang, Y., Lu, X.: Intelligent design of shear wall layout based on graph neural networks. *Adv. Eng. Inform.* **55**, 101886 (2023). <https://doi.org/10.1016/j.aei.2023.101886>
11. Zhao, P., Liao, W., Huang, Y., Lu, X.: Intelligent beam layout design for frame structure based on graph neural networks. *J. Build. Eng.* **63**, 105499 (2023). <https://doi.org/10.1016/j.jobe.2022.105499>
12. Li, M., Liu, Y., Wong, B.C., Gan, V.J., Cheng, J.C.: Automated structural design optimization of steel reinforcement using graph neural network and exploratory genetic algorithms. *Autom. Constr.* **146**, 104677 (2023). <https://doi.org/10.1016/j.autcon.2022.104677>
13. Kraus, M.A., Taras, A.: Physik-informierte Künstliche Intelligenz zur Berechnung und Bemessung im Stahlbau. *Stahlbau* **89**, 824–832 (2020). <https://doi.org/10.1002/stab.202000074>
14. Fisch, R., Stecker, E., Kraus, M.A.: Maschinelles Lernen beim Entwurf und der Bemessung von Stahlrahmenhallen. *Stahlbau* **92**, 332–344 (2023). <https://doi.org/10.1002/stab.20200054>
15. Chang, K.-H., Cheng, C.-Y: Learning to simulate and design for structural engineering. *Int. Conf. Mach. Learn.* 1426–1436 (2020)
16. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks (2017)
17. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks (2018)



Neural Network Based Defect Classification in Real-Field Rail Inspection Data Augmented by Simulated Defects

Olm Georg^(✉)

Department of Civil Engineering, Technische Universität Berlin, Hardenbergstraße, Berlin,
Germany
georg.olm@tu-berlin.de

Abstract. Railway systems are an essential backbone of our society, facilitating transporting people and large quantities of goods across long distances. As the shift from cars to trains gains momentum, their reliability becomes even more essential. To ensure the efficiency of this transportation system, it is necessary to plan and execute regular inspection measures. In Germany, this testing is typically carried out using Ultrasonic sensor data from probes mounted on inspection trains.

The research presented in this paper is based on collaborative work between partners from different research fields and inspection practitioners. The aim is to leverage real-world Ultrasonic field data from inspection runs and augment them with simulated rail defects that would otherwise occur too infrequently to enable meaningful Deep Learning analysis. Given the spatial and temporal characteristics of the data, modeling approaches that can effectively capture both elements by employing convolutional and recurrent neural network layers are compared. Combining these two layers offers a suitable solution for identifying and classifying defects in typical sensor data from routine inspections. Evaluation of the Model usability evaluation happens on indicators reflecting the practical usability of such. Every approach will be validated against essential metrics, such as AUC in binary prediction or True and False Positive Rate for classification. Considerations for the Probability of Detection (POD) are also made, as this metric is essential to evaluate Non-Destructive Testing methods.

Keywords: NDT · Neural Network · Defect Simulation

1 Introduction

Artificial Intelligence is a wide field that plays an elemental role in various engineering disciplines. Its potential to improve railway inspection and maintenance has been researched in review studies, for example by Tang et al. [1] or Nakhaee et al. [2]. They observe a strong increase in publications and a large variance of applied methods. However, despite significant theoretical advancements and proof of concept studies based on laboratory experiments, the practical implications have been limited due to insufficient data and real-world validation.

Our research aims to bridge this gap by validating the effectiveness of Deep Learning models on real-world field data, which was obtained during actual inspection runs conducted by maintenance trains at speeds of up to 80 km/h in the German railway network. One of the significant challenges in this research is the need for a reliable ground truth, necessary for developing comprehensive Neural Network models, tasked with artefact classification within railway systems. This means measurement accurate identification of drillings, cracks, or other defects. The current evaluation of the rail is done on a broader scale, which does not necessarily demand such high detection resolution. This obliges a refinement of the evaluation data to create highly expressive labels.

This research uses Deep Learning models to process rail systems' Ultrasonic Testing (UT) image data. The primary target is to build a model that can accurately interpret and handle regular and noisy rail data, facilitating the effective maintenance of railway systems and increasing the level of automation in the process.

On top to the vast amount of actual field measurements, this research uses especially simulated defects. The goal is to tackle the scarcity of defects observed in the field. As the maintenance process wants to remove and prevent this type of occurrence in the rail, it is difficult to gather enough cases to sufficiently train a classifier neural network, even in a large dataset. We aim to construct models and systems capable of sophisticated rail condition monitoring, transcending beyond conventional defect detection to predictive and preemptive maintenance strategies.

2 Overview

There has been a strong increase in the utilization of research on the applicability of Machine Learning and AI. Different Review Papers provide a valuable overview of the current state of research and their applied methods. We can generally see an extensive application across all possible ML/AI methods in the domain of railway systems by Tang et al. [1]. Nakhaei et al. [2] specialized their review around rail track maintenance. They see ML methods as the primary means of conquering defect detection in railway maintenance. However, both reviews point to the general problem of imbalanced datasets, the need for proper labeling, and the low number of defects to train an AI classifier effectively. They also stress the need to use and publish openly available rail track data to increase reproducibility and trustworthiness in the research. One way to resolve the lack of defects is by using augmentation methods or synthesizing defect data to complement possible training datasets. Uhlig et al. [3] reviewed existing research on how this is already utilized for machine learning on ultrasonic non-destructive testing. Due to its massive potential to overcome the main problem of missing defect data a strong increase in conducted research can be seen, although it remains only a small fraction of the overall research. They also stretch the importance of a proper match between actual measurements and synthetic data to generalize the model's predictive value.

Overall, the problem of having enough field measurements and defect data for meaningful ML and AI research appears evident and is an overarching topic in the research field.

Some research groups, such as Ha et al. [4] and Mahajan and Banerjee [5], create faulty specimens in the laboratory to generate respective data. This approach, by design,

is limited in its ability to create variance in the data and might differ from what would have been measured in the real field. Li et al. [6] collected a limited number of UT measurements on a real track, which were used to finetune an EfficientNet-b7 model. Although the research is limited by its small amount of data and differs from efficient real-world measurements, it resulted in performance values that can serve as a comparison for future models.

Our Project AIFRI aims to solve these issues in many ways. This research uses maintenance data from regular inspection runs in Germany across around 944 km from 38 inspection runs with two rail sides each. It is the same data that Deutsche Bahn evaluates to identify points of further examination. This provides the possibility to train a model with enough real field data, but it still has the same lack of meaningful defect data. Therefore, our cooperation partner, the Federal Institute for Materials Research and Testing (BAM), simulates ultrasonic measurements that represent defects in rails and can capture a large variance in defect appearance. This serves as a complement to the actual field data in model training. Implementing both data into a Deep Learning pipeline can lead to an answer on how to create better models for detecting and classifying defects for railway maintenance.

3 Research Method

3.1 Data Collection

The measurement process currently employed in the German railway system fulfills the requirements based on ISO 16729-1 [7] and contains 8 UT-Probes, which measure under different angles (-75°, -55°, -35°, 0°, 0°, 35°, 55°, 75°). Every probe is supposed to detect reflectors in different positions. Armbruster et al. [8] describe the system employed on Deutsche Bahn maintenance trains. The goal is to combine different probe signals and create a coherent image for manually evaluating the testing data. Due to the combination of multiple probe signals, we have properties that can be identified as spatial because they represent an image across the rail. Additionally, the results can be interpreted as sequential probe measurements along the rail, just like a time series.

Simulated Ultrasonic signals are created by BAM using the CIVA 2023 software for modeling and simulation. The method is currently discussed by respective domain experts in Germany and standardization institutions [9]. It follows the standards DIN EN 16729 -1 [7] and has been prepared to be as similar to the field measurements as possible.

3.2 Model Comparison

Specific layer structures in Neural Networks can capture individual pattern. Typically used are convolutional neural networks (CNN) for spatial data or recurrent neural networks (RNN) for temporal data [10]. As described earlier, the data in this research can be interpreted as either. This means different methods and architectures to model field data and synthetic simulations need to be considered. To examine which architecture fits our use case best I compare architectures suitable for learning the data's temporal and

spatial properties in one combined model. These will additionally be compared to the two basic constructions of plain LSTM and CNN networks.

A typical combination of both elements is a ConvLSTM Layer—with LSTM as a special RNN- typically used to capture temporal dependencies between images [11]. However, the goal is to examine how both properties can be extracted from the same 2D array, why other architectures need to be compared. The most promising architecture will be increased in depth to improve the predictive performance. This means an increase of overall trainable parameters, either by increasing layers or units per layer.

Besides comparing plain LSTM and CNN, I will examine model performance on the following architectures. For a **Linear Combination**, the input data goes to one kind of layer first and then serves as input for the next kind of layer. For this, we have the option *a*) with LSTM first and CNN second or *b*) CNN first and LSTM second. Both layers can also be arranged in **parallel** *c*). The temporal and spatial properties will be trained in their respective layer and concatenated afterward before both trained patterns lead to the output layer. A typical but computationally expensive way to improve learning from 2-dimensional arrays with CNN is introducing an **Inception** layer block into the model *d*). The combination is like the parallel variant but can improve pattern extraction by introducing multiple convolutional kernels with different sizes. A simplistic overview of the model architectures to be compared can be seen in Fig. 1. Those combinations of multiple layer types can be observed across domains [12–14].

An additional level of robustness is brought into the models by adding batch normalization after the CNN Layer and a dropout rate of 0.2 to the LSTM layer. Because ultrasonic measurements are not limited to one direction, LSTM layers are implemented bidirectionally, training patterns in both forward and backward directions. This is done to prevent the model from overfitting and performing poorly on the test dataset.

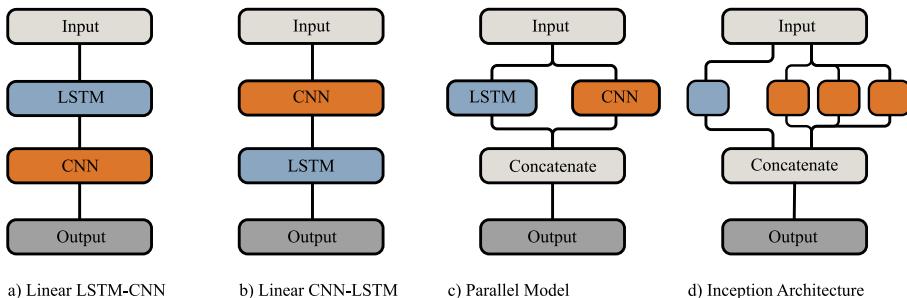


Fig. 1. Overview of model architecture schematic

3.3 Further Consideration

To examine the performance properly, I do a rough 60/20/20 train/test/valid split of the data. Splitting the data happens semi-randomly to guarantee as strict separation between splits. A completely random separation by label is not meaningful, because they sometimes are very close together, can overlap and cause data leakage. I split

the data based on inspection runs and simultaneously keep the types and classes evenly balanced in between. Data leakage is a serious issue, leading to unrealistically optimistic model performance when the model gets premature access to information in the test or validation dataset during training.

The first performance comparison is based on the in-training loss on the testing data set. This describes the progress of optimizing the training loss on data that does not explicitly serve as training data. The gap between loss in training and testing data shows whether the training process improves general model performance or simply overfits the training data. Final evaluations need to consider a more specific set of metrics that capture the real-world considerations. They are based on a third validation data set that could not influence the training process.

3.4 Metrics

Due to the complex consideration of the model's real-world use and the multivariate output, it is necessary to consider multiple metrics to evaluate its performance. Following, I will describe the most important aspects to evaluate the models fit to its intended use case.

Area Under the Curve (AUC) is a critical metric for binary detection tasks. It measures a binary classifier's overall performance by comparing False Positive Rate and False Negative Rate at different decision thresholds. 1 stands for perfect predictability, and 0.5 can be considered random decision-making. This metric is helpful to observe the capability to differentiate between an artefact and normal rail without explicitly classifying a specific type. As the most significant part of our data is recorded under real field circumstances, it is likely that a lot of noise can cause a high False Positive Rate.

The capability of type prediction can be monitored in a confusion matrix. Classification of each type will be compared against its true value. This can be used to identify if specific types are more correctly or falsely classified than others. Tackling real-world problems, it is crucial to compare how many defects are correctly classified, as missed defects can cause massive problems in the railway and would fail the concept of Ultrasonic Testing in rail maintenance. Therefore, I will look at the **True Positive** and **False Positive rates** of defect classification.

Following this Matrix, it is also insightful to look at the overall classification performance in terms of a binary confusion matrix, which only differentiates between artefact and normal rail. Understanding how many artefacts are correctly identified (**True Positive Rate**) or incorrectly identified (**False Positive Rate**) and cause additional work. These metrics can serve as an extension to the ROC but do not have a threshold to adjust for in a type-based binary classification.

Prediction error of defect characteristics: Further metrics, which will be essential to understanding the model but are of minor importance for now, are the error of how well the model predicts properties such as length, height, and angle of defects.

POD is one of the main target metrics in this project. The project's purpose is not only to create a NN for maintenance purposes but also to improve the maintenance system as a whole. In Non-Destructive Testing Systems, it is typical to look at the Probability of Detection to validate the usability of a system. Ultimately, it considers the interaction of all aspects of a maintenance system, i.e., the testing system, automated detection

algorithm, evaluators, and intrinsic detection capabilities [15]. Therefore, considering the final model's capabilities to improve the POD is essential.

With this focus, we conduct a comparative performance evaluation of Deep Learning Architectures with an introduction of simulated defects into real-field Ultrasonic measurements.

4 Data & Results

The data used in this research resembles a 2D array of greyscale values, also known as P-Data from Ultrasonic testing, in a resolution of 3x3 mm per pixel on a pixel array of size 115x64. This represents 115 measurements along the rail distance every 3mm and 64 measurements for every 3mm rail height. A visualization can be seen in Fig. 2. The goal is to predict values that describe each measurement along the rail in the driving direction. The primary information of interest is listed in Table 1, with the value range in squared brackets.

The value range of raw data is 0–255 as integer values, resulting from combining multiple probes into a coherent visual. Simulations create a very clear signal without any background noise. To create simulations closer to real data, background noise has been randomly sampled from the field measurements and added to all values without a clear reflection signal. Variations of simulations range from randomly sampled noise to different sizes of defect and three different rail profiles, i.e., profiles 49E5, 54E4, and 60E2.

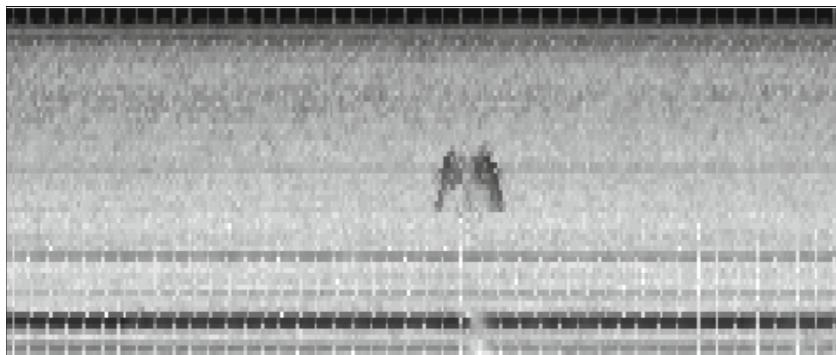


Fig. 2. Sample image of a drilling measured in real field

4.1 Data Preparation

Model training is executed on a set of ultrasonic data containing regular artefacts, simulated defects, and normal rail. In order to ensure a concurrent label structure across inspection runs and simulated data, I relabeled the field data in a specific app, based on the R Shiny framework. In order to choose labels that are useful for training, I decided to

filter for the artefacts that are clearly identifiable and comparable to the simulated data. This contains drillings and welds as artefact types.

The target variables the model is trained to predict can be seen in Table 1. In conclusion, we receive a model output array of shape 115x5. Multiple outputs demand multiple loss functions to be optimized. The loss functions for each variable can be seen in Table 1. The highest priority is detecting artefacts, so I introduce a callback in the training routine that monitors the learning progress based on the loss function of **is_artefact** on the testing data. It stops after 20 epochs of no improvement and returns the model weights corresponding to the best-performing version. Type accuracy is also a vital loss function for our task. However, the remaining class imbalance makes monitoring accuracy per se a challenging condition.

Table 1. Label types and loss

| Variable | Value type | Description | Loss Function |
|-------------|-----------------|--|---------------------------|
| Is_artefact | Integer [0,1] | Is there an artefact at this point of measurement | Binary cross entropy |
| Type | String | What type of indication is located at this point | Categorical cross entropy |
| Height | Integer [0–30] | What is the height of the artefact | MSE |
| Length | Integer [3–24] | If there is a defect at the artefact, what is its length | MSE |
| Angle | Integer [0–180] | What is the angle, a defect extends in the rail | MSE |

4.2 Pre-Processing

In order to train a model efficiently, it is generally advised to normalize the data. Depending on the activation function that is used in the modeling process, the raw data will be normalized between 0–1 in case of using the relu activation in CNN or –1 to 1 in case of tanh activation in LSTM layers. Target variables, except the type variable, will be normalised between 0–1. The type variable is a string and will be one-hot-encoded.

Pixel-wise prediction is likely to cause false results that are unimportant when looking at longer sequences as a whole. For example, if in a 115-value long sequence, ten values can be predicted with a correct type and two values cause a false negative, it would make no difference in practice. Usually, an evaluator can skim through multiple values and look at a sequence as a whole. Therefore, an evaluation based on a complete sequence, in addition to the pixel-wise classification, will be done.

The simulated data is always a 999 mm long sequence containing a drilling with a crack. The drilling is always located at the sequence center point. This representation deviates from real field data, where artefacts can occur everywhere in the measurement. To increase the variability of the data and robustness of the model, each standard sequence

will be transformed into multiple smaller arrays of shape 115x64 pixels by a moving window with a stride of 10 pixels. To prevent the model from being trained on the order of those sequences, they will be shuffled between all sequences created per input batch. Additionally, the order of input data will be shuffled after each epoch. The total amount of sequences and their corresponding types, assigned by priority of detection, can be seen in Table 2. To maintain structure parity between simulated and field data, the same windowing happens with measured data.

Table 2. No. sequences by type

| Artefact Type | # Total Sequences | # Validation Sequences |
|----------------|-------------------|------------------------|
| normal | 123,030 | 29,042 |
| other | 3,923 | 631 |
| weld | 26,962 | 6,080 |
| drilling_crack | 1,211 | 286 |
| drilling | 122,736 | 29,804 |

4.3 Model Results

From the metrics, we can compare how each model performs on the main tasks, first to classify defects and second not to cause additional false positive signals for the evaluators, which would cause additional workload. The results can be seen in Table 3. Some considerations need to be mentioned upfront. The classification of defects, in general, was so reliable that the immediate occurrence of a defect caused a detection. As no defects, independent of length, were missed, no real POD could be calculated and is therefore not considered in the table. Defect characteristics, such as angle, length, and height are not evaluated. No model was able to capture a meaningful pattern on that characteristic and predictions were close to randomness. To keep the resulting metrics table clear, they will not be considered here. Further consideration on this topic will be part of the discussion.

CNN-LSTM and LSTM architectures have the lowest loss in the training process, indicated by the loss on the test data set. Both models achieve similarly good results at the AUC on the measurement stepwise prediction of artefact occurrence, with the LSTM model in a very narrow lead. The same model can also classify all sequences containing defects with only six false positives from 65,843 classified sequences. Only the much more complex Inception architecture can deliver comparable results. Considering all artefacts, the Inception architecture has the best True Positive Rate with 0.952, just before the parallel architecture and the less complex LSTM model. The best accuracy on artefact detection has the CNN-LSTM model with 0.797. This aligns with what could be achieved with more complex models in different research [6]. In this experiment, LSTM-CNN and CNN models failed to predict the respective patterns and defaulted to output only normal rail. This might be due to a misconfiguration in the model architecture.

Table 3. Comparison of Model Metrics

| Model | Test-Loss | AUC | Defect TPR | Defect FPR | Artefact TPR | Artefact FPR | Artefact Accuracy |
|-----------|--------------|--------------|------------|------------|--------------|--------------|-------------------|
| LSTM | 0.145 | 0.959 | 1 | (6) | 0.917 | 0.381 | 0.785 |
| CNN | 0.332 | 0.5 | 0 | 0 | 0 | 0 | 0.441 |
| LSTM-CNN | 0.332 | 0.5 | 0 | 0 | 0 | 0 | 0.441 |
| CNN-LSTM | 0.137 | 0.957 | 0.965 | 0.002 | 0.815 | 0.225 | 0.797 |
| Parallel | 0.179 | 0.931 | 0.734 | 0 | 0.922 | 0.403 | 0.779 |
| Inception | 0.197 | 0.888 | 1 | (2) | 0.952 | 0.505 | 0.750 |

Brackets indicate explicit values, instead of rates, in case the result is still 0 after rounding at three digits

Future research can focus on rebuilding this architecture or switch to a different output, as CNN networks are usually able to achieve better results.

Overall, the LSTM model seems to be the most promising. It delivers the best results at AUC and detects all defect sequences while maintaining a very low FPR. It is not the best in classifying artefacts, but usually performs closely behind the much more complex architectures. As its results are either best or closely behind while being less complex, it should be chosen as a model to deepen research on this kind of data. The combination of LSTM and CNN layers in a neural network can increase performance in some cases but not in all. As it doesn't seem to prove a significant improvement across the board, it is reasonable to choose the model with less complexity and continue experimentation with a deeper version of that model.

4.4 Further Exploration

Following the previous chapter, it can be useful to take a closer look at the LSTM model's analytics during training and prediction. In Fig. 3 the loss curves on training and testing for the most important variables across epochs are visualised. A very fast overfitting can be identified when focusing on the loss of is_artefact and type_loss. Longer training epochs appear to not increase the performance. Implementing robustness during training might enhance the overall model performance.

In Table 4, the structure of the LSTM model shows the considerations and depth of each layer. For every model, three trainable layers of LSTM or CNN were configured to achieve a minimal meaningful depth for training. Increasing the trainable parameters can increase the model performance when robustness during training is considered.

5 Discussion

5.1 Performance and Limitations

Introducing simulated defects into real-field ultrasonic measurements made it possible to create Neural Networks to classify simulated defects and artefacts from the rail. A separate hold-out set of simulated data was correctly classified on a sequence level,

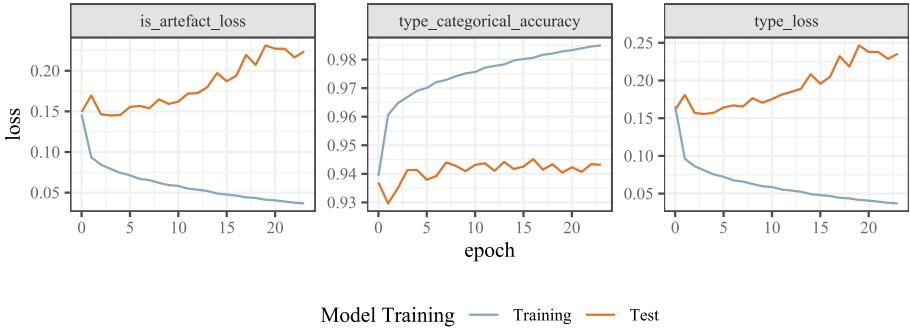


Fig. 3. Loss on training/testing data across epochs

Table 4. Model Structure and Parameter

| Layer | Output Shape | Param # |
|-----------------------|------------------|---------|
| Input | (None, 115, 64) | 0 |
| Rescaling | (None, 115, 64) | 0 |
| Bidirectional-LSTM | (None, 115, 230) | 165600 |
| Bidirectional-LSTM | (None, 115, 230) | 318320 |
| Bidirectional-LSTM | (None, 115, 230) | 318320 |
| Dense | (None, 115, 115) | 26565 |
| Output: Is_artefact | (None, 115, 1) | 116 |
| Output: Type | (None, 115, 6) | 696 |
| Output: z_artefact | (None, 115, 1) | 116 |
| Output: a_defect | (None, 115, 1) | 116 |
| Output: length_defect | (None, 115, 1) | 116 |

Total parameter: 829,965; Trainable parameter 829,965; Non-trainable parameter: 0

see Table 3. All defects could be successfully identified. However, the performance on classifying simulated defects appears too good. I expected to see a larger number of defect misclassifications. This is reasonable to assume, as the simulation data has been split into smaller sequences naively. Some sequences should contain very little of the actual defect and therefore are likely to be misclassified. It is probable that simulations from the validation set were still too similar to simulations in the the training data and not similar enough to the remaining field data. This Is why an overfitting on training data led to bad generalisation to the validation data, despite the effort to create multiple distinct variations with noise sampled from the field data.

Further generalization test for the model are needed to investigate its real-world performance. It needs to be examined whether the model can differentiate between simulations or real artefacts, i.e., is the model trained to detect defects or detect simulated data? There are multiple ways to test for that. First, defects from the real field need to

be classified in a validation set, without explicitly being in the training data. The small number of defects, that was found by regular inspections should be used as a specific validation holdout. They very likely are not enough to contribute as training material but are very valuable to assess generalization capabilities. Second, it is necessary to simulate regular drillings and welds without defect. For this case, we have a lot of real-world data to train, test and validate and it creates the possibility to directly compare the performance on simulation and actual measurements. It is not possible to make a claim about the model's capability of generalization, based on the current data, as types are not mixed between simulations and real field data. They are either completely simulated or completely measured. Additionally, it needs to be stated that the amount of simulated data is still very small. More variations, higher noise levels, and different types of defects should be added to the training to create a more comprehensive overview of its capabilities.

Misclassification can be observed for other classes. When summarizing into general classes as artefacts/defects and normal rail, we can observe an accuracy of 0.785. The rate of false positives, which can be interpreted as unnecessary noise from the evaluation, is 0.381, and 8.3 percent are missed sequences, which would have contained a drilling or a weld that was not correctly classified, given by the inverse of True Positive Rate. The reasons for misclassification still need clarification, and a deeper look into the data is required. A possible reason could be the data quality. Real-world data suffers from noise and difficult measurements. Outside influence such as manual adjustment of probe sensitivity or partial coupling loss can heavily influence and distort the measurements. It seems promising to take a closer look at these factors, because there might be a connection between noise and the quality of classification. A quantifiable noise threshold, above which predictions are much more likely to fail, could lead to adjusting the measurement process and possibly improve data quality.

POD, which is an important metric in Non-Destructive Testing, could not be meaningfully calculated based on the current defect classifications. As every defect was detected, the POD was 1, independent of the defect's characteristics. I already suggested the addition of more variation and different noise levels into simulated defects. Simulations always contain very clear signals that are supposed to be identified easily. Implementing multiple steps of stronger noise in the simulation could make the defects harder to detect and enable the determination of a meaningful POD for bad measurement environment.

5.2 Future Research

The current state of the research leaves many options for possible improvements. Besides the limitations mentioned in chapter 5.1 which focus on data variation and examining the question of the models generalization capabilities, there are additional factors that are woth to be researched and introduced in the models architecture and training pipeline.

Next to the greyscale-/P-Values used in this research, B-Gate values are available and can be implemented into the modeling process. They are in a range of 0 or 1 and describe the exceeding of a pre-defined value from the measurements. This way, the model could be trained to focus more on critical and high values and reject noise as

unimportant. This is viable opportunity, as both datasets are usually apparent after an inspection run and should be examined.

In the next optimization step, the still apparent class imbalance could be tackled by introducing well-elaborated sample weights in the training process. This can put a greater focus on underrepresented classes. However, it might only have a small effect on this research because the minor group of defects is already reliably detected. Nevertheless, it would make for a more robust training pipeline and allow for arbitrary addition of new data.

Proper Hyperparameter tuning might further enhance the performance. Parameters like the dropout rate, number of layers, or number of units per layer can influence the predictive power of the model. However, we are limited with the number of units per layer because the model output is time-distributed and expects a specific shape of the previous layer. Additional modeling practices should be examined to see whether they further improve the performance of promising models. Residual steps ensure that no input data gets lost in the training process and can be added to a simple LSTM model. LLMs' popularity and architecture provide another opportunity that should be explored. Adding foundational LLM elements, such as embedding or attention layers to the architecture, could improve performance. All options need to be carefully introduced and researched so they do not negatively influence the architecture's performance.

Another common practice is not to build every model from scratch, mainly if the available data is limited, like the simulation data in this research. Utilizing pre-trained models, like Yolo, and finetuning them on a small amount of UT-Data allows for a very efficient approach to detect defects and should be considered for future research. These points should be explored in future research.

6 Conclusion

It can be concluded that modeling UT Data, even though they have temporal and spatial properties, does not necessarily demand incorporating both in their architecture. Using mainly LSTM-Layer in the model structure provides a comparatively proper solution to represent that type of data. Additionally, this research finds that even adding a small number of simulated defects enables the model to detect those defects reliably. This effectively helps to create a modeling pipeline that conquers the problem of defect scarcity in automated railway maintenance. It can be used to improve the primary metrics of Non-Destructive Testing and contribute to a more efficient railway maintenance system.

Nevertheless, this study is still narrow in its applicability, as the number of simulations and the variety of defect types need to increase to make a stronger statement about its predictive power. Further, generalization checks are necessary to evaluate whether the results are valid in the real world or if they are only working in an encapsulated modeling environment. Additional model and data exploration are needed to identify cases when the model correctly predicts or fails. This is needed to make more elaborated suggestions about its use and data quality implications.

Future work should build on this research and work out its limitations by increasing the number of simulations and introducing generalization checks. Afterwards, more robust and better-understood models can be introduced into the railway maintenance

process. This research provides a meaningful starting point for these goals. It enables future research to narrow its scope and focus on improving the data quality, synthetic defects, and their interaction in the modeling pipeline. This should result in better insight into how a model might differentiate between a simulation and real field defects rather than examining too much about the modeling architecture itself.

Acknowledgements. This research is part of the project AIFRI, which is funded by the German Federal Ministry of Digital and Transport as part of the mFUND innovation initiative with a total of € 2.416 million.

References

1. Tang, R., De Donato, L., Besinovic, N., Flammini, F., Goverde, R.M., Lin, Z., Liu, R., Tang, T., Vittorini, A.V., Wang, Z.: A literature review of Artificial Intelligence applications in railway systems. *Transp. Res. Part C: Emerg. Technol.* **140**, 103 679, (2022)
2. Chenariyan Nakhaee, M., Hiemstra, D., Stoelinga, M., van Noort, M.: The recent applications of machine learning in rail track maintenance: a survey, in Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification: Third International Conference, RSSRail 2019, Lille, France, June 4–6, 2019, Proceedings 3, Springer, pp. 91–105 (2019)
3. Uhlig, S., Alkhasli, I., Schubert, F., Tschöpe, C., Wolff, M.: A review of synthetic and augmented training data for machine learning in ultrasonic non-destructive evaluation. *Ultrasonics* **134**, 107041 (2023)
4. Ha, J.M., Seung, H.M., Choi, W.: Autoencoder-based detection of near-surface defects in ultrasonic testing. *Ultrasonics* **119**, 106637 (2022)
5. Mahajan, H., Banerjee, S.: A machine learning framework for guided wave-based damage detection of rail head using surface-bonded piezo-electric wafer transducers. *Mach. Learn. Appl.* **7**, 100216 (2022)
6. Li, W., Wang, J., Qin, X., Jing, G., Liu, X.: Artificial intelligence-aided detection of rail defects based on ultrasonic imaging data. *Proc. Inst. Mech. Eng., Part F: J. Rail Rapid Transit* **238**(1), 118–127 (Jan. 2024)
7. Railway applications—Infrastructure—Non-destructive testing on rails in track—Part 1: Requirements for ultrasonic inspection and evaluation principles, German Institute for Standardization, Standard (Nov. 2016)
8. Armbruster, R., Heckel, T., Fenger, S.: Die gläserne Schiene-Fortgeschrittene Ultraschall Schienenprüfung. *ZfP-Zeitung* **121**, 65–68 (2010)
9. Zhang, T., Heckel, T.: Virtuelle Referenzschienen nach DIN EN 16729–1 für die Ultraschall-prüfung von verlegten Eisenbahnschienen. DGZfP Jahrestagung 2023, Friedrichshafen, May, Germany (2023). <https://www.ndt.net/?id=28645>
10. Chollet, F.: Deep learning with python, 2nd edn. Manning (2021)
11. Tayeh, T., Aburakhia, S., Myers, R., Shami, A.: An attention-based ConvLSTM autoencoder with dynamic thresholding for unsupervised anomaly detection in multivariate time series. *arXiv [cs.LG]* (2022)
12. Cho, M., Ha, J., Park, C., Park, S.: Combinatorial feature embedding based on CNN and LSTM for biomedical named entity recognition. *J. Biomed. Inform.* **103**, 103381 (2020)
13. Islam, M.M., Islam, M.Z., Asraf, A., Al-Rakhami, M.S., Ding, W., Sodhro, A.H.: Diagnosis of COVID-19 from X-rays using combined CNN-RNN architecture with transfer learning. *BenchCouncil Trans. Benchmarks, Stand. Eval.* **2**(4), 100088 (2022)

14. Karim, F., Majumdar, S., Darabi, H., Harford, S.: Multivariate LSTM-FCNs for time series classification. *Neural Netw.* **116**, 237–245 (Aug. 2019)
15. Rentala, V.K., Kanzler, D., Fuchs, P.: POD evaluation: the key performance indicator for NDE 4.0. *J. Nondestruct. Eval.* **41**(1), 20 (Feb. 2022)



Federated Learning in Infrastructure Predictive Models: A Case Study of Utah's Culverts

Pouria Mohammadi¹(✉), Abbas Rashidi¹, and Sadegh Asgari²

¹ University of Utah, Salt Lake City, UT, USA

pouria.mohammadi@utah.edu

² Merrimack College, North Andover, MA, USA

Abstract. With the rapid advancements in technology and data analytics, transportation agencies are keen on adopting innovative solutions to enhance their infrastructure management plans. Given this, condition prediction models have received more attention due to their impact on the effective management of inspection resources. Machine learning (ML) algorithms are typically employed in the development of these prediction models. Traditional centralized ML models often face challenges related to data privacy, transferability, and integration from multiple sources. This paper proposed an innovative approach for infrastructure condition prediction models by leveraging Federated Learning (FL), a decentralized ML paradigm. To illustrate the proposed approach, we selected culverts in Utah as a case study. In addition to the Utah culvert inventory, we obtained further data from inventories in five other states of the US. We presented a comparative analysis of two federated models—Federated Proximal and Federated Averaging—as well as a centralized ML model. Our findings highlight the efficacy of the proposed FL-based models in enhancing prediction accuracy while ensuring data privacy and reducing data transmission overheads. Furthermore, Utah was able to leverage insights from other states' inventories through Federated Learning since it faced a data deficiency regarding culvert inspection records.

Keywords: Federated Learning · Culvert · Machine Learning · Federated Proximal · Federated Average

1 Introduction

Artificial Intelligence (AI) methods, such as Machine Learning (ML), are becoming more common and influential in the management of transportation infrastructure. These methods enable transportation agencies to make the most of their resources for maintaining and managing transportation infrastructures [1]. For instance, ML-driven predictive analytics assists transportation agencies to proactively address infrastructure issues before they fail, thereby reducing failure costs and downtime. This proactive approach to infrastructure management extends the lifespan of transportation infrastructures and significantly enhances the user experience by ensuring smoother and safer travel [2]. By strategically utilizing AI in transportation asset management, we can create a more efficient, sustainable, and user-centric system.

Despite AI's potential in transportation infrastructure management, agencies face significant challenges in adopting these methods, specifically regarding limited inventory data. Various factors contribute to the lack of availability of local datasets, including low budgets, insufficient inspectors, and the lack of a centralized management system. Moreover, most transportation systems were not initially designed with modern data collection, resulting in gaps in historical data [3]. For instance, the state of Utah has limited inspection records for its culvert inventory due to the lack of a comprehensive management system [2]. Limited datasets negatively affect the performance of the predictive models developed using ML algorithms. Additionally, it avoids leveraging advanced ML algorithms to improve the results of predictive models.

Addressing the challenge of limited infrastructure data for using ML in transportation infrastructure management involves several strategic approaches such as enhancing data collection methods, collaborating with other transportation agencies, and leveraging existing data through advanced analytics. The practical and common strategy for transportation agencies is to obtain and merge inventory data from other agencies into their existing dataset [4]. This collaborative data sharing can significantly enrich the datasets available for ML models, providing a more comprehensive view of transportation infrastructure needs and usage patterns. Data sharing, however, faces three major obstacles. The first one is the need for standardized data formats and protocols. The second one, which is the most significant one, is the legal concerns related to privacy regulations. Data sharing is hindered by regulations such as the General Data Protection Regulation (GDPR) [5] and the California Consumer Privacy Act (CCPA) [6], which place strict guidelines and prohibitions on sharing data. The last obstacle is transportation agencies' disinclination to share data due to competitive interests or administrative hurdles [7].

Researchers have investigated novel methodologies to address limited infrastructure data and data sharing obstacles. One such breakthrough was the development of federated learning (FL) by Google in 2016 [8]. This state-of-the-art ML technique was designed to preserve data privacy and tackle data scarcity issues. Initially built to improve text prediction on a broad range of Android devices, FL complies with privacy standards, including the GDPR and CCPA. In FL, a central server coordinates a network of decentralized entities—such as mobile devices or local servers from various organizations—to collaborate in the learning process. With this method, collective model training can be performed without directly exchanging raw data. Entities locally train a model that the central server provides and only send back model updates. The server then aggregates these updates to enhance the overall model's performance. As a result, this process is suitable for maintaining data privacy while leveraging the collective insight of all contributors [9].

Previous studies have demonstrated the efficacy of FL in various fields, showcasing its excellent performance. Despite this, the application of FL in scenarios with limited infrastructure data remains relatively unexplored. Our research aims to investigate FL's effectiveness in predictive analysis, particularly for infrastructures with limited historical data. To conduct this investigation, we chose Utah's culvert inventory, which contains only 272 records, as a case study. This choice assesses FL's potential in addressing the challenges faced by the Utah Department of Transportation (UDOT) and other transportation agencies with similar limitations. Our research introduces an FL-based

approach, allowing UDOT to utilize data from other state DOTs while maintaining data privacy and security. To achieve this, our research involved a comprehensive review of prior studies, gathering and preprocessing historical culvert inspection data, and implementing two FL models using Federated Averaging (FedAvg) and Federated Proximal (FedProx) algorithms. We then compared the performance of these decentralized FL models against a local model developed exclusively with Utah's 272 culvert records. This approach provides UDOT with a more robust model for efficiently assessing culvert conditions and planning inspections.

2 Literature Review

FL has recently attracted the attention of researchers from a variety of fields. These studies primarily target data privacy, communication overhead, and data scarcity challenges. Similar to other fields, civil engineering has benefited from FL. In this section, we first discuss FL applications in civil engineering and then we examine the progress made and methods developed for improving the efficiency of culvert inspection programs.

2.1 FL's Application in Civil Engineering

FL is capable of revolutionizing civil engineering through the creation of ML models that are not only more effective and accurate but also maintain the privacy of data. For instance, Li et al. [10] leveraged FL in their study to address privacy and security issues when collecting construction workers' personal image data for occupational health and safety (OHS) monitoring. They developed a framework known as federated smart work packaging (FedSWP), incorporating federated transfer learning to preserve the personal image data of construction workers. By developing a hybrid deep neural network model made up of Multi-Task Cascaded Convolutional Neural Networks, MobileNet, and Long Short-Term Memory, they enhanced the FedSWP framework. Their application of FedSWP for facial fatigue monitoring in crane operators demonstrated its effectiveness in providing customized safety alerts and health recommendations. The study emphasized FedSWP's broader applicability in various construction OHS monitoring cases. In another significant study, Moretti et al. [11] introduced a federated open data model for digital twins (DTs) in the built environment, focusing on overcoming interoperability challenges in DT applications. They developed three-tiered process modeling to align asset information standards with service expectations during the life cycle of assets. The validity of this model was demonstrated through its application in developing a building-level DT data model for the West Cambridge Campus. This federated data model proved useful in facilitating DT-based asset management at both building and built environment scales.

In another study, Saputra et al. [12] presented a novel methodology based on FL for estimating energy demand in electric vehicle networks. Their Federated Energy Demand Learning (FEDL) approach significantly enhanced estimation results in this domain. The algorithm they developed, utilized by a charging station provider (CSP), demonstrated remarkable accuracy for energy demand across charging stations. FEDL effectively minimized communication overhead and guaranteed the privacy of electric

vehicle users by allowing charging stations to share model updates, not actual raw data, with the CSP. This method surpassed traditional ML algorithms in terms of performance. In another application of FL, Khalil et al. [13] focused on maintaining privacy in ML for industrial Internet of Things (IoT) environments. Fed-NN is an FL algorithm built around a neural network model that accurately predicts worker thermal comfort and secures the worker's privacy. When tested on an actual dataset, Fed-NN achieved an 80% accuracy rate, outperforming the support vector machine and multiple linear regression models. This success indicated Fed-NN's potential for reliable and privacy-sensitive thermal comfort predictions in industrial IoT settings. Following these studies, future advancements in FL technology can lead to a more digitalized civil industry.

2.2 Culvert Condition Prediction Models

Several predictive analytics models have been developed to enhance transportation agencies' culvert inspection programs. Tatari et al. [14] conducted a study to create a regression model using Artificial Neural Networks (ANN) aimed at predicting the condition of culverts in Ohio. This model, however, was constrained by its reliance on a small dataset from Ohio DOT, containing just 39 culvert records and nine distinct features. The high feature-to-data ratio was a notable drawback of this study. As part of another endeavor, Stoner et al. [15] developed models using ANNs and logistic regression to predict culvert conditions, utilizing a much larger dataset from the South Carolina DOT. There were 8,000 culverts in the dataset. These models accounted for various culvert types and ten different categories of defects. Notably, these models were successful at multiclass classification even without considering the age of the culverts. However, the applicability of these studies is somewhat limited, as they were based on only Ohio state's culvert data.

In addition, Gao and Elzarka [16] created a binary classification model for culvert condition assessment using the decision tree algorithm. The model was trained on 11 different culvert features like material, shape, and span. The authors applied this model to a substantial dataset of 12,400 culverts from Ohio, achieving a classification accuracy of 75%. However, the model's binary classification nature and its reliance on data from a single state were identified as constraints of the research. In a separate research project, Mohammadi et al. [2] evaluated the effectiveness of five well-known multiclass classification ML algorithms for predicting the condition of Utah's culverts. These algorithms included random forest, decision tree, support vector machine, k-nearest neighbor, and ANN. Their study analyzed a dataset of 2,555 culvert records collected from four states, each data row featuring environmental and physical attributes. The random forest algorithm emerged as the top performer in all evaluation metrics. Despite its insights, the study encountered challenges concerning data privacy, as it involved combining data from multiple state inventories.

Although earlier studies have developed various models related to culvert condition prediction, they failed to offer solutions for states with limited datasets, like Utah, that lack an effective predictive model to enhance culvert inspection programs. These previous models often relied on binary classification algorithms and were often constrained to data from a single state's inventory. Considering the demonstrated applications of FL in civil engineering and its successful results, our research proposes a more robust

and generalized model utilizing the FL technique. This model aims to improve upon the limitations of previous culvert condition prediction models, particularly in terms of data scarcity and diversity, while adhering to strict data confidentiality and privacy standards.

3 Methodology

Our research primarily aims to tackle the issue of data scarcity in the infrastructure asset inventories of transportation agencies. We chose the culvert inventory of Utah with 272 data rows as our case study. The lack of comprehensive culvert inspection records in Utah's inventory creates a significant gap, hindering effective decision-making in terms of culvert inspection and maintenance [17, 18]. Consequently, our approach involves compiling culvert data from the inventories of other states in the US, thereby providing a more informative dataset. We collected data from Vermont, Ohio, New York, Massachusetts, and Colorado state culvert inventories. After data collection, we preprocessed the data, which includes removing outliers, sampling the data, normalizing the data, and combining the data features. Then, we did not fuse them since we wanted to develop two FL models. Therefore, we trained and tested the base ML models on separate preprocessed datasets as part of FL model development. A reliable model must be validated and tested during this step. Figure 1 shows the FL model developed in this study, which is a privacy-preserving decentralized approach.

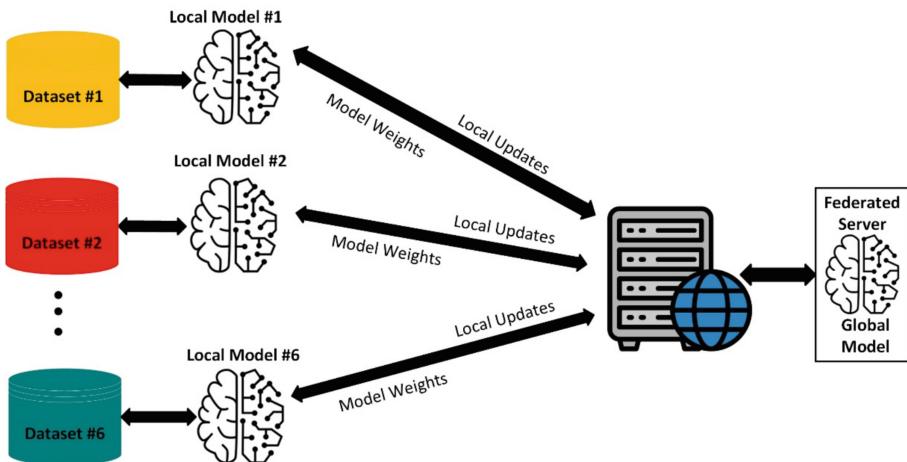


Fig. 1. The suggested FL framework for culvert condition prediction

3.1 Preprocessing of Data

In order to expand Utah's culvert dataset, we gathered data from five additional states: New York (1050 data rows), Massachusetts (417 data rows), Colorado (766 data rows), Vermont (3884 data rows), and Ohio (1851 data rows). This selection was based on

the accessibility of data and the resemblance of these states' culvert inventory features to those of Utah. Given the different inspection protocols used across these states, a preprocessing step was necessary for the features in the datasets. Also, we modified the target labels from these states to align with UDOT's culvert rating criteria. In these additional states, the rating scales were 9-point, 7-point, and 10-point, which must be converted into UDOT's 5-point rating scale. The process included a detailed comparison of the parameters of culvert inspection among the five states and Utah. The conversion of their rating scales to Utah's system is detailed in Table 1.

Table 1. Conversion of rating scales to Utah's

| Utah | Minor Defects | Moderate Defects | Significant Defects | Major Defects | Critical Defects |
|-----------------|---------------|------------------|---------------------|---------------|-------------------|
| Colorado & Ohio | 9 & 8 | 7 | 6 | 4 & 5 | 1 & 2 & 3 |
| Vermont | Excellent | Good & Fair | Poor | Critical | Urgent & Closed |
| New York | 7 | 6 & 5 | 4 | 3 | 0 & 1 & 2 |
| Massachusetts | 9 & 8 | 7 | 6 | 5 | 0 & 1 & 2 & 3 & 4 |

As a way of making the datasets more detailed and developing a more generalized ML model that can be applied in various states, we have added environmental features such as soil pH into these six datasets. We obtained these data from the Web Soil Survey website [19]. As illustrated in Table 2, the final dataset consisted of ten features categorized into physical and environmental.

Table 2. Final dataset features and label

| | | | | | |
|------------------------|---------------------------|---------------------|------------------------------|----------------|-------------------------|
| Environmental Features | Soil pH | Soil drainage class | Soil electrical conductivity | Soil Moisture | Flooding frequency |
| Physical Features | Culvert installation year | Culvert structure | Culvert shape | Culvert length | Culvert inspection date |
| Label | Culvert condition rating | | | | |

3.2 Federated Learning

FL represents a modern methodology in distributed learning, enabling the development of a global model by aggregating locally trained models from multiple entities without

exchanging the actual data they hold. This aggregation typically occurs by averaging the parameters or gradients of these local models to update the global model, ensuring data confidentiality and efficiency [20]. In our research, we employed the Flower [21], a cutting-edge open-source framework designed to streamline the development and assessment of FL models. Flower offers the flexibility to pick from a range of ML algorithms and optimization methods, aligning with specific research requirements. We selected ANN as our core ML algorithm, considering its robust predictive abilities, capability for multiclass classification problems, and excellent integration with the Flower framework.

The process of model aggregation in FL, or model fusion, is crucial to constructing robust global models by combining models trained on decentralized entities. This approach addresses privacy concerns as actual data remains on the local devices [22]. The most common aggregation algorithm is Federated Averaging (FedAvg), which involves averaging the parameters of local models (like weights and biases) to update the global model. FedAvg is simple and effective but assumes that all local datasets are identically distributed, which is often not the case in real-world scenarios [9, 23].

On the other hand, Federated Proximal (FedProx) is an extension of FedAvg designed to handle heterogeneity in data—the situation where data across devices vary significantly. FedProx adds a proximal term to the local optimization problem, which helps in dealing with the statistical challenges posed by non-IID (independently and identically distributed) data. This term penalizes the divergence of the local model updates from the global model, encouraging local models to stay closer to the global model. While FedProx generally offers better stability and performance in heterogeneous data environments compared to FedAvg, it can be slightly more complex to implement and may require careful tuning of additional hyperparameters [24].

As previously stated, we developed two horizontal FL models using the FedProx and FedAvg algorithms for aggregating the model parameters as well as a centralized ANN model based on Utah's culvert dataset. The horizontal FL paradigm, also known as data parallelism, focuses on cases where multiple entities are given equal features but varying subsets of data to work with [25]. We applied hold-out cross-validation to provide a precise estimate of our ML models' performance on unseen culvert data. Using this method, the dataset is randomly split into training and testing sets. Afterward, the model is trained on the training set and evaluated on the testing set [26]. We took 80% of the dataset for training and reserved 20% for testing. There are, however, some adjustments that need to be made when using this method in the FL framework. Instead of a centralized evaluation, each participant entity assesses the global model independently based on its own hold-out data. The metrics of this assessment, such as accuracy or F1 score, are averaged based on their weight (size of each entity test set) to determine the global model's overall performance. For further understanding, the pseudo-code of the developed FL framework is presented below.

Algorithm 1: Federated learning

| | |
|---------------|--|
| Input: | Local datasets, the strategy of the framework, including the number of communication rounds, the fraction of entities on each round, the number of local epochs, learning rate, optimization function, and local mini-batch size |
|---------------|--|

(continued)

(continued)

| | | | | | | | | | | | | | | | |
|--|---|---|--|---|--|---|---|--|---|--|---|--|---|---|--|
| | Output: Updated parameters of the global ML model | | | | | | | | | | | | | | |
| | Initialize global neural network model: <i>global_model</i> | | | | | | | | | | | | | | |
| | repeat for the number of communication rounds: | | | | | | | | | | | | | | |
| | <table border="1"> <tr> <td>Distribute <i>global_model</i> to the selected entities</td> </tr> <tr> <td>Initialize two empty lists: <i>model_updates</i>, <i>model_performances</i></td> </tr> <tr> <td>for each entity in selected entities:</td> </tr> <tr> <td> <table border="1"> <tr> <td>Train local ML model: <i>local_model</i> using the entity's local train set</td> </tr> <tr> <td>Calculate the difference between <i>local_model</i> and <i>global_model</i> weights: <i>weight_diff</i></td> </tr> <tr> <td>Calculate the performance of <i>local_model</i> on the entity's local test set: <i>entityy_metrics</i></td> </tr> <tr> <td>Append <i>weight_diff</i> to <i>model_updates</i> and <i>entityy_metrics</i> to <i>model_performances</i></td> </tr> </table> </td> </tr> <tr> <td>Aggregate model_updates using FedAvg or FedProx algorithm:</td> </tr> <tr> <td> <table border="1"> <tr> <td>Calculate the weighted average of all <i>weight_diffs</i> in <i>model_updates</i>: <i>avg_weight_diff</i></td> </tr> <tr> <td>Calculate the weighted average of all <i>entityy_metrics</i> in <i>model_performances</i></td> </tr> </table> </td> </tr> <tr> <td>Update <i>global_model</i> by applying <i>avg_weight_diff</i></td> </tr> <tr> <td>Deploy <i>global_model</i> for making predictions or further fine-tuning</td> </tr> </table> | Distribute <i>global_model</i> to the selected entities | Initialize two empty lists: <i>model_updates</i> , <i>model_performances</i> | for each entity in selected entities: | <table border="1"> <tr> <td>Train local ML model: <i>local_model</i> using the entity's local train set</td> </tr> <tr> <td>Calculate the difference between <i>local_model</i> and <i>global_model</i> weights: <i>weight_diff</i></td> </tr> <tr> <td>Calculate the performance of <i>local_model</i> on the entity's local test set: <i>entityy_metrics</i></td> </tr> <tr> <td>Append <i>weight_diff</i> to <i>model_updates</i> and <i>entityy_metrics</i> to <i>model_performances</i></td> </tr> </table> | Train local ML model: <i>local_model</i> using the entity's local train set | Calculate the difference between <i>local_model</i> and <i>global_model</i> weights: <i>weight_diff</i> | Calculate the performance of <i>local_model</i> on the entity's local test set: <i>entityy_metrics</i> | Append <i>weight_diff</i> to <i>model_updates</i> and <i>entityy_metrics</i> to <i>model_performances</i> | Aggregate model_updates using FedAvg or FedProx algorithm: | <table border="1"> <tr> <td>Calculate the weighted average of all <i>weight_diffs</i> in <i>model_updates</i>: <i>avg_weight_diff</i></td> </tr> <tr> <td>Calculate the weighted average of all <i>entityy_metrics</i> in <i>model_performances</i></td> </tr> </table> | Calculate the weighted average of all <i>weight_diffs</i> in <i>model_updates</i> : <i>avg_weight_diff</i> | Calculate the weighted average of all <i>entityy_metrics</i> in <i>model_performances</i> | Update <i>global_model</i> by applying <i>avg_weight_diff</i> | Deploy <i>global_model</i> for making predictions or further fine-tuning |
| Distribute <i>global_model</i> to the selected entities | | | | | | | | | | | | | | | |
| Initialize two empty lists: <i>model_updates</i> , <i>model_performances</i> | | | | | | | | | | | | | | | |
| for each entity in selected entities: | | | | | | | | | | | | | | | |
| <table border="1"> <tr> <td>Train local ML model: <i>local_model</i> using the entity's local train set</td> </tr> <tr> <td>Calculate the difference between <i>local_model</i> and <i>global_model</i> weights: <i>weight_diff</i></td> </tr> <tr> <td>Calculate the performance of <i>local_model</i> on the entity's local test set: <i>entityy_metrics</i></td> </tr> <tr> <td>Append <i>weight_diff</i> to <i>model_updates</i> and <i>entityy_metrics</i> to <i>model_performances</i></td> </tr> </table> | Train local ML model: <i>local_model</i> using the entity's local train set | Calculate the difference between <i>local_model</i> and <i>global_model</i> weights: <i>weight_diff</i> | Calculate the performance of <i>local_model</i> on the entity's local test set: <i>entityy_metrics</i> | Append <i>weight_diff</i> to <i>model_updates</i> and <i>entityy_metrics</i> to <i>model_performances</i> | | | | | | | | | | | |
| Train local ML model: <i>local_model</i> using the entity's local train set | | | | | | | | | | | | | | | |
| Calculate the difference between <i>local_model</i> and <i>global_model</i> weights: <i>weight_diff</i> | | | | | | | | | | | | | | | |
| Calculate the performance of <i>local_model</i> on the entity's local test set: <i>entityy_metrics</i> | | | | | | | | | | | | | | | |
| Append <i>weight_diff</i> to <i>model_updates</i> and <i>entityy_metrics</i> to <i>model_performances</i> | | | | | | | | | | | | | | | |
| Aggregate model_updates using FedAvg or FedProx algorithm: | | | | | | | | | | | | | | | |
| <table border="1"> <tr> <td>Calculate the weighted average of all <i>weight_diffs</i> in <i>model_updates</i>: <i>avg_weight_diff</i></td> </tr> <tr> <td>Calculate the weighted average of all <i>entityy_metrics</i> in <i>model_performances</i></td> </tr> </table> | Calculate the weighted average of all <i>weight_diffs</i> in <i>model_updates</i> : <i>avg_weight_diff</i> | Calculate the weighted average of all <i>entityy_metrics</i> in <i>model_performances</i> | | | | | | | | | | | | | |
| Calculate the weighted average of all <i>weight_diffs</i> in <i>model_updates</i> : <i>avg_weight_diff</i> | | | | | | | | | | | | | | | |
| Calculate the weighted average of all <i>entityy_metrics</i> in <i>model_performances</i> | | | | | | | | | | | | | | | |
| Update <i>global_model</i> by applying <i>avg_weight_diff</i> | | | | | | | | | | | | | | | |
| Deploy <i>global_model</i> for making predictions or further fine-tuning | | | | | | | | | | | | | | | |

4 Results and Discussion

Challenges in infrastructure management, such as limited infrastructure data and concerns about data confidentiality, have hindered the full implementation of ML. To overcome these obstacles, we suggested using FL instead of centralized learning methods. As a case study, we explored the use of FL in developing models for predicting the condition of culverts, focusing on Utah state. Due to Utah's small culvert dataset, we recommended adopting data from five other DOTs through the FL methodology. Accordingly, we created and compared three models: two FL models (FedAvg and FedProx) using culvert data from six states and one centralized ANN model (Utah-ANN) using Utah's limited culvert dataset. To evaluate these models, we used metrics such as accuracy, precision, recall, F1 score, and total loss. Figure 2 shows that the FL-based models—FedProx and FedAvg—outperformed the Utah-ANN model, with significant differences in precision and accuracy. The enhancement of Utah-ANN accuracy by 30 percent implies that integrating local parameters from several entities into the global ANN model leads to greater generalization and robustness of a predictive model. The FedAvg model showed a significant improvement over a centralized model developed on a limited dataset, but FedProx proved that the algorithm used for FL is critical and can increase the level of reliability. Furthermore, FedProx performed better than FedAvg in predicting instances belonging

to a class. The FedProx and FedAvg models offer additional advantages besides the improvement of performance metrics that are extremely important when it comes to data privacy.

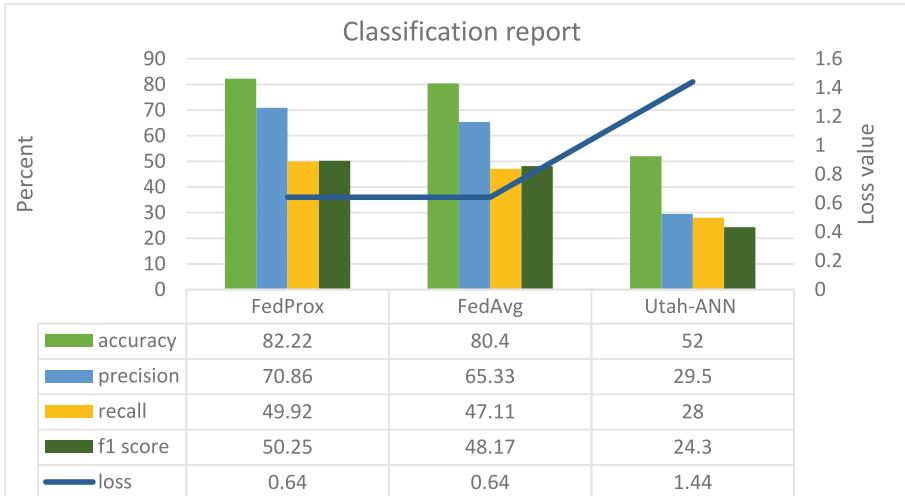


Fig. 2. Analyzing classification reports of FedProx, FedAvg, and Utah-ANN models

The FedProx model exhibited superior performance compared to FedAvg when applied to the six states' datasets due to its enhanced ability to handle non-IID data across entities in the FL environment. While FedAvg assumes that the data is identically distributed among all entities, FedProx accommodates the statistical diversity inherent in real-world data by introducing a proximal term to the objective function. This term reduces the skew in updates that can occur due to data heterogeneity, thereby stabilizing the learning process. Also, it reduces the impact of local updates that deviate too far from the global model due to unique local datasets. In other words, FedProx makes sure that the aggregated global model benefits from a more balanced and representative update from each entity. Because FedProx addresses the common challenges posed by data distribution and variability in decentralized datasets, it can achieve higher performance metrics.

While the FedProx model exhibited commendable efficacy in predicting the condition of culverts, it did not match the performance of the centralized ML model presented by Mohammadi et al. [4]. They utilized a fused dataset from four distinct culvert inventories to enhance the predictive accuracy of ML models. Centralized models generally yield better outcomes because they benefit from the fusion of diverse datasets into a unified database but at the cost of violating data privacy. Nonetheless, improvements to the predictive performance of FL-based models present a dual advantage, negating the need to sacrifice privacy for accuracy. Organizations can thus reap the rewards of a precise model while maintaining the confidentiality of their data.

5 Conclusion

This research introduces a new method to overcome the challenge of data scarcity in the development of efficient predictive models for transportation infrastructure management. We investigated culvert predictive models in Utah as a case study. Our findings demonstrated that it is possible to accurately predict culvert conditions for the UDOT while protecting the data privacy, security, and access rights of other DOTs. By employing the FL framework, we developed two FL-based models—FedProx and FedAvg—which outperformed the local ANN model developed on Utah’s culvert dataset. Notably, the FedProx model exhibited a 30% higher accuracy than the Utah-ANN model. By modifying the objective function, FedProx more effectively handles the diversity of datasets in distributed learning compared to FedAvg, leading to better results. This improvement is particularly noticeable in the precision and accuracy metrics, where FedProx outperforms FedAvg. FL’s unique approach allows each state to enhance the collective global model while keeping its culvert inventory data private, aligning with the increasing focus on data privacy and security. Although the centralized model utilizing aggregated data from multiple states showed superior performance compared to the FedProx and FedAvg models, its violation of data privacy still remains an issue, limiting its application. Our results indicate that enhancing the predictive capabilities of FL-based models enables transportation agencies like UDOT to access highly accurate models without compromising data confidentiality. Therefore, we recommend the application of optimization algorithms to improve FL-based models or the incorporation of more state culvert datasets for broader generalizability in future studies. This study also paves the way for further exploration into FL’s potential, especially in situations where both data scarcity and privacy are significant concerns.

References

1. Iyer, L.S.: AI enabled applications towards intelligent transportation. *Transp. Eng.* **5**, 100083 (2021)
2. Mohammadi, P., Rashidi, A., Malekzadeh, M., Tiwari, S.: Evaluating various machine learning algorithms for automated inspection of culverts. *Eng. Anal. Bound. Elem.* **148**, 366–375 (2023). <https://doi.org/10.1016/j.enganabound.2023.01.007>
3. Sinha, K.C., Labi, S., Agbelie, B.R.D.K.: Transportation infrastructure asset management in the new millennium: continuing issues, and emerging challenges and opportunities. *Transp. A Transp. Sci.* **13**(7), 591–606 (2017)
4. Mohammadi, P., Rashidi, A., Asgari, S.: An improved hybrid XGBoost model for culvert inspection using swarm intelligence algorithms. *Comput. Civ. Eng.* 2023, pp. 100–108 (Jan. 25, 2024). <https://doi.org/10.1061/9780784485224.013>
5. and repealing directive (general data protection regulation). Regulation of the European Parliament and of the council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, “EU,” Dec. 2018. https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en (accessed Apr. 22, 2023)
6. California Legislative Information: *California Consumer Privacy Act*. 2018, pp. 1–59. Accessed: Apr. 22, 2023 [Online]. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5

7. Figueiredo, A.S.: Data sharing: convert challenges into opportunities. *Front. public Heal.* **5**, 327 (2017)
8. Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated learning: Strategies for improving communication efficiency. *arXiv Prepr. arXiv1610.05492* (2016)
9. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics, pp. 1273–1282 (2017)
10. Li, X., lin Chi, H., Lu, W., Xue, F., Zeng, J., Li, C.Z.: Federated transfer learning enabled smart work packaging for preserving personal image information of construction worker. *Autom. Constr.* **128**(March), 103738 (2021). <https://doi.org/10.1016/j.autcon.2021.103738>
11. Moretti, N., Xie, X., Merino Garcia, J., Chang, J., Kumar Parlikad, A.: Developing a federated data model for built environment digital twins. In: Computing in Civil Engineering 2021, pp. 613–621 (2021)
12. Saputra, Y.M., Hoang, D.T., Nguyen, D.N., Dutkiewicz, E., Mueck, M.D., Srikantheswara, S.: Energy demand prediction with federated learning for electric vehicle networks. *IEEE Glob. Commun. Conf. (GLOBECOM)* **2019**, 1–6 (2019)
13. Khalil, M., Esseghir, M., Merghem-Boulahia, L.: A federated learning approach for thermal comfort management. *Adv. Eng. Informatics* **52**(February), 101526 (2022). <https://doi.org/10.1016/j.aei.2022.101526>
14. Tatari, O., Sargand, S.M., Masada, T., Tarawneh, B.: Neural network approach to condition assessment of highway culverts: case study in Ohio. *J. Infrastruct. Syst.* **19**(4), 409–414 (2013). [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000139](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000139)
15. Stoner, M., Pang, W., Piratla, K.: Predicting culvert deterioration using physical and environmental time-independent variables. *J. Pipeline Syst. Eng. Pract.* **10**(4), 04019035 (2019). [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000409](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000409)
16. Gao, C., Elzarka, H.: The use of decision tree based predictive models for improving the culvert inspection process. *Adv. Eng. Informatics* **47**(May 2020), 101203 (2021). <https://doi.org/10.1016/j.aei.2020.101203>
17. McGrath, T.J., Beaver, J.L.: Condition assessment of Utah highway culverts and determination of culvert performance measures (2004)
18. Mohammadi, P., Sherafat, B., Rashidi, A.: Developing a culvert inspection manual and estimating culverts' deterioration curve, inspection frequency and service life for UDOT. Department of Transportation, Utah (2023)
19. Web Soil Survey: <https://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx> (accessed May 18, 2023)
20. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: challenges, methods, and future directions. *IEEE Signal Process. Mag.* **37**(3), 50–60 (2020). <https://doi.org/10.1109/MSP.2020.2975749>
21. Beutel, D. J., et al.: Flower: a friendly federated learning research framework (2020). <http://arxiv.org/abs/2007.14390>
22. Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., Piccialli, F.: Model aggregation techniques in federated learning: a comprehensive survey. *Futur. Gener. Comput. Syst.* (2023)
23. Mohammadi, P., Rashidi, A., Asgari, S.: Improving culvert condition prediction models using federated learning: the case study of Utah (2024)
24. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. *Proc. Mach. Learn. Syst.* **2**, 429–450 (2020)
25. Li, L., Fan, Y., Tse, M., Lin, K.-Y.: A review of applications in federated learning. *Comput. Ind. Eng.* **149**, 106854 (2020)

26. Gondia, A., Siam, A., El-Dakhakhni, W., Nassar, A.H.: Machine learning algorithms for construction projects delay risk prediction. *J. Constr. Eng. Manag.* **146**(1), 4019085 (2020)



Dimensionality Reduction in Structural Reliability Analysis

I.-Tung Yang^(✉) and Jonathan Aloysius Budiman

Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan
ityang@mail.ntust.edu.tw

Abstract. Engineering structures, such as bridges and buildings, are impacted by uncertainties. Risk management, therefore, necessitates the prediction of failure probability to assess the structure's safety in the presence of uncertainties. Several methods, namely Monte Carlo Simulation (MCS), First-order Reliability Method (FORM), and Second-order Reliability Method (SORM), have been proposed to assess structural reliability. Yet, these methods encounter difficulties in dealing with high-dimensional problems as the required computation time grows exponentially. This study develops an integrated framework, Dimension-reduction Reliability Analysis (DRRA), to provide a remedy. To assess the structural reliability, DRRA adopts a deep learning technique to reduce the dimensionality of input variables for the active-learning reliability method AK-MCS. DRRA involves training an autoencoder and a deep feedforward network to map the training samples from a high-dimensional space to a lower-dimensional latent representation. Then, AK-MCS is employed to estimate structural reliability in the latent space. DRRA is demonstrated by modeling a practical high-dimensional case. The performance is compared with traditional AK-MCS (without dimensionality reduction) and another existing method: HDDA-GP. The comparison results confirm that the DRRA framework achieves better accuracy and higher efficiency than previous work. The application of the DRRA framework across various design cases will further demonstrate its overall performance.

Keywords: Reliability Analysis · Dimensionality reduction · Machine Learning

1 Introduction

The response of an engineering structure is influenced by a wide range of factors. The purpose of structural reliability analysis is to evaluate the safety of a structure considering uncertainties that can lead to deviations from the expected state and potential negative consequences. Sampling-based techniques like Monte Carlo simulation (MCS) [9], important sampling (IS) [8] and directional sampling (DS) [2] have been developed to perform robust reliability analysis for practical engineering problems involving nonlinear performance functions. Although there have been advancements in variance-reducing techniques, simulation-based methods still demand a significant number of

time-consuming function evaluations to estimate small probabilities. To address this issue, surrogate models have been considered as an alternative to address computational costs in engineering problems.

One commonly used surrogate modelling technique is Kriging or Gaussian Process Regression (GPR), which is advantageous due to its stochastic nature that allows for the development of adaptive techniques. The adaptive or active-learning technique involves iteratively identifying the most informative and influential samples to enhance the surrogate model. The active-learning method combined with Kriging and Monte Carlo simulation (AK-MCS) has been widely adopted because it significantly reduces the number of function evaluations compared to MCS [4]. However, surrogate models, including Kriging, face significant computational challenges when predicting model parameters using high-dimensional data [7].

One effective solution for high-dimensional problem analysis is to reduce the dimensionality of input variables. Several methods have been developed as dimensionality reduction techniques, such as t-Distributed Stochastic Neighbor Embedding (t-SNE) [12], univariate dimension-reduction method (UDRM) [11], and High-Dimensional Model Representation (HDMR) [3]. Recently, Li and Wang [6] introduced a high-dimensional data abstraction (HDDA) framework as a dimensionality reduction method for high-dimensional reliability analysis. The framework includes training an autoencoder as a failure-informant network and a deep feedforward network (DFN) to plot the high-dimensional variables to low-dimensional representations.

The objective of this study is to propose an effective approach for assessing the reliability of high-dimensional systems. Therefore, this study combines a dimensionality reduction technique with a reliability assessment method in a combined framework to provide accurate reliability analysis without computationally expensive simulations. For the dimensionality reduction method, the HDDA framework developed by Li and Wang [6] is employed by training two Artificial Neural Network (ANN): an autoencoder that compresses the high-dimensional input variables into a latent space with low dimensionality and distinguishable limit states, and a deep feedforward network (DFN) that plots the high-dimensional input variables to the latent space from the autoencoder. For the reliability analysis method, AK-MCS is used to predict the system's probability of failure. AK-MCS is chosen because it can estimate the failure probability of a system with fewer training samples compared to traditional MCS, while still providing highly accurate results [4].

2 Methods

In proposed DRRA Framework, there are two main stages in the framework: the reduction of dimensionality using ANN and the reliability estimation using the AK-MCS method. The ANN reduces the dimensionality of the problem before AK-MCS predicts the reliability of the low-dimensional system, wherein accuracy is enhanced by enriching the Kriging model using Active-learning Function. The framework will stop once convergence of the reliability is achieved.

2.1 Artificial Neural Network (ANN)

The present study adopts the approach proposed by Li and Wang [6] for constructing a High-Dimensional Data Abstraction (HDDA) framework via the utilization of artificial neural networks (ANNs). The graphical representation of the HDDA framework, portrayed in Fig. 1, consists an autoencoder that takes failure-related information to yield a latent space for reducing dimensionality, and a deep feedforward network tasked with mapping high-dimensional input variables to their low-dimensional latent space.

As illustrated in Fig. 1, the autoencoder neural network is trained by combining high-dimensional system input variables and outputs. This process yields a set of low-dimensional latent variables representation of the actual high-dimensional limit state function. Furthermore, the deep feedforward network is trained using the high-dimensional system input variables while the target values are set to be the latent variables derived from the autoencoder.

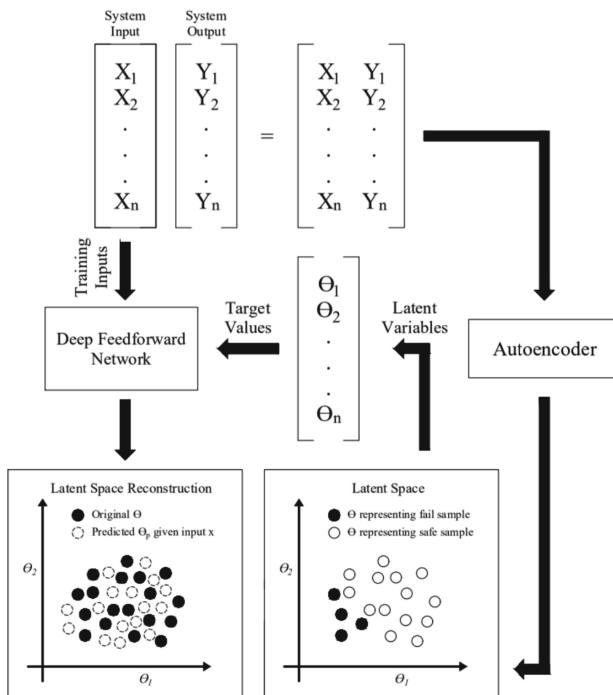


Fig. 1. HDDA Framework [6]

An autoencoder neural network consists of an encoder and a decoder and seeks to recover a low-dimensional representation from a high-dimensional system. Ensuring that the number of nodes in the latent layer is less than the number of nodes in the input layer, it is possible to produce a compressed representation of the input layer in the latent layer, also known as latent space. The introduction of autoencoder can be referred to [1].

The Deep Feedforward Network (DFN) is used to directly plot the high-dimension input variables into the latent space to create a set of low-dimension latent variables. Latent space acquired from the autoencoder training is considered as the target value for the DFN to project its training inputs. The target values are set to be the latent space θ_t , resulting in the predicted training inputs in the latent space: the latent variables θ_p . Therefore, DFN projects the training inputs X_t to the latent space θ_t as θ_p , and the dimensionality of θ_p is lower than X_t .

2.2 Active Learning Reliability Method Combining Kriging and Monte Carlo Simulation (AK-MCS)

The idea of AK-MCS is to approximate the limit state function, focusing on the MCS population instead of the whole space. It carries out Monte Carlo Simulation (MCS) without analyzing the entire training samples. Initial samples for the Kriging model are chosen from the MCS sampling space, after which the optimal point is determined using a learning function. After the model has been adequately trained, the failure probability may be calculated using MCS. This approach results in an effective evaluation as it avoids evaluating the limit state function in regions with a low probability of occurrence. Given $n_{\widehat{G}}$ represents the number of predicted limit state function and n_{MCS} is the number of MCS samples, the probability of failure can be assessed by using Eq. (1).

$$\widehat{P}_f = \frac{n_{\widehat{G} \leq 0}}{n_{MCS}} \quad (1)$$

The effectiveness of an adaptive Kriging model is contingent on its capacity to enhance its predictive accuracy. The learning function is an indicator that quantifies the amount of information that a candidate samples would contribute to the Kriging model if incorporated. U-function is one of the most widely used learning functions in AK-MCS [4].

$$U = \left| \mu_{\widehat{G}} \right| / \sigma_{\widehat{G}} \quad (2)$$

where $\mu_{\widehat{G}}$ and $\sigma_{\widehat{G}}$ is the mean and standard deviation of the Kriging predictor respectively. The U -function indicates the probability of misclassifying the prediction's failure state. The closer $\mu_{\widehat{G}}$ to zero and the greater $\sigma_{\widehat{G}}$, the smaller U becomes, indicating a greater risk of error. The sample with the lowest U value is therefore considered the best candidate for model addition. Once the minimum U value exceeds or equals 2, it is presumed that the model adequately represents the actual limit state function.

However, it is demonstrated that value-based stopping conditions are too conservative in the majority of cases [10]. Furthermore, due to the high number of variances caused by the dimensionality-reduction method towards the prediction of the sampling space, the value-based stopping conditions become irrelevant. So, a new stopping condition based on the convergence of the system's reliability is proposed. Considering $P_{f_{avg}}$ to be the average failure probability of the previous 4 iterations, P_{f_n} is the failure probability

of the n^{th} iteration, and Δ_t as the convergence threshold, the stopping condition can be expressed as:

$$|P_{f_n} - P_{f_{avg}}| \leq \Delta_t \quad (3)$$

2.3 Dimension-Reduction Reliability Analysis (DRRA) Framework

Figure 2 Illustrates the flowchart of the proposed DRRA Framework whereas the steps are as follows:

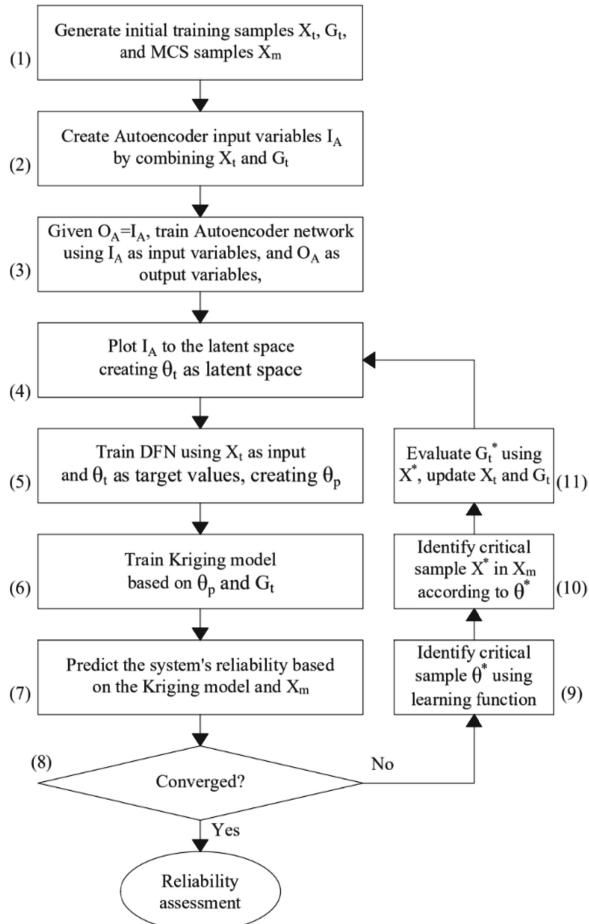


Fig. 2. DRRA Framework Flowchart

1. Generate MCS samples X_m where a nts number of training samples X_t are randomly chosen and their respective limit state function G_t are evaluated.

2. X_t and G_t are combined to be the input variables I_A for training the Autoencoder. G_t is normalized to be one scale higher than X_t to help the autoencoder generate a distinguishable failure area in the latent space.
3. The autoencoder network is trained using the input variables I_A and the output variables O_A in which $O_A = I_A$.
4. Latent space θ_t is created by mapping I_A .
5. DFN is trained using X_t as input and θ_t as target values. As a result, DFN produces θ_p : the predicted latent variables of X_t .
6. The Kriging model is trained using θ_p and G_t as the basis variables.
7. Predict the limit states of X_m using the Kriging model, and assess the reliability of the system using Eq. (1).
8. Following several iterations of incorporating critical samples into the training input variables X_t , an evaluation of reliability is conducted to determine if convergence has been achieved. When the failure probability has reached convergence, the system's reliability is considered to be accurately assessed.
9. In cases where the failure probability has not yet converged, the learning function is employed to identify the critical sample within the latent space, denoted as θ^* .
10. The corresponding critical sample, denoted as X^* , is selected from the MCS sample pool X_m . This sample is then evaluated using its limit state function and subsequently added to the training samples.
11. Assess the limit state function of the critical sample X^* , generating G_t^* . Subsequently, incorporate both X^* and G_t^* into the current training samples X_t and limit state function G_t . This includes generating new latent variables to facilitate the ongoing process of dimensionality reduction and reliability analysis. This iterative process persists until the failure probability reaches convergence.

3 Case Study

The parameter for DRRA Framework is set as follows. Autoencoder consists of an input layer, a two-dimensional hidden layer, and an output layer. The number of neurons in the hidden layer is considered to represent the number of latent variables. The input layer will apply *zero-center* normalization technique to alleviate redundancy, and the hidden layer is set to be the actual latent variables of the autoencoder. The output layer is made to be the same as the input layer, which consists of the combination of the training samples X_t with its respective response G_t . In order to help distinguish the failed with the success samples, X_t will be scaled in the range of 0 to 1 while the limit state function G_t is set to be one scale higher: from 0 to 10. Deep Feedforward Network (DFN) has one input layer, two hidden layers with 20 neurons in each hidden layer, and the output layer. While training the DFN, latent space obtained from the hidden layer of the autoencoder is considered the target value, while the output layer of the DFN is taken as the predicted latent variables of the training samples X_t . Unlike the autoencoder, the input layer of the DFN utilizes z-score data normalization rather than *zero-center*. The number of neurons and layers within the autoencoder and DFN are set to be exactly the same as Li and Wang [6]'s HDDA-GP approach, whereas the rest of the parameters are chosen according to preliminary result to ensure accurate result because it is not mentioned in previous study.

For training the Kriging model, zero-order polynomial regression function and Gaussian correlation function are adopted, and a sample is classified as a failure when the limit state function of the system is less than zero. In order to verify the performance of the proposed DRRA Framework, the result is compared to the traditional AK-MCS method as well as HDDA-GP approach proposed by Li and Wang [6]. To ensure the results are not biased, all models are run 10 times, and the results are presented as the mean value from the 10 runs. The true solution is taken from the reliability assessed by using conventional MCS because it has been recognized as a relatively robust reliability assessment tool, providing accurate estimation results regardless of the nonlinearity and irregularity of the limit state function [13].

The case study is a 35-dimension truss structure problem, shown in Fig. 3, which combines a high-dimensional problem and a practical case study. The truss is composed of 16 bars, each measuring 30 cm in length. The structure is subjected to three external lateral loads: F1, F2, and F3, applied to nodes 2, 7, and 8, respectively, along the horizontal direction. Meanwhile, nodes 1, 3, and 5 are fixed. In this study, the self-weight of the structure is disregarded, and the joints are treated as having two degrees of freedom each. Specifically, 35 normally distributed random variables are considered for the design variables, as outlined in Table 1. For this case study, 175 initial training samples X_t have been employed to assess the system's reliability across a sampling space X_m of 106. By configuring the number of neurons in the latent layer of the autoencoder to 2, the problem's dimensionality is effectively reduced to 2. As the 35-dimensional truss structure problem is relatively simple, further efforts will be directed towards solving more complicated cases.

Table 1. Characteristic of random variables for truss structure

| Random variables | Description | Distribution | Mean Value | Std | Unit |
|------------------|-----------------------------------|--------------|------------|-----|-----------------------|
| x_{1-16} | Cross-sectional area for bar 1–16 | Normal | 3 | 0.3 | 10^{-4} m^2 |
| x_{17-32} | Young's modulus for bar 1–16 | Normal | 70 | 5 | GPa |
| x_{33-35} | External load F_{1-3} | Normal | 10 | 1 | kN |

A MATLAB code for the finite element method (FEM) is developed to analyze the truss structure, calculating the horizontal and vertical displacements at each node. Consequently, the limit state function for this case study is as follows:

$$G(\mathbf{x}) = \delta_{all} - \delta_{max} \quad (4)$$

where δ_{all} is the allowed displacement, and δ_{max} is the maximum displacement of the nodes. For this case, δ_{all} is set to be 1.75 mm according to [6]. Table 2 shows the comparative results of the 35-dimension truss structure problem. The accurate MCS generates a reliability of 0.9762 by using 10^6 samples, and it took 265.67 s to run. In this particular case study involving a simple 16 bar truss problem, the modeling of the

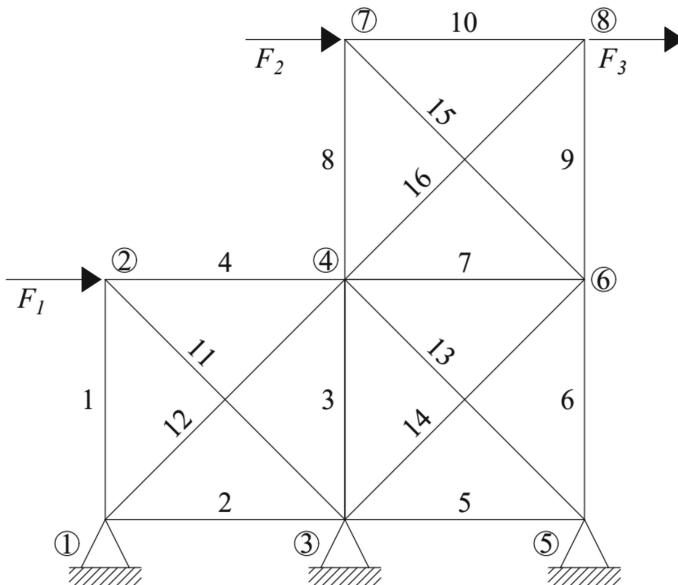


Fig. 3. Truss Structure [6]

structure using finite element method is simple enough to be written in MATLAB. Thus, MCS shows superior computational efficiency as it does not have to do the simulation using another structural modeling software. AK-MCS produces an average reliability of 0.9998 with an error of 2.418%. The iteration process of AK-MCS is terminated after 375 iterations due to having no improvement on the learning function value. Therefore, the results provided by AK-MCS are considered invalid. HDDA-GP produces a reliability of 0.9800 with an error of 0.389%. The time utilized by HDDA-GP was not reported previously. The DRRA framework produces an average reliability of 0.9785, so the error is only 0.237%, and the time utilized is 1875.03 s. In this specific case study, the DRRA framework showcases a superior accuracy by 0.15% when compared to the HDDA-GP approach. Moreover, the DRRA framework exhibits enhanced efficiency in terms of sample usage, achieving an efficiency improvement of 5.52%.

Table 2. Comparative results of 35-dimension truss structure problem

| Methods | Reliability | Error (%) | Total Evaluated Samples | Time (s) |
|----------------|-------------|-----------|-------------------------|----------|
| MCS | 0.9762 | – | 10^6 | 265.67 |
| AK-MCS | 0.9998* | 2.418 | 375* | 2580.26* |
| HDDA-GP | 0.9800 | 0.389 | 241 | – |
| DRRA Framework | 0.9785 | 0.237 | 228 | 1875.03 |

Note Invalid results because of non-convergence

4 Conclusion

The paper introduces the Dimension-reduction Reliability Analysis (DRRA) Framework, a novel approach designed to address high-dimensional reliability challenges. This framework leverages Artificial Neural Networks (ANN) as a dimensionality-reduction tool alongside AK-MCS for precise and efficient reliability assessment. By employing ANN for dimensionality reduction, AK-MCS accurately predicts system reliability without resource-intensive simulations.

To demonstrate its robustness, the DRRA framework tackles a 35-dimensional truss structure problem, representing a practical and complex scenario. Results showcase the framework's exceptional accuracy in predicting system reliability using a substantially smaller sample size in high-dimensional problems. Notably, the DRRA framework outperforms the HDDA-GP approach, exhibiting a 0.15% improvement in accuracy and a 5.52% enhancement in efficiency.

This innovative approach not only achieves superior accuracy but also significantly reduces computational requirements, a critical advantage in addressing high-dimensional cases. The integration of ANN and AK-MCS within the DRRA framework presents a promising solution for accurately predicting system reliability while mitigating the computational burden associated with complex systems. The application of the DRRA framework across various design cases will further demonstrate its overall performance.

References

1. Baldi, P.: Deep learning in science. Cambridge University Press (2021)
2. Bjerager, P.: Probability integration by directional simulation. *J. Eng. Mech.* **114**(8), 1285–1302 (1988)
3. Chowdhury, R., Rao, B.N.: Assessment of high dimensional model representation techniques for reliability analysis. *Probabilistic Eng. Mech.* **24**(1), 100–115 (2009). <https://doi.org/10.1016/j.probengmech.2008.02.001>
4. Echard, B., Gayton, N., Lemaire, M.: AK-MCS: an active learning reliability method combining Kriging and Monte Carlo Simulation. *Struct. Saf.* **33**(2), 145–154 (2011). <https://doi.org/10.1016/j.strusafe.2011.01.002>
5. Echard, B., Gayton, N., Lemaire, M., Relun, N.: A combined Importance Sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models. *Reliab. Eng. & Syst. Saf.* **111**, 232–240 (2013). <https://doi.org/10.1016/j.ress.2012.10.008>
6. Li, M., Wang, Z.: Deep learning for high-dimensional reliability analysis. *Mech. Syst. Signal Process.* **139**, 106399 (2020). <https://doi.org/10.1016/j.ymssp.2019.106399>
7. Liu, Y., Li, L., Zhao, S., Zhou, C.: A reliability analysis method based on adaptive Kriging and partial least squares. *Probabilistic Eng. Mech.* **70**, 103342 (2022). <https://doi.org/10.1016/j.probengmech.2022.103342>
8. Melchers, R.: Radial importance sampling for structural reliability. *J. Eng. Mech.* **116**(1), 189–203 (1990)
9. Metropolis, N., Ulam, S.: The monte carlo method. *J. Am. Stat. Assoc.* **44**(247), 335–341 (1949)
10. Moustapha, M., Marelli, S., Sudret, B.: Active learning for structural reliability: Survey, general framework and benchmark. *Struct. Saf.* **96**, 102174 (2022). <https://doi.org/10.1016/j.strusafe.2021.102174>

11. Rahman, S., Xu, H.: A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics. *Probabilistic Eng. Mech.* **19**(4), 393–408 (2004). <https://doi.org/10.1016/j.probengmech.2004.04.003>
12. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(11). (2008)
13. Zhang, X., Wang, L., Sørensen, J.D.: REIF: a novel active-learning function toward adaptive Kriging surrogate models for structural reliability analysis. *Reliab. Eng. & Syst. Saf.* **185**, 440–454 (2019). <https://doi.org/10.1016/j.ress.2019.01.014>



An Artificial Intelligent Design System for Shear Wall Structures with Large Language Model Controlling Generation and Optimization

Si-zhong Qin¹, Hong Guan², Wen-jie Liao¹, Yi Gu¹, Zhe Zheng¹, Hong-jing Xue³, and Xin-zheng Lu¹⁽⁾

¹ Key Laboratory of Civil Engineering Safety and Durability of China Education Ministry,
Department of Civil Engineering, Tsinghua University, Beijing, China
luxz@tsinghua.edu.cn

² School of Engineering and Built Environment, Griffith University, Queensland, Australia

³ Beijing Institute of Architectural Design Co., Ltd., Beijing, China

Abstract. Artificial intelligent design technology for shear wall structures holds promise in enhancing design efficiency and addressing the tedious and repetitive nature of conventional design work. This technology has experienced rapid development in recent years. However, based on deep learning, the existing design methods for shear wall structures face challenges in simulating the entire process of real structural design. As such, improvement in terms of safety (mechanical performance) and cost-effectiveness (material consumption) becomes necessary. In response to these issues, this study proposes an intelligent design system for shear wall structures based on a large language model (LLM). The system utilizes the LLM as the core controller, to facilitate interaction with engineers and interpretation of their language descriptions into computer-executable code. The core controller then calls upon corresponding structural generation and optimization methods to improve the design outcomes. Additionally, during the process of structural optimization, the system integrates mechanical performance, material utilization, and empirical rules to simulate the engineer's structural adjustment workflow, thereby improving the quality of design. The proposed system showcases its proficiency in successfully executing the comprehensive conversion process of architectural drawings, structural schemes, and analysis models, achieving an automated and highly efficient design approach for shear wall structures. Analysis and verification of multiple cases demonstrate that this system can improve design speed by more than 30 times (approximately 1 h as opposed to 30 h using traditional methods), while ensuring the safety and economic viability of the proposed design solutions. As a result, this study provides valuable insights for the realization of automated structural design.

Keywords: Intelligent Design · Shear Wall Structure · Large Language Model · Generative AI · Structural Optimization

1 Introduction

Shear wall structures are commonly used for residential buildings. However, design of such structures generally takes a large amount of repetitive work, making it rather inefficient and time-consuming. Therefore, creating an intelligent design system for shear wall structures would improve the efficiency of the design process and elevate the level of intelligence in civil engineering practice.

The design process for shear wall structures typically involves four phases: conceptual design, structural modeling, structural optimization, and structural drawing preparation [1]. Previous studies have only focused on the specific phases within the whole process. However, recent advancements in artificial intelligence, such as generative adversarial networks (GANs) [2–6] and graph neural networks (GNNs) [7, 8], have enabled the automatic generation of multiple design options for shear wall structures in conceptual design. Structural modeling and corresponding analyses are essential to ensure building safety and to prevent structural failures, such as building collapses [9–11]. For structural modeling, in particular, deep learning techniques can be used to quickly create architectural and structural models [12, 13]. Furthermore, parametric modeling is commonly used for structural analysis and optimization, aiming to minimize material usage while meeting performance requirements. Yet, integrating structural generation and optimization processes into a unified system presents a challenge, and simulating real-world design processes and ensuring safety (mechanical performance) and cost-effectiveness (material consumption) also remain complex tasks.

Large language models (LLMs) have brought new possibilities for solving such problems [14, 15]. LLMs can comprehend and extract intended meanings from text, which allows them to perform complex tasks with simple instructions. For example, they can convert design requirements given by structural engineers into code and automatically carry out specific actions for designing shear wall structures. Research teams like Auto-GPT [16] and HuggingGPT [17] are using LLMs as the main controllers to accomplish difficult tasks. These controllers firstly plan tasks using methods like chain of thought [18] or tree of thought [19], find the right steps to take, and then correct mistakes by reflecting on their own work [20]. Finally, they choose the right tools or methods to use. Similarly, in structural design, we can use an LLM as the main controller to handle the whole process of coming up with design schemes, optimizing the structure, and creating design drawings.

Therefore, this study aimed to develop an intelligent design system for shear wall structures with LLM controlling generation and optimization. Section 2 offers a thorough introduction to the proposed system and its associated layers, outlining the interconnections between different modules employed during distinct design stages, including an LLM controller, structural generation module, and structural optimization module. Section 3 presents a case study to demonstrate the efficacy of the system. Lastly, research findings are summarized in Sect. 4.

2 Method

In this paper, an intelligent design system for shear wall structures powered by LLMs is introduced. As shown in Fig. 1, the system's framework comprises three primary components: data, control, and output layers. The data layer stores architectural drawings and essential data associated with structural design. The control layer utilizes an LLM controller to carry out design, adjustment, and optimization tasks. The output layer showcases the final design outcomes and establishes the corresponding structural analysis model.

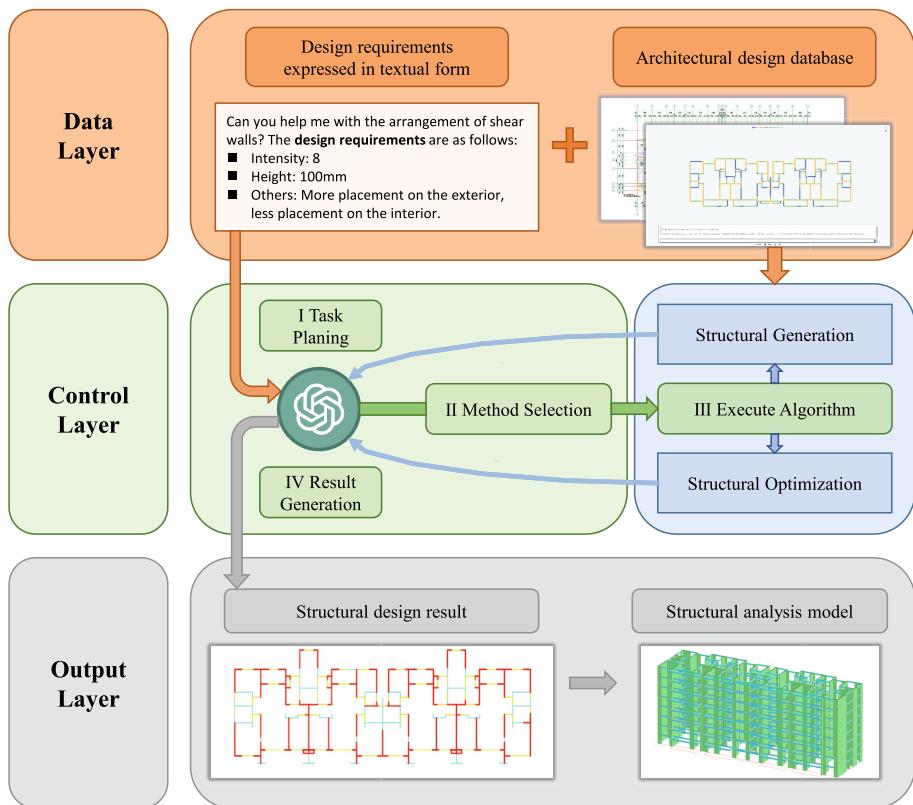


Fig. 1. Framework of the intelligent design system for shear wall structures driven by LLMs and generative AI.

The data layer comprises two primary categories of information. The first category comprises design requirement texts to be directly entered into the control layer by the users. The second category includes architectural design drawings for the users to upload in commonly used CAD file formats. These CAD drawings are subsequently converted into JSON format data using the backend system and are saved in a database. This conversion allows for their visual representation and streamlines subsequent processing.

The control layer is centered around the LLM controller, drawing inspiration from the task-processing approach employed by HuggingGPT [17]. The process is shown in Fig. 2. Initially, the text outlining the design requirements is analyzed to plan subsequent tasks and provide the necessary parameters. Using the parsing results as a guide, suitable methods are chosen, such as structural generation and optimization. The system draws on architectural drawings from a database, and the chosen algorithms, i.e., GANs or GNNs for structural generation and the two-stage method in Fig. 3 for structural optimization are executed to produce the desired design outcomes. Through an iterative process of interaction between the user and the controller, the design results are continuously refined, ultimately yielding an optimized structural design solution that meets the specified engineering requirements.

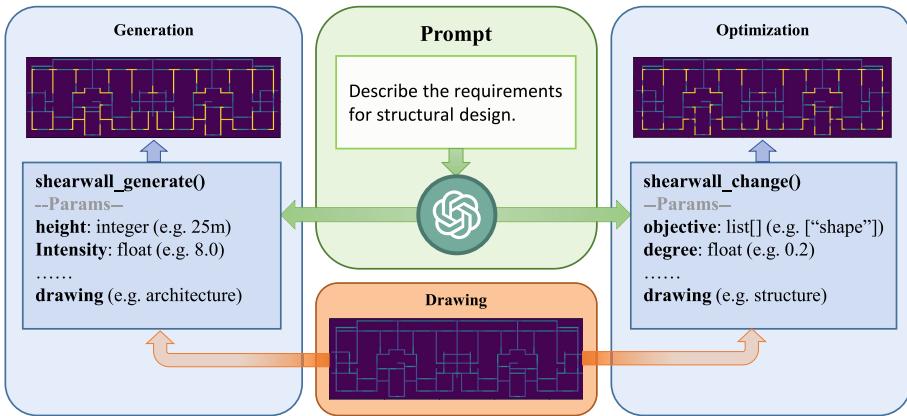


Fig. 2. Methods invoked by the LLM controller.

The control layer comprises three modules: LLM controller, structural generation module, and structural optimization module.

For successful completion of the aforementioned tasks utilizing the LLM controller, it is essential to employ a prompt engineering approach [21]. Inspired by the implementation of AutoGPT [16], the LLM is directed through carefully designed prompts to accomplish the desired objective. By providing each function with its appropriate name and parameters, prompts are used to guide the LLM in generating a JSON-formatted response. Additionally, the LLM is utilized to verify and rectify the JSON-formatted results, thereby ensuring their accuracy.

The structural generation module offers various techniques for creating shear wall structure schemes, utilizing advanced methods like GANs and GNNs [22]. Each method presents distinctive advantages. GAN methods are known for their robustness, providing reliable results. On the other hand, GNN methods excel in capturing intricate design details, leading to superior design effects. Engineers can choose the most suitable generative method based on their specific requirements and the application scenarios they are working with.

For optimization, this study proposes a three-level, two-stage optimization method based on topology, pattern, and size. Topology refers to the arrangement of structural components with their connections. Pattern primarily refers to the shape characteristics of a component. Size denotes the component's cross-sectional dimension and length. To ensure a balance between quality and efficiency, a two-stage optimization approach is utilized, as depicted in Fig. 3.

In the first stage, the optimization process starts by leveraging engineers' experience and design specifications. This helps in optimizing the structure at the topology and pattern levels and leads to an initial determination of the structural scheme. Each node and half of its edges form a point-edge combination, which serves as the variable for optimization. In addition, a set of empirical metrics were established to indicate the optimization direction, based on the work of Qin et al. [23]. Due to the discrete characteristics of the optimization variables, a widely utilized genetic algorithm (GA) was employed for the purpose of optimization.

Next, precise calculations are employed to make adjustments to the pattern and optimize the sizes of the components. These calculations take into account mechanical and material characteristics, resulting in an improved and safer structural scheme. The optimization variables include the beam height, beam width, and wall length. Parametric modeling of the shear wall structure can be performed using the GAMA module of the YJK V5.2.1 structural modeling and analysis software. The final results are obtained through an iterative optimization process utilizing GAMA's online learning algorithm.

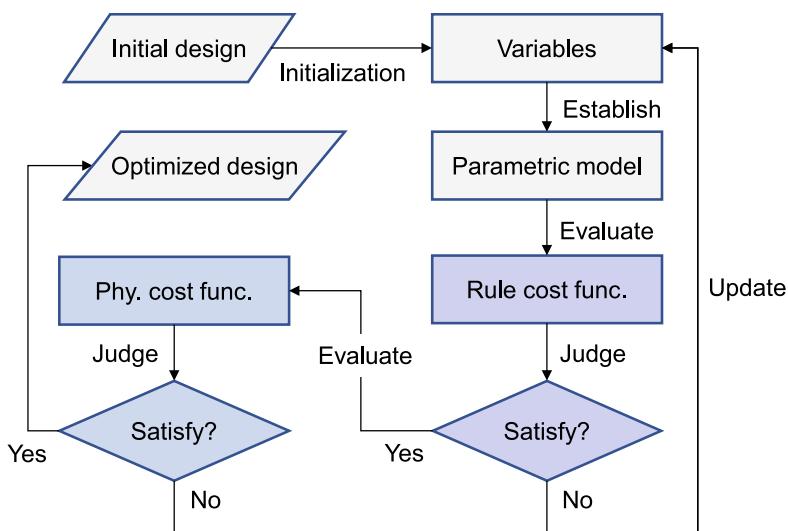


Fig. 3. Flowchart of the two-stage optimization process.

Ultimately, the design outcomes are conveyed through the output layer, enabling a seamless conversion into precise CAD drawings. Furthermore, it has the capability to be converted into a structural analysis model by utilizing parametric modeling methods for subsequent analysis and verification.

3 Case Study

3.1 Basic Information

The case study relates to a shear wall structure building consisting of eight stories, each with a height of 2900 mm and a floor area of 552 m². The seismic intensity is 8 degrees. The characteristic period at the site is 0.4 s. The building layout is shown in Fig. 4.

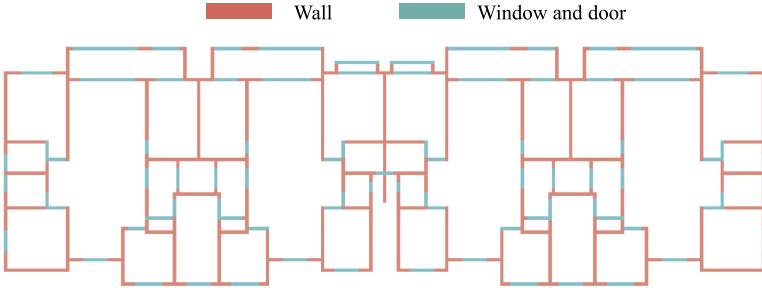


Fig. 4. Architectural layout of the case study building.

3.2 Structural Scheme Generation and Optimization Under LLM Controller

In this case study, the ChatGPT model served as the core controller. The generation algorithm utilized was StructGAN [2], while the optimization algorithm followed the two-stage approach. Figure 5 illustrates the question-answer outcomes and the modifications made to the shear wall structure during a two-round dialogue.

3.2.1 Generation

According to Fig. 6, it is evident that the initial round of dialogue triggered the utilization of structural generation, specifically the function “*shearwall_generate (height = 23.2, intensity = 8.0)*”. This function incorporated architectural drawings and the LLM analysis parameters as inputs. Through the implementation of a generative design approach, an initial design for the shear walls was formulated. The resulting design was then shared with the engineer, who proceeded to suggest additional adjustments based on the structural layout.

3.2.2 Optimization

During the second phase of the dialogue, the focus was on identifying suitable optimization targets based on the engineer’s modification requirements. To begin with, adjustments were made to the shear wall layout using the Stage 1 optimization method. Specifically, the function called was “*shearwall_change(objective = [“shape”, “corner”], degree = 0.5)*”. In this context, the term “*objective*” referred to the optimization goals, which encompassed the shapes of the shear walls and the support of corner points. The

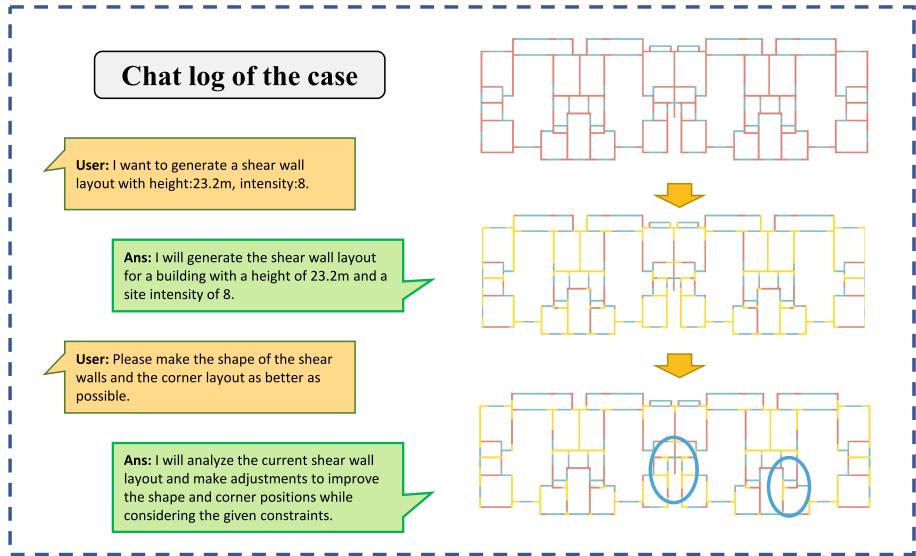


Fig. 5. Intelligent design system for the case study shear wall structure.

value of “*degree*” was set to 0.5, indicating the relative importance assigned to these two metrics in the optimization objective. The remaining weight of 0.5 was allocated to the *IoU* metric, which measures the difference between the optimized and the original schemes.

Stage1 The optimization objectives in this scenario encompassed *IoU*, *S_{shape}* index for shear wall shape, and *S_{corner}* index for corner support. A weight of 0.5 was assigned to *IoU* based on the “*degree*” parameter in the LLM invocation. Additionally, the weights for the shear wall shape and corner support objectives were determined using the entropy weight method [23, 24]. Consequently, the optimized objective function can be represented by Eq. (1).

$$F_1 = 1 - (0.5 \times IoU + 0.304 \times S_{shape} + 0.196 \times S_{corner}) \quad (1)$$

The optimization was conducted utilizing a genetic algorithm with the optimization objective *F₁*. The process exhibited a swift convergence rate, and Stage 1 was completed within a total of 459 s. Building upon this outcome, Stage 2 involved further optimization to determine the structure’s pattern and size.

Stage2 First, the shear wall groups were organized into clusters using the *k*-means algorithm. To avoid challenges caused by the high number of variables, it is important to choose a reasonable value for *K* in the *k*-means algorithm. In this case, we opted for *K* = 8 for Stage 2 optimization. The clustering outcomes for the shear wall groups can be observed in Fig. 6. Note that the shear wall groups were categorized into 8 classes according to their shapes, lengths, and positions.

The optimization process utilized an online learning method facilitated by YJK-GAMA, with Stage 2 lasting for a total of 59 min.

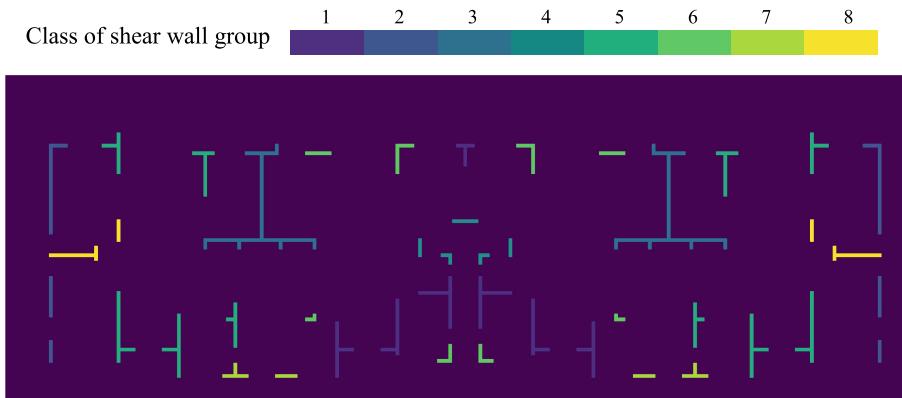


Fig. 6. Results of shear wall clustering based on k-means algorithm.

3.3 Results and Discussion

By comparing the evaluation scores and drawings of each stage between the engineer's design and the traditional method of direct optimization, the following results were obtained, as presented in Table 1.

Table 1. Evaluation results of mechanical performance and material consumption.

| Design scenario | Steel reinforcement consumption (t) | Concrete consumption (m^3) | Total material cost ($\times k$ CNY) | Maximum story drift ratio |
|-----------------------|-------------------------------------|--------------------------------|---------------------------------------|---------------------------|
| Designed by engineers | 129.74 | 1203.28 | 879.94 | 1/1269 |
| Generated by AI | 126.73 | 1395.22 | 925.49 | 1/1908 |
| After Stage 1 | 124.29 | 1269.42 | 877.99 | 1/1596 |
| After Stage 2 | 119.25 | 1186.32 | 832.90 | 1/1237 |
| Traditional method | 121.20 | 1313.26 | 878.78 | 1/1414 |

The results reveal that during Stage 1, substantial improvements were made in the arrangement of shear walls, with a notable increase in the utilization of L-shaped and T-shaped walls in Fig. 5. This initial optimization effort resulted in a preliminary reduction in material usage. In Stage 2, the optimization process shifted its focus towards refining specific dimensions, resulting in a further decrease in material consumption. Notably, when compared to the conventional direct optimization approach, both the empirical evaluation and mechanical performance–material consumption evaluation methods demonstrated noticeable enhancements. Furthermore, the empirical evaluation and mechanical performance outcomes aligned closely with the engineer's design,

while concurrently achieving an overall reduction in material consumption. However, the current design outcomes still have certain flaws in terms of details. Some shear walls do not meet the requirement of local symmetry, and a few shear walls have lengths that are too short. It can be enhanced in subsequent work by incorporating additional optimization metrics.

Finally, a comparative analysis of process time was conducted on typical design scenarios. According to the study conducted by senior engineers from various renowned design institutes in China [25], the preliminary design process for a shear wall structure with an approximate floor area of 500 m² necessitates approximately 20 h, and on top of this, the structural modeling requires around 10 h. Additionally, a considerable amount of time is often dedicated to subsequent adjustments and optimization following the preliminary design phase. In contrast, the proposed intelligent design system only requires approximately 1 h, resulting in an efficiency enhancement of approximately 30 times. Furthermore, the system can be implemented through natural language dialogues, thereby significantly reducing the learning threshold.

4 Conclusion

This paper presents a novel approach to enhance the design process of shear wall structures through the integration of LLMs and generative AI. By incorporating simple interactions, the system achieves an efficient structural design, resulting in improved automation and convenience. As a core controller of the system, LLM translates design and optimization requirements expressed in natural language into computer-executable code. Generative AI algorithms are then utilized to generate the initial layout, while the optimization objectives are controlled to refine the structure details. Through iterative interactions, valuable references for the future development of generic intelligent design systems are obtained. Following are the main conclusions derived from this study:

- (1) The intelligent design system proposed for shear wall structures is based on LLMs. This system integrates structural generation and optimization methods, with the LLM serving as the central component for interactive control. Through this approach, the entire structural design process can be automated interactively. By adopting this system, not only can the safety (mechanical performance) and cost-effectiveness (material consumption) of the design outcome be ensured, but the design efficiency can also be improved by approximately 30 times.
- (2) In this study, a two-stage optimization methodology is proposed, which is based on the consideration of empirical rules, mechanical performance, and material consumption. The objective of this methodology is to facilitate the entire process of AI intelligent generation and optimization. By employing this method, a notable reduction in material consumption can be achieved, while simultaneously ensuring that the structural design adheres to the mechanical constraints. Consequently, the optimization efficiency is also enhanced.
- (3) A well-designed prompt and comprehensive training in professional terminology were conducted in the domain of structural design for the LLM model. The successful application of the LLM method substantially reduces the difficulties of using the intelligent design system.

Acknowledgment. This work is supported by Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park Z231100005923043.

References

- Chi, H.-L., Wang, X.Y., Jiao, Y.: BIM-enabled structural design: Impacts and future developments in structural modelling, analysis and optimisation processes. *Arch Computat Methods Eng.* **22**, 135–151 (2015). <https://doi.org/10.1007/s11831-014-9127-7>
- Liao, W.J., Lu, X.Z., Huang, Y.L., Zheng, Z., Lin, Y.Q.: Automated structural design of shear wall residential buildings using generative adversarial networks. *Autom. Constr.* **132**, 103931 (2021). <https://doi.org/10.1016/j.autcon.2021.103931>
- Liao, W.J., Huang, Y.L., Zheng, Z., Lu, X.Z.: Intelligent generative structural design method for shear wall building based on “fused-text-image-to-image” generative adversarial networks. *Expert Syst. Appl.* **210**, 118530 (2022). <https://doi.org/10.1016/j.eswa.2022.118530>
- Lu, X.Z., Liao, W.J., Zhang, Y., Huang, Y.L.: Intelligent structural design of shear wall residence using physics-enhanced generative adversarial networks. *Earthq. Engng. Struct. Dyn.* **51**, 1657–1676 (2022). <https://doi.org/10.1002/eqe.3632>
- Zhao, P.J., Liao, W.J., Huang, Y.L., Lu, X.Z.: Intelligent design of shear wall layout based on attention-enhanced generative adversarial network. *Eng. Struct.* **274**, 115170 (2023). <https://doi.org/10.1016/j.engstruct.2022.115170>
- Feng, Y.T., Fei, Y.F., Lin, Y.Q., Liao, W.J., Lu, X.Z.: Intelligent generative design for shear wall cross-sectional size using rule-embedded generative adversarial network. *J. Struct. Eng.* **149**, 04023161 (2023). <https://doi.org/10.1061/JSENDH.STENG-12206>
- Zhao, P.J., Liao, W.J., Huang, Y.L., Lu, X.Z.: Intelligent design of shear wall layout based on graph neural networks. *Adv. Eng. Inform.* **55**, 101886 (2023). <https://doi.org/10.1016/j.aei.2023.101886>
- Zhao, P.J., Liao, W.J., Xue, H.J., Lu, X.Z.: Intelligent design method for beam and slab of shear wall structure based on deep learning. *J. Build. Eng.* **57**, 104838 (2022). <https://doi.org/10.1016/j.jobe.2022.104838>
- Lu, X., Lu, X.Z., Guan, H., Ye, L.P.: Collapse simulation of reinforced concrete high-rise building induced by extreme earthquakes. *Earthq. Eng. Struct. Dynam.* **42**(5), 705–723 (2013). <https://doi.org/10.1002/eqe.2240>
- Li, Y., Lu, X.Z., Guan, H., Ye, L.P.: An improved tie force method for progressive collapse resistance design of reinforced concrete frame structures. *Eng. Struct.* **33**(10), 2931–2942 (2011). <https://doi.org/10.1016/j.engstruct.2011.06.017>
- Lu, X.Z., Lin, K.Q., Li, Y., Guan, H., Ren, P.Q., Zhou, Y.L.: Experimental investigation of RC beam-slab substructures against progressive collapse subject to an edge-column-removal scenario. *Eng. Struct.* **149**, 91–103 (2017). <https://doi.org/10.1016/j.engstruct.2016.07.039>
- Zhou, X.H., et al.: Semi-automatic generation of shear wall structural models. *Structures* **51**, 42–54 (2023). <https://doi.org/10.1016/j.istruc.2023.03.031>
- Li, T., Shu, B., Qiu, X.J., Wang, Z.Q.: Efficient reconstruction from architectural drawings. *Int. J. Comput. Appl. Technol.* **38**, 177–184 (2010). <https://doi.org/10.1504/IJCAT.2010.034154>
- Zheng, Z., Lu, X.Z., Chen, K.Y., Zhou, Y.C., Lin, J.R.: Pretrained domain-specific language model for natural language processing tasks in the AEC domain. *Comput. Ind.* **142**, 103733 (2022). <https://doi.org/10.1016/j.compind.2022.103733>
- Zheng, Z., Chen, K.Y., Cao, X.Y., Lu, X.Z., Lin, J.R.: LLM-FuncMapper: Function identification for interpreting complex clauses in building codes via LLM (2023). <https://arxiv.org/abs/2308.08728v1>

16. AutoGPT: The official Auto-GPT website. <https://news.agpt.co/>
17. Shen, Y.L., Song, K.T., Tan, X., Li, D.S., Lu, W.M., Zhuang, Y.T.: HuggingGPT: solving AI tasks with ChatGPT and its friends in hugging face (2023). <http://arxiv.org/abs/2303.17580>
18. Wei, J., Wang, X.Z., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., Zhou, D.: Chain-of-thought prompting elicits reasoning in large language models (2023). <http://arxiv.org/abs/2201.11903>
19. Yao, S.Y., Yu, D., Zhao, J., Shafran, I., Griffiths, T.L., Cao, Y., Narasimhan, K.: Tree of thoughts: deliberate problem solving with large language models (2023). <http://arxiv.org/abs/2305.10601>
20. Yao, S.Y., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y.: ReAct: synergizing reasoning and acting in language models (2023). <http://arxiv.org/abs/2210.03629>
21. White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., Schmidt, D.C.: A prompt pattern catalog to enhance prompt engineering with ChatGPT (2023). <http://arxiv.org/abs/2302.11382>
22. Liao, W.J., Lu, X.Z., Fei, Y.F., Gu, Y., Huang, Y.L.: Generative AI design for building structures. *Autom. Constr.* **157**, 105187 (2024). <https://doi.org/10.1016/j.autcon.2023.105187>
23. Qin, S.Z., Liao, W.J., Lin, Y.Q., Lu, X.Z.: An efficient assessment method for intelligent design results of shear wall structure based on mechanical performance, material consumption, and empirical rules. *Eng. Mech.* (2023). <https://doi.org/10.6052/j.issn.1000-4750.2023.05.0360>
24. Zou, Z.H., Yun, Y., Sun, J.N.: Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. *J. Environ. Sci.* **18**, 1020–1023 (2006). [https://doi.org/10.1016/S1001-0742\(06\)60032-6](https://doi.org/10.1016/S1001-0742(06)60032-6)
25. Fei, Y.F., et al.: Integrated schematic design method for shear wall structures: A practical application of generative adversarial networks. *Buildings* **12**, 1295 (2022). <https://doi.org/10.3390/buildings12091295>



Efficient and Rapid Optimal Chiller Loading Using a Simplified Two-Stage Algorithm

Mohamed S. Kandil¹, Gary Chang², and J. J. McArthur¹(✉)

¹ Department of Architectural Science, Toronto Metropolitan University, Toronto, ON, Canada
jjmcarthur@torontomu.ca

² H.H. Angus and Associates Limited Consulting Engineers, Toronto, ON, Canada

Abstract. Efficient management of chiller systems in large buildings is key to energy savings in HVAC operations, with the potential to significantly reduce greenhouse gas emissions. Traditional Optimal Chiller Loading (OCL) strategies, which keep all chillers running continuously, are not always energy-efficient and can lead to higher maintenance costs. To improve this, OCL has been approached through advanced optimization techniques, allowing for chillers to be selectively switched on or off. While metaheuristic methods currently dominate research in solving OCL problems, they often suffer from complex hyperparameter tuning, slow processing times, and inherent stochasticity in their results. To address these issues, we introduce a new algorithm within the deterministic algorithms category, employing the sequential least squares quadratic programming (SLSQP) solver, known for its proficiency in handling constrained nonlinear programming. Our approach employs a two-stage strategy to effectively overcome the inherent limitations of standard optimization solvers when dealing with the discrete decision-making elements in chiller system management. This deterministic approach not only enhances the reliability of the solutions but also ensures consistent performance. After extensive testing and validation across various case studies, our algorithm has demonstrated notable efficiency. It not only computes at a fast rate but also consistently achieves optimal results rapidly, often within a split second.

Keywords: Optimal Chiller Loading · Chiller Sequencing · Energy Saving

1 Introduction

The increasing reliance on mechanical cooling systems for comfortable indoor climates has led to a significant rise in energy consumption and greenhouse gas emissions. Heating, Ventilation, and Air Conditioning (HVAC) systems, particularly in large buildings, are notably energy-intensive, accounting for 38% of building consumption, which is equivalent to 12% of the final energy use [1]. This underscores the urgency for energy optimization and conservation strategies to mitigate global warming and achieve economic savings. A critical focus area in this context is the optimization of cooling systems in large-scale infrastructures, such as buildings, data centers, or industrial complexes,

where multiple chillers are used to manage varying cooling loads. Running multiple chillers simultaneously in these settings is not always energy-efficient, especially during periods of lower cooling demand [2]. A more effective approach involves chiller staging or sequencing, combined with optimal chiller loading strategies, to selectively activate the required number of chillers and distribute loads efficiently. This yields a system that not only meets the immediate cooling needs but does so in a manner that is both environmentally and financially sound, underpinning the sustainability of large-scale cooling operations.

The strategic approach of chiller staging and optimal chiller loading involves complex decision-making. The challenge encompasses both selecting (a discrete choice problem) which chillers to run (chiller sequencing) and determining (a continuous optimization problem) how to distribute the load among them (optimal chiller loading). This dual problem is traditionally addressed through various optimization techniques. Initially, deterministic methods like the Lagrangian method [3, 4] and Dynamic Programming [5] were used. However, they often faced convergence issues, especially at low demand, leading to a shift towards stochastic optimization techniques. These include Genetic Algorithms (GA) [6], Differential Evolution (DE) [7], Particle Swarm Optimization (PSO) [8], Simulated Annealing (SA) [9], and Improved Firefly Algorithm (IFA) [10], and other metaheuristic algorithms. These algorithms, though useful for finding near-optimal solutions in complex and multimodal landscapes, do not guarantee the optimal solution. They can also be computationally intensive and may yield different results upon each run due to their stochastic nature, making them less reliable for consistent application in operational systems.

In our study, we introduce a streamlined two-stage Sequential Least Squares Quadratic Programming (2S-SLSQP) method tailored for chiller system optimization. This method leverages the strengths of SLSQP in handling nonlinear objectives and constraints and simplifies the problem by relaxing binary decisions, thereby enhancing the efficiency of the solution search. Our 2S-SLSQP algorithm uses a two-step process: it begins by relaxing the complex binary decisions, quickly identifying a broad solution space, and then refines the search for an optimal or nearly optimal solution. This approach is notably faster and more reliable than the more complex and time-consuming metaheuristic algorithms, making it highly suitable for practical chiller management scenarios. The deterministic nature of the 2S-SLSQP, coupled with its methodical navigation through the search space, ensures compliance with operational constraints and accuracy in performance. This makes it a particularly advantageous approach for optimizing part-load ratios in chiller systems, thereby enhancing energy efficiency and reducing operational costs.

The structure of this paper is laid out in the following manner: Sect. 2 introduces the methodology, encompassing the formulation of the mathematical optimization problem, the strategy proposed in this study, and a comprehensive overview of the case studies employed for validating and testing this strategy. Section 3 is dedicated to presenting and comparing the results of the proposed strategy with the conventional SLSQP approach. This is followed by an in-depth discussion in Sect. 4. The paper concludes with Sect. 5, where the final conclusions are drawn.

2 Methodology

The methodology section outlines the approach of this study, starting with mathematical optimization modeling to set the foundational context for our proposed strategy. At its core is the introduction of the two-stage Sequential Least Squares Quadratic Programming (2S-SLSQP) strategy, explained through detailed pseudo-code to illustrate its step-by-step implementation. To evaluate the effectiveness of 2S-SLSQP, the section includes two case studies involving 4-chiller and 6-chiller systems, demonstrating the strategy's adaptability and robustness in varied scenarios. These case studies are key to proving the practical relevance of the 2S-SLSQP approach in real-world settings.

2.1 Mathematical Optimization Modeling

Optimal chiller loading (OCL) aims to allocate the cooling demand efficiently across multiple chillers, ensuring that each chiller operates at its most energy-efficient point. This method not only maximizes the performance of each chiller but also leads to reduced overall power consumption. The OCL problem can be modeled as a mathematical optimization task. The primary goal is either to minimize the combined energy usage of the chiller system or to maximize their coefficient of performance (COP). In this context, the part-load ratio (PLR) for each chiller, which shows the proportion of the present load relative to its maximum design capacity, is taken as the decision variables. The power consumption for a centrifugal chiller can be described by the following cubic polynomial equation [6]:

$$P_{chiller,i} = a + b \times PLR_i + c \times PLR_i^2 + d \times PLR_i^3 \quad (1)$$

where $P_{chiller,i}$ represents the power consumed by the i^{th} chiller. The coefficients a, b, c , and d correspond to the performance metrics of the chiller, while PLR_i is the part-load ratio of the i^{th} chiller. The primary objective is to minimize the total power consumption across all chillers, given by:

$$\min_{PLR} J = \left(\sum_{i=1}^N P_{chiller,i}(PLR_i) \right) \quad (2)$$

where J is the objective function, and N stands for the total number of chillers. The total cooling output from all chillers in the system must meet the system's cooling demand which can be formulated as:

$$\sum_{i=1}^N PLR_i \times Q_i = Q_{demand}, \quad \text{for } i = 1, 2, 3, \dots, N \quad (3)$$

where Q_i represents each chiller's refrigerating capacity while Q_{demand} is the cooling load. For stable operation of chillers, the PLR should neither be too low nor exceed its capacity. Thus, the PLR is limited between allowable lower and upper bounds:

$$PLR_i^{min} \leq PLR_i \leq PLR_i^{max} \quad (4)$$

where PLR_i^{\min} and PLR_i^{\max} denote the minimum and maximum part-load ratios for each chiller, respectively. However not all chillers need to be activated simultaneously to satisfy lower cooling demands. Instead, a subset of the chillers can be activated to meet the cooling needs. By using only, the necessary number of chillers, energy efficiency can be improved, and operational costs can be reduced. It allows for more flexibility in adjusting the cooling capacity to match the actual cooling load. Therefore, we can rewrite (4) by introducing a binary decision variable U_i to determine the ON/OFF status of each chiller:

$$PLR_i^{\min} \times U_i \leq PLR_i \leq PLR_i^{\max} \times U_i \quad (5)$$

The logic here is:

- If chiller i is OFF ($U_i = 0$), then PLR_i is forced to be 0.
- If chiller i is ON ($U_i = 1$), PLR_i can vary between its minimum and maximum values.

2.2 Proposed Strategy

Incorporating the binary decision variable U_i into the constraints, as seen in (5), facilitates determining the operational status (ON/OFF) of each chiller. Nevertheless, this inclusion turns the problem into a mixed-integer nonlinear programming (MINLP) problem, demanding a specialized, robust solver. Given the inherent limitations of MINLP solvers, such as their computational complexity, scalability issues, and lengthy solution times, we propose an alternative representation in our study. This representation retains the core idea of the model while addressing these challenges:

$$PLR_i^{\min} \leq PLR_i \leq PLR_i^{\max} \quad \text{or} \quad PLR_i = 0 \quad (6)$$

Effectively, constraint defined by (6) is equivalent to (5), but without the need for a binary decision variable, offering a more streamlined expression. Furthermore, this reformulation of the constraints effectively decouples the problem into two distinct modes or stages. Each of these stages can then be addressed using a non-linear programming (NLP) solver, streamlining the overall computational process. This is exemplified in the implementation of our 2S-SLSQP algorithm, which optimizes the PLRs of a chiller system. The pseudocode for this algorithm is presented in Fig. 1. Given the chiller coefficients and the required cooling load, it iteratively addresses the underlying NLP challenge. Our algorithm ensures dynamic adjustment of inequality constraints, thereby facilitating the identification of the optimal solution.

The first stage commences with a relaxed search space, wherein the PLRs can vary from 0 to 1. During this stage, the algorithm seeks a global optimal solution by minimizing the objective function J , subject to power consumption and operational constraints. If the resulting PLRs from this stage satisfy the relaxed constraints, delineated in (6), and thus ensure all chillers operate within their feasible region or remain off, the obtained solution is deemed potentially global optimal. If Stage (1) fails to satisfy all constraints, the algorithm transitions to Stage (2), which imposes tightened bounds on the PLRs based on their respective minimum values. The search space is thus confined, and the algorithm now searches for a local optimal solution, starting from the PLRs obtained

from Stage (1). This stage effectively handles the scenario where the relaxation of constraints in Stage (1) results in an infeasible solution with respect to the original problem. Upon completion of both stages, a validation and selection process are conducted. If the solution from Stage (1) is feasible and offers a lower cost than the Stage (2) solution, it is selected as the optimal set of PLRs, considering that it might represent a global optimum. Otherwise, the algorithm selects the local optimal solution from Stage (2).

In the practical implementation of the 2S-SLSQP algorithm for chiller system optimization, we leveraged the SLSQP algorithm provided by the SciPy library in Python. The development and testing of this algorithm were performed on a robust computing setup that included a machine powered by a 6-core Intel Core i7 processor coupled with 16 GB of RAM and operated on a Windows 11 environment.

2.3 Case Studies

The first case is for a four-chiller system used by a hotel in Taipei, Taiwan [6], which includes coefficients of power consumption function and rated capacities presented in the referenced Table 1. The range of PLR_i for these chillers is established between 0.3 and 1.0. Notably, the power consumption profiles for the chillers are generally convex. However, an exception is observed with the second chiller, which demonstrates non-convex behavior when operating below a PLR_i of 0.5.

In the second case, a hospital in Kaohsiung, Taiwan, utilizes a system with six chillers. The specific coefficients and nominal capacities of these chillers are recorded in Table 2 [11]. The PLR_i vary, with a minimum of 0.3 for chillers numbered #1 to #4, and a higher minimum of 0.5 for chiller units #5 and #6. For all the chiller units, the maximum PLR_i is uniformly set at 1.0. A significant observation is that the power consumption curves for these chillers are not convex for certain segments of their operational range, except for chiller units #1 and #5 which do not display this characteristic.

3 Results

The analysis begins by comparing the conventional SLSQP algorithm, used as a reference point, with the innovative 2S-SLSQP method. Standard SLSQP results, which are the outcomes of applying the original SLSQP to the constrained OCL problem, are limited to determining the best chiller load assuming all chillers are online. Conventional SLSQP also corresponds to the stage (2) optimization illustrated in the pseudocode, see Fig. 1. This approach, while functional, lacks energy efficiency. On the other hand, the newly introduced 2S-SLSQP is designed to address both aspects of the problem: chiller staging and optimal chiller loading. By doing so, it not only fulfills immediate cooling requirements but also achieves this in an environmentally friendly and cost-effective manner, enhancing the sustainability of large-scale cooling systems. To validate the efficiency of 2S-SLSQP, both algorithms were applied to the two case studies detailed in Sect. 2.3.

Algorithm 1 2S-SLSQP for Chiller Optimization

```

1: Input:
2:  $a_i, b_i, c_i, d_i$ : Chiller coefficients for  $i = 1, \dots, N$ 
3:  $Q^i$ : Cooling capacity for chiller  $i$ 
4:  $Q_{\text{demand}}$ : Required cooling load
5:  $PLR_i^{\min}$ : Minimum part-load ratio for chiller  $i$ 
6:  $PLR_0$ : Initial part-load ratio guess
7:
8: Output:
9:  $PLR_{opt}$ : Optimized part-load ratios
10:  $P_{\min}$ : Minimized total power consumption
11:
12: Initialization:
13:  $PLR \leftarrow PLR_0 = 0.5$ 
14:  $LB_i \leftarrow PLR_i^{\min}$  for  $i = 1, \dots, N$ 
15:  $B_i \leftarrow [0, 1]$  for Stage (1) optimization
16:  $B_i^{con} \leftarrow [LB_i, 1]$  for Stage (2) optimization
17:
18: Procedure:
19: // Stage (1) Optimization:
20: Solve (2) subject to (3) using bounds  $B_i$ 
21:  $(PLR_{stage1}, Cost_{stage1}) \leftarrow$  Solution
22:
23: // Stage (2) Optimization:
24: Solve (2) subject to (3) using bounds  $B_i^{con}$ 
25: Start from  $PLR_{stage1}$ 
26:  $(PLR_{stage2}, Cost_{stage2}) \leftarrow$  Solution
27:
28: // Validation and Selection:
29: if  $PLR_{stage1}$  satisfies Eq. (6) and  $Cost_{stage1} <$ 
    $Cost_{stage2}$  then
30:    $PLR_{opt} \leftarrow PLR_{stage1}$ 
31:    $Cost_{opt} \leftarrow Cost_{stage1}$ 
32: else
33:    $PLR_{opt} \leftarrow PLR_{stage2}$ 
34:    $Cost_{opt} \leftarrow Cost_{stage2}$ 
35: end if
36:
37: return  $PLR_{opt}, Cost_{opt}$ 

```

Fig. 1. Pseudocode for the proposed 2S-SLSQP algorithm.

Table 1. Data for Case (1) Four-Chiller System

| Chiller | a | b | c | d | Capacity (RT) |
|---------|--------|---------|----------|---------|---------------|
| #1 | 104.09 | 166.57 | -430.13 | 512.53 | 450 |
| #2 | -67.15 | 1177.79 | -2174.53 | 1456.53 | 450 |
| #3 | 384.71 | -779.13 | 1151.42 | -63.20 | 1000 |
| #4 | 541.63 | 413.48 | -3626.5 | 4021.41 | 1000 |

Table 2. Data for Case (1) Six-Chiller System

| Chiller | a | b | c | d | Capacity (RT) |
|---------|---------|---------|----------|--------|---------------|
| #1 | 57.2 | 329.73 | 0.05 | 7.85 | 550 |
| #2 | 50.09 | 419.28 | -123.8 | 76.36 | 550 |
| #3 | -76.29 | 1226.94 | -709.37 | 296.93 | 1000 |
| #4 | -72.56 | 1100.42 | -145.77 | -137.1 | 1000 |
| #5 | 69.39 | 620.62 | -28.24 | 59.33 | 1000 |
| #6 | -186.18 | 1817.08 | -1755.59 | 847.43 | 1000 |

In the first case study with a four-chiller configuration, the SLSQP and 2S-SLSQP algorithms were applied, and the simulation results are tabulated in Table 3. For high-load conditions, specifically between 60% and 90% of the system's maximum cooling capacity, both algorithms displayed similar behavior with all chillers in operation, resulting in identical power consumption values. Conversely, under lower demand conditions (below 60% of maximum capacity), the 2S-SLSQP algorithm demonstrated a more energy-efficient approach by deactivating chillers not required to meet the cooling load. Specifically, at a load of 1450 RT, the 2S-SLSQP algorithm reduced the Total Power consumption to 820.07 kW, a savings of 8.65% compared to the standard SLSQP's consumption of 897.59 kW. At a reduced load of 1160 RT, the savings were even more pronounced, with the 2S-SLSQP consuming 651.07 kW, which constitutes a 23.39% reduction in comparison to the SLSQP's 849.99 kW.

In the second case study, a six-chiller system, the operational efficiency of SLSQP and 2S-SLSQP was evaluated at different load levels and summarized in Table 4. At the system's maximum load of 4080 RT, the SLSQP algorithm maintained all chillers in operation with a total power consumption of 3003.85 kW while the 2S-SLSQP algorithm optimized the system by switching off chiller #4, resulting in a slightly reduced total power consumption of 2982.15. As the demand decreases, the 2S-SLSQP algorithm begins to exhibit its optimization capability by selectively deactivating chillers. At 3570 RT, it shuts down chiller #4, reducing the Total Power to 2618.39 kW, which is a 0.71% energy saving compared to the SLSQP's 2637.28 kW. This trend becomes more prominent at lower loads. For example, at 2040 RT, the 2S-SLSQP deactivates three chillers (#3, #4, and #6), achieving a Total Power of 1493.85 kW, which is a significant

8.69% reduction from the SLSQP's 1636.01 kW. The most considerable savings are observed at the load of 1530 RT, where 2S-SLSQP turns off four chillers (#2, #3, #4, and #6) operating at a much-reduced power consumption of 1101.15 kW. This power usage represents approximately two-thirds of the SLSQP's consumption, translating to a substantial energy saving of approximately 29.84%.

The 2S-SLSQP strategy consistently maintains a low runtime in both case studies. For instance, in the first case study, at the highest load of 2610 RT, the runtime remains as low as 0.011 s, and it maintains this same runtime of 0.011 s at the lowest load of 1160 RT. Similarly, in the second case study, the runtime at the highest load of 4080 RT is an impressively short 0.010 s, and at the lowest load of 1530 RT, it is a mere 0.009 s. These runtime results underscore the remarkable computational efficiency of the 2S-SLSQP strategy.

The performance of the proposed 2S-SLSQP algorithm was compared with several established optimization algorithms: Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Differential Evolution (DE), Improved Firefly Algorithm (IFA), and Differential Cuckoo Search Approach (DCSA). The evaluation was conducted across the two case studies with varying load demands. Due to the stochastic nature of certain benchmarking algorithms, we are unable to directly compare their optimal results. Instead, we employ the following metrics for our comparative analysis: minimum (Min), maximum (Max), mean, and standard deviation (SD).

For the four-chillers system, the 2S-SLSQP algorithm consistently achieved optimal power consumption values with a standard deviation of zero at multiple loads, indicating its deterministic nature as shown in Table 5. At 2610 RT, 2320 RT, and 2030 RT loads, it matched the lowest recorded power consumptions of 1857.30 kW, 1455.66 kW, and 1178.14 kW, respectively, demonstrating its reliability and precision. At intermediate loads of 1740 RT and lower loads of 1450 RT and 1160 RT, the 2S-SLSQP algorithm continued to maintain optimal performance with power consumptions of 998.53 kW, 820.07 kW, and 651.07 kW, each with zero variability.

In the second case study, the 2S-SLSQP algorithm's performance was closely benchmarked against other optimization algorithms across different load conditions as shown in Table 6. At the highest load of 4080 RT, both 2S-SLSQP and PSO achieved the lowest power consumption at 2982.15 kW. However, the 2S-SLSQP demonstrated a distinct advantage with an SD of 0.00, indicating absolute consistency in its results, compared to PSO's SD of 13.61, which suggests a slight variability in its outcomes. For a load of 3570 RT, the IFA recorded the lowest power consumption at 2611.16 kW. The 2S-SLSQP algorithm delivered a marginally higher result with a negligible difference of approximately 0.28%. This small deviation is practically insignificant, especially considering that the 2S-SLSQP maintained an SD of 0.00, ensuring predictable and stable results. At a load of 3060 RT, PSO led with the lowest recorded power consumption of 2226.46 kW and an SD of 16.42. The 2S-SLSQP algorithm's result was almost identical, with a power consumption of 2227.88 kW. Although the difference is minor at about 0.06%, the 2S-SLSQP's zero standard deviation once again underscores its deterministic nature. The load of 2550 RT presented a unique situation where the IFA achieved the best performance with a power consumption of 1853.92 kW, albeit with a higher SD of 26.96. In contrast, the 2S-SLSQP's power consumption was higher than all benchmarks,

Table 3. Optimal chiller loading results for Case (1) four-chillers system.

| Load (RT) | Chiller | SLSQP | | | 2S-SLSQP | | | |
|-----------|---------|--------|------------------|------------------|----------|------------------|------------------|----------|
| | | Status | PLR _i | Total Power (kW) | Status | PLR _i | Total Power (kW) | time (s) |
| 2610 | #1 | ON | 0.99 | 1857.3 | ON | 0.99 | 1857.30 | 0.011 |
| | #2 | ON | 0.91 | | ON | 0.91 | | |
| | #3 | ON | 1 | | ON | 1.00 | | |
| | #4 | ON | 0.76 | | ON | 0.76 | | |
| 2320 | #1 | ON | 0.83 | 1455.66 | ON | 0.83 | 1455.66 | 0.019 |
| | #2 | ON | 0.81 | | ON | 0.81 | | |
| | #3 | ON | 0.9 | | ON | 0.90 | | |
| | #4 | ON | 0.69 | | ON | 0.69 | | |
| 2030 | #1 | ON | 0.73 | 1178.14 | ON | 0.73 | 1178.14 | 0.018 |
| | #2 | ON | 0.74 | | ON | 0.74 | | |
| | #3 | ON | 0.72 | | ON | 0.72 | | |
| | #4 | ON | 0.65 | | ON | 0.65 | | |
| 1740 | #1 | ON | 0.6 | 998.53 | ON | 0.60 | 998.53 | 0.020 |
| | #2 | ON | 0.66 | | ON | 0.66 | | |
| | #3 | ON | 0.56 | | ON | 0.56 | | |
| | #4 | ON | 0.61 | | ON | 0.61 | | |
| 1450 | #1 | ON | 0.53 | 897.59 | ON | 0.61 | 820.07 | 0.017 |
| | #2 | ON | 0.3 | | OFF | 0.00 | | |
| | #3 | ON | 0.49 | | ON | 0.57 | | |
| | #4 | ON | 0.59 | | ON | 0.61 | | |
| 1160 | #1 | ON | 0.3 | 849.99 | OFF | 0.00 | 651.07 | 0.011 |
| | #2 | ON | 0.3 | | OFF | 0.00 | | |
| | #3 | ON | 0.35 | | ON | 0.56 | | |
| | #4 | ON | 0.54 | | ON | 0.60 | | |

standing at 1959.44 kW. Despite this suboptimal result, the 2S-SLSQP maintained an SD of 0.00, reinforcing its consistent output. The reasons behind this particular outcome will be explored in the discussion section. For lower loads of 2040 RT and 1530 RT, the DE algorithm achieved the lowest power consumption figures of 1482.34 kW and 1100.94 kW, respectively, with notable SDs of 37.45 and 42.72. The 2S-SLSQP's results were slightly higher at 1493.85 kW and 1101.15 kW for each respective load, yet with the significant benefit of an SD of 0.00, illustrating the predictability and reliability of the proposed algorithm despite a negligible marginal difference in performance of about 0.77% and 0.02%.

Table 4. Optimal chiller loading results for Case (2) six-chillers system.

| Load
(RT) | Chiller | SLSQP | | | | | | 2S-SLSQP | | | | | |
|--------------|---------|--------|----|------------------|------|------------------|--------|----------|------------------|------|------------------|----------|--|
| | | Status | | PLR _i | | Total Power (kW) | Status | | PLR _i | | Total Power (kW) | Time (s) | |
| 4080 | #1 #4 | ON | ON | 1.00 | 0.30 | 3003.85 | ON | OFF | 1.00 | 0.00 | 2982.15 | 0.010 | |
| | #2 #5 | ON | ON | 0.92 | 0.84 | | ON | ON | 1.00 | 1.00 | | | |
| | #3 #6 | ON | ON | 1.00 | 0.88 | | ON | ON | 1.00 | 0.98 | | | |
| 3570 | #1 #4 | ON | ON | 1.00 | 0.30 | 2637.28 | ON | OFF | 1.00 | 0.00 | 2618.39 | 0.010 | |
| | #2 #5 | ON | ON | 1.00 | 0.96 | | ON | ON | 0.85 | 0.76 | | | |
| | #3 #6 | ON | ON | 0.30 | 0.91 | | ON | ON | 0.94 | 0.86 | | | |
| 3060 | #1 #4 | ON | ON | 1.00 | 0.30 | 2284.47 | ON | OFF | 1.00 | 0.00 | 2227.88 | 0.007 | |
| | #2 #5 | ON | ON | 0.76 | 0.66 | | ON | ON | 1.00 | 1.00 | | | |
| | #3 #6 | ON | ON | 0.30 | 0.84 | | OFF | ON | 0.00 | 0.96 | | | |
| 2550 | #1 #4 | ON | ON | 1.00 | 0.30 | 1959.44 | ON | ON | 1.00 | 0.30 | 1959.44 | 0.008 | |
| | #2 #5 | ON | ON | 0.63 | 0.55 | | ON | ON | 0.63 | 0.55 | | | |
| | #3 #6 | ON | ON | 0.30 | 0.50 | | ON | ON | 0.30 | 0.50 | | | |
| 2040 | #1 #4 | ON | ON | 0.50 | 0.30 | 1636.01 | ON | OFF | 1.00 | 0.00 | 1493.85 | 0.012 | |
| | #2 #5 | ON | ON | 0.30 | 0.50 | | ON | ON | 0.99 | 0.94 | | | |
| | #3 #6 | ON | ON | 0.30 | 0.50 | | OFF | OFF | 0.00 | 0.00 | | | |
| 1530 | #1 #4 | ON | ON | 0.30 | 0.30 | 1569.29 | ON | OFF | 1.00 | 0.00 | 1101.15 | 0.009 | |
| | #2 #5 | ON | ON | 0.30 | 0.50 | | OFF | ON | 0.00 | 0.98 | | | |
| | #3 #6 | ON | ON | 0.30 | 0.50 | | OFF | OFF | 0.00 | 0.00 | | | |

4 Discussion

Our proposed 2S-SLSQP algorithm provides a structured approach to tackling the complexities of the MINLP problem inherent in the optimal chiller loading process, without the need for binary decision variables. Tailored to address the nuances of chiller systems, the algorithm strategically navigates the search space for the PLRs. It initially attempts to capture the global optimum by solving a relaxed version of the MINLP problem, thereby simplifying the intricate interdependencies of the system components. Subsequently, the search space is methodically tightened, converging on a solution that ensures both operational efficiency and feasibility in real-world applications. This two-stage process, refined for the chiller loading problem, allows for a more precise and practical optimization, aiming to deliver both energy efficiency and cost savings.

In both case studies, the 2S-SLSQP algorithm's performance was remarkable, consistently achieving optimal or near-optimal results with negligible deviations from the best-reported outcomes of other established algorithms. Notably, its deterministic nature led to

Table 5. Comparison of Optimal Results with Selected Benchmarks for Case (1)

| Method | Load = 2610 RT | | | | | | Load = 2320 RT | | | | | | Load = 2030 RT | | | | | |
|-----------|----------------|---------|---------|-------|----------------|---------|----------------|-------|----------------|---------|---------|-------|----------------|-----|------|----|-----|-----|
| | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max |
| GA [6] | 1862.18 | — | — | — | 1457.23 | — | — | — | 1183.80 | — | — | — | — | — | — | — | — | — |
| LGM [3] | 1857.30 | 1864.17 | — | — | 1455.66 | 1461.05 | — | — | 1178.14 | 1182.50 | — | — | — | — | — | — | — | — |
| PSO [12] | 1857.30 | — | 1857.43 | 0.04 | 1455.66 | 1462.32 | 1462.34 | 20.03 | 1178.14 | 1180.54 | 1178.14 | 0.00 | — | — | — | — | — | — |
| DE [7] | 1857.30 | 1858.57 | 1857.43 | 0.40 | 1455.66 | 1455.66 | 1455.66 | 0.00 | 1178.14 | 1178.14 | 1178.14 | 0.00 | — | — | — | — | — | — |
| DCSA [13] | 1857.30 | 1857.40 | 1857.32 | 0.02 | 1455.67 | 1458.81 | 1455.81 | 0.53 | 1178.14 | 1199.50 | 1181.07 | 4.80 | — | — | — | — | — | — |
| 2S-SLSQP | 1857.30 | 1857.30 | 1857.30 | 0.00 | 1455.66 | 1455.66 | 1455.66 | 0.00 | 1178.14 | 1178.14 | 1178.14 | 0.00 | — | — | — | — | — | — |
| Method | Load = 1740 RT | | | | | | Load = 1450 RT | | | | | | Load = 1160 RT | | | | | |
| | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max |
| GA [6] | 1001.62 | — | — | — | 907.72 | — | — | — | 856.30 | — | — | — | — | — | — | — | — | — |
| LGM [3] | 998.53 | 1002.22 | — | — | 904.62 | 907.97 | — | — | 849.99 | 853.13 | — | — | — | — | — | — | — | — |
| PSO [12] | 998.53 | 1005.36 | 1005.36 | 5.71 | 820.07 | 847.93 | 826.52 | 10.88 | 651.07 | 691.19 | 667.12 | 19.65 | — | — | — | — | — | — |
| DE [7] | 998.53 | 1009.20 | 1000.21 | 3.66 | 820.07 | 821.28 | 820.19 | 0.38 | 651.07 | 655.63 | 651.53 | 1.44 | — | — | — | — | — | — |
| DCSA [13] | 1008.24 | 1074.55 | 1038.13 | 25.72 | 825.72 | 897.06 | 838.05 | 17.43 | 652.16 | 794.25 | 713.17 | 44.02 | — | — | — | — | — | — |
| 2S-SLSQP | 998.53 | 998.53 | 998.53 | 0.00 | 820.07 | 820.07 | 820.07 | 0.00 | 651.07 | 651.07 | 651.07 | 0.00 | — | — | — | — | — | — |

Table 6. Comparison of Optimal Results with Selected Benchmarks for Case (2)

| Method | Load = 4080 RT | | | | | Load = 3570 RT | | | | | Load = 3060 RT | | | | | |
|----------|----------------|---------|---------|---------|----------------|----------------|---------|---------|----------------|---------|----------------|-------|-----|-----|------|----|
| | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD |
| GA [6] | 3002.48 | 3080.04 | 3051.41 | 16.51 | 2652.79 | 2733.84 | 2690.37 | 15.61 | 2281.31 | 2391.24 | 2335.34 | 20.26 | | | | |
| PSO [12] | 2982.15 | 3053.28 | 3004.75 | 13.61 | 2617.29 | 2675.91 | 2645.47 | 17.19 | 2226.46 | 2315.43 | 2250.23 | 16.42 | | | | |
| DE [7] | 3008.38 | 3106.96 | 3065.37 | 13.81 | 2645.28 | 2784.49 | 2718.82 | 33.35 | 2242.21 | 2429.14 | 2390.57 | 51.94 | | | | |
| IFA [10] | 2984.21 | 3053.09 | 3013.61 | 14.22 | 2611.16 | 2678.05 | 2661.03 | 11.91 | 2245.24 | 2347.62 | 2294.24 | 25.38 | | | | |
| 2S-SLSQP | 2982.15 | 2982.15 | 0.00 | 2618.39 | 2618.39 | 2618.39 | 0.00 | 2227.88 | 2227.88 | 2227.88 | 0.00 | | | | | |
| Method | Load = 2550 RT | | | | | Load = 2040 RT | | | | | Load = 1530 RT | | | | | |
| | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD | Min | Max | Mean | SD |
| GA [6] | 1913.82 | 2029.16 | 1977.87 | 24.85 | 1518.30 | 1651.93 | 1601.51 | 23.27 | 1158.77 | 1284.01 | 1217.25 | 24.37 | | | | |
| PSO [12] | 1913.82 | 2029.16 | 1977.87 | 24.85 | 1518.30 | 1651.93 | 1601.51 | 23.27 | 1158.77 | 1284.01 | 1217.25 | 24.37 | | | | |
| DE [7] | 1838.67 | 1854.14 | 1842.69 | 6.82 | 1482.34 | 1664.87 | 1525.33 | 37.45 | 1100.94 | 1321.54 | 1159.24 | 42.72 | | | | |
| IFA [10] | 1853.92 | 1971.80 | 1926.22 | 26.96 | 1486.83 | 1617.98 | 1547.35 | 26.91 | 1107.35 | 1215.08 | 1170.71 | 26.48 | | | | |
| 2S-SLSQP | 1959.44 | 1959.44 | 0.00 | 1493.85 | 1493.85 | 1493.85 | 0.00 | 1101.15 | 1101.15 | 1101.15 | 0.00 | | | | | |

a standard deviation of 0.00 across all tested scenarios, reflecting its robustness and reliability. This is particularly advantageous in real-world operational environments, where energy management systems rely on frequent, regular optimization intervals—commonly every 15 min—to accommodate varying load conditions. The 2S-SLSQP algorithm excelled within this framework, demonstrating exceptional convergence speeds that far exceeded the requirements of such operational frequencies.

The rapid convergence of the 2S-SLSQP algorithm, documented at fractions of a second across both case studies, offers direct economic advantages. By significantly reducing the computational resources and time needed to arrive at optimal solutions, the algorithm cuts down the operational costs associated with the computational infrastructure needed for running such complex optimizations. Furthermore, the lower energy requirements for computation dovetail with the goals of environmental sustainability, reducing the carbon footprint associated with the optimization processes. This dual benefit—economic and ecological—positions the 2S-SLSQP algorithm as a superior alternative to traditional metaheuristic algorithms, which are typically more computationally demanding and slower to converge.

Nevertheless, the algorithm did exhibit sensitivity to the initial guess at a load of 2550 RT in case study 2, producing a suboptimal result when compared with other benchmarks. This outlier highlights the importance of incorporating a multi-start strategy in future research endeavors, where the optimization process is repeatedly executed with diverse initial guesses to guarantee the identification of the most optimal outcome.

Looking ahead, there is potential to enhance the execution speed of the 2S-SLSQP algorithm further by leveraging parallel computing techniques. This could reduce already minimal convergence times, translating to even greater economic and environmental efficiency. Additionally, the robustness and adaptability of the algorithm will be evaluated through a broader range of case studies, further validating its effectiveness across different operational scenarios and system scales.

5 Conclusion

The research presented in this paper introduces the 2S-SLSQP algorithm as a robust and efficient solution to the chiller loading optimization problem, a complex MINLP challenge characteristic of large-scale cooling systems. Our findings from two comprehensive case studies demonstrate the algorithm's capability to achieve optimal or near-optimal solutions consistently, with the added advantage of a deterministic output and unparalleled speed in convergence.

Significantly, the 2S-SLSQP algorithm's ability to converge rapidly, often in just fractions of a second, makes it highly compatible with the dynamic nature of modern energy management systems. This efficiency translates into significant cost savings and reduced energy demands for the computational processes, thereby contributing to the sustainability goals of the broader energy system.

While the algorithm displayed superior performance across most tested scenarios, it showed a degree of sensitivity to initial conditions at certain load levels, particularly at 2550 RT. This insight indicates an area for further refinement, suggesting the adoption of a multi-start strategy to ensure robustness across all possible scenarios. Future

developments could also explore the benefits of parallel computing to enhance the algorithm's execution speed even further. The effectiveness of the 2S-SLSQP algorithm represents a notable step forward in optimization techniques for energy systems, contributing positively to the field. Its economic and environmental benefits, coupled with its operational reliability, position it as a valuable tool for practitioners and researchers alike. As we move forward, expanding the range of case studies and exploring advanced computational techniques will continue to elevate the potential of this algorithm, driving innovation in energy management optimization.

Acknowledgments. This research was funded by the Natural Science and Engineering Research Council (Canada) [ALLRP-580958-22] and H. H. Angus and Associates Consulting Engineers Ltd.

References

1. González-Torres, M., Pérez-Lombard, L., Coronel, J.F., Maestre, I.R., Yan, D.: A review on buildings energy information: Trends, end-uses, fuels and drivers. *Energy Rep.* **8**, 626–637 (2022)
2. Liquid-Chilling Systems. In: ASHRAE Handbook—HVAC Systems and Equipment (2020)
3. Chang, Y.-C.: A novel energy conservation method—optimal chiller loading. *Electric Power Systems Research* **69**(2–3), 221–226 (2004)
4. Chang, Y.-C., Chan, T.-S., Lee, W.-S.: Economic dispatch of chiller plant by gradient method for saving energy. *Appl. Energy* **87**(4), 1096–1101 (2010)
5. Chang, Y.-C.: An outstanding method for saving Energy-optimal chiller operation. *IEEE Trans. Energy Convers.* **21**(2), 527–532 (2006)
6. Chang, Y.-C., Lin, J.-K., Chuang, M.-H.: Optimal chiller loading by genetic algorithm for reducing energy consumption. *Energy and Buildings* **37** (2) (2005)
7. Lee, W.-S., Yi-Ting Chen, Y. K.: Optimal chiller loading by differential evolution algorithm for reducing energy consumption. *Energy and Buildings* **43** (2–3), 599–604 (2011)
8. Ardakani, A.J., Ardakani, F.F., Hosseiniyan, S.: A novel approach for optimal chiller loading using particle swarm optimization. *Energy and Buildings* **40**(12), 2177–2187 (2008)
9. Chang, Y.-C.: An innovative approach for demand side management—optimal chiller loading by simulated annealing. *Energy* **31**(12), 1883–1896 (2006)
10. Coelho, L. d. S., Mariani, V. C.: Improved firefly algorithm approach applied to chiller loading for energy conservation. *Energy and Buildings* **59**, 273–278 (2013)
11. Lo, C.-C., Tsai, S.-H., Lin, B.-S.: Economic dispatch of chiller plant by improved ripple bee swarm optimization algorithm for saving energy. *Appl. Therm. Eng.* **100**, 1140–1148 (2016)
12. Lee, W.-S., Lin, L.-C.: Optimal chiller loading by particle swarm algorithm for reducing energy consumption. *Appl. Therm. Eng.* **29**(8–9), 1730–1734 (2009)
13. Coelho, L. d. S., Klein, C. E., Sabat, S. L., Mariani, V. C.: Optimal chiller loading for energy conservation using a new differential cuckoo search approach. *Energy* **75**, 237–243 (2014)



Gray-Box Modeling of an Energy Recovery Ventilator Using Streamed Data

Neda Adimi and J. J. McArthur^(✉)

Department Architectural Science, Toronto Metropolitan University, Toronto, ON, Canada
j.jmcARTHUR@torontomu.ca

Abstract. This paper presents the development of a complex emulator to support the energy prediction and controls optimization for an energy recovery ventilator. The emulator comprises gray-box models of several components, namely the enthalpy recovery wheel, heating and cooling coils, and supply and exhaust fans. For each constituent model, a physics-based model was created and tuned with streamed data from a test building. The emulator was tuned using data from the primary energy recovery ventilator serving as the primary ventilation system for a multi-use academic building and tested using the two parallel units of the same type to test its validity and the generalizability of the tuned parameters. This paper presents the full process for creating this model, including the building automation system data streaming method, pre-processing algorithms implemented to permit online learning, physics-based model development, and parameter optimization to tune unknown coefficients using the streamed data. Results demonstrate that the streaming and tuning method was effective to rapidly tune the emulator and that minimal effort was required to re-tune for previously-unseen pieces of equipment. Further, because of the modular nature of the emulator, its constituent models have significant potential to support gray-box modeling of fan coil units and air handling units, thus providing significant value for creating digital twins of air-side HVAC systems.

Keywords: ERV Efficiency · Grey-box Models · HVAC Models

1 Introduction

Precise predictions of energy performance in Heating, Ventilation, and Air-Conditioning (HVAC) systems are crucial for optimizing building energy system efficiency and minimizing energy consumption. Until now, several sophisticated predictive modeling approaches for HVAC systems have been introduced. These prediction models can be broadly classified into three types: physics-based model predictions (white-box), data-driven model predictions (black-box), and hybrid models (grey-box) to articulate the behavioral characteristics of HVAC systems [1–5]. Reviewing HVAC system modeling methods highlights the benefits of gray-box models, which build on physical relationships and enhance or complete these with data-driven modeling. This approach strikes a balance by combining theoretical principles with empirical data to achieve a more comprehensive and adaptable model [6, 7].

Numerous prior studies have focused on hybrid modeling approaches for HVAC systems. For example, Afram and Janabi-Sharifi (2015) developed gray-box models for residential HVAC subsystems. The model, replicating the existing HVAC system, serves as a tool to assess advanced controllers and energy conservation strategies prior to implementation [1]. Another research creates gray-box models for a Canadian residential HVAC system with various subsystems. Achieving a strong fit of 70–88% using measured data from a house during both winter and summer. The models rely on energy balance equations and distinct parameter sets for each season [2]. Another study [3] tackles challenges in current white-box and gray-box models for residential air conditioner energy prediction, presenting a simplified gray-box steady-state model validated through field experiments. The model accurately forecasts cooling capacity and electrical power input with normalized root mean square errors of 2.3% and 0.87%, respectively [3]. The other paper suggests a grey-box model for air-handling units, representing AHU elements as products of static and dynamic gains. This approach combines theoretical modeling, experimental parameter identification, and optimization to address temperature and humidity control nonlinearities [4].

This brief review reveals several shortcomings in previous works. The most notable limitation is the absence of comprehensive models that incorporate both the building and the entire ERV system with detailed components. Many studies focus on specific elements such as cooling coils, heating coils, or AHUs, rather than considering the holistic HVAC system. Additionally, most studies evaluated the residential buildings. Some models exhibit simplifications that eliminate crucial features, and in certain cases, simplifying assumptions are employed to mitigate the complexity of thermal interactions. These simplifications pose challenges in obtaining reliable analytical models.

This addresses this research gap by developing a comprehensive gray-box model for the HVAC system, modeling of each sub-component, in a complex mixed-use building situated in ASHRAE zone 5, Canada. The proposed approach involves developing a grey-box (hybrid) model that seamlessly combines streamed thermal data from building automation system (BAS_sensors of test building with the various components constituting the complete HVAC system. This hybrid identification method incorporates physical functions based on thermal equations, encompassing all potential known & unknown variables or features that impact the entire system. Thus, this paper presents the full process for creating this model, including the BAS data streaming method, pre-processing algorithms implemented to permit online learning, physics-based model development, and parameter optimization using both linear regression and deep learning approaches to tune unknown coefficients using the streamed data.

2 Methodology

The methodology used for this research consists of two key phases: (1) data collection and pre-processing and (2) emulator development. The latter are discussed in detail in this section and the real equipment used to create the emulator is presented. The detailed emulator development and parameter estimation are discussed in Sect. 3.

The equipment emulated is one of three parallel energy recovery ventilators (ERVs) serving the academic portion of the mixed-use building. Each ERV (see Fig. 1) consists of an energy recovery wheel (“HRW”), heating and cooling coils, a humidifier (heating only), supply and exhaust fans, and filters.

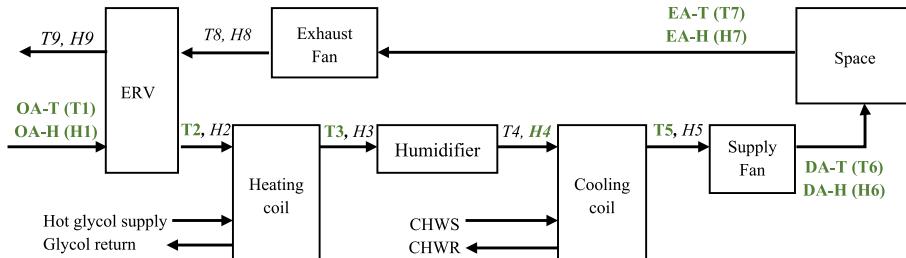


Fig. 1. Schematic of one of the ERVs showing the various interconnections with other building systems, Bold green for known text and Black Italic for unknown values

In order to maintain the integrity of the firewalls protecting the campus during data collection, a pair or virtual machines is used for data streaming. The first is installed on the BAS network, stores data in a local buffer, and streams it via a one-way VPN to the second through a, which again buffers and formats the streamed data and sends it to a cloud data lake where it is stored in a time-series database. Data is streamed from all available BAS data points with missing values imputed using a backward and forward method and outliers removed. A relational database in the data lake supports queries of this streamed data. Data was pre-processed to remove erroneous values and divide the ERV data into three operational periods: occupied heating mode, occupied cooling mode, and unoccupied modes where air is recirculated in the building.

For the purpose of prediction, there are six known or easily estimated values: return air conditions (22°C , 30% RH in heating; 24°C , 60% RH in cooling), discharge air temperature and relative humidity (setpoints), and outdoor air conditions (from weather forecasts). Other sensors have been installed that measure the real-time values of T_2 , T_3 , H_4 , T_5 , and T_9 ; however, these values are not readily predictable in advance. These were inferred through the creation of a series of sub-models, as illustrated in Fig. 2.

3 Results: Emulator Development

The following subsections present the derivation of the governing equations and show the inferred model performance for each device.

3.1 Energy Recovery Wheel

The first – and last – step in the modeling process is to determine the energy transfer at the energy recovery wheel (“HRW”). The output air temperature and humidity at the wheel are a function of its input air conditions. The effectiveness – and thus the degree

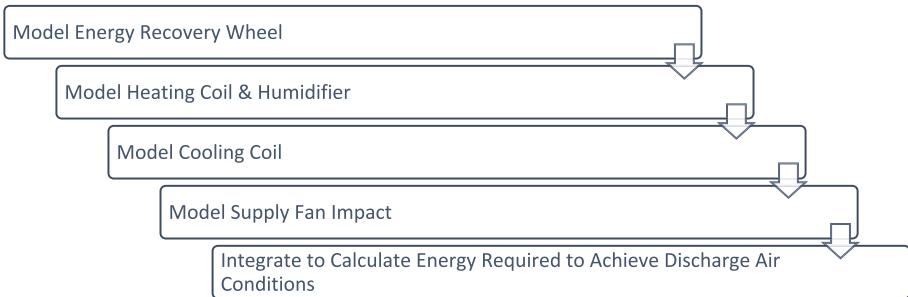


Fig. 2. Emulator development through a series of sub-models.

of heat transfer – of an energy recovery wheel varies with flow rate, wheel speed, and inlet temperatures, however at the AHRI testing points, they are known. Further, there are temperature sensors at both discharges, permitting supervised learning to be used to train a regression model.

The entering air conditions at the wheel were calculated as follows. Outdoor air temperature and humidity are predictable known values, as are return air conditions. A standard fan pick-up of 1°C was used to estimate the air temperature increase across the exhaust fan and because this is a sensible heat gain, the change in humidity is assumed to be zero. Recognizing that absolute humidity varies with temperature and is distinct from relative humidity at each temperature, the next step involved precise humidity determination at each stage. This was achieved by initially converting the relative humidity of outdoor and returned air (H_1 and H_8) into moisture content values (W_1 and W_8 ¹) using a psychrometric chart. Subsequent calculations for each component are then carried out based on absolute humidity and temperature. Additionally, as per the as-built shop drawings, the fans were balanced such that the airflow through the exhaust fan is 100% of the supply fan airflow, which was noted from the data to be operating at a constant speed (~20,800 l/s).

First, in our analysis of the variables used to predict discharge air temperature and relative humidity, a correlation matrix is utilized to evaluate the strength and direction of linear relationships among these continuous variables. Figure 3 illustrates a distinct negative correlation between outdoor air temperature (T_1) and the return air difference from outdoor air temperature ($T_8_minus_T_1$). This indicates that an increase in one variable corresponds to a predictable decrease in the other. However, for the remaining variables, linear regression did not reveal robust associations.

The observed correlation also extends to relative humidity, where H_1 shows a comparable negative correlation with the return humidity difference ($H_8_minus_H_1$). Similar to the temperature correlation mentioned earlier, an increase in H_1 corresponds to a predictable decrease in $H_8_minus_H_1$. These consistent patterns provide valuable insights into the interconnected dynamics of our studied parameters. Moreover, the discharge air humidity (H_2_Calc), calculated by assuming that the heat recovery wheel efficacy is constant as implied by the manufacturer specifications, exhibits a robust linear correlation of 0.93 with the return air humidity.

¹ kg_water/kg_dry_air.

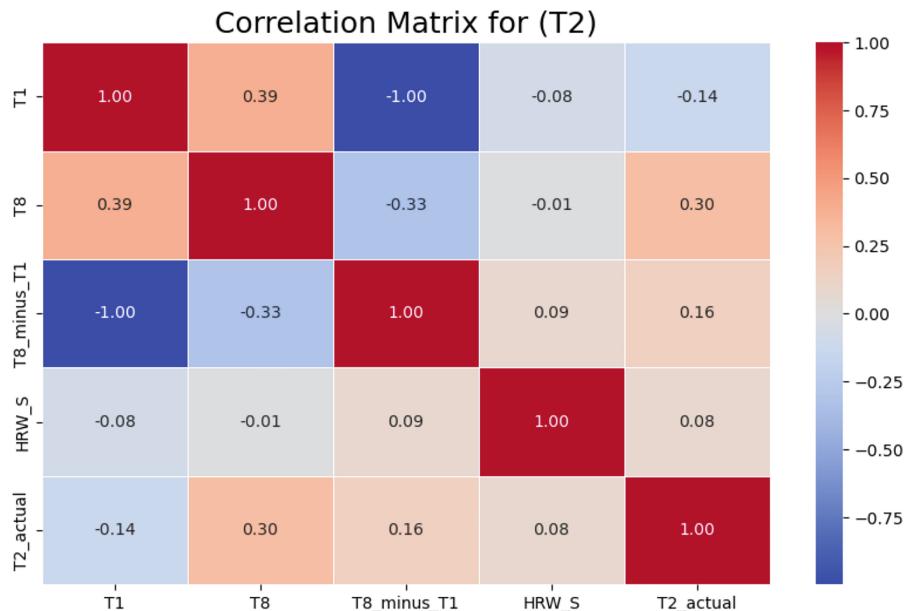


Fig. 3. Correlation Matrix of discharge air temperature.

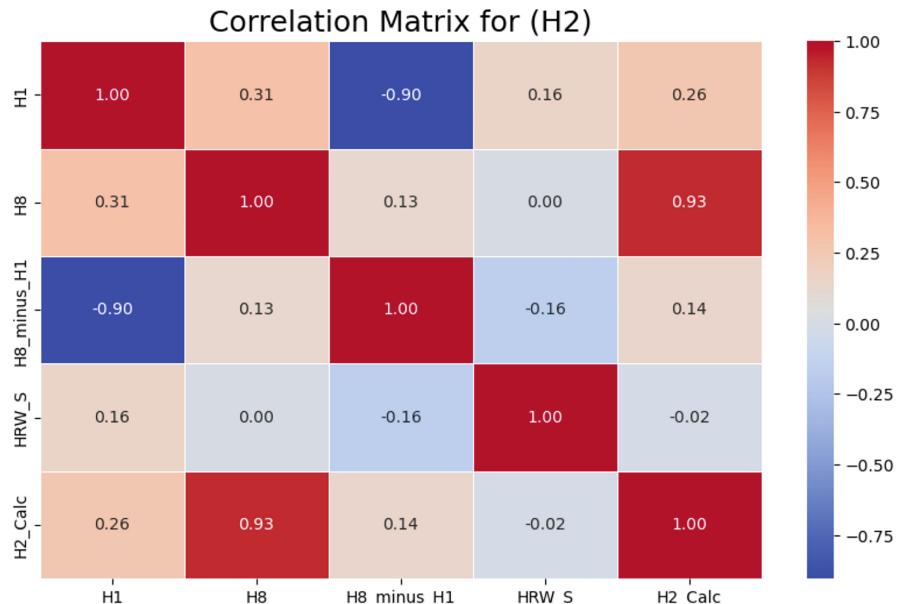


Fig. 4. Correlation Matrix of discharge air Humidity.

The correlation matrix indicates a lack of linear regression among other variables. Consequently, in response to the unsatisfactory performance observed in linear regression, a non-linear regression approach, specifically implementing a tree forest model, is employed.

The application of the Random Forest Model to regression problems, specifically in predicting continuous outcomes such as T2 shown in Fig. 5, has yielded promising results. Evaluation of the model's performance showcases compelling metrics: the Mean Squared Error (MSE) is notably low at 0.0414, signifying minimal average squared differences between predicted and actual values. Additionally, the R-squared (R2) value is remarkable at 0.9827, indicating the model's adeptness in explaining approximately 98.27% of the variance in T2. This high R2 underscores the model's robust fit to the dataset. Furthermore, an examination of the residual plot emphasizes the high accuracy of the temperature prediction model. There exists a robust correlation between the model's predictions and the actual results. However, it reveals the presence of an outlier in the data. To address this, it is required to conduct data validation and transformation techniques again, aiming to enhance the dataset's accuracy and reliability for subsequent analyses. The same analysis of forest regression model and residual time series is also conducted for H2, shown in Fig. 6.

The Random Forest Model exhibits exceptional accuracy in predicting H2, as evidenced by a low Mean Squared Error (MSE) of 0.1965 and a high R-squared (R2) value of 0.9937, reflecting its robust performance in explaining the variance in H2. The residual plot pattern for H2 suggests a clustering towards the center, indicating accuracy in the predicted model despite the presence of outliers. Similarly, for T2, further investigation is needed to better understand underlying trends or features.

3.2 Heating Coil

A 40% propylene glycol system serves the heating coil, which provides only sensible heat, thus following Eqs. 1, at steady state the resultant outlet temperature (T2) is calculated as:

$$T2 = T3 + (\dot{m}_{40\%PG} \cdot Cp_{40\%PG}) (HWS - T - HWR - T) / (\dot{m}_{air} \cdot Cp_{air}) \quad (1)$$

3.3 Humidification:

A steam injection humidifier modulates to maintain the return air humidity at a setpoint based on an outdoor air reset (Eq. 2) and operates only during the heating season. This humidification has been approximated as isothermal and thus T3 = T4 and all change in enthalpy is related to the increase in relative humidity. Based on a humidification efficiency Δh from the shop drawings, the energy needed for this humidification is calculated as:

$$Q_{steam} = (\dot{m}_{air} \cdot h_{air}) / h \quad (2)$$

Unfortunately, there is no monitoring on the steam system and thus the heat injection at the humidifier is unknown. To overcome this challenge, the relationship between discharge and return air humidity was calculated in the winter season (Fig. 7).

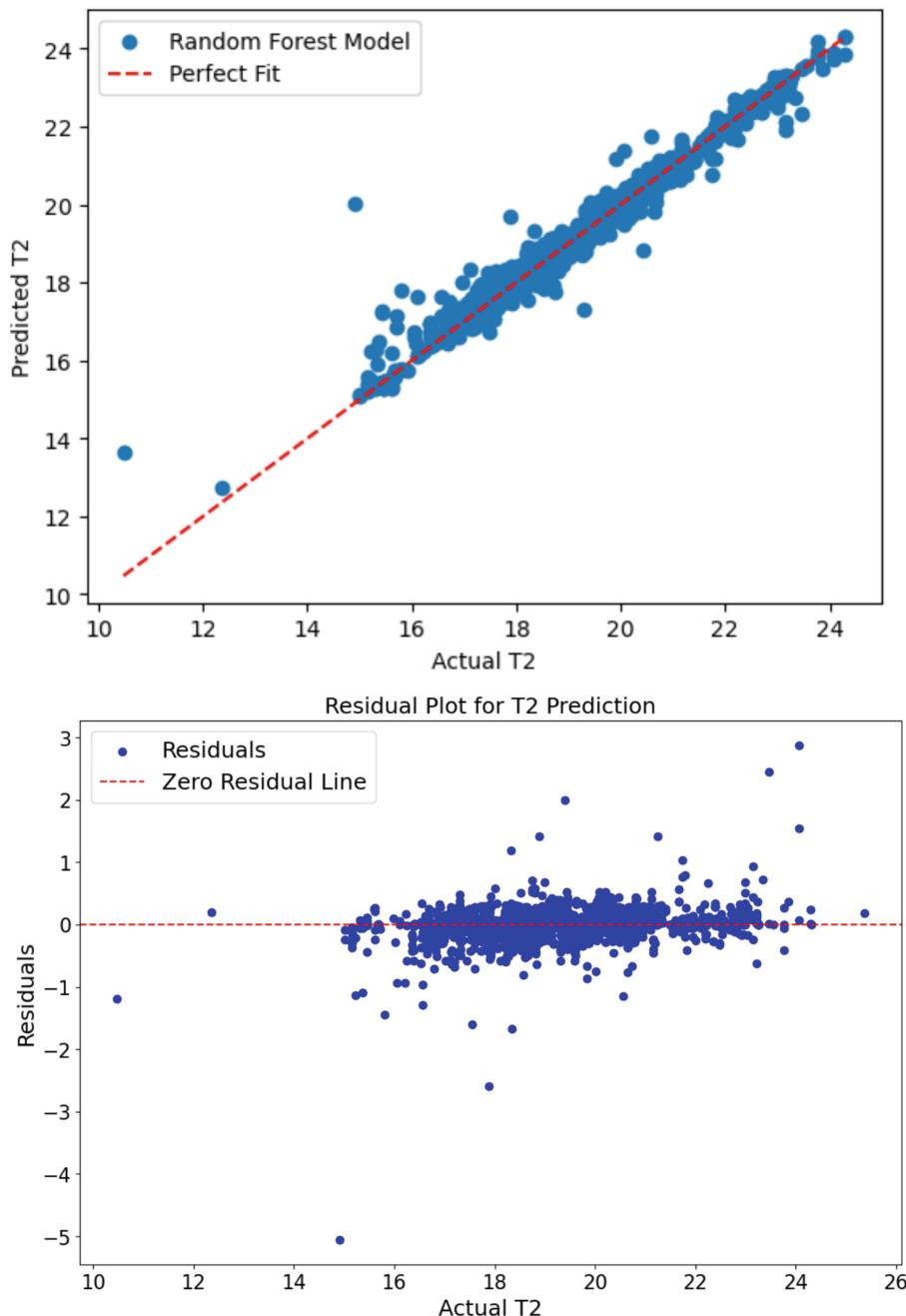


Fig. 5. Predicted vs actual wheel discharge temperatures (top) and residual plot (bottom).

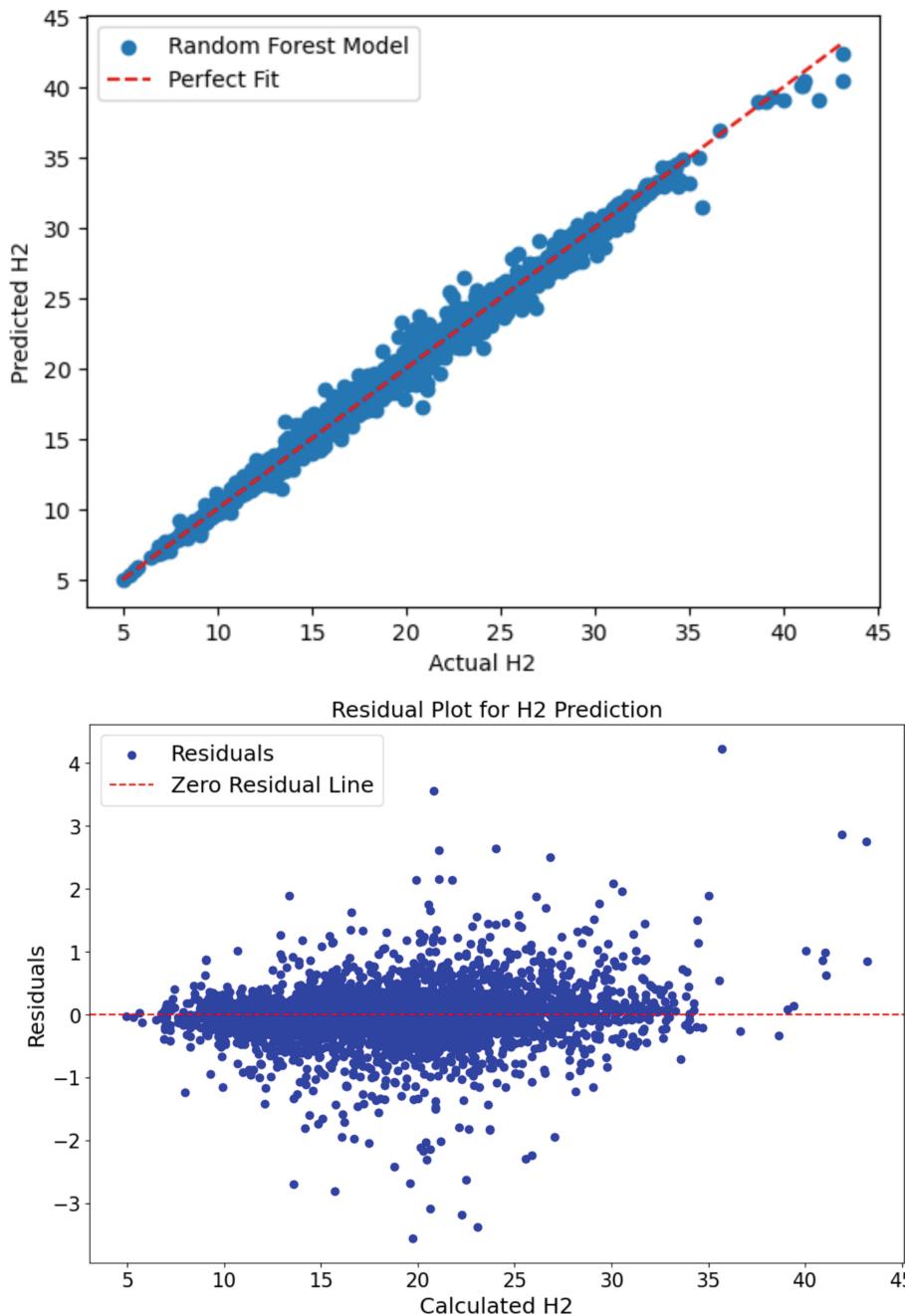


Fig. 6. Predicted vs actual wheel discharge Relative humidity (top) and residual plot (bottom).

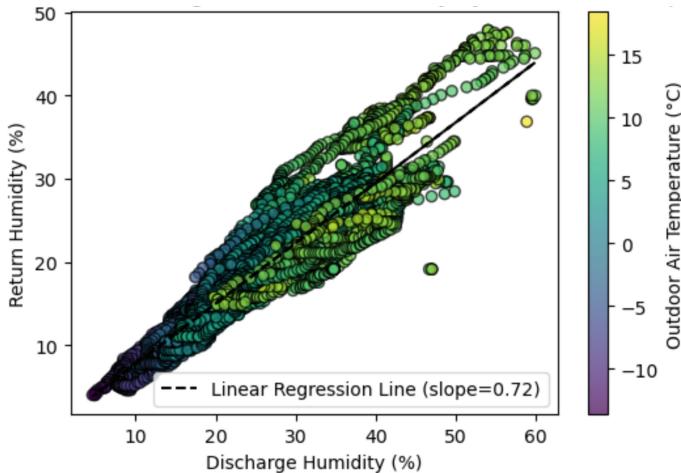


Fig. 7. Relationship between discharge and return humidity as a function of outdoor air temperature.

This linear regression analysis provides valuable insights into the strength and nature of the relationship between discharge and return air humidity, with the high correlation coefficient (0.93) and positive slope (0.72) indicating a significant positive association. The lack of a distinct pattern based on outdoor air temperature suggests that this relationship holds consistently across various temperature conditions. Therefore, H3 is the function of discharge humidity (H5), return humidity (H8), outdoor temperature (T1) and Humidifier output (HUM_O). Also, correlation analysis were employed, revealing positive linear associations between H3 and both H5 and T1. However, to discern the connection between H3 and Humidifier Output (HUM_O), further analysis is deemed necessary (Fig. 8).

Investigating the equations and correlations of the Heating coil and humidifier is crucial for winter performance analysis. The Leaving Air system study focuses on the predicting T2 (off-wheel supply temperature) by stepping back from T6 (discharge temperature) and H6 to H2. The Supply Fan pick-up increases T6 by approximately 1°C over T5, allowing the latter to be estimated. In the absence of cooling coil activity during winter (so T5 = T4), the given an isothermal humidification process (typical for the steam humidifier present in this system), T5 can be assumed equal to T3. The heating coil emulator predicts T2 then as a function of T3 as per Eq. 1.

With respect to actual humidity, neither the cooling coil (inactive) nor supply fan pick-up impact the water content of the air between points 4 and 6 and thus W4 and W6 lie on the same moisture content line, with minimal impact on relative humidity ($H_6 > H_4$) due to the small temperature change. The linear regression analysis between discharge and return air humidity consistently indicates that W3 consistently falls below W4. Therefore, in the absence of any temperature change in the humidifier, it can be inferred that H3 is lower than H4. The heating coil, being sensible only, also has negligible impact on the relative humidity.

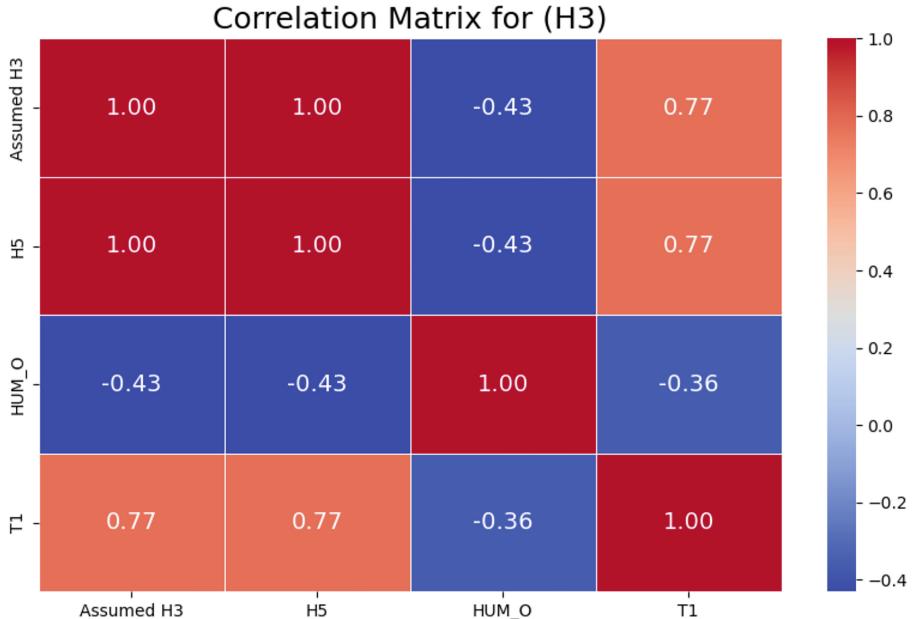


Fig. 8. Correlation Matrix of H3.

3.4 Cooling Coil Emulator

To compute the cooling coil performance, the total heat transferred by the air from the cooling coil must be equivalent to the total energy and heat acquired by the water (Eq. 3), indicating the process of cooling the air.

$$Q_{total,air} = Q_{total,water} \quad (3)$$

The fluid heating is sensible only (Eq. 4), while the air cooling is both sensible and latent (Eq. 5) thus:

$$Q_{total,water} = \dot{m}_{water} \cdot Cp_{water} \cdot \Delta T_{water} \quad (4)$$

So:

$$Q_{total,air} = \dot{m}_{air} \cdot Cp_{air} \cdot (h_{in,air} - h_{out,air}) \quad (5)$$

Given that all variables for the water heat energy (Q_{water}) are known; from above equations the entering temperature (T_4) into the cooling coil can be determined using the rearranged equation:

$$h_{in} = h_{out} + (Q_{air}) / (\dot{m}_{air} \cdot h_{air}) \quad (6)$$

To express this equation into temperatures, the specific heat capacity relationship is applied:

$$Q_{s,air} = \dot{m}_{air} \cdot Cp_{air} \cdot (T_{in,air} - T_{out,air}) \quad (7)$$

By rearranging the terms in Eq. (6) and incorporating the specific heat capacity relationship, the Eq. (8) allows explicitly relate the inlet and outlet temperatures in the context of the enthalpy balance.

$$hin = hout + \frac{\dot{m}_{air}.C_{pair}.(T_{inair} - T_{outair})}{\dot{m}_{air}.h_{air}} \quad (8)$$

The leaving coil temperature was assumed to be at the dew point. After the cooling coil, the only element impacting air conditions is the supply fan which, like the exhaust fan, was assumed to increase the air temperature by 1°C. The leaving water content at discharge, h_{out} is a known quantity, and will be equal to that at the discharge to the coil, thus providing the coil leaving temperature conditions. To calculate the cooling load, however, the heat transfer is idealized as a perfectly sensible cooling to dewpoint₁ (T_x), followed by perfectly latent cooling to the final dewpoint. This intermediate temperature is a variable carried through the emulator.

4 Discussion and Conclusions

This paper presents the methodology for creating a model that incorporates real-time data from a BAS, applies pre-processing for online learning, develops a physics-based model, and optimizes parameters using traditional linear regression approaches.

The study acknowledges limitations related to simplifying assumptions, especially concerning the cooling coil. Despite these constraints, the results from the discharge air emulator (T2, H2) demonstrate a notable achievement with a low Mean Squared Error (MSE) and a high R-squared (R2). This collective success underscores the effectiveness of the Random Forest Model in accurately predicting both T2 and H2. Also, the observed trends in the residuals within T2 and H2 model prompts further investigating the causes behind these residuals and making informed adjustments to enhance the model's accuracy and reliability. To ascertain the presence of seasonality, a direct examination of the data for periodic patterns at specific time intervals is necessary. Additionally, an assessment of potential overfitting or underfitting is essential to comprehensively understand the observed trend in the residuals.

Future research endeavors will focus on addressing these issues through finding new features and deep learning approaches. The ultimate objective of this ongoing research is to develop an emulator capable of adapting and enhancing its performance with new data, enabling deployment in new buildings for online energy optimization.

Acknowledgments. This research was funded by the Natural Science and Engineering Research Council [ALLRP-544569 and CREATE 510284-2018] and FuseForward Solutions Group.

References

1. Afram, A., Janabi-Sharifi, F.: Gray-box modeling and validation of residential HVAC system for control system design. *Appl. Energy* **137**, 134–150 (2015)

2. Afram, A., Fung, A.S., Janabi-Sharifi, F., Raahemifar, K.: Development of an accurate gray-box model of ubiquitous residential HVAC system for precise performance prediction during summer and winter seasons. *Energy and Buildings* **171**, 168–182 (2018)
3. Ding, L., Wang, G., Song, L.: Development, calibration, and validation of a novel gray-box energy model for residential split air conditioners. *Energy and Buildings* **295** (2023)
4. Ghiaus, C., Chicinas, A., Inard, C.: Grey-box identification of air-handling unit elements. *Control. Eng. Pract.* **15**(4), 421–433 (2007)
5. Afram, A., Janabi-Sharifi, F.: Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods. *Energy and Buildings* **94**, 121–149 (2015)
6. Stoffel, P., Maier, L., Kümpel, A., Schreiber, T., Müller, D.: Evaluation of advanced control strategies for building energy systems. *Energy and Buildings* **280** (2023)
7. Homod, R. Z.: Review on the HVAC System Modeling Types and the Shortcomings of Their Application. *Journal of Energy* (2013)



Near Real-Time Steel Rust Recognition Using Transformer-Based Convolutional Neural Networks

Shahinuzzaman^{1(✉)}, Po-Han Chen^{1,2}, and Luh-Maan Chang²

¹ Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC H3G 2W1, Canada
s_sujon@live.concordia.ca

² Department of Civil Engineering, National Taiwan University, Taipei, Taiwan

Abstract. Steel is an integral part of the infrastructure. Nowadays, steel has become a popular material in infrastructure due to its reduced processing time and easy fitting. Rust damage is one of the major causes of the degradation of steel infrastructure, and effective monitoring of rust is crucial for evaluating the damage to steel structures. The purpose of this research is to develop a near real-time rust recognition method for steel structure health monitoring using transformers-based Convolutional Neural Networks, e.g., SegFormer. The proposed transformer-based CNN rust recognition method is a simple, efficient, and powerful solution for semantic segmentation, which utilizes transformers and lightweight multilayer perceptron (MLP) decoders to reduce processing time and enhance accuracy. Conventional manual rust inspection is time-consuming, sometimes dangerous, and prone to human mistakes. The transformer-based CNN method could tackle these problems and minimize the costly manpower requirement for inspection. Most deep learning models for semantic segmentation were developed solely based on convolutional architecture and required explicit positional encoding. Consequently, the resolution variance between the testing and training phases led to reduced performance. In contrast, SegFormer is a hybrid architecture that integrates components from both transformers and convolutional neural networks (CNNs) and does not require positional encoding. In this research, the proposed method adopts the “light PyTorch” backend to significantly reduce the required processing time and develop a near real-time rust recognition method. It is able to quickly process rust segmentation at the image pixel level, and provide fast rusting deterioration assessment of steel structures.

Keywords: Rust recognition · Semantic segmentation · transformers deep learning

1 Introduction

Deep learning techniques, especially Convolutional Neural Networks (CNN), have been widely applied in the construction sector in recent years for infrastructure health monitoring [1–3]. Also, vision-based image processing methods have received much attention in infrastructure surface defect inspection (i.e., rust and cracks) since the late 1990s [4–7]. Rusting of steel structures is a significant challenge for engineers worldwide [8], and timely and accurate monitoring of rust damage on steel surfaces is essential to ensure the structural integrity, safety and longevity of steel facilities [9] [10]. Manual rust inspection is labor-intensive, time-consuming, and sometimes dangerous [11]. In Canada, there are many steel bridges, including the Quebec Bridge, which holds the record of the world's longest cantilever bridge span and is affected by rusting due to lack of proper monitoring [12]. Deep learning models have demonstrated remarkable capabilities in defect recognition and could well serve the purpose of facility surface inspection and monitoring [13, 14].

In the expansive field of computer vision and image processing applications, a multitude of methods for image segmentation facilitate the precise identification and delineation of target objects within figures. Image segmentation could be done based on edge, threshold, region, neural network, cluster and hybrid, and so on [15]. In this research, we utilized a hybrid deep learning model that integrates neural networks and transformer architecture. The main objective of this paper is to introduce transformer-based neural networks that are powerful for image segmentation with less processing time and reduced complexity. Semantic segmentation is common in computer vision and enables numerous downstream operations. It is suitable for image segmentation for it produces pixel-wise prediction rather than image-position prediction [16].

1.1 Related Work

Chen et al. [17] developed a new method that combines a support vector machine (SVM) and Fourier transforms to efficiently detect rust in various paint colours, including brown and red, while also addressing the issue of non-uniform illumination. They conducted a comparative analysis with their prior work involving the box and ellipse-based (ANFIS) [18] for bridge coating assessment. Katsamenis et al. [19] developed a U-Net convolutional neural network for corrosion localization and identification of rust grade on steel surfaces based on ISO8501–1. They employed a cutting-edge global-local data processing technique, a novel approach in this field that expands beyond solely utilizing local data processing methods. First, they focused on the localization of the corrosion region. In the second layer, they used a data projection approach for refining the region of the defect object and in the third layer, rust classification based on grade their proposed SLPAC U-Net showed remarkable improvement in comparison with the K-means and GMM algorithms. Anna and Oleg [14] proposed a real-time steel surface defect recognition based on CNN. Their research showcased competitive performance among various segmentation models, revealing a minimum processing time of 3.06 s per image using DeepLabV3 (ResNet34). Ihor Konovalenko et al. [21] conducted an experimental verification that confirmed the optimality of using the pre-trained ResNet50 model as a classification network for detecting steel surface flaws. To tackle the problem of

imbalanced data samples, they utilized binary focus loss, resulting in the successful identification of faults on steel surfaces.

In the recent year 2024, Wang et al. [20] proposed a deep learning algorithm for detecting carbon steel corrosion based on the degree of corrosion, using a modified version of YOLOv7. They incorporated the Convolutional Block Attention Module (CBAM) technique to enhance the significance of the corrosion region in complex scenarios involving corroded carbon steel. The performance of their model was compared using confusion matrix data with other deep models such as SSD, Faster RCNN, and the original YOLOv7. Fu et al. [21] proposed a robust and faster lightweight deep learning model based on SqueezeNet for steel surface defect classification, and they found that their model's efficiency exceeded 100 frames per second (fps).

2 Methodology

2.1 Deep Neural Network Model

Many robust and precise automated visual recognition techniques were developed with neural networks that are capable of processing multidimensional data [22], especially convolutional neural networks (CNN) handle image recognition and pattern identification well due to their capacity to identify the space invariance of shapes in images [22] [23]. There are many types of CNN models for image segmentation such as U-Net, YOLO, segment anything model (SAM), SegFormer, and so on. In this paper, we adopted a SegFormer neural network model (Fig. 1), a simple and efficient yet powerful semantic segmentation framework that unifies transformers with lightweight multilayer perceptron (MLP) decoders [16]. The used multi-layer perceptron (MLP) decoder gathers information from various layers, and integrates both local attention and global attention to rendering potent representations [16]. Nowadays, global-local data processing architectures are utilized in medical image analysis over local data processing due to their accuracy [19]. Furthermore, we employed a lightweight model backend known as “Light PyTorch” to reduce processing time. Sourget et al. [24] proved that SegFormer-B0 obtained better results than U-Net in medical image segmentation, and required less training time than the original U-Net due to its light architecture.

2.2 Experimental Data Sets

In this paper, datasets were collected and annotated using the tool “Roboflow.” The auto-annotated smart polygon tools were used to reduce time, and the labels' positions were manually checked during the annotation process. The dataset contains images from different sources, e.g., iPhone-14-pro-max, DSLR cameras, and websites. Finally, all annotated masks were verified by a senior engineer with more than eight years of relevant experience. In this paper, we use a total of 50 rust images (Fig. 2) with the size of 512x512 pixels.

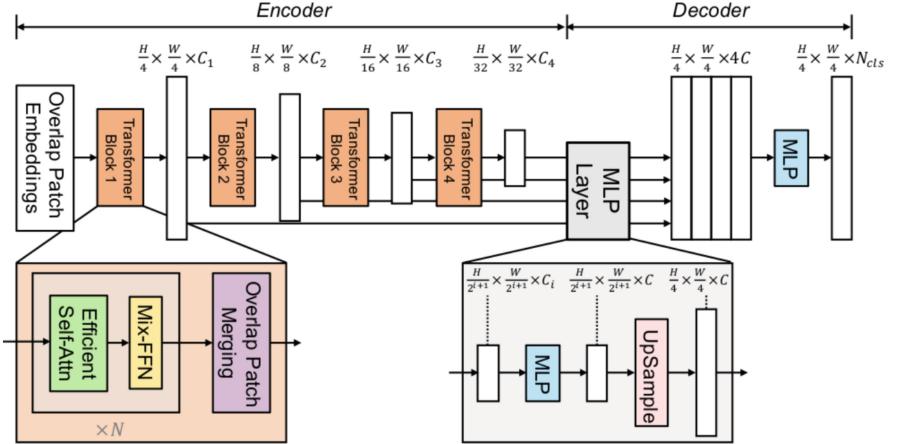


Fig. 1. The proposed SegFormer framework [16].



Fig. 2. Original rust image on steel used in this study.

Among the image dataset, 80% of them are for training, 20% for validation, and 10% for testing. To improve the learning process, the data augmentation technique was applied [19]. We augmented our training datasets using a positional augmentation technique. Each image rotated 90°, 180°, and 270°, as well as applied horizontal and vertical flipping. Additionally, we adjusted brightness between -20% and + 20%, applied image blur up to 2.5 pixels, and incorporated Gaussian noise up to 5% of pixels (Fig. 3). Hence, we increased the size of dataset from 50 to 120.

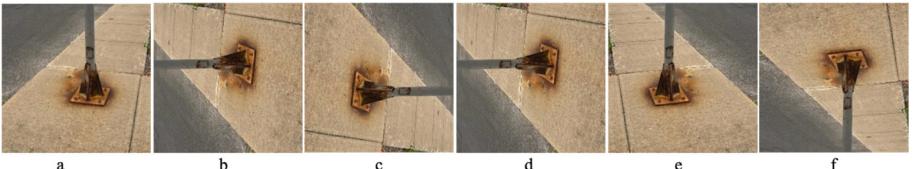


Fig. 3. Images generated with data augmentation technique, (a) original image, (b-d) image rotated by 90°, 180°, 270° (e-f) image flipping by horizontally and vertically.

2.3 Data Processing with Deep Neural Networks Models

We used the SegFormer deep learning transformer-based CNN model, with TensorFlow and Pytorch (Light version) as the backend. The deep learning model has two main characteristics, a hierarchical transformer encoder and a lightweight all-multi-layer perceptron (MLP) decoder, to fuse the multi-level features to produce final semantic segmentation masks. RGB three-channel images ($H * W * 3$) are divided into patches of 4x4 pixels. The usage of patches of 16x16 pixels is employed rather than in the case of the Vision Transformer (ViT). For dense prediction, smaller patches are a preference. As input, a hierarchical transformer encoder extracts multi-level features at {1/4, 1/8, 1/16, 1/32} of the original image resolution, and for output, passes the multi-level features to the all-MLP decoder to predict the segmentation mask at a $H/4 \times W/4 \times N_{cls}$ resolution, where N_{cls} is the number of categories [16]. The training and evaluation of this deep learning model were conducted using the NVIDIA Tesla T4 GPU available through Google Colab. During the data preprocessing step, the input of three-channel RGB images were resized to a resolution of 512x512 pixels. Additionally, we set up 100 epochs with an early stopping criterion, specifically with a patience of 10, to prevent overfitting. Moreover, we utilized mini-batches of 4x4 pixels in this study (Fig. 4).

3 Result & Discussion

3.1 Accuracy

A confusion matrix was used to evaluate deep learning models on a rust image data sample, where we considered precision, accuracy, recall, dice coefficient (F1 Score), and intersection over union (IoU). A total of 120 rust images including augmentation were used for training with the SegFormer deep learning model. We partitioned the entire image dataset into three subsets: 80% for training, 20% for validation, and the remaining 10% for testing. The precision (p), recall (r), accuracy score, F-measure (the harmonic mean of precision and recall) and intersection over union (IoU) are defined in Eqs.(1)-(5) [25]. To visualize the accuracy and loss graphs of the deep learning model, along with the associated data, we installed TensorBoard and gathered all the necessary data (Table 1) using the following equations.

Precision: The precision score is the number of true positive results divided by the number of all positive results.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

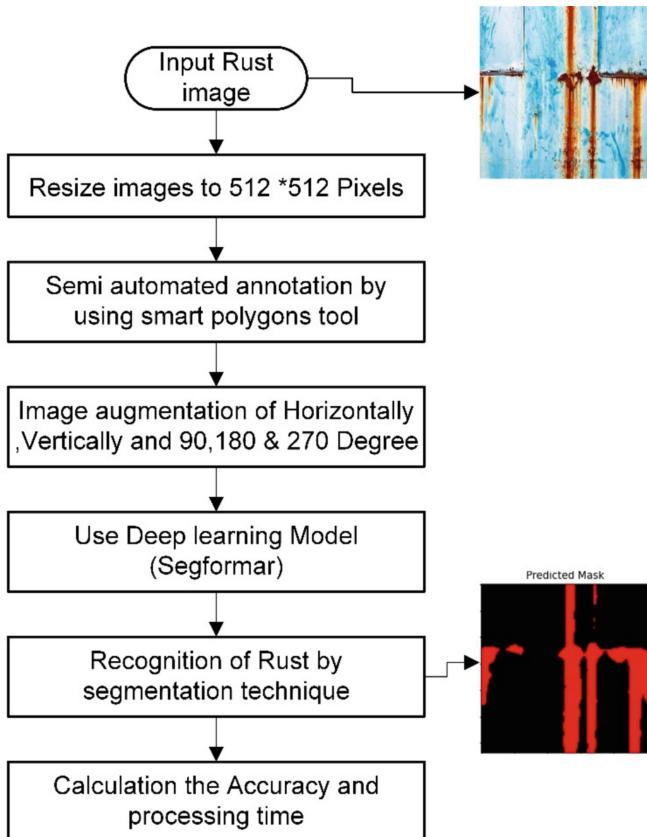


Fig. 4. Process of rust recognition by deep learning model

Recall: The recall score, also referred to as sensitivity or true positive rate, measures the ratio of true positive results to the total number of samples.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Accuracy score: The accuracy, calculated as the count of correct predictions (comprising both correct positive and negative predictions), is divided by the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (3)$$

Dice Coefficient (F1 Score): The dice coefficient is derived by considering the precision and recall of a prediction and assessing the overlap between the predicted segmentation and the ground truth.

$$Dice(F1) = \frac{2 * TP}{2 * TP + FN + FP} \quad (4)$$

Intersection over Union (IoU): The intersection over Union (IoU) is the area of the intersection over the union of the predicted segmentation and the ground truth.

$$\text{Intersection over Union (IoU)} := \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \quad (5)$$

where

True Positive (TP) refers to the count of samples correctly predicted as positive;

True Negative (TN) represents the number of samples correctly predicted as negative;

False Positive (FP) indicates the count of samples incorrectly predicted as positive; and

False Negative (FN) denotes the number of samples incorrectly predicted as negative.

Table 1. Model accuracy value

| Precision (p) % | Recall @ % | Accuracy % | Dice Coefficient (F1 Score) % | Mean Intersection over Union (mIoU) % | Processing time per image (s) |
|-----------------|------------|------------|-------------------------------|---------------------------------------|-------------------------------|
| 96.97 | 96.88 | 97.22 | 96.58 | 91.68 | 1.54 |

The confusion matrix is one of the most effective options for evaluating a deep learning model's accuracy [26]. Table 1 shows our model's performance metrics: an outstanding overall accuracy of 97.22%, mean intersection over union (mIoU) of 91.68%, and an impressively fast processing time of 1.54 s per image.

As depicted in Fig. 5, our best training loss stands at 4.45%, while the validation loss is 14.67%. These results indicate that the training accuracy is satisfactory. However, there is room for improvement in the validation loss, potentially due to the limited number of images in the validation datasets.

In Fig. 6, the deep learning model's accuracy performance graph illustrates a training mean accuracy of 97.22%. In contrast, the validation mean accuracy stands at 92.80%. These results suggest that the model performs well on the training dataset and reasonably well on the validation dataset, demonstrating good generalization to unseen external data. In Fig. 7, we present three segmented images from our test dataset, displaying comparisons between the original images, the ground truth images created using semi-automated annotation tools, and the predicted masks generated by our proposed deep model.

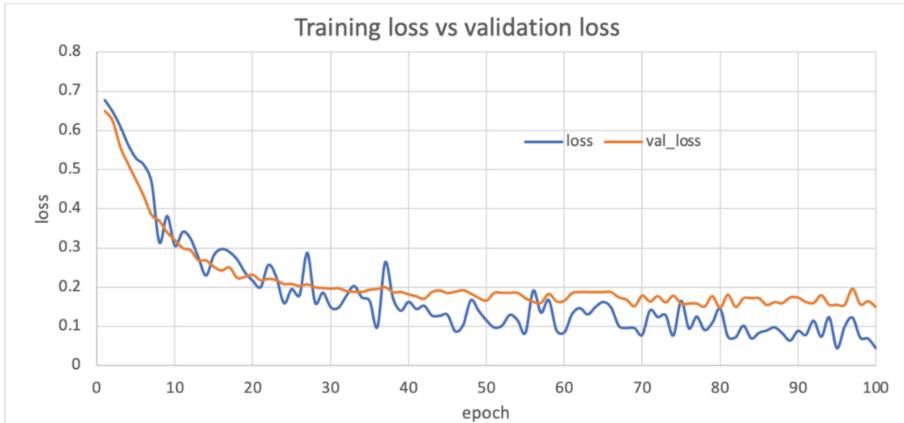


Fig. 5. Training and validation loss vs. number of epochs.

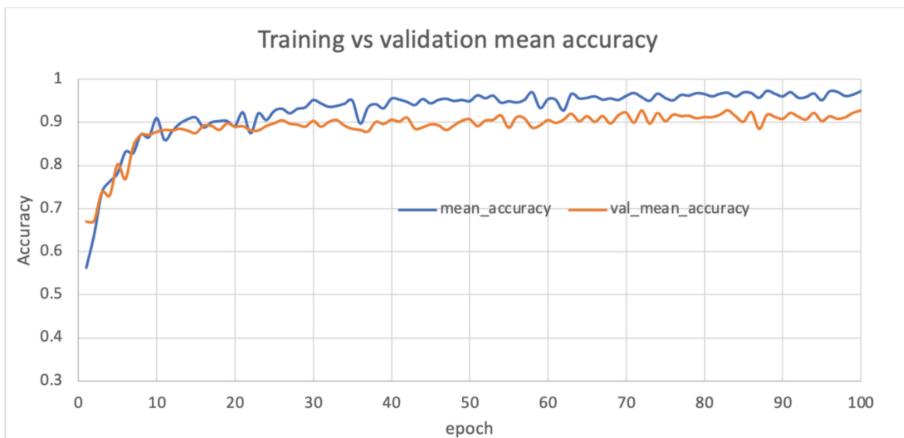


Fig. 6. Train and validation mean accuracy vs number of epochs.

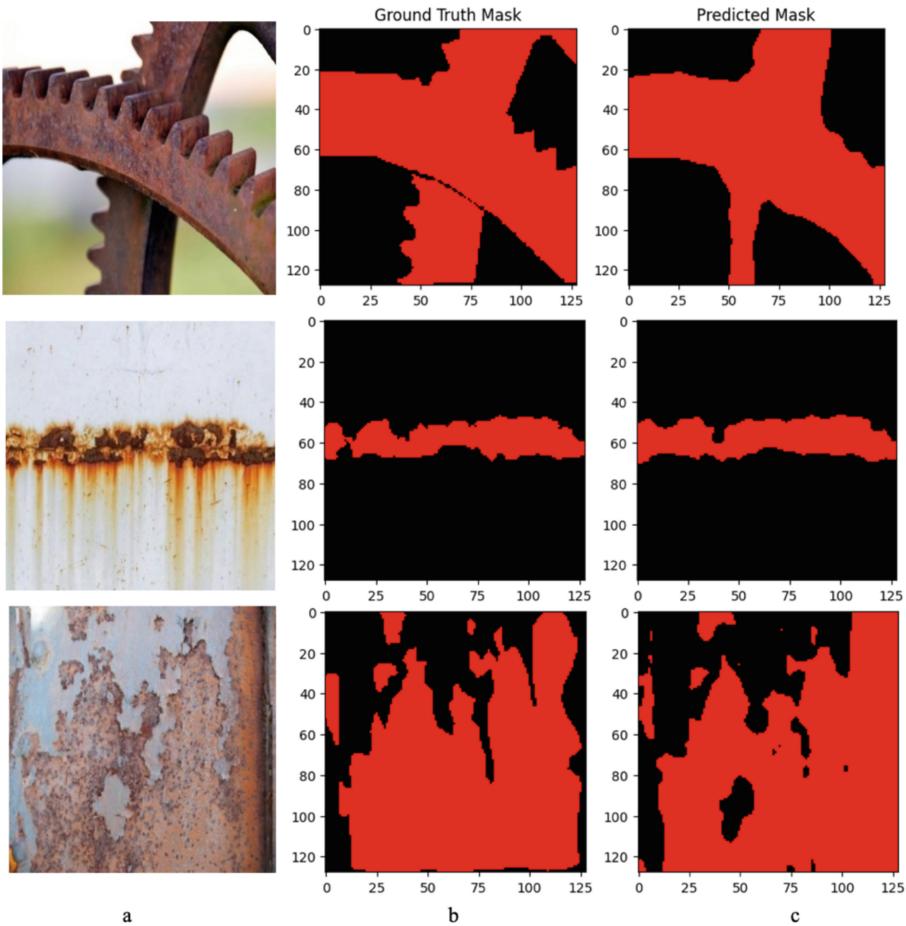


Fig. 7. (a) Original image, (b) Ground truth image, (c) Rust segmentation image.

4 Conclusion

In this paper, we present a deep learning transformer-based approach for efficiently recognizing rust on steel surfaces with improved accuracy and reduced processing time. The objective was to conduct pixel-level semantic segmentation of the rust image. The underlying aim was to significantly decrease processing time to achieve near real-time performance. To accomplish this, we implemented a transformer-based CNN model known as “SegFormer.” “SegFormer” is an efficient, simple, enough powerful semantic segmentation deep learning model which unifies lightweight multilayer perceptron (MLP) decoders with Transformer encoders. We collected our image datasets from diverse sources, including iPhones, Cameras, websites and more. A total of 50 original images were utilized in this model. The data contain several types of obstructions, including lighting conditions, rust color, and severity of rust which could be making the segmentation/classification process challenging. Assessing the model’s confusion matrix,

including precision (p), recall (r), accuracy score, F1-measure, and mean intersection union (mIoU), we obtained promising results that appear to be sufficiently effective for robust rust image recognition in practical applications. The proposed deep learning model demonstrates that precision (p), recall (r), accuracy score, F1-measure score, and mean intersection union (mIoU) are 96.97%, 96.88%, 97.22%, 96.58%, and 91.68%, respectively. Additionally, the model achieves an average accuracy of 97.22%. Our error analysis results indicate that training and validation errors are 4.45% and 14.67%, respectively. Furthermore, we tested this model using an external rust image and observed a processing time of 1.54 s per image, marking a novel achievement in this field. These findings could have significant potential for practical applications in infrastructure health and monitoring industries, especially steel infrastructure.

References

1. Rahman, A., Wu, Z.Y., Kalfarisi, R.: Semantic Deep Learning Integrated with RGB Feature-Based Rule Optimization for Facility Surface Corrosion Detection and Evaluation (2021). [https://doi.org/10.1061/\(ASCE\)](https://doi.org/10.1061/(ASCE)1010-0283.0002009)
2. Akinoshio, T. D., et al.: Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering* **32** (Nov 2020). <https://doi.org/10.1016/j.jobe.2020.101827>
3. Taye, M. M.: Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers* **12**(5). MDPI May 01, 2023. <https://doi.org/10.3390/computers12050091>
4. Jiang, S., Wu, Y., Zhang, J.: Bridge coating inspection based on two-stage automatic method and collision-tolerant unmanned aerial system. *Autom Constr* **146** (Feb 2023) <https://doi.org/10.1016/j.autcon.2022.104685>
5. Dhanachandra, N., Manglem, K., Chanu, Y. J.: Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, Elsevier, 764–771(2015). <https://doi.org/10.1016/j.procs.2015.06.090>
6. Shen, H.K., Chen, P.H., Chang, L.M.: Automated steel bridge coating rust defect recognition method based on color and texture feature. *Autom. Constr.* **31**, 338–356 (2013). <https://doi.org/10.1016/j.autcon.2012.11.003>
7. Lee, S., Chang, L.M., Skibniewski, M.: Automated recognition of surface defects using digital color image processing. *Autom. Constr.* **15**(4), 540–549 (Jul 2006). <https://doi.org/10.1016/j.autcon.2005.08.001>
8. Sajid, H.U., Kiran, R.: Influence of corrosion and surface roughness on wettability of ASTM A36 steels. *J. Constr. Steel Res.* **144**, 310–326 (May 2018). <https://doi.org/10.1016/j.jcsr.2018.01.023>
9. Zhang, Y., Ayyub, B.M., Fung, J.F.: Projections of corrosion and deterioration of infrastructure in United States coasts under a changing climate. *Resilient Cities and Structures* **1**(1), 98–109 (Mar 2022). <https://doi.org/10.1016/j.rcns.2022.04.004>
10. Angst, U.M.: Challenges and opportunities in corrosion of steel in concrete. *Materials and Structures/Materiaux et Constructions* **51** (1) (Feb 2018). <https://doi.org/10.1617/s11527-017-1131-6>
11. Liu, C. Y., Chou, J. S.: Bayesian-optimized deep learning model to segment deterioration patterns underneath bridge decks photographed by unmanned aerial vehicle. *Autom Constr* **146** (Feb 2023). <https://doi.org/10.1016/j.autcon.2022.104666>

12. National Trust of Canada: <https://nationaltrustcanada.ca/nt-endangered-places/the-quebec-bridge>. Accessed Sep 28, 2023. [Online]. <https://nationaltrustcanada.ca/nt-endangered-places/the-quebec-bridge>
13. Guo, Z., Tian, Y., Mao, W.: A Robust Faster R-CNN Model with Feature Enhancement for Rust Detection of Transmission Line Fitting. *Sensors (Basel)* **22**(20) (Oct. 2022). <https://doi.org/10.3390/s22207961>
14. Litvintseva, A., Evstafev, O., Shavetov, S.: Real-time Steel Surface Defect Recognition Based on CNN. In: IEEE International Conference on Automation Science and Engineering, IEEE Computer Society, pp. 1118–1123. (Aug 2021). <https://doi.org/10.1109/CASE49439.2021.9551414>
15. Shen, H.K., Chen, P.H., Chang, L.M.: Human-visual-perception-like intensity recognition for color rust images based on artificial neural network. *Autom. Constr.* **90**, 178–187 (Jun 2018). <https://doi.org/10.1016/j.autcon.2018.02.023>
16. Xie, E., Wang, W., Yu, Z., et al.: SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers (May 2021) [Online]. Available: <http://arxiv.org/abs/2105.15203>
17. Chen, P.H., Shen, H.K., Lei, C.Y., Chang, L.M.: Support-vector-machine-based method for automated steel bridge rust assessment. *Autom. Constr.* **23**, 9–19 (May 2012). <https://doi.org/10.1016/j.autcon.2011.12.001>
18. Chen, P.-H., Asce, A.M., Yang, Y.-C., Chang, L.-M., Asce, M.: Box-and-Ellipse-Based ANFIS for Bridge Coating Assessment. *J. Comput. Civ. Eng.* (2010). <https://doi.org/10.1061/ASC/ECP.1943-5487.0000041>
19. Katsamenis, I., Doulamis, N., Doulamis, A., et al.: Simultaneous Precise Localization and Classification of metal rust defects for robotic-driven maintenance and prefabrication using residual attention U-Net. *Autom Constr* **137** (May 2022). <https://doi.org/10.1016/j.autcon.2022.104182>
20. Wang, Q., Gong, H., Fu, Z., Zhang, D.: Automatic detection of carbon steel corrosion degree based on image identification. *Comput Mater Sci* **233** (Jan 2024). <https://doi.org/10.1016/j.commatsci.2023.112717>
21. Fu, G., et al.: A deep-learning-based approach for fast and robust steel surface defects classification. *Opt. Lasers Eng.* **121**, 397–405 (Oct 2019). <https://doi.org/10.1016/j.optlaseng.2019.05.005>
22. Sahu, M., Dash, R.: A survey on deep learning: Convolution neural network (cnn). In Smart Innovation, Systems and Technologies. Springer Science and Business Media Deutschland GmbH, 317–325 (2021). https://doi.org/10.1007/978-981-15-6202-0_32
23. Wang, Y., Ahsan, U., Li, H., Hagen, M.: A Comprehensive Review of Modern Object Segmentation Approaches. Foundations and Trends in Computer Graphics and Vision **13**(2–3). Now Publishers Inc., 111–283 (2022). <https://doi.org/10.1561/0600000097>
24. Sourget, T., Hasany, S. N., Mériadeau, F., Petitjean, C.: Can SegFormer be a True Competitor to U-Net for Medical Image Segmentation?, 111–118 (2024). https://doi.org/10.1007/978-3-031-48593-0_8
25. Huang, I. F., Chen, P. H.: Automated steel bridge coating rust defect recognition method based on U-net fully convolutional networks. In: 2nd IEEE International Conference on Architecture, Construction, Environment and Hydraulics 2020, ICACEH 2020, Institute of Electrical and Electronics Engineers Inc., 18–21(Dec 2020). <https://doi.org/10.1109/ICACEH51803.2020.9366258>
26. Elamparo, P. A. A., Telan, E. C. M., Ibarra, J. B. G.: Rust Detection on Galvanized Iron Sheets using Convolutional Neural Network and Mask R-CNN. In: 2023 IEEE 14th Control and System Graduate Research Colloquium, ICSGRC 2023 - Conference Proceeding. Institute of Electrical and Electronics Engineers Inc., 39–44 (2023). <https://doi.org/10.1109/ICSGRC57744.2023.10215403>



The Cost of Modelling Silicon-Based Materials with Machine Learning Potential

Karim Zongo¹, Laurent Karim Béland², and Claudiane Ouellet-Plamondon¹(✉)

¹ École De Technologie Supérieure, Quebec University, Montreal, QC H3C 1K3, Canada

karim.zongo.2@ens.etsmtl.ca,

claudiane.ouellet-plamondon@etsmtl.ca

² Queen's University, Kingston, ON K7L 3N6, Canada

laurent.beland@queensu.ca

Abstract. The emergence of machine learning potential has revolutionized material modelling, yet its development involves thousands of quantum mechanical calculations. These calculations, along with potential training and molecular dynamic simulations, require substantial computational resources. In this study, we aim to elucidate the resource consumption involved in developing machine learning interatomic potential for silicon-based materials modelling. The consumption data presented here are extracted from the Digital Research Alliance of Canada account, automatically saved over four calendar years by their internal computation system. This model is the start of a model aiming to represent construction materials.

Keywords: Silicon · Moment tensor potential · Parallel computing

1 Introduction

Multiscale modelling serves as a cornerstone in civil engineering, enabling a comprehensive understanding of materials at the atomic level and aiding in risk mitigation strategies. Silicon and silica (SiO_2) play integral roles in various civil engineering materials such as aggregates, cement, and concrete [1]. Their interaction with water is crucial to the long-term durability of infrastructure. Despite the valuable insights offered by first principles calculations based on density functional theory, their computational demands pose significant challenges in terms of cost and scalability [2]. In contrast, machine learning (ML)-based material modelling leverages extensive databases derived from quantum mechanical calculations [3]. ML force fields, developed through optimization procedures, encode geometric and ab initio properties without explicitly representing electronic structures [4, 5]. However, the computational demands of such methodologies necessitate access to high-performance computer clusters, surpassing the capabilities of standard computing resources. In practice, the development of the ML potential involves three main steps: database generation, model training, and molecular dynamics simulations [6, 7] (or coupling with other modelling engines such as ARTn). Every step involves the consumption of computational resources. The database generations involved thousands of calculations. Estimating computation consumption at each step becomes very

complicated due to the extensive computational time required for database generations, training, including database optimization. This article delves into an analysis of resource consumption during the development of a machine learning potential database, utilizing the Digital Alliance cluster in Canada [8]. The cluster, distributed across Canada, enables the use of parallel computing techniques. The computation was conducted across five distinct clusters: Cedar, Graham, Beluga, Narval, and Niagara. Data collection spanned four calendar years, from 2020 to 2023, within the Alliance account. Every time a job is initiated on the Digital Research Alliance of Canada cluster, the consumption is automatically logged by the compute job monitoring system and transmitted to the user's account. Notably, at the onset of data collection, Narval had not yet been integrated into the Alliance cluster system. These data correspond to consumption recorded by Digital Research Alliance of Canada for each client. By comprehensively analyzing resource consumption across various modelling techniques, this article aims to optimize computational workflows and enhance efficiency in civil engineering material modelling endeavours. Through the integration of advanced computational techniques and high-performance computing resources, researchers can advance our understanding of material properties, leading to safer and more resilient infrastructure designs.

2 Computational Methods

This section details respectively our ML model (the moment tensor potential), the database preparation with density functional theory and the main computing procedure.

2.1 Moment Tensor Potential

The moment tensor potential (MTP) formalism is based on the partitioning property of potential energy surface (PES) within a certain medium mainly non-polarized or non-charged medium in which the PES can be approximated as the contribution of individual atomic energies given by their local environments [9]. The MTP model linearly expands the potential energy $V_i(r)$ of an atom i into a complete set of basis functions $B_\beta(r)$.

$$V_i(r) = \sum_{\beta} c_{\beta} B_{\beta}(r) \quad (1)$$

The basis functions $B_\beta(r)$ are determined from the scalar tensorial contraction of the local atomic environment descriptors give as follow:

$$M_{\mu,\vartheta}(r) = \sum_j f_{\mu}(r_{ij}) \underbrace{r_{ij} \otimes \dots \otimes}_{v \times} \quad (2)$$

$M_{\mu,\vartheta}(r)$ is a tensor of rank ϑ enclosing the radial distribution ($f_{\mu}(r_{ij})$) and the angular information.

$(r_{ij} \otimes \dots \otimes r_{ij})$ of the local environment of the central atom i . In practice, the contraction of $M_{\mu,\vartheta}(r)$ to a scalar is formulated by $n \times n$ symmetric matrix β which off-diagonal elements indicates the number of dimensions to be contracted between tensors M_{lm} and M_{pq} . The indexes $\mu \wedge \vartheta$ enables the bounding of basis size for practical implementation and allow the estimation of computational complexity and efficiency. This mathematical description of atomic interactions is implemented in MTP code [10].

2.2 Generation of a Database on Parallel Computing

The fitting of the MTP model mainly relies on a database from the first-principle calculation. Thus, in this work a database consisting of Si, SiO₂, O and H₂O structures was generated using quantum mechanical calculations in the framework of density functional theory. The procedure for creating the database is illustrated in Fig. 1 below. Several modelling softwares were used for manipulating, conversion, and visualization of atomic structures. Packages like XCrySDen [11], VESTA [12], Atomsk [13], and OVITO [14] were employed for these purposes. These software tools are set up on local computers and do not account for resource consumption. On the other hand, Quantum ESPRESSO (QE) [15] is utilized specifically for electronic structure calculations. The QE package, along with LAMMPS [16] and MTP code [10], are solely responsible for the resource consumption presented in this work. Firstly, the QE package is employed to generate the database. Subsequently, the MTP code is utilized to train the model, followed by the utilization of the LAMMPS code to perform molecular dynamics simulations. All these codes are highly parallelized, maximizing computational efficiency. We prepare the input files for jobs on Digital Research Alliance of Canada, previously called Compute Canada, from our local computer, as depicted in Fig. 2. Additionally, we detail the job monitoring and performance analysis in Fig. 2. This performance analysis enables us to adjust the parallelization scheme for optimal calculations.

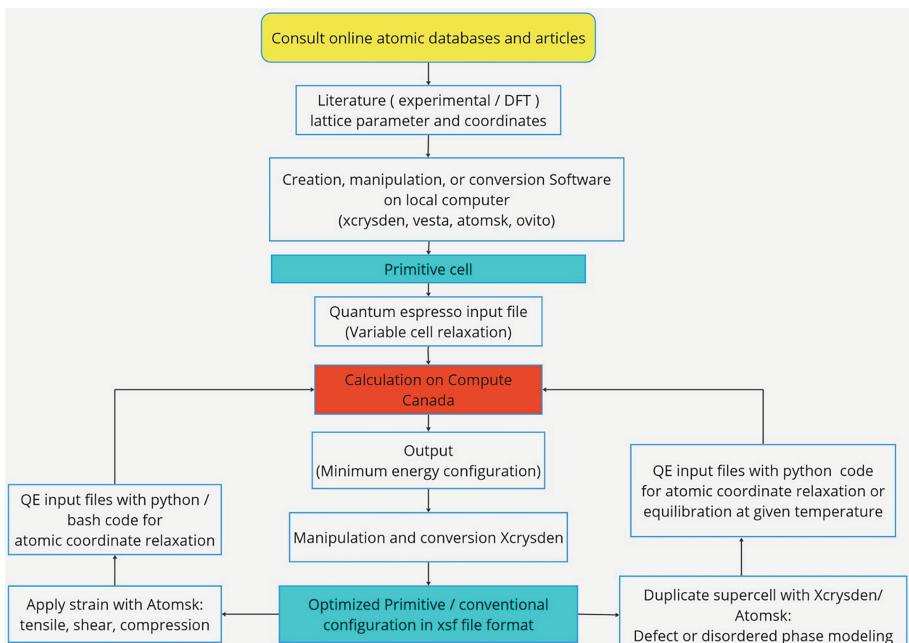


Fig. 1. Creating a database of Si and SiO₂ configurations: A Blueprint for database creation by combining two or more modelling software such as Xcrysden, VESTA, Atomsk, OVITO, and Quantum ESPRESSO.

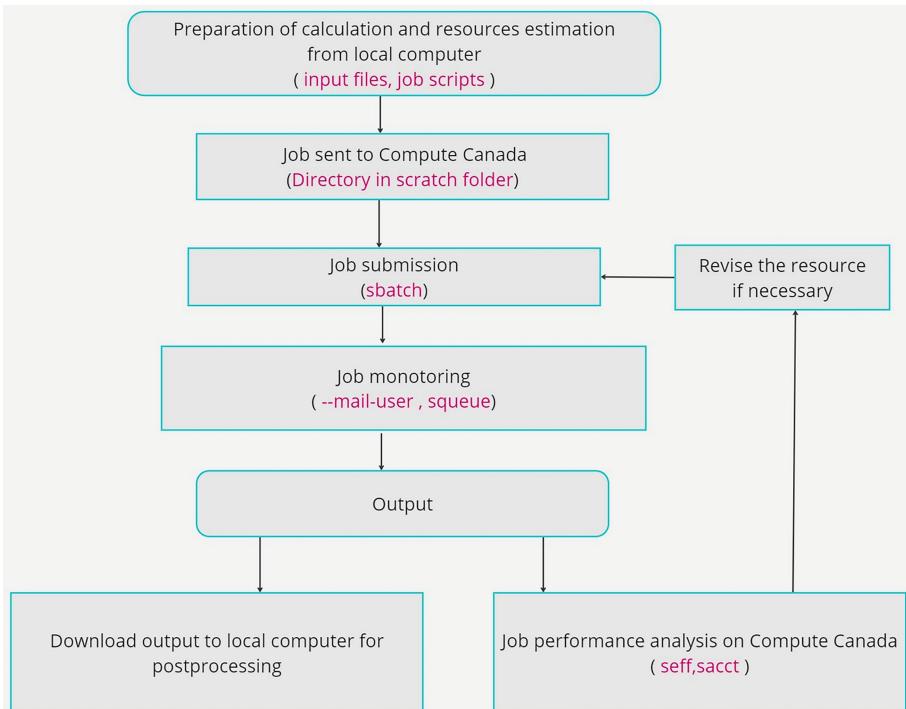


Fig. 2. Migration of computation files and management, monitoring of computations on the Digital Research Alliance of Canada cluster.

3 Results

This section presents the database generated using the QE package, along with the training of machine learning models and the resulting consumption. We generated approximately 7000 configurations comprising Si, SiO₂, and O structures. This number corresponds to approximately 10,000 single-point DFT calculations, as the selections were made from ab initio trajectories and variable cell relaxations. The table below presents the summary of the database (Table 1).

Table 1. Comprehensive database summary.

| Compounds | Number of configurations | Fraction used for the implementation of the potential (%) |
|------------------|--------------------------|---|
| Si | 3900 | 13 |
| SiO ₂ | 3500 | 14 |
| O | 100 | 20 |
| H ₂ O | 2500 | - |

Another significant contributor to computational resource consumption is the training of the machine learning (ML) model. The training process begins with optimizing the database. Figure 4 provides an illustration of the cost function during the training of the MTP model using level 26. Since the parameters are initially randomly initialized, finding the optimal model and data set for atomistic modelling necessitates several rounds of training. Consequently, the training and database optimization are iteratively repeated until a satisfactory model is obtained, which can then be utilized for molecular dynamics simulations (Figs. 3 and 5).

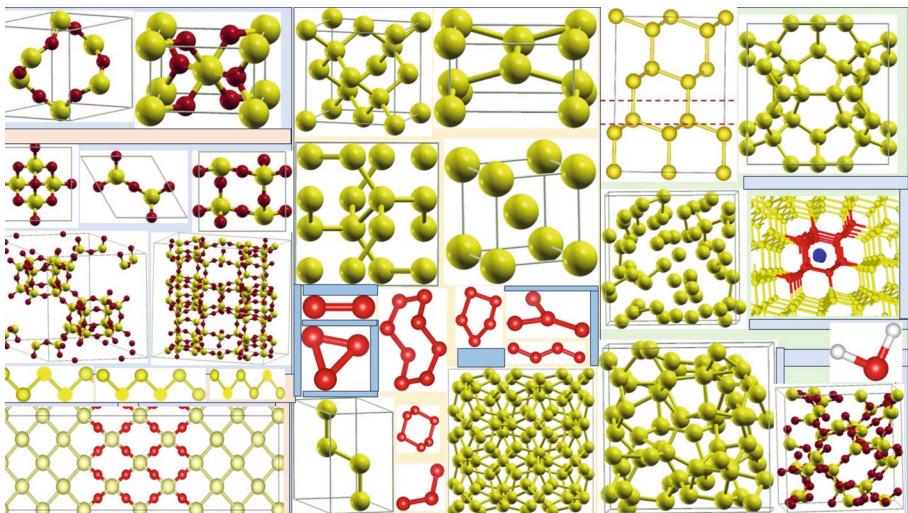


Fig. 3. Atomic structures in the database: Liquid, amorphous, defects of silicon and silica, oxygen molecules, and water molecules. The configurations are labeled with energies, forces, and stress values.

The costs associated with database generation, training, and molecular dynamics are depicted in Fig. 5. The left figure illustrates the consumption for each year. Notably, the year 2021 exhibits the highest consumption. During this period, quantum mechanics calculations were exclusively performed, along with preliminary testing and training of the MTP model. In 2022 and 2023, consumption levels are nearly identical. During these years, classical molecular dynamics simulations were conducted concurrently with the generation of water databases. Notably, in 2022, the database underwent refinement, and the MTP model was trained. In 2020, we used fewer resources as our focus was primarily on theoretical research activities. In terms of cluster usage, Cedar computers were predominantly employed for computations. Following closely, the Narval cluster also saw significant utilization. This can be attributed to the high number of computer nodes available on both Cedar and Narval, resulting in minimal waiting times. Additionally, the node configurations are noteworthy. For example, the nodes on Narval boast 64 and 48 cores each.

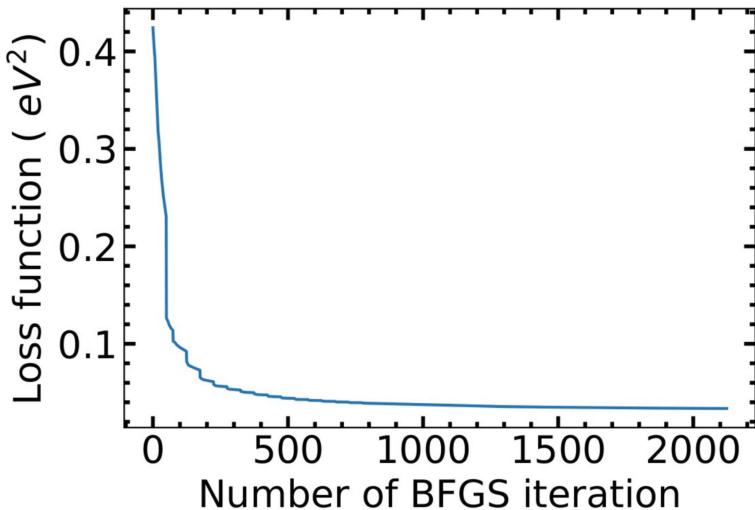


Fig. 4. Evolution of the cost function during training on Digital Research Alliance of Canada Cluster

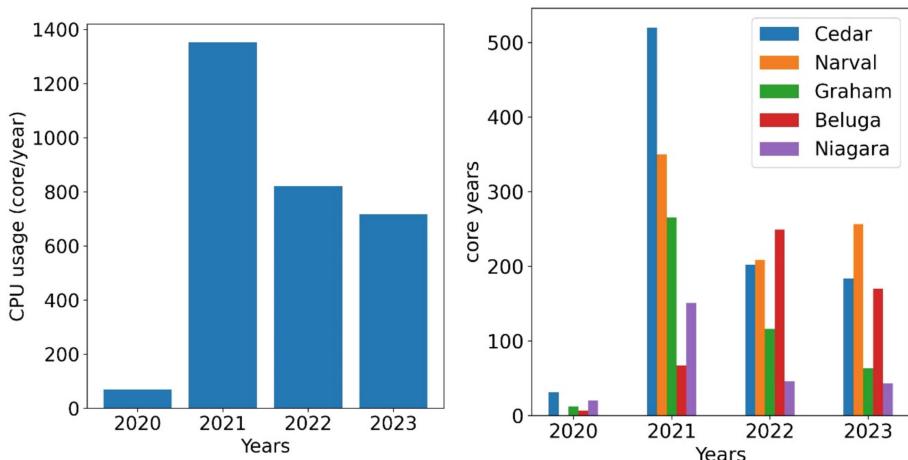


Fig. 5. Resource consumption for our modelling projects over four calendar years.

Cedar boasts nodes with 24, 32, and 48 cores each, offering a range of computational power. Additionally, Narval surpasses Cedar in speed and efficiency when using our modelling packages. Other clusters, such as Graham and Beluga, also played significant roles in supporting our research endeavours.

4 Conclusion

In conclusion, our study provides valuable insights into resource consumption during the development of machine learning potential for material modelling. The overall consumption can be attributed to the extensive calculations performed to explore a wide range of material structures. It's worth noting that, due to a lack of experience and a lack of adherence to proper methodologies at the outset of our research, some of the calculations may have been unnecessary. However, the resulting configurations can still be valuable for future research endeavours. This highlights the importance of experience and adherence to established practices in minimizing resource consumption in computational research. By shedding light on the computational demands involved in this process, researchers can better understand and manage resource utilization, ultimately leading to more efficient and sustainable modelling practices. In summary, by implementing strategies such as refining algorithms, optimizing code, and effectively leveraging high-performance computing resources, researchers can significantly enhance the efficiency and sustainability of their modelling endeavours.

Acknowledgement. The authors thank the Digital Research Alliance of Canada, Canada Research Chair Program (award CRC-2019-00074), ACFAS mobility scholarship, Nuclear Waste Management Organization of Canada (NWMO), and NSERC to support the research.

References

1. Karlina, A.I., Karlina, Y.I., Gladikh, V.A.: Analysis of Experience in the Use of Micro-and Nanoadditives from Silicon Production Waste in Concrete Technologies. *Minerals* **13**(12), 1525 (2023)
2. Kohn, W., Sham, L.J.: Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**(4A), A1133 (1965)
3. Behler, J., Parrinello, M.: Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**(14), 146401 (2007)
4. Zongo, K., Béland, L.K., Ouellet-Plamondon, C.: First-principles database for fitting a machine-learning silicon interatomic force field. *MRS Advances* **7**(2), 39–47 (2022)
5. Zongo, K., et al., A unified moment tensor potential for silicon, oxygen, and silica. arXiv preprint [arXiv:2311.15170](https://arxiv.org/abs/2311.15170) (2023)
6. Bartók, A.P., et al.: Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**(4), 041048 (2018)
7. Zuo, Y., et al.: Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**(4), 731–745 (2020)
8. https://docs.alliancecan.ca/wiki/Technical_documentation
9. Shapeev, A.V.: Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**(3), 1153–1173 (2016)
10. Novikov, I.S., et al.: The MLIP package: moment tensor potentials with MPI and active learning. *Machine Learning: Science and Technology* **2**(2), 025002 (2020)
11. Kokalj, A.: XCrySDen—a new program for displaying crystalline structures and electron densities. *J. Mol. Graph. Model.* **17**(3–4), 176–179 (1999)
12. Momma, K., Izumi, F.: VESTA: a three-dimensional visualization system for electronic and structural analysis. *J. Appl. Crystallogr.* **41**(3), 653–658 (2008)

13. Hirel, P.: Atomsk: A tool for manipulating and converting atomic data files. *Comput. Phys. Commun.* **197**, 212–219 (2015)
14. Stukowski, A.: Visualization and analysis of atomistic simulation data with OVITO—the Open Visualization Tool. *Modell. Simul. Mater. Sci. Eng.* **18**(1), 015012 (2009)
15. Giannozzi, P., et al.: QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **21**(39), 395502 (2009)
16. Thompson, A.P., et al.: LAMMPS—a flexible simulation tool for particle-based materials modelling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022)



Autonomous Crack Segmentation Based on Segment Anything Model

Ghodsiyeh Rostami^(✉), Po-Han Chen, and Yang Wang

Concordia University, Montreal, QC H3G 1M8, Canada
{rose.rostami,pohan.chen,yang.wang}@concordia.ca

Abstract. The infrastructure of Canada, including bridges, roads, and buildings, is aging. It is of paramount importance to evaluate the state of infrastructure to ensure its serviceability and prevent catastrophic failures. One of the most common defects in concrete structures is crack, which develops from the surface of the structure to deeper parts of it. Early detection of cracks is vital for in-time maintenance planning. The traditional methods of inspection are resource-intensive (labor and equipment), and time-consuming, making the procedure expensive and difficult. Therefore, there is a need for fast, automated, and cost-effective inspection methods. Recent advancements in Deep Learning and Computer Vision have resulted in significant progress in defect detection algorithms for visual identification of defects and extracting further data for severity assessment. However, the majority of deep learning-based crack detection models do not demonstrate generalization to conditions of real-world practices as they are bound to limited datasets and relatively small models. Thus, to develop a more robust crack detection model, this study introduces Crack Large Model (CLM) which leverages large-scale semantic segmentation models, specifically, Segment Anything Model (SAM) for crack segmentation. Through transfer learning, SAM is fine-tuned on a diverse and challenging crack dataset including 11k pairs of crack images and masks. The comparison results of CLM with popular segmentation models (DeepLabv3+ and DeepCrack) have shown its promising performance in multi-scale and multi-level semantic segmentation of different types of cracks, which makes it a potential practical method for crack damage assessment in the near future.

Keywords: Crack Detection · Vision Foundation Model · Segment Anything Model

1 Introduction

The constructed facilities exhibit patterns of aging during their operation lifecycle. Crack is the most common defect developing from the surface of the structure to the deeper parts resulting in structure failure. Early detection of cracks is of utmost importance in infrastructure health assessment and in-time maintenance plans [1]. Prevalent manual inspection methods for crack detection are labor and equipment-intensive making the process unsafe, expensive, and time-consuming [2]. In the past decades, automated

visual inspection methods have been investigated to enhance the efficiency and safety of inspection processes [3]. Such methods require equipped robotic devices, such as Unmanned Aerial Vehicles (UAVs), with imaging systems for data acquisition and analysis to detect occurring damages (e.g., cracks) and extract information for further severity assessments.

Recently, with the advancements in computer vision and deep learning, crack recognition algorithms have been extensively explored to enable automated visual inspections [4]. Outperforming traditional image processing techniques, Deep Neural Networks (DNNs) have shown promising results in identifying crack regions in natural images. Specifically, Convolutional Neural Networks (CNNs) have been widely studied to learn and extract crack features from inspection images and identify crack regions at the pixel level, known as semantic segmentation. Dung and Anh's research work was among the first studies on classifying crack images pixel by pixel through DNNs. They proposed a crack semantic segmentation network using a Fully Convolutional Network (FCN) for crack density evaluation. They trained their network on a dataset of 500 images of 227×227 -pixels in size and quality illumination conditions. Their experiment achieved 90% scores in both F1-score and precision [5]. Ren et al. worked on improving the crack segmentation efficiency and robustness by developing CrackSegNet, a deep FCN by incorporating dilated convolution, pyramid pooling, and skip connections. Their dataset contained 919 images with 512×512 pixels. CrackSegNet was compared with U-Net and a conventional image processing method on the same dataset. CrackSegNet achieved F1-score and IoU of 74.55% and 59.06%, respectively, while U-Net scored 63.09% and 47.64%, and the conventional method scored 25.79% and 14.46% [6]. Fan et al., developed a deep residual CNN, Parallel ResNet, for crack segmentation in noised images. Their model performance on two popular datasets, CrackTree200 and CFD, was 93.08% and 95.63% F1-score, respectively [7]. Yu et al. proposed a two-stage crack segmentation (image classification followed by segmentation) using GoogLeNet, a sliding window, and DeepLabv3+. They used three different datasets in terms of difficulty and resolution of images. The model performed best on the simple and high-resolution dataset where cracks were obvious and vivid [8]. Although the high accuracy achieved by CNNs, the developed CNN-based crack detection models perform poorly in practice. This is due to 1) limited datasets (i.e., size or level of difficulty) being used in the previous studies, and 2) the inherent bias of CNNs towards local features and patterns, which limits them in globally understanding a scene and, therefore, learning to distinguish crack artifacts against a variety of challenging conditions in the background (e.g., material imprints).

To overcome the limitations of CNNs, Vision Transformers (ViT) emerged in 2020 by Dosovitskiy et al. which employ self-attention operations to capture global patterns and long-range dependencies [9]. Limitedly investigated in crack recognition studies, ViTs have shown stronger generalization capabilities [10]. Wang and Su developed SegCrack based on an encoder-decoder architecture, implementing a hierarchical Transformer as the encoder and a top-down pathway with lateral connections as the decoder. They trained SegCrack on a dataset of 2735 images and compared it with four CNNs, namely, FCN, UPerNet, EMANet, and BiSeNet2. Their results demonstrated the superiority of

SegCrack over CNNs by achieving F1-score and mIoU of 96.05% and 92.63%, respectively [11]. However, to reduce the computational costs, they reduced the sequence length in the self-attention module, which can result in losing information [12]. Asadi Shamsabadi et al. developed TransUNet for crack segmentation which incorporates both CNN and ViT in the encoder. They trained TransUNet on two separate datasets each containing 520 and 2200 images with 448×448 pixels in size. According to their comparison results with UNet and Deeplabv3 +, TransUNet performed best by achieving 0.73 and 0.88 dice scores on each of the datasets [10]. Nevertheless, using CNNs alongside ViT still imposes the same limitations of the CNN on the model by focusing on local features, and losing global information. Wu et al. presented Pixel Crack Transformer Network (PCTNet) to achieve a balance between transformers' computational cost and accuracy. PCTNet employs a hierarchical structure with a Cross-Scale PatchEmbedding Layer and Dual Attention Transformer Block. They collected 5398 images with 896×896 pixels (Crack896) for training and employed Cross Entropy Loss as their loss function. Yet, their dataset lacks diversity in terms of background. By comparing their network with DeepLabV3 +, U-Net, SegNeXt, BiFormer, PVTv2, and Swin Transformer, PCTNet outperformed the others in Recall, F1-score, and mIoU of 93.77%, 94.8%, and 90.53%, respectively [12].

Leveraging Vision Transformers, large-scale vision foundation models have been introduced as a paradigm shift in the field of deep learning. They are defined as large models trained on broad datasets showing high zero-shot performance which can be adapted to downstream tasks [13]. As generalization is of utmost importance for crack detection models in practical inspection operations (where a variety of scenes in terms of illumination, texture background, and similar-to-crack objects may occur in the captured images), vision foundation models can be adopted for crack detection problems. One of the state-of-the-art foundation models for segmentation is Segment Anything Model (SAM) which was recently released by Meta AI. It has been trained on over 11M images and has demonstrated promising results in semantic segmentation with zero-shot generalization [14]. It is a promptable segmentation model and is composed of three main components, namely, image encoder (based on ViT), prompt encoder, and mask decoder. Output image and prompt embeddings from the image encoder and prompt encoder are passed to the mask decoder to generate masks. SAM has demonstrated promising results in semantic segmentation with zero-shot generalization. Nevertheless, experiments have shown SAM is not capable of detecting cracks (**Fig. 1**). To this end, Ge et al. developed CrackSAM by fine-tuning SAM on a large crack segmentation dataset using two finetuning methods, adapter and low-rank adaptation (LoRA). They used adapter and LoRA to fine-tune image encoder while preserving pre-trained weights, prompt encoder, and mask decoder. Using LoRA, CrackSAM achieved IoU of 64.88% on the dataset. However, the type of input prompts was unclear.

This study introduces Crack Large Model (CLM) to achieve a robust crack segmentation model by adopting SAM through transfer learning. As demonstrated in **Fig. 2**, we developed a segmentation network based on SAM components and leveraged transfer learning techniques to train the network. We use the largest available crack dataset for training SAM including 11k pairs of images and masks. Comparative experiments have been conducted by finetuning DeepLabv3 +, a popular CNN segmentation model, on

the same dataset [15]. Moreover, DeepCrack, a CNN specifically designed for crack segmentation is also trained on this dataset [16]. In the following sections, our dataset, methodology, and results are further expanded.

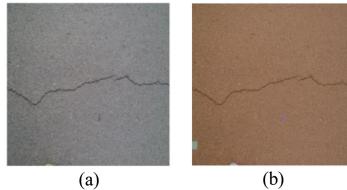


Fig. 1. Example output from the SAM automatic mask generator. (a) is a sample input image fed into the model, and (b) is the corresponding prediction of the SAM model on the input image, showing the limitation of SAM in detecting cracks.

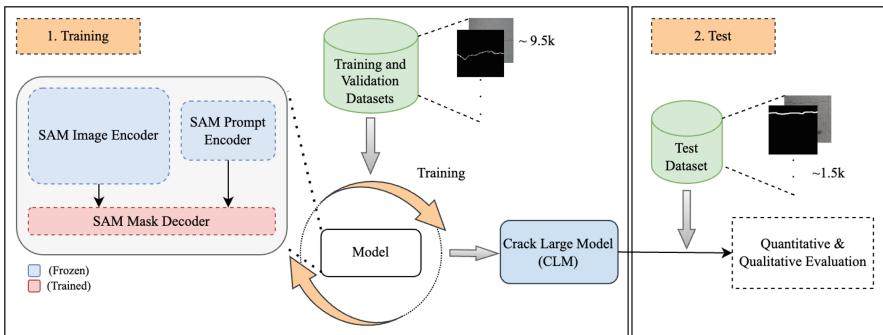


Fig. 2. Diagram illustrating the framework of the study, which follows a two-step process for training and testing the SAM model. In the training phase (left), the SAM model's Image Encoder and Prompt Encoder remain frozen, while the Mask Decoder is trained on a dataset of approximately 9.5k samples (training and validation). The model undergoes iterative optimization during this phase. In the test phase (right), the trained model, referred to as the Crack Large Model (CLM), is evaluated on a test dataset of 1.5k samples through both quantitative and qualitative assessments.

2 Methodology

2.1 Dataset

The dataset is the largest crack segmentation dataset publicly available on Kaggle¹. It comprises 11 well-established sub-datasets including 11282 images with corresponding binary masks (see **Table 1** and **Fig. 3**). Samples are provided in **Fig. 4**. Images and masks are 448 × 448-pixels in size. According to **Table 1**, the majority of the dataset is composed of Rissbilder (33.88%), Crack500 (29.81%), and non-crack (12.51%) categories which are concrete, stone, and concrete and tile cracks, respectively.

¹ <https://www.kaggle.com/datasets/lakshaymiddha/crack-segmentation-dataset>.

The dataset is considered a challenging dataset due to the varying image textures, background, illumination, and scene coverage. Specifically, Rissbilder contains images with low illumination, thin cracks, and material imprints in the background. Crack500 includes stone cracks with highly textured backgrounds which makes the segmentation challenging.

Table 1. Statistics of the crack segmentation dataset

| Dataset Category | Number of images | Percentage of total |
|----------------------|------------------|---------------------|
| CFD [17] | 102 | 0.90 |
| Crack500 [18] | 3363 | 29.81 |
| CrackTree [19] | 206 | 1.83 |
| DeepCrack [16] | 521 | 4.62 |
| Eugen Muller | 55 | 0.49 |
| Forest | 118 | 1.05 |
| GAPS384 [20] | 509 | 4.51 |
| Non-Crack | 1411 | 12.51 |
| Rissbilder | 3822 | 33.88 |
| Sylvie | 185 | 1.64 |
| Volker | 990 | 8.78 |
| <i>Total</i> | <i>11282</i> | <i>100</i> |

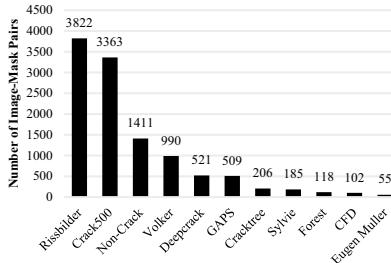


Fig. 3. Bar chart showing the number of image-mask pairs across different categories of the crack dataset. Rissbilder has the highest count at 3822, followed by Crack500 with 3363, and Non-Crack with 1411. Other sub-datasets include Volker (990), Deepcrack (521), GAPS (509), Cracktree (206), Sylvie (185), Forest (118), CFD (102), and Eugen Muller (55). The chart highlights the significant variation in dataset sizes.

2.2 Models

Inspired by [21], we developed Crack Large Model (CLM) based on SAM’s image encoder, prompt encoder, and mask decoder (see **Fig. 5**). CLM generates image embeddings through its image encoder and generates bounding boxes over the input image as

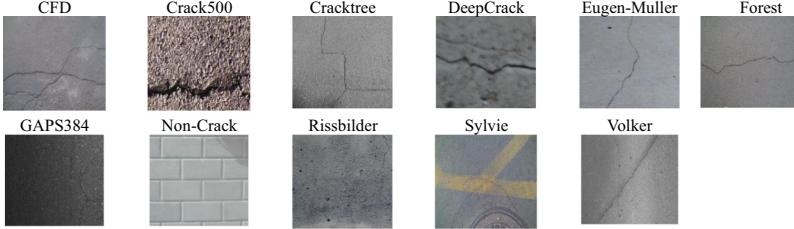


Fig. 4. A grid of twelve images showing samples from different categories of the crack dataset. Each image is labeled with its corresponding category: CFD, Crack500, Cracktree, DeepCrack, Eugen-Muller, Forest, GAPS384, NanoCrack, Rissbilder, Sylvie, and Volker. The images display various range of surface materials, such as concrete, brick, and asphalt, highlighting the dataset's variability in textures and crack patterns.

segmentation prompts. Prompts are passed to the prompt encoder module, and the outputs of the image and prompt encoders are further directed to the mask decoder. Finally, the mask decoder generates binary crack masks.

As mentioned earlier, SAM has been trained on 11M images with 1.1B Masks. To compare the capability of such a model, DeepLabv3 + with ResNet-50 as the backbone, pre-trained on a subset of the COCO dataset (including 330k images) has been finetuned on the same crack dataset. Moreover, DeepCrack, a CNN-based segmentation network designed by [16] for crack segmentation, was also trained on the dataset [15].

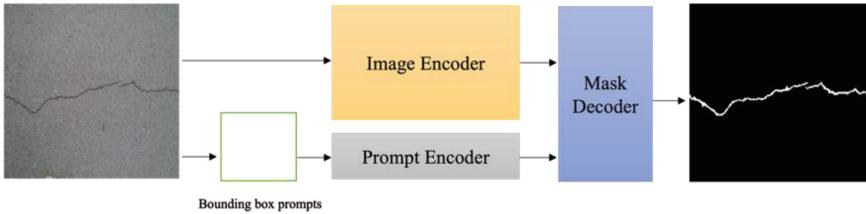


Fig. 5. A flowchart illustrating the schematic architecture of the Crack Large Model (CLM). On the left, a sample crack image serves as input to the model. The Image Encoder processes the image to generate embeddings, while corresponding bounding boxes are fed into the Prompt Encoder. The outputs from both encoders are then passed to the Mask Decoder, which predicts the final binary segmentation map, where crack pixels appear in white and the background in black.

2.3 Training and Evaluation

For training, the dataset was randomly split into train, validation, and test sets with 0.7, 0.15, and 0.15 ratios. For CLM's training, images were resized to 1024×1024 pixels, as the image encoder expects 1024×1024 pixels images. Moreover, the images were normalized with a mean and standard deviation of 0.25 and 0.5, respectively. For DeepLabv3 + and DeepCrack training, images were normalized and resized to 128×128 pixels.

For loss function, binary cross-entropy loss, dice loss, and combined binary cross-entropy and dice loss were examined (see Eqs. 1, 2) [4, 22]. Trainings were performed using the Adam optimizer with different learning rates for 20 epochs.

$$\text{BinaryCrossEntropy Loss} = - \sum_{i=1}^N [g_i \log(p_i) + (1 - g_i) \log(1 - p_i)] \quad (1)$$

$$\text{DiceLoss} = 1 - \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (2)$$

where N is the number of pixels, p_i and g_i are the predictions and ground truths, respectively.

To evaluate the performance of each experiment, metrics based on confusion matrix and objects' similarly were used. Specifically, precision, recall, F1-score, and AUC (Area Under the Curve ROC) were considered, for which calculations are done based on the models' True Positive, True Negative, False Positive, and False Negative values (Eq. 3–5) [4]. Due to the hardware's limitations, these metrics were calculated on 12% of the test set. Additionally, Dice Score was used to measure the similarity of the models' predicted masks and the ground truth masks according to Equation 6. During inference, images were resized to 1024×1024 pixels as input to all the trained models.

$$\text{Precision}(Pr) = \frac{TP}{TP + FP} \quad (3)$$

$$\text{DiceScore} = \frac{2|p \cap g|}{p + g} \quad (4)$$

$$\text{Recall}(Re) = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1-score}(F1) = \frac{2 \times Pr \times Re}{Pr + Re} \quad (6)$$

2.4 Results and Discussion

According to our experiments, the best loss function was the combination of binary cross-entropy loss and dice loss. Quantitative and qualitative results on our test dataset are given in **Table 2** and **Fig. 6**. It should be noted that this is the first large-scale training for crack semantic segmentation without using data augmentation or image processing techniques to enhance output masks. Considering the difficulty of the dataset, CLM achieves dice score of 0.52 which is higher than that of DeepLabv3 + (0.30) and DeepCrack (0.20). Overall, CLM outperforms the other two models in all metrics which shows its high capability in differentiating crack and non-crack pixels against a variety of texture backgrounds and similar objects. Also, CLM segmentation outputs are more continuous than the other two models, and the overall crack shape is partitioned in an almost perfect way.

Deeplabv3 + and DeepCrack, however, exhibit relatively similar poor performance on this dataset. Comparing Deeplabv3 + and DeepCrack, DeepCrack is able to better identify the crack regions by generating a higher number of false positives which results in notably lower precision than that of DeepLabv3 + (0.39 vs 0.47). On the other hand, Deeplabv3 + demonstrates better numerical performance (precision and F1-score of 0.47 and 0.27) than DeepCrack while it is unable to fully highlight crack pixels.

The results demonstrate the superiority of vision foundation models in extracting global contextual information and differentiating crack regions against any type of background (i.e., concrete, stone, asphalt). On the contrary, small networks whether pre-trained on smaller databases or trained from scratch have difficulty in segmenting cracks against complicated backgrounds. This suggests that crack semantic segmentation is a challenging problem and requires scene understanding rather than local pattern recognition solely based on geometry, color, or texture.

Table 2. Evaluation results of CLM, DeepLabv3 +, and DeepCrack on test set.

| Model | Dice score | Precision | Recall | F1-score | AUC |
|-------------|------------|-----------|--------|----------|------|
| CLM | 0.52 | 0.58 | 0.57 | 0.57 | 0.96 |
| DeepLabv3 + | 0.30 | 0.47 | 0.19 | 0.27 | 0.78 |
| DeepCrack | 0.20 | 0.39 | 0.19 | 0.25 | 0.78 |

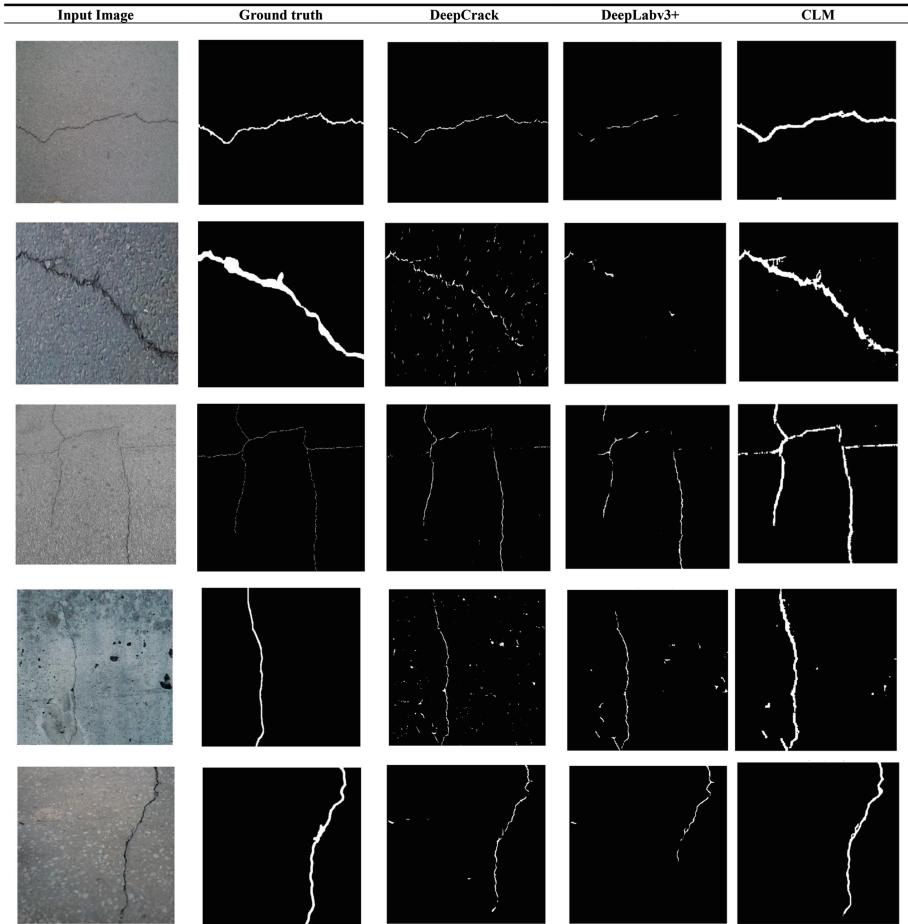


Fig. 6. A grid of images comparing crack detection methods on sample images of the test dataset. The first column shows input images of crack surfaces. The second column displays the ground truth, highlighting actual crack locations in white. The third column shows results from the DeepCrack method, the fourth from DeepLabv3+, and the fifth from the Crack Large Model (CLM). Each method's effectiveness in detecting cracks is visually compared, with varying levels of accuracy and noise in the detected crack patterns.

3 Conclusion

This study presented Crack Large Model (CLM) based on the state-of-the-art vision foundation model, Segment Anything Model (SAM), to achieve a robust and human-like crack semantic segmentation algorithm for real-world infrastructure inspections. CLM was trained on the largest crack segmentation dataset (11k pairs of images and masks) and the results were compared with two convolutional neural networks, namely, DeepLabv3+ and DeepCrack. CLM achieved a dice score of 52% on the dataset while DeepLabv3+ and DeepCrack reached 30% and 20% dice scores, respectively. Comparison results

demonstrated the superiority and robustness of foundation models in learning multi-scale and multi-level features of cracks against a variety of challenging backgrounds, and our developed Crack Large Model demonstrates a promising future for concrete crack segmentation and recognition. Further experiments will be conducted on leveraging foundation models for practical inspection demands.

References

1. Munawar, H.S., Hammad, A.W.A., Haddad, A., Soares, C.A.P., Waller, S.T.: Image-Based Crack Detection Methods: A Review. *Infrastructures* **6**(8), 115 (Aug 2021). <https://doi.org/10.3390/infrastructures6080115>
2. S. Halder and K. Afsari, "Robots in Inspection and Monitoring of Buildings and Infrastructure: A Systematic Review," *Appl. Sci.*, vol. 13, no. 4, Art. no. 4, Jan. 2023, <https://doi.org/10.3390/app13042304>
3. Agnisarman, S., Lopes, S., Chalil Madathil, K.: A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection. *Autom. Constr.* **97**, 52–76 (Jan 2019). <https://doi.org/10.1016/j.autcon.2018.10.019>
4. Ai, D., Jiang, G., Lam, S.-K., He, P., Li, C.: Computer vision framework for crack detection of civil infrastructure—A review. *Eng. Appl. Artif. Intell.* **117**, 105478 (Jan 2023). <https://doi.org/10.1016/j.engappai.2022.105478>
5. Dung, C.V., Anh, L.D.: Autonomous concrete crack detection using deep fully convolutional neural network. *Autom. Constr.* **99**, 52–58 (Mar 2019). <https://doi.org/10.1016/j.autcon.2018.11.028>
6. Ren, Y., et al.: Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Constr. Build. Mater.* **234**, 117367 (Feb 2020). <https://doi.org/10.1016/j.conbuildmat.2019.117367>
7. Fan, Z., Lin, H., Li, C., et al.: Use of Parallel ResNet for High-Performance Pavement Crack Detection and Measurement. *Sustainability* **14** (3) (Jan 2022). <https://doi.org/10.3390/su14031825>
8. Yu, J.-C., Yi, T.-H., Zhang, S.-H., Li, H.-N., Wang, Y.-F., Mei, X.-D.: Automatic Quantitative Identification of Bridge Surface Cracks Based on Deep Learning. *J. Perform. Constr. Facil.* **37**(1), 04022072 (Feb 2023). <https://doi.org/10.1061/JPCFEV.CFENG-4238>
9. Dosovitskiy, A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* (Jun 03, 2021). <https://doi.org/10.48550/arXiv.2010.11929>
10. Asadi Shamsabadi, E., Xu, C., Rao, A.S. et al. Vision transformer-based autonomous crack detection on asphalt and concrete surfaces. *Autom. Constr.* **140**, 104316 (Aug 2022). <https://doi.org/10.1016/j.autcon.2022.104316>
11. Wang, W., Su, C.: Automatic concrete crack segmentation model based on transformer. *Autom. Constr.* **139**, 104275 (Jul 2022). <https://doi.org/10.1016/j.autcon.2022.104275>
12. Wu, Y., Li, S., Zhang, J., Li, Y., Li, Y., Zhang, Y.: Dual attention transformer network for pixel-level concrete crack segmentation considering camera placement. *Autom. Constr.* **157**, 105166 (Jan 2024). <https://doi.org/10.1016/j.autcon.2023.105166>
13. Bommasani, R., et al. On the Opportunities and Risks of Foundation Models. *arXiv* (Jul 12, 2022). <https://doi.org/10.48550/arXiv.2108.07258>
14. Kirillov, A., et al.: Segment Anything. *arXiv* (Apr. 05, 2023). Accessed Jul. 24, 2023. [Online]. Available: <http://arxiv.org/abs/2304.02643>
15. Chen, L.-C., Papandreou, G., Schroff, F., et al.: Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* (Dec. 05, 2017). Accessed Dec. 01, 2023. [Online]. Available: <http://arxiv.org/abs/1706.05587>

16. Liu, Y., Yao, J., Lu, X., Xie, R., Li, L.: DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing* **338**, 139–153 (Apr 2019). <https://doi.org/10.1016/j.neucom.2019.01.036>
17. Shi, Y., Cui, L., Qi, Z., Meng, F., Chen, Z.: Automatic Road Crack Detection Using Random Structured Forests. *IEEE Trans. Intell. Transp. Syst.* **17**(12), 3434–3445 (Dec 2016). <https://doi.org/10.1109/TITS.2016.2552248>
18. Zhang, L., Yang, F., Daniel Zhang, Y., et al.: Road crack detection using deep convolutional neural network. In 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA: IEEE, 3708–3712 (Sep 2016). <https://doi.org/10.1109/ICIP.2016.7533052>
19. Zou, Q., Cao, Y., Li, Q., Mao, Q., Wang, S.: CrackTree: Automatic crack detection from pavement images. *Pattern Recognit. Lett.* **33**(3), 227–238 (Feb 2012). <https://doi.org/10.1016/j.patrec.2011.11.004>
20. Eisenbach, M., et al.: How to get pavement distress detection ready for deep learning? A systematic approach. In 2017 International Joint Conference on Neural Networks (IJCNN), 2039–2047 (May 2017). <https://doi.org/10.1109/IJCNN.2017.7966101>
21. Ma, J., He, Y., Li, F., et al.: Segment Anything in Medical Images. arXiv (Jul. 17, 2023). Accessed Aug. 19, 2023. [Online]. Available: <http://arxiv.org/abs/2304.12306>
22. Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv (Jun. 15, 2016). Accessed Nov. 29, 2023. [Online]. Available: <http://arxiv.org/abs/1606.04797>



Building the Blueprint for AI-Powered Compliance Checking: Analyzing ChatGPT-4 & Gemini by Question Category in Engineering Regulations

Entesar Al Nama^(✉) and Maqsood Mahmud

College of Business Administration, University of Bahrain, Zallaq, Bahrain
entesarjasim@gmail.com

Abstract. This paper forms a part of an in progress PhD research, exploring the transformative potential of question-and-answer (Q&A) models within the context the law that regulates the practice of engineering professions (law) in the Kingdom of Bahrain. Driven by recent advancements in Machine Learning (ML) and large language models like ChatGPT-4 and Gemini. Q&A systems offer exceptional promise to improve access to crucial regulatory information and empower informed decision-making for engineers. The researchers delve into the history of Q&A models and the key role of word transformation and embeddings in their accuracy. By analyzing data pertaining to regulatory requirements outlined in the aforementioned law, the researchers conducted a comprehensive comparison of Chat GPT-4 and Gemini focusing on five user-relevant question types: factual, critical thinking, hypothetical, open-ended, and comparison.

The performance analysis of ChatGPT-4 and Gemini across different question types indicates that both models performed effectively in factual questions. ChatGPT-4 achieved 93.6% correct answers, with 4.3% partially correct, while Gemini achieved 89.4% correct and 6.4% partially correct. In critical thinking tasks, ChatGPT-4 demonstrated higher accuracy, with 93.6% correct answers, compared to Gemini's 89.3% correct. For hypothetical scenarios, both models performed well, with ChatGPT-4 at 91.5% correct and Gemini at 87.2%. In open-ended discussions, ChatGPT-4 achieved 100% accuracy, whereas Gemini had 89.4% correct. However, both models showed room for improvement in conducting full comparisons. Further analysis of these partially correct responses will provide deeper insights into model performance and areas for improvement.

Based on these results, the researchers argue for the strategic choice of fine-tuning ChatGPT-4 for a Q&A model tailored to supporting engineers in navigating the intricacies of the law and empower them with deeper insights into regulatory compliance, nuanced perspectives on applying legal mandates, and improved decision-making throughout their professional practice.

This research marks a significant step in harnessing the power of AI for improved knowledge access and informed decision-making within the specific domain of engineering regulations in Bahrain. Further research recommends exploring domain-specific fine-tuning techniques and integration with collaborative platforms holds immense potential for unlocking the full potential of AI

in streamlining compliance processes and enhancing overall efficiency within the engineering profession.

Large Language Models (LLMs) can be leveraged to create a Q&A model for Engineering Regulations. One way to achieve this is by converting natural language (NL) requirements into machine-readable requirements using LLMs. This can be accomplished by creating a requirements table from free-form NL requirements and identifying boilerplate templates for different types of requirements based on linguistic patterns. By utilizing these language models together, the standardization of requirements can be achieved. To demonstrate the effectiveness of this approach, the researchers employed requirements from the law, with respect to regulating the practice of engineering professions in the Kingdom of Bahrain.

Keywords: large language models (LLMs) · ChatGPT-4 vs Gemini for engineering · engineering regulations · decision-making in engineering · fine-tuning ChatGPT-4 for engineering law · improving engineering practice with AI · natural language processing (NLP) · machine learning (ML) · deep learning · word embeddings · transformer architecture · architecture · engineering and construction (AEC) industry · Bahrain · Middle East

1 Introduction

1.1 Background and Significance

Current methods for extracting and transforming information in rule-based systems for Automated Compliance Checking (ACC) can handle complex requirements. However, when applied to different types of regulatory documents, they become inflexible. To facilitate downstream tasks within ACC systems, such as information matching and compliance reasoning, it is essential to develop a more flexible approach for automatically processing and interpreting requirements [1]. Compliance with law governing the practice of engineering is vital for ensuring the quality, safety, and sustainability of engineering projects. Engineering projects often involve multiple stakeholders, such as clients, contractors, regulators, and the public, who have different interests and expectations. [2] explored how a common data environment (CDE) could enhance stakeholder communications by reducing time wastage and minimizing the risk of data loss. Complex construction projects are streamlined by a well-executed CDE, which improves stakeholder collaboration throughout the project lifecycle. Non-compliance with the law or the code of ethics can result in serious consequences for the project and the stakeholders, such as legal liabilities, financial losses, reputational damages, and environmental impacts. Therefore, engineers should be aware of the laws and regulations that apply to their projects and follow the best practices and standards of their profession. Regulations include natural subjective requirements help designers (architects and engineers) to make decisions without limiting their design solutions too much. These requirements are utilized by designers and are understood through abductive processes during the decision-making phase [3].

Compliance aids, such as project management tools, data security controls, training programs, and professional associations, can help engineers to comply with the law and

enhance their professional competence and accountability. Compliance checking and monitoring are also important processes to ensure that the project meets the legal and ethical requirements and to identify and correct any deviations or issues. It is acknowledged that the problems with the manual compliance checking process can be effectively resolved by using the ACC process [4]. Projects guided by stakeholders are executed with reduced conflict and greater success [5, 6].

In the rapidly evolving field of artificial intelligence, Question and Answering (Q&A) models have become instrumental in various domains, providing users with immediate access to information and expert-like responses. With the advent of sophisticated language models, such as Gemini and GPT-4, there is a burgeoning potential for AI-driven assistance in highly specialized fields, including engineering and law. LLM improves design team and client communication. Conversational interfaces let engineers and architects ask questions, get design advice, and discuss project requirements, fostering an iterative design process [7]. A Q&A model is a natural language processing (NLP) system that can answer questions based on a given text or data source. A Q&A model can be beneficial for ensuring compliance with the law governing the practice of engineering in the Kingdom of Bahrain, as it can provide quick and accurate answers to queries related to the legal framework, regulations, and standards that apply to engineering projects in the country. For example, a Q&A model can help engineers and contractors to understand the requirements and procedures for obtaining licenses and approvals from the authority in charge of licensing and regulating engineering offices and professionals in Bahrain [8]. The Kingdom of Bahrain, with its progressive stance on technological integration, presents a unique landscape for the deployment and evaluation of such AI models. By using Prompt Engineering—a process that involves directly processing text documents in natural language—LLMs can improve ACC automation by providing customization and integrating domain knowledge through fine-tuning [4].

1.2 Research Objectives

This study represents the first stage of a multi-part research initiative aimed at creating a Q&A model to assist Bahrain's engineering community in understanding and navigating the law with respect to Regulating the Practice of Engineering Professions in the Kingdom of Bahrain. This initial phase focuses on evaluating the accuracy and suitability of two large language models (LLMs): Gemini and ChatGPT-4.

The primary objective is to identify the LLM best suited for building a reliable and effective Q&A model for the law. To achieve this, the researchers will:

- Utilize a diverse dataset of questions covering various formats (factual, comparative, hypothetical, critical thinking, open-ended).
- Evaluate each LLM's performance based on accuracy, ability to handle legal concepts and terminology, and understanding of the law's nuances.
- Consider practical factors like computational efficiency and future development potential.

While further research will focus on fine-tuning the chosen LLM and building the Q&A model, this initial evaluation forms a crucial foundation. By selecting the most

suitable LLM, we ensure a solid base for developing a tool that can expedite regulatory navigation and instill greater confidence in compliance-related decision-making for Bahrain's engineering community. Furthermore, the researchers anticipate setting a benchmark for the utility of AI in specialized sectors, providing a measurable template for future AI deployments that are sensitive to the unique compliance requirements of the AEC industry. Figure 1 illustrates the steps in phase 1 of creating a Q&A Model concerning the Practice of Engineering Professions in the Kingdom of Bahrain.

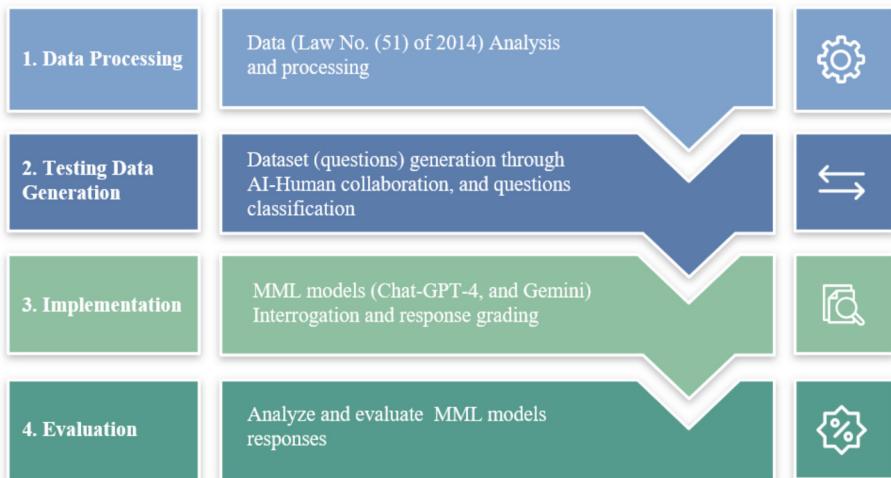


Fig.1. Phase 1 of Creating a Q&A Model Concerning the Practice of Engineering Professions in the Kingdom of Bahrain

2 Related Works

2.1 Evolution of Q&A Models

In the early 1960s, Q&A Systems were created. The central database or knowledge system of BASEBALL and LUNAR, two of the first Q&A systems, allowed them to succeed. BASEBALL, which was created to provide answers to inquiries on the US baseball league for a whole year, handled inquiries like "where did each club play on July 7?" with ease. The geological examination of rocks recovered by the Apollo lunar mission was covered by LUNAR. These previous systems emphasized "closed domains," in which all questions must pertain to the particular area or have a restricted vocabulary. The knowledge that both BASEBALL and LUNAR systems possessed was a major factor in their respective successes. Over the following two decades, open domain systems with an emphasis on information retrieval were developed [9]. The development of Q&A models based computer programs that try to provide users with answers to their inquiries, typically in natural language, began in the late 1990s with the introduction of internet services like Ask Jeeves and Google Answers. These services depended on

volunteers or human specialists to respond to inquiries from users, frequently in exchange for payment or other benefits (Q&A Software, 2023).

Nowadays Q&A models are natural language processing systems designed to respond in natural language to questions from people. Open-domain and closed-domain are the two primary categories into which they can be divided. Open-domain Q&A models can answer inquiries on any topic, using general knowledge sources such as the web or Wikipedia. Closed-domain Q&A models can use domain-specific knowledge sources, such as databases, papers, or ontologies, to provide answers to questions on certain subjects.

Later on, Q&A models developed into increasingly automated and complex systems that analyzed questions and produced answers from a variety of sources, including databases, documents, and the internet, using methods like machine learning, information retrieval, natural language processing, and knowledge representation. These models include Yahoo! Answers, Quora, and Stack Exchange, where users can post questions and receive answers on a variety of subjects and domains (Q&A Software, 2023). Moreover, there are several sorts of Q&A models based on the structure of the questions and the sources of the responses. Factoid Q and A models, for instance, respond to queries like "What is the capital of France?" and "Who is the creator of Harry Potter?" that have a single factual response. In addition, non-factoid Q&A models address topics like "Why is the sky blue?" and "What is the best method to learn a new language?" that call for more in-depth analysis, justification, or opinion. In addition, there are closed-domain Q and A models that are focused on a particular domain or topic, like sports, history, or medicine, and open-domain Q and A models that can respond to queries from any domain or topic [11].

Many researchers suggest that Q&A models have made significant progress in recent years, especially with the advances in deep learning and natural language processing. However, there are still many challenges and opportunities in the field, such as data quality and availability, domain knowledge and complexity, user needs and preferences, and technology integration and innovation.

2.2 Foundations of Language Understanding

Word embeddings are a powerful tool in natural language processing (NLP) that can be used to represent words as vectors of real numbers. These vectors capture the semantic and syntactic meaning of words, allowing machine learning models to perform various NLP tasks such as sentiment analysis, machine translation, and question answering. [12] defines embedding as a list of floating-point numbers that represents a vector. The relatedness between two vectors is measured by their distance. When the distance between two vectors is small, it indicates high relatedness, while large distances suggest low relatedness.

The quest for a universal representation of text is central to the automated processing of natural languages. The development of pretrained text embeddings such as word2vec or GloVe has been a significant breakthrough in this area. While supervised models have consistently outperformed unsupervised models over the years, models based on unsupervised learning have become more prevalent in recent years. These models do not require specially labeled datasets, but can learn from existing or automatically generated

large corpora of texts, taking full advantage of deep learning [13, 14]. In Q&A models, word embeddings can be used to represent the question and the candidate answers, allowing the model to identify the most relevant answer. For example, in a typical Q&A model, the input question is first converted into a vector using a word embedding model such as Word2Vec or GloVe. The candidate answers are then also converted into vectors using the same word embedding model.

Each word embedding algorithm has its own set of advantages and disadvantages, including early algorithms such as Word2Vec. As an illustration, Word2Vec is a well-known approach to word embedding generation that predicts the context of a word by analyzing its adjacent words via neural network [15]. Although these models perform well in natural language processing tasks, they ignore word order and lose some of the text's semantic information [16].

2.3 Q&A Models in Architecture, Engineering and Construction (AEC) Industry

Q&A models have emerged as pivotal tools within the Architecture, Engineering, and Construction (AEC) domain, addressing critical activities such as knowledge management and information retrieval. The AEC sector's data complexity, coupled with the challenges of non-standardized information models and natural language query ambiguity, necessitates the deployment of advanced algorithms and methodologies. These include machine learning, information extraction, and semantic parsing to navigate and resolve the sector's inherent complexities (Ren & Zhang 2021; Zheng et al., 2022).

2.4 Regulatory Compliance Through Q&A Models

Q&A models serve a vital function in querying about AEC regulations, such as construction codes and environmental guidelines, which are essential yet complex and dynamic in nature. By providing precise and relevant answers, these models assist stakeholders in understanding and adhering to regulations,

thereby ensuring project sustainability and safety. Overcoming challenges such as regulation formalization, data alignment, and model validation remains a critical area of focus for enhancing these systems' efficacy (European Commission, 2021, 2022).

2.5 Methodological Advances in ML for q&A Models

The application of Machine Learning (ML) methodologies in developing Q&A models for the AEC industry has significantly augmented project efficiency and efficacy. Key methodologies encompass:

- **Natural Language Processing (NLP):** Employed for the extraction and analysis of information from textual data, with applications demonstrated in automated construction specification reviews (Moon et al., 2022).
- **Feature Selection and Extraction:** This critical step in ML involves identifying and extracting salient features from data, as evidenced by [22] work on filter feature selection.

- **Model Training and Evaluation:** The process involves training ML models on datasets to discern patterns, followed by evaluation to ascertain performance, as shown by [23]
- **Model Optimization:** Post-training, models are optimized to enhance performance. [23] utilized genetic algorithms for this purpose in predicting construction project durations.
- **Model Deployment and Application:** [24] in their model deployment process, they gathered data from various sources, including BIM models, building regulations, and user queries. This data was then processed using natural language techniques to extract relevant information. The processed data was stored in a knowledge base, which combined a relational database and an ontology. To respond to user queries, they utilized a bidirectional encoder (BERT) model, matching queries with relevant data from the knowledge base and providing answers in natural language. This approach aimed to enhance conversational AI systems in the AEC industry by ensuring efficient and accurate information retrieval.

2.6 Comparative Studies of LLMs (ChatGPT and Gemini) in Practice

2.6.1 Chat GPT

ChatGPT is an AI language model created by OpenAI in 2022. It is based on neural network architecture and can process and generate responses for any sequence of characters that make sense, including different spoken languages, programming languages, and mathematical equations. ChatGPT is trained using a large dataset and fine-tuning techniques, which contribute to its impressive language generation capabilities. It can understand and respond to a wide range of user queries and prompts, making it versatile in conversational interactions. ChatGPT has the ability to handle complex and nuanced conversations, providing detailed and informative responses (Ahmed et al., n.d.; Cretu, 2023).

ChatGPT-4 is a newer and improved version of ChatGPT, it was introduced in March 2023 with more advanced language generation capabilities and a larger number of parameters, it has the ability to generate high-quality and coherent responses to human queries. Previous research has demonstrated that ChatGPT-4 exhibits impressive generation capabilities when compared to other existing models. However, there has been limited quantitative analysis of Chat GPT-4's ability to understand the content it generates. Chat GPT-4 has limitations in handling paraphrase and similarity tasks, but excels in inference tasks and outperforms all BERT models by a significant margin. Chat GPT-4 also performs comparably to BERT on sentiment analysis and question-answering tasks, indicating its competence in these domains [27, 28].

2.6.2 Gemini (Formerly Known as Bard)

Google's AI chatbot Bard underwent a significant transformation and was renamed to Gemini. This change occurred on February 7, 2024. Gemini is a conversational generative artificial intelligence chatbot developed by Google. Initially based on the LaMDA family of large language models (LLMs), it was later upgraded to PaLM and then to

Bard. Bard was developed as a direct response to the meteoric rise of OpenAI's ChatGPT, and was released in a limited capacity in March 2023 to lukewarm responses before expanding to other countries in May. Bard is an artificially intelligent language model (LLM) developed by Google AI. It is trained on a massive dataset of text and code, which allows it to generate human-quality text, translate languages, write different kinds of creative content, and answer questions in an informative way. Bard has the ability to handle complex and nuanced conversations, providing detailed and informative responses. It can understand and respond to a wide range of user queries and prompts, making it versatile in conversational interactions [26, 29].

A recent study by Ars Technica compared ChatGPT-4 and Google Bard in seven categories: dad jokes, argument dialog, mathematical word problems, summarization, factual retrieval, creative writing, and coding. The study found that ChatGPT-4 outperformed Google Bard on inference tasks, while Google Bard was better at handling back-and-forth voice dialogue on mobile devices [30]. Another article by Techopedia compared the two chatbots in a head-to-head output test and explored their major differences [31]. Zapier also compared the two models and found that Bard is built for research, while ChatGPT-4 is the better writer [32]. While these studies provide some insight into the capabilities of each model, it's important to note that the specific strengths and weaknesses of each model will depend on the use case and the specific requirements of the bot. Both ChatGPT-4 and Bard can be used to create custom chatbots using their APIs, but the specific capabilities of each model will depend on the use case and the specific requirements of the bot [30, 31].

2.7 Comparing Between ChatGPT and Gemini (Bard) Performance

A recent study by Ars Technica compared ChatGPT-4 and Google Bard in seven categories: dad jokes, argument dialog, mathematical word problems, summarization, factual retrieval, creative writing, and coding. The study found that ChatGPT-4 outperformed Google Bard on inference tasks, while Google Bard was better at handling back-and-forth voice dialogue on mobile devices [30]. Another article by Techopedia compared the two chatbots in a head-to-head output test and explored their major differences [31]. Zapier also compared the two models and found that Bard is built for research, while ChatGPT-4 is the better writer [32]. While these studies provide some insight into the capabilities of each model, it's important to note that the specific strengths and weaknesses of each model will depend on the use case and the specific requirements of the bot. Both ChatGPT-4 and Bard can be used to create custom chatbots using their APIs, but the specific capabilities of each model will depend on the use case and the specific requirements of the bot [30, 31].

[33] examine the performance of GPT-3.5, GPT-4, and Bard models, by performing a thorough technical evaluation on different reasoning tasks across eleven distinct datasets. The paper provides empirical evidence showcasing the superior performance of ChatGPT-4 in comparison to both ChatGPT-3.5 and Bard in zero-shot setting throughout almost all evaluated tasks. While the superiority of GPT-4 compared to GPT-3.5 might be explained by its larger size and NLP efficiency, this was not evident for Bard. The study also demonstrate that the three models show limited proficiency in Inductive, Mathematical, and Multi-hop Reasoning Tasks.

ChatGPT4 and Bard have been compared in terms of their finetuning for a Q&A model. The understanding ability of ChatGPT4 was evaluated on the GLUE benchmark and compared with four representative BERT-style models. It was found that ChatGPT4 falls short in handling paraphrase and similarity tasks, but outperforms all BERT models on inference tasks. Additionally, ChatGPT4 achieves comparable performance to BERT on sentiment analysis and question-answering tasks. By combining advanced prompting strategies, the understanding ability of ChatGPT4 can be further improved. A comprehensive survey comparing the capabilities and features of ChatGPT and Bard positions ChatGPT as a leading model with exceptional performance [27]. In a study evaluating the ability of large language models to answer anesthesia-related queries, ChatGPT was found to have lengthier, intellectual, and effective responses compared to Bard. However, Bard responses were more conversational and easier to read. Computational sentiment analysis showed that Bard had significantly greater polarity scores than ChatGPT [26].

Most recent papers suggest that ChatGPT-4 is a more capable LLM than Bard in terms of generating high-quality and coherent responses to human queries, but more research is needed to compare their performance on other NLP tasks and domains. [33] examine the performance of GPT-3.5, GPT-4, and Bard models, by performing a thorough technical evaluation on different reasoning tasks across eleven distinct datasets. [33] provide empirical evidence showcasing the superior performance of ChatGPT-4 in comparison to both ChatGPT-3.5 and Bard in zero-shot setting throughout almost all evaluated tasks. While the superiority of GPT-4 compared to GPT-3.5 might be explained by its larger size and NLP efficiency, this was not evident for Bard. The study also demonstrate that the three models show limited proficiency in Inductive, Mathematical, and Multi-hop Reasoning Tasks.

2.8 Future Trajectory of ML-Based Q&A Models

The current GPT models have limitations in processing large and complex text documents, requiring careful prompt organization and fine-tuning [4]. However, looking forward, ML-based Q&A models are poised to significantly impact AEC project management. By harnessing the capabilities of NLP and ML, these models provide extensive decision support across various AEC facets, from planning to safety and productivity. The continued evolution of ML technologies promises even more groundbreaking Q&A applications in the AEC industry.

3 Methodology

3.1 Research Approach

The methodology section of this study serves as a blueprint for assessing the capabilities of advanced Large Language Models (LLMs) – specifically ChatGPT-4 and Gemini – against the backdrop of the law, which governs the practice of engineering professions. This assessment is anchored by a robust dataset, derived from the aforementioned 45 legal articles and enriched through AI-assisted question generation. The methodology unfolds through a series of meticulously structured steps: from the nuanced construction

of the dataset to the strategic interrogation of the LLMs, and the analytical grading of their responses. Each phase is designed to not only quantify the performance of each model using precise metrics but also to discern their suitability for real-world application within the engineering legal framework.

The following sections provide a clear explanation of the systematic approach used in this study, highlighting the thoroughness and detailed scope of the comparative analysis.

3.2 Dataset Description

The dataset for this research is a legal document that regulates the practice of engineering professions in the Kingdom of Bahrain. It is a PDF file that can be accessed online, and it contains 45 articles that address various aspects of practicing the engineering professions in the Kingdom.

The dataset is analyzed according to five characteristics: volume, variety, velocity, veracity, and value. The analysis shows that the dataset is small, but it has both structured and unstructured data that need further processing and annotation. The dataset is also static, but it may require periodic updates to reflect the changes in the law or its interpretation. Moreover, the dataset is reliable and accurate, but it may pose some challenges in terms of ambiguity or contradiction in different contexts and scenarios. The dataset is valuable for the research, as it provides relevant information that can be used to generate realistic and meaningful questions and answers. The research methodology involves using natural language processing techniques to extract information from the document and generate questions and answers based on the content and purpose of the document. The research evaluation measures the performance and accuracy of the machine learning model using various metrics and benchmarks.

The 5Vs—Volume, Velocity, Variety, Veracity, and Value—are crucial characteristics of big data. These 5Vs guide the design, training, and deployment of Q&A models, ensuring they perform effectively in real-world scenarios [34, 35]. Table 1 gives a summary of dataset description:

3.3 Dataset Generation Through AI-Human Collaboration

The study embarked on a collaborative approach, utilizing both AI capabilities and human expertise to curate a dataset of 235 diverse questions. This dataset was meticulously constructed from 45 articles of law (including 2 articles with 2 parts), distributed among 11 clauses, with each article contributing five questions of distinct types. These question types were carefully chosen to encapsulate various analytical dimensions: factual, open-ended discussion, comparison, critical thinking, and hypothetical scenarios, ensuring comprehensive coverage of potential query spectrums.

3.4 Question Generation and Classification

Each question was designed to elicit responses that would not only test the AI models' knowledge retrieval capabilities but also their reasoning, inferencing, and contextual understanding. The questions were calibrated to ensure a minimum quality threshold, with the anticipation of varying degrees of complexity and interpretability.

Table 1. Characteristics of the Dataset for Q&A Generation

| Dataset | Size | Format / at | Content | The 5 V's of Big Data | | | Value |
|----------------------------------|-------|--------------|---|--|--------------------------|---|--|
| | | | | Volume | Velocity | Variety | |
| Bahrain Engineering Practice Law | Small | PDF / online | 45 articles on regulating the practice of engineering professions | Small, may need augmentation or advanced NLP | Static, may need updates | Structured and unstructured, need processing and annotation | Valuable, provide relevant info for Q&A generation |

3.5 AI Model Interrogation and Response Grading

Two advanced language models, ChatGPT-4 and Gemini, were subjected to the dataset questions. The responses were meticulously evaluated and graded on a scale where ‘0’ denoted an incorrect answer, ‘0.5’ indicated a partially correct response, and ‘1’ signified a fully correct answer. An integral part of this assessment also involved noting whether the models could supplement their answers with additional information or reference their responses to preceding queries, which provided insight into their contextual awareness and knowledge integration capabilities.

3.6 Evaluation Metrics and Analysis

Accuracy, precision, recall, and F1 score are commonly used evaluation metrics in machine learning. Accuracy measures the overall correctness of a classifier’s predictions. Precision measures the proportion of true positive predictions out of all positive predictions, while recall measures the proportion of true positive predictions out of all actual positive instances. F1 score is the harmonic mean of precision and recall, providing a balanced measure of a classifier’s performance. These metrics are important for assessing the effectiveness of binary, multi-class, and multi-labelled classifiers. While accuracy and F1 score are widely used, it is important to consider other metrics such as ROC/AUC and Kappa statistics for a comprehensive evaluation [36–38].

To quantify the proposed models’ performance, standard evaluation metrics such as precision, recall, and F1 score were employed. Precision measures the accuracy of the provided answers, recall assesses the models’ ability to retrieve all relevant instances, and the F1 score, being the harmonic mean of precision and recall, offers a balanced view of the models’ accuracy in both retrieving relevant information and omitting irrelevant details. The inclusion of recall is crucial, as it evaluates the completeness of the models’ answers, ensuring they are not only correct when they provide an answer but also that they provide an answer whenever it is appropriate to do so.

3.7 Objective Comparison for Model Suitability

The heart of this methodology lies in establishing an empirical comparison between ChatGPT-4 and Gemini. This comparison is not merely a performance showdown but a structured analysis to determine which model is best suited for developing a Q&A model. The ultimate aim is to discern the most apt AI model that aligns with the stringent requirements of engineering practice compliance within the Kingdom of Bahrain.

4 Research Findings

The Research Results chapter presents a comprehensive analysis of the comparative performance of ChatGPT-4 and Gemini, two advanced Large Language Models, in the context of engineering law compliance. This chapter systematically unveils the efficacy of each model across various question types, dissecting their ability to provide accurate and contextually rich responses. Key findings include a nuanced examination of model

precision, revealing inherent strengths and areas necessitating improvement. The ensuing sections detail the methodologies employed and the resulting implications of each model's capabilities, as informed by a dataset derived from the law.

Detailed below the calculated metrics and the analysis per question type, with the note that 1 point was given for each correct answer, 0.5 points for each partially correct answer, and 0 points for each incorrect answer, the partially correct answers were categorized alongside the correct answers for metrics calculation purposes:

4.1 Assessment of Overall Performance

The assessment reveals that both ChatGPT-4 and Gemini exhibit commendable performance across a spectrum of question types. Notably, each system often achieved full marks in their responses, underscoring their robust answering capabilities.

- **ChatGPT-4:** Achieved an accuracy of 93.2%, precision of 93.9%, recall of 99.1%, and F1-score of 096.5%.
- **Gemini:** Achieved an accuracy of 85.1%, precision of 88.9%, recall of 95.2%, and F1-score of 91.9%.

4.2 Analysis by Question Type

The performance of ChatGPT-4 and Gemini per question type analysis reveals the followings:

- **Factual Questions:** Both models performed very well. Gemini achieved 89.4% (42/47) correct answers and 6.4% (3/47) partially correct answers. ChatGPT-4 achieved 93.6% (44/47) correct answers and 4.3% (2/47) partially correct answers. Only one factual error was made ChatGPT-4, and two by Gemini.
- **Critical Thinking Questions:** ChatGPT-4 showed high accuracy, with Gemini achieving 89.3% (42/47) correct answers and 4.3% (2/47) partially correct answers. ChatGPT-4 achieved 93.6% (44/47) correct answers and 4.3% (2/47) partially correct answers.
- **Hypothetical Questions:** Both models achieved high accuracy. ChatGPT-4 achieved 91.5% (43/47) correct answers and 8.5% (4/47) partially correct answers. Gemini achieved 87.2% (41/47) correct answers and 8.5% (4/47) partially correct answers. Only one error was made by Gemini, but overall, both models demonstrated good understanding and ability to handle hypothetical scenarios.
- **Open-ended Questions:** ChatGPT-4 showed very high accuracy, with Gemini achieving 89.4% (42/47) correct answers and 6.4% (3/47) partially correct answers. ChatGPT-4 achieved 100% (47/47) correct answers.
- **Comparative Questions:** Both models showed less ability to conduct full comparisons. Gemini achieved 70.2% (33/47) correct answers and 27.7% (13/47) partially correct answers. ChatGPT-4 achieved 87.2% (41/47) correct answers and 12.8% (6/47) partially correct answers. Only one error was made by Gemini, highlighting its ability to conduct better comparisons with more training.

4.3 Consistency Across Clause Titles

Examination of various clause titles suggested uniform performance, suggesting that the understanding and response generation by both systems is consistent, irrespective of the subject matter.

4.4 Consideration of Additional Information

Gemini's responses included "additional information" in 22 answers comparing to 14 answers provided by ChatGPT-4, signaling Gemini's propensity to provide detailed answers and further context, a feature noted in the evaluation comments.

4.5 Identification of Areas for Improvement

Situations where either model scored below full points suggest opportunities for further refinement in their answering strategies, highlighting potential areas for model enhancement.

5 Conclusion

In conclusion, the comparative analysis of ChatGPT-4 and Gemini in the domain of engineering law compliance yields valuable insights into their respective strengths and areas for improvement. While both models demonstrate commendable performance across various question types, nuanced differences emerge in their precision and ability to handle specific scenarios. ChatGPT-4 exhibits superior accuracy in critical thinking and comparison tasks, whereas Gemini excels in providing additional contextual information. Consistency in performance across different clause titles underscores the robustness of both systems in comprehending diverse legal contexts. However, areas where either model falls short suggest avenues for refining their answering strategies and enhancing overall performance. Moving forward, leveraging these insights can inform the ongoing development and optimization of language models for specialized legal domains, ultimately contributing to more effective and reliable legal compliance solutions. These findings, coupled with ChatGPT-4's accessible API and other functionalities, make it a strong contender for developing an effective Q&A model for Bahrain's engineering community.

As the research advances, the focus will shift to applying ChatGPT-4 in the creation of a Q&A model designed to support the engineering industry's decision-making processes and compliance with legal standards. The next steps will involve real-world implementation to rigorously test the model's performance, with the goal of optimizing stakeholder decisions and enhancing the quality and safety of engineering projects. The culmination of this research will not only contribute to the academic understanding of LLMs' capabilities but will also guide the practical selection of an AI-driven Q&A model. The selected model is expected to enhance the adherence to regulatory compliance in the engineering sector, thereby upholding the standards of quality, safety, and sustainability.

References

1. Al Nama, E., Alalawi, A.: The Adoption of Automated Building Code Compliance Checking Systems in the Architecture, Engineering, and Construction Industry (2023). Accessed: May 14, 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/10050930>
2. Fronk, A.: The challenges of Digital Building Permit. In: F. Noardo and G. Malacarne (eds) Digital Building Permit: A State of Play, pp. 17–18 (2021). Accessed: Nov. 25, 2022. [Online]. Available: <https://eu4dbp.net/events-dbp/i-eunet4dbp-international-workshop-on-digital-building-permit-1eunet4dbpws/>
3. Soliman-Junior, J., et al.: Automated compliance checking in healthcare building design. Autom Constr. **129** (Sep 2021). <https://doi.org/10.1016/j.autcon.2021.103822>
4. Liu, X., Li, H., Zhu, X.: A GPT-based method of Automated Compliance Checking through prompt engineering.
5. Miller, D., Oliver, M.: Engaging Stakeholders for Project Success. Engaging Stakeholders for Project Success.
6. Voinov, A., Bousquet, F.: Modelling with stakeholders. Environ Model Softw. **25**(11), 1268–1281 (Nov 2010). <https://doi.org/10.1016/J.ENVSOFT.2010.03.007>
7. Rane, N.L., Choudhary, S.P., Rane, J.: Integrating ChatGPT, Bard, and Leading-Edge Generative Artificial Intelligence in Architectural Design and Engineering: Applications, Framework, and Challenges. International Journal of Architecture and Planning **3**(2), 92–124 (Sep 2023). <https://doi.org/10.51483/IJARP.3.2.2023.92-124>
8. CRPEP: The Council for Regulating the Practice of Engineering Professions. Accessed: Feb 13, 2024. [Online]. Available: <http://www.crpep.bh/>
9. Taylor, J.: What Are Question-Answering Systems?. Accessed: Nov. 05, 2023. [Online]. Available: <https://lucidworks.com/post/what-are-question-answering-systems/>
10. Q&A Software. Wikipedia (2023). Accessed: Nov. 05, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Q%26A_software
11. Bhatia, R.: History in the Revolution of QSAR: A Review (2023). Accessed: Nov. 05, 2023. [Online]. Available: <https://www.pharmatutor.org/articles/history-revolution-of-qsar-quantitative-structural-activity-relationship>
12. OpenAI: Embeddings. Accessed: Nov. 14, 2023. [Online]. Available: <https://platform.openai.com/docs/guides/embeddings/what-are-embeddings>
13. Koroteev, M.V.: BERT: A Review of Applications in Natural Language Processing and Understanding (2021). Accessed: Nov. 17, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/BERT%3A-A-Review-of-Applications-in-Natural-Language-Koroteev/879eaab2275a364549809560b42f0fef357ebcce>
14. Ma, X. Wang, Z., Ng, P.: Universal Text Representation from BERT: An Empirical Study (Oct 2019), [Online]. Available: <http://arxiv.org/abs/1910.07973>
15. Mikolov, T., Chen, K., Corrado, G., et al.: Efficient Estimation of Word Representations in Vector Space (Jan 2013), [Online]. Available: <http://arxiv.org/abs/1301.3781>
16. Chu, Y., Cao, H., Diao, Y., et al.: Refined SBERT: Representing sentence BERT in manifold space. Neurocomputing **555** (Oct 2023). <https://doi.org/10.1016/J.NEUCOM.2023.126453>
17. Ren, R., Zhang, J.: A New Framework to Address BIM Interoperability in the AEC Domain from Technical and Process Dimensions. Advances in Civil Engineering (2021). <https://doi.org/10.1155/2021/8824613>
18. Zheng, Z., Lu, X.-Z., Chen, K.-Y. et al.: Pretrained Domain-Specific Language Model for General Information Retrieval Tasks in the AEC Domain (2022). Accessed: Nov. 07, 2023. [Online]. Available: https://www.researchgate.net/publication/359130279_Pretrained_Domain-Specific_Language_Model_for_General_Information_Retrieval_Tasks_in_the_AEC_Domain

19. European Commission, “Standard Contractual Clauses for Data Transfers between EU and non-EU Countries
20. European Commission: Regulatory Framework Proposal on Artificial Intelligence
21. Moon, S., Lee, G., Chi, S.: Automated system for construction specification review using natural language processing. *Advanced Engineering Informatics* **51** (Jan 2022). <https://doi.org/10.1016/j.aei.2021.101495>
22. Nematzadeh, H., Enayatifar, R., Mahmud, M., Akbari, E.: Frequency based feature selection method using whale algorithm. *Genomics* **111**(6), 1946–1955 (Dec 2019). <https://doi.org/10.1016/j.ygeno.2019.01.006>
23. Guo, H., Lin, J.-R., Yu, Y.: Intelligent and Computer Technologies’ Application in Construction. *Buildings* **13**(3), 641 (Feb 2023). <https://doi.org/10.3390/buildings13030641>
24. Saka, A.B., Oyedele, L.O., Akanbi, L.A.: Conversational artificial intelligence in the AEC industry: A review of present status, challenges and opportunities. *Advanced Engineering Informatics* **55**. Elsevier Ltd. (Jan 01, 2023). <https://doi.org/10.1016/j.aei.2022.101869>
25. Cretu, C.: How Does ChatGPT Actually Work? An ML Engineer Explains. Scalable Path. Accessed: Feb. 04, 2024. [Online]. Available: <https://www.scalablepath.com/machine-learning/chatgpt-architecture-explained>
26. Ahmed, I., Kajol, M., Hasan, U., et al.: ChatGPT vs. Bard: A Comparative Study (2023). Accessed: Feb. 05, 2024. [Online]. Available: https://www.researchgate.net/publication/371799069_ChatGPT_vs_Bard_A_Comparative_Study
27. Z. Qihuang *et al.*, “Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT,” 2023, <https://doi.org/10.48550/arXiv.2302.10198>
28. Jacobi, O.: Exploring architectures and capabilities of foundational LLMs. Aporia. Accessed: Feb 04, 2024. [Online]. Available: <https://www.aporia.com/learn/exploring-architectures-and-capabilities-of-foundational-lm/>
29. Knight, W., Goode, L.: Bard (chatbot). Wikipedia. Accessed: Feb 05, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Bard_%28chatbot%29
30. Edwards, B.: ChatGPT vs Google Bard: Which is better? We put them to the test. Accessed: Feb 05, 2024. [Online]. Available: <https://arstechnica.com/information-technology/2023/04/clash-of-the-ai-titans-chatgpt-vs-bard-in-a-showdown-of-wits-and-wisdom/>
31. Keary, T.: Bard vs. ChatGPT: Which is Better & What’s the Difference? Techopedia (2024). Accessed: Feb 05, 2024. [Online]. Available: <https://www.techopedia.com/bard-vs-chatgpt-which-is-better>
32. Lau, J.: Bard vs. ChatGPT: What’s the difference? (2024). Accessed: Feb 05, 2024. [Online]. Available: <https://zapier.com/blog/chatgpt-vs-bard/>
33. López Espejel, J., Ettifouri, E.H., Yahaya Alassan, M.S., et al.: GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. *Natural Language Processing Journal* **5**, 100032 (Dec 2023). <https://doi.org/10.1016/j.nlp.2023.100032>
34. Paiva, A.R., Tasiden, T.: (12) United States Patent (2013). Accessed: Feb 17, 2024. [Online]. Available: <https://patentimages.storage.googleapis.com/53/6a/7f/940e0ac09c2cd1/US8412651.pdf>
35. Abdelkader, B., Ibrahim, B.: Contribution to decision-making in the big data industry based on the multiparametric similarity measure for Pythagorean fuzzy sets. *Journal of Logic and Computation* **33** (3), 517–535 (April 2023). Accessed: Feb. 17, 2024. [Online]. Available: <https://academic.oup.com/logcom/article-abstract/33/3/517/6647540?redirectedFrom=fulltext>
36. Silhavy, R., Silhavy, P.: Artificial Intelligence Application in Networks and Systems. In: Proceedings of 12th Computer Science On-line Conference 2023 **3** (2023). Accessed: Feb. 17, 2024. [Online]. Available: <https://doi.org/10.1007/978-3-031-35314-7>

37. Jude Chukwura Obi: A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences* **8**(1), 308–314 (Feb 2023). <https://doi.org/10.30574/wjaets.2023.8.1.0054>
38. Riyanto, S., Imas, S. S., Djatna, T., et al.: Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. *International Journal of Advanced Computer Science and Applications; West Yorkshire* **14** (6) (2023). Accessed: Feb. 17, 2024. [Online]. Available: <https://www.proquest.com/openview/0aee35afdb3b2075ac884a6ec288dbbc/1?pq-origsite=gscholar&cbl=5444811>



Data Mining and Feature Selection in Large Hospital Central Heating Plant

Marjan FatehiJananloo, Helen Stopps, and J. J. McArthur^(✉)

Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada

j.jmcARTHUR@torontomu.ca

Abstract. In the face of anthropogenic climate change, with buildings contributing to roughly one-third of CO₂ emissions, optimizing boiler operations is essential. This is particularly pertinent in hospitals in cold climates, where high heating loads are driven by the need for 100% outdoor air in ventilation systems, coupled with strict temperature and humidity controls. A major challenge lies in existing facilities, where older equipment often lacks integration with the Building Automation System (BAS), having few accessible measurement points outside their proprietary boiler controllers. Addressing this, our paper introduces a data mining and feature selection approach to aid the development of grey-box emulators and optimization algorithms for complex, existing central heating plants. We utilize a real-world building to develop a dataset reflective of typical plants. This includes analyzing BAS points, weather data, and variables from physical models, leading to a dataset for feature selection and engineering. The study finds Random Forest Regression slightly superior to Lasso Regression for feature selection, effectively capturing the dynamics of the central heating plant and boiler system. We discuss the key features for predicting Return Water Temperature and forecasting Thermal Power, vital for assessing heating system efficiency and performance. Our results show that Random Forest Regression adeptly predicts Thermal Power and Return Water Temperature with the selected features, despite some challenges in predicting extreme temperatures. Testing the model's accuracy without significant predictors like VFD Speed and Flow Rate highlights the predictive value of Hour of the Day and Day-type for occupancy patterns. Future research will leverage these features to develop online optimization models for the central plant system. The goal is to simultaneously reduce CO₂ emissions and energy costs in the studied hospital.

Keywords: Feature Engineering · Random Forest Regression · Lasso Regression · Heating Plant Modeling · Energy Efficiency

1 Introduction

1.1 Global Trends in Energy Efficiency

There is currently a worldwide effort to decrease energy consumption and reduce emissions production (Liang et al. 2023). Buildings account for about 40% of total energy consumption in developed nations (Tian et al. 2020). Compared to other types of buildings of similar size, hospitals have markedly higher energy requirements (Bawaneh et al. 2019). The continuous functioning of heating, ventilation, and air conditioning (HVAC) systems in hospitals is essential and constitutes a major portion of their energy consumption (Coccagna et al. 2017). In the context of hospital energy management, particular attention must be paid to the efficiency of boilers. Boilers are a critical component of hospital HVAC systems, responsible for providing the necessary heating to ensure a comfortable and sterile environment (Shen et al. 2019).

Typically, HVAC equipment in mixed-buildings like hospitals are designed to meet peak demand, leading to scenarios where they operate inefficiently under normal conditions (Fraile et al. 2014). The role of commissioning services is often to mitigate this overdesign by conducting tests, balancing systems, and establishing equipment setpoints at different stages post-construction. However, most mixed-use buildings do not benefit from continuous commissioning. They are typically equipped with a set of operating setpoints established during the design phase and initial occupancy, but these often remain unchecked during the prolonged operation of the building (Stock et al. 2021).

The advent of Artificial Intelligence (AI) in building management systems offers a transformative solution to these challenges. By utilizing AI and machine learning algorithms, it is possible to analyze vast amounts of operational data, enabling more intelligent and adaptive control of boiler systems (Nemitaiah et al., 2023). This approach goes beyond traditional reactive maintenance, offering predictive insights that can optimize performance, reduce energy consumption, and extend equipment life (Saloux et al. 2023). The application of AI in this domain involves continuous monitoring, decisions on predictive maintenance and fault detection, along with controlling boilers, especially in the context of intensive peak demand (Nemitaiah et al., 2023). Data-driven predictive maintenance and fault detection in boilers utilize machine learning algorithms and various AI methods to examine sensor data from the boiler, predicting the necessity for maintenance or intervention (Prabhu and Chaudhary 2021).

This paper presents the feature selection methods and findings to support the optimization of a central heating plant for a case study hospital in ASHRAE Climate Zone 5. This contribution is important for IT in building engineering as it uses real-world operational data obtained from a building automation system, thus enabling its replication across the majority of existing buildings. By focusing on the central plant and its impact on the hospital as a whole, this study provides a comprehensive view of energy usage and demand patterns without the need for additional boiler instrumentation. Such a perspective is vital for developing effective energy management strategies in hospitals.

1.2 Background

Recent studies have demonstrated the effectiveness of AI and machine learning in enhancing boiler operations. For instance, the implementation of model predictive control (MPC) strategies using heating load forecasting and boiler performance models has shown potential in reducing greenhouse gas emissions in multi-boiler systems (Saloux et al. 2023). Similarly, the application of evolutionary computing and intelligent optimization algorithms such as genetic algorithms (GAs) or support vector machines (SVMs) has been identified as a promising direction for boiler combustion optimization, particularly in energy saving and emission reduction (Liang et al. 2023). The importance of data reliability and the appropriate selection of AI models are emphasized in studies focusing on boiler optimization for efficient and clean energy production (Nemitallah et al. 2023). This is complemented by the innovative use of Digital Twins in the power industry, which illustrates the potential of AI in learning from operational data to enhance the economic efficiency of power plants (Aiki et al. 2018).

In order to develop appropriate AI models for boiler plant optimization, feature selection is critical. This supports the use of historical process data for real-time monitoring of boiler efficiency and the identification of process states based on variable ranking, which has been shown to be effective in performance enhancement (Nikula et al. 2016). However, the challenge of incorporating design variables in the development of energy-efficient models has also been addressed, highlighting the importance of identifying key design features critical for predicting energy consumption (Tian et al. 2020). In many real-life situations, many variables are not measured and thus must be inferred or calculated using physical principles. As a result, boiler optimization for existing plants is particularly challenging. In this paper, we focus on identifying and evaluating key features for accurate thermal power and return water temperature prediction in the central heating plant of a large hospital, using machine learning techniques.

In this study, we integrate a comprehensive analysis of data from Building Automation Systems (BAS), weather patterns, and physical models to develop a predictive model. The primary objective of this research is not only to enhance our understanding of hospital boiler operations but also to provide practicable insights for energy management. By doing so, we aim to contribute to the creation of a digital twin for boilers, a sophisticated tool that can significantly optimize heating systems in healthcare facilities. The anticipated outcome is a model that predicts essential operational parameters with high accuracy and adapts to varying data availability, making it a versatile solution for a wide range of hospital settings. This research is anticipated to establish a foundation for more sustainable and cost-effective energy practices in healthcare facilities.

2 Datasets Overview and Boiler Plant Description

A case study hospital in British Columbia was used for this study and access to the BAS system was granted to permit real-time and limited trended data to be used for model development. The hospital campus is served by a central plant, providing chilled water, heating hot water, and process steam. Heating hot water is provided by two hot water condensing boilers (B04 and B05) and a back-up steam heat exchanger (HE-3) (see **Error! Reference source not found.**) on a variable flow primary loop. This loop is

served by a triplex pump system (P01 to P03) and interfaces with a number of secondary loops to the various hospital campus buildings. Limited gas submetering is also available, grouping the steam boilers alongside the condensing boilers together on a single meter. The full set of BAS shop drawings were reviewed to understand the heating system control sequences and these informed the feature selection, as noted in the following section.

To further challenge model creation, it was determined through discussions with the controls contractor that several measured points were unreliable, most notably the individual flow rate meters for the two boilers. These could not be installed at suitable locations due to site constraints and could not even be relied upon for relative measurements. Fortunately, it was confirmed that the master boiler controller runs both boilers at equal flow in high heating demand situations, so these flows could be assumed equal when both boiler status indicators were on (Fig. 1).

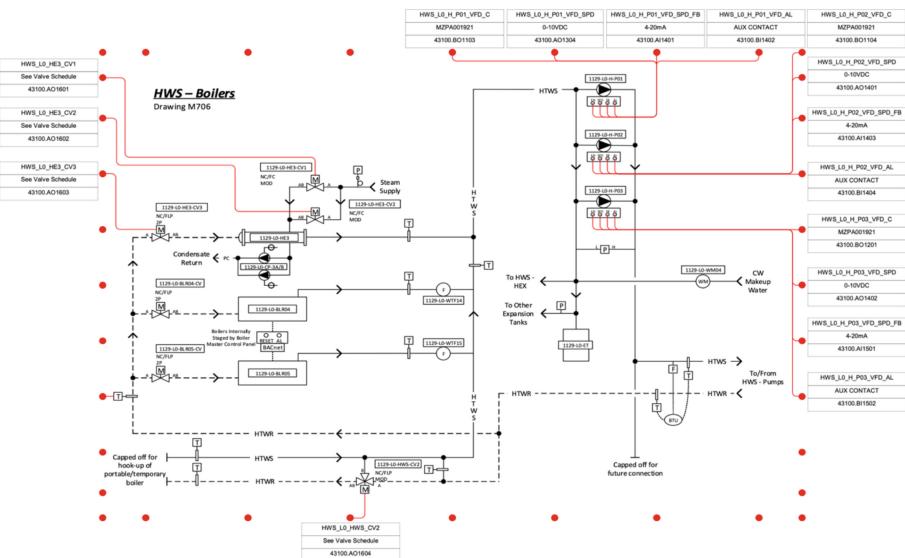


Fig. 1. Schematic from BAS shop drawings for heating hot water system serving the various hospital buildings.

3 Methodology

In this section, we outline the systematic approach taken to analyze the hospital's heating system. Our methodology encompasses various stages, from data preprocessing to feature selection and model development. Each stage is designed to extract meaningful insights from the data, ensuring a comprehensive understanding of the system's performance and dynamics.

3.1 Data Preprocessing

In the preliminary phase of data preprocessing, we addressed the inconsistencies in sample times within our raw data. Our objective was to standardize these timestamps, initially captured at approximately 10-min intervals, to maintain temporal uniformity across the dataset. This standardization was crucial for subsequent analyses and was executed by employing a Python round function, rounding each sample time to its nearest 10-min mark. Notably, this adjustment had minimal impact as most of the heating system measurements were recorded within a 600-s timeframe of each other. Subsequently, we delved into categorizing operational schedules of the hospital's various departments. This step involved defining day-type categories, such as weekdays, weekends, and holidays, along with their specific time segments within the daily 24-h cycle. This categorization is a fundamental aspect of feature engineering in temporal data analysis, recognizing the significant influence of day types on pattern formation (Zor, Çelik, Timur, & Teke, 2020). Integrating this information as a feature in our dataset allowed us to infer hospital occupancy trends more accurately, which are pivotal in understanding energy consumption patterns.

The dataset initially lacked direct measurements of energy generated by the boilers. To address this, the sensible heat transfer from the boilers to the primary loop was calculated using the specific heat formula, as indicated in Eq. 1:

$$Q = MC\Delta T \quad (1)$$

where Q represents the energy transferred (in joules), M is the mass flow rate of water (in kilograms per second), C is the specific heat capacity of water (approximately 4.186 J/gram °C), and ΔT is the temperature difference between the supply and return water (in degrees Celsius). A critical aspect of our analysis involved understanding the delay in return water temperature's response to changes in supply water temperature. To determine this time delay, we applied Dynamic Time Warping (DTW), as a technique widely used for aligning sequences with temporal shifts (Hsu et al. 2015). Our analysis revealed that, given the 10-min interval of our data, a zero-delay assumption was appropriate for our calculations, as reflected in Eq. 1.

For missing values within our dataset, we implemented a forward-fill method, which is a common technique in time series data where subsequent values are used to fill gaps (Kamalov and Sulieman, 2021). This approach is particularly effective in maintaining continuity in time-indexed data. An additional challenge we encountered was the impact of daylight-saving time changes on our dataset. To maintain consistency and ensure the integrity of our time series data, we developed a specific correction mechanism. This mechanism involved adjusting the timestamps of our data points to account for the one-hour shift. During the spring forward transition, we added an hour to the dataset to compensate for the lost time, while in the autumn backward shift, we removed the duplicate hour to correct the overlap. These adjustments ensured that our time series data remained continuous and accurately reflected the actual hours of operation, despite the time shifts.

Finally, to ensure comparability between the different scales of our data points, we applied the Min-Max Scaling method. This normalization involved scaling each feature

to a range between 0 and 1, as described in Eq. 2:

$$\text{Normalized Value} = \frac{\text{Value} - \text{Min}}{\text{Max} - \text{Min}} \quad (2)$$

This scaling was essential for harmonizing the different units and magnitudes of our variables, enabling a more balanced and unbiased analysis. In line with this processed dataset, we utilized specific key variables critical and available to our study. Table 1 outlines these data points, providing clarity on their role within the datasets:

Table 1 Overview of Key Data Points in the Heating System Analysis

| Data Point | Description |
|---|---|
| Return Water Temperature (RT) | Temperature of water returning to the boilers |
| Supply Temperature (ST) | Temperature of the supply water |
| Flow Rate (FR) | The overall flow rate of water in the system |
| Average Supply Temperature (AVS) | The average temperature of the water supplied by the boilers |
| Pump VFD Speed (VFD) | Speed of the Variable Frequency Drives for Pumps 1, 2, and 3 |
| Outdoor Air Temperature (OAT) | Air temperature outside the hospital |
| Hour of the Day | Time of the day categorized into 24-h segments |
| Day-Type | Categorization of days into weekday, weekend, or holiday |
| Thermal Power (THP) | Calculated thermal power output of the boilers |
| Humidity (H) | Ambient humidity levels, which can influence heating efficiency |

An important aspect of our dataset involves the Variable Frequency Drive (VFD) speeds for the triplex pump system, consisting of Pumps P01, P02, and P03. In our analysis, we focused on the total VFD speed, combining the speeds of all three pumps. This decision was informed by the operational pattern of these pumps, where they were alternated biweekly with one pump always in standby mode. In high heating demand situations, the master boiler controller ensures that both active pumps run at equal flow, allowing us to assume a cumulative flow rate for our calculations. By aggregating the speeds of P01, P02, and P03, we could more accurately reflect the overall contribution of the pumping system to the heating process, despite the limitations in individual flow rate measurements for each pump. Finally, we applied one-hot encoding to the ‘Daytype’ feature. This transformation converted the categorical data, representing different types of days, into a numerical format, making it more conducive for analysis in machine learning models.

3.2 Feature Selection and Model Development

Following the completion of our data preprocessing phase, we shifted our focus to analysis, which included several key steps. Initially, the data was filtered into three distinct scenarios based on operational conditions: one scenario with only Boiler 4 active, another with only Boiler 5 active, and a third scenario where both boilers were operational. This allowed us to examine the heating system's performance under varied conditions. It was noted that the heat exchanger was not in use as a backup during the data collection period, thereby excluding it from these scenarios.

For the critical task of feature selection, we employed two machine learning models: Lasso Regression and Random Forest. These models were specifically chosen for their capacity to identify the most relevant features while mitigating issues like overfitting. In the next phase of our methodology, we concentrated on creating predictive models for return water temperature and thermal power. This step involved selecting the most effective model from our feature selection process. The chosen model demonstrated superior performance in handling the complexities of predicting these specific parameters. This step was crucial in demonstrating the effectiveness of our feature selection process.

4 Results

Following the steps outlined in our methodology, this section presents the key findings of our analysis. We focus on illustrating the predictive accuracy of our models in the different operational scenarios of the boilers, as well as the overall performance of different algorithms. The results shed light on the practical implications of our feature selection, particularly highlighting how the variables like Return Water Temperature and Thermal Power behave under varying operational conditions. These findings are instrumental in not only validating the effectiveness of our modeling approach but also in enhancing our understanding of the hospital's heating system from a data-driven perspective.

4.1 Comparative Analysis of Feature Selection Algorithms

In this part of our Results section, we explore a comparative analysis of the feature selection performance by two machine learning algorithms: Lasso Regression and Random Forest. The effectiveness of these algorithms was evaluated based on their mean square error (MSE) across three distinct operational scenarios of the boilers, focusing on two key target variables: Return Water Temperature and Thermal Power. This comparison provides a clear understanding of each model's accuracy in feature selection and prediction. Although both algorithms demonstrated valuable insights, the Random Forest model showed a slightly superior performance. Hence, we will present a detailed view of the features identified as most significant by this algorithm. Table 2 summarizes the Mean Square Error of our two proposed methods, offering a quantitative perspective on their performance across different scenarios and target variables.

Table 2 Comparative Mean Square Error (MSE) of Lasso Regression and Random Forest Regression

| Scenario | Target Variable | LR (MSE) | RF (MSE) |
|----------------------------|-------------------------------|------------------------|------------------------|
| Boiler 4 Active | Return Water Temperature (C°) | 0.0042 C° ² | 0.0011 C° ² |
| Boiler 4 Active | Thermal Power (kJ) | 0.0050 kJ ² | 0.0012 kJ ² |
| Boiler 5 Active | Return Water Temperature (C°) | 0.0074 C° ² | 0.0033 C° ² |
| Boiler 5 Active | Thermal Power (kJ) | 0.0044 kJ ² | 0.0011 kJ ² |
| Both Boilers Active | Return Water Temperature (C°) | 0.0041 C° ² | 0.0018 C° ² |
| Both Boiler Active | Thermal Power (kJ) | 0.0083 kJ ² | 0.0022 kJ ² |

4.2 Feature Importance in Random Forest Regression Model

Following the comparative analysis of Lasso Regression and Random Forest algorithms, we now focus on the feature selection results from the Random Forest model. Given the similarity in the boiler systems, the findings for Boiler 4 provide a representative understanding of the feature importance, which closely mirrors the patterns observed in other scenarios. Table 3 presents the coefficients of each feature in the Random Forest Regression model, highlighting their relative importance in predicting Return Water Temperature and Thermal Power. This approach enables us to extract key insights within the scope of our study, offering a detailed view of the most influential factors affecting the heating system's performance, specifically under the operational conditions of Boiler 4.

Table 3 Feature Importance derived from Random Forest Regression for Boiler 4

| Feature | Importance in Predicting
Return Water Temperature | Importance in Predicting
Thermal Power |
|--------------------------------|--|---|
| Outdoor Air Temperature | 0.5339 | 0.4102 |
| Flow Rate | 0.2410 | 0.3802 |
| Pump VFD Speed | 0.1716 | 0.3684 |
| Supply Temperature | 0.1073 | 0.2045 |
| Humidity | 0.0749 | 0.0240 |
| Hour of the Day | 0.0497 | 0.0240 |
| Weekday | 0.0035 | 0.0016 |
| Weekend | 0.0032 | 0.0015 |
| Holiday | 0.0008 | 0.0007 |

4.3 Analysis of Model Accuracy Using Selected Features

Expanding upon the insights gained from feature importance in the Random Forest model, this subsection evaluates the accuracy of our predictive model for Return Water Temperature. Initially, we incorporated all identified features, which demonstrated commendable accuracy. Notably, the VFD speed, closely related to the flow rate, emerged as a significant predictor, reflecting the occupancy pattern of the hospital. For instance, higher VFD speeds during peak hospital hours indicate increased heating demand. However, recognizing the potential limitation in data availability, we also evaluated the model's performance by selectively excluding VFD speed and flow rate. This scenario simulates conditions where such detailed operational data might not be accessible, relying instead on more commonly available variables like Hour of the Day and Day type (Weekday, Weekend, Holiday). Although this modified model exhibited a reduction in accuracy, it still provided valuable predictive capability, underlining the robustness of the model under varied data availability scenarios. The following table presents the Mean Square Error (MSE) of our model in different scenarios, reflecting the impact of feature inclusion on model accuracy (Table 4).

Table 4 MSE of Predictive Model in Different Scenarios

| Scenario | MSE ($^{\circ}\text{C}^2$) with All Features | MSE ($^{\circ}\text{C}^2$) Without VFD Speed and Flow Rate |
|----------------------------|--|--|
| Boiler 4 Active | 0.0011 | 0.0020 |
| Boiler 5 Active | 0.0033 | 0.0044 |
| Both Boilers Active | 0.0018 | 0.0026 |

4.4 Visualization of Seasonal Predictions

To complement our quantitative analysis, we have included graphical representations that illustrate the model's predictive accuracy across various seasonal periods (Fig. 2). These figures, each representing a different season with a unique color scheme, demonstrate the model's predictive consistency and accuracy in relation to actual temperature readings. The results for Boiler 4 are presented, reflecting a representative pattern of the heating system's behavior which is consistent with trends observed in other operational scenarios. Upon analyzing these figures, it is evident that while the model achieves a general consistency with the actual temperatures, there are noticeable deviations at the extremes. Specifically, the model appears less accurate at predicting the higher and lower ranges of the return water temperature, as indicated by the disparities between the actual and predicted values during these periods. Despite these outliers, the overall fit of the model remains robust across different seasons, without a visible disparity in performance between them. The figures do not show a clear differentiation between seasons, suggesting that the model's performance is uniformly reliable irrespective of seasonal variations. The y-axis of the figures denotes the return water temperature, and the x-axis

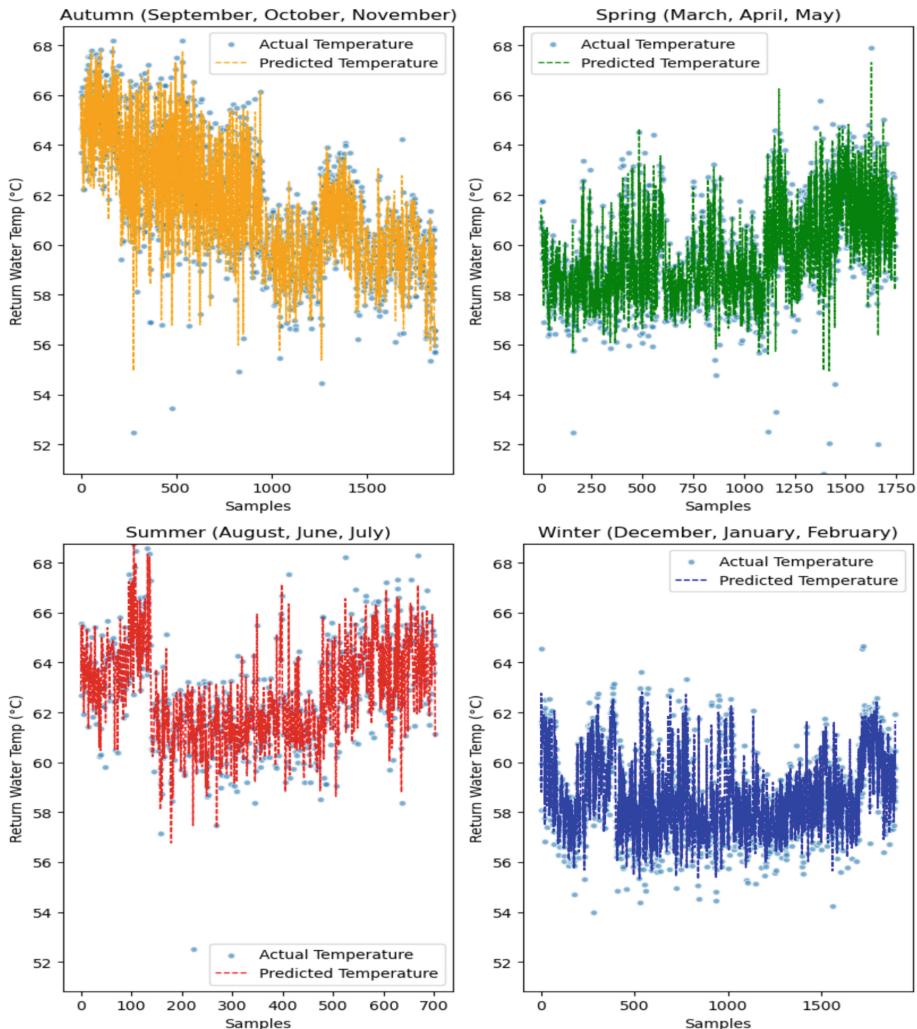


Fig. 2. Seasonal Variation in Return Water Temperature for Boiler 4 - Comparison of Actual and Predicted Values. Each panel represents a different season as denoted by color: autumn in orange, spring in green, summer in red, and winter in blue. The y-axis measures the return water temperature ($^{\circ}\text{C}$), while the x-axis corresponds to sequential sample data points. The dashed lines represent the predicted temperatures from the model, illustrating the model's performance against the actual recorded temperatures represented by solid dots.

represents a sequence of sample data points, captured at regular intervals. It's essential to clarify that the x-axis does not directly correspond to specific times or months, but rather to a sequence of data points that follow the temperature trends through the seasons. These visualizations underscore the model's strength in capturing the cyclical nature of the heating system's operation, yet also highlight areas where the predictive accuracy could be further refined, particularly for temperatures at the extremities of the range.

5 Conclusion

This study provides valuable insights into optimizing heating systems within hospital environments, highlighting the potential of machine learning algorithms to create effective models based on the limited data. By employing feature selection techniques and model development processes, we have identified critical variables that influence the operation of hospital boilers, particularly focusing on the central heating plant of a case study hospital in British Columbia. The comparative analysis of feature selection algorithms revealed that Random Forest outperformed Lasso Regression in terms of mean square error, underscoring its suitability for modeling complex systems like hospital HVAC.

A pivotal aspect of our research involved evaluating model performance with and without the inclusion of VFD Speed and Flow Rate. The results indicate that while these features are highly informative—reflecting the hospital's occupancy patterns—predictive accuracy can still be maintained with alternative data. Specifically, Hour of the Day and Daytype have proven to be effective proxies for occupancy, enabling meaningful predictions of Return Water Temperature even in the absence of detailed operational data. This adaptation underscores the versatility of our model to accommodate data constraints, providing a reliable alternative for facilities where comprehensive data collection is not feasible. The visualizations of seasonal predictions for Boiler 4 highlighted the model's general accuracy but also revealed difficulties in predicting extreme temperature values. This finding suggests an area for future research to develop models that can more accurately predict temperature at these critical ranges, which are important for system efficiency and reliability.

In conclusion, this study exemplifies the practical application of AI techniques in the realm of building engineering, offering a scalable solution for energy management in healthcare facilities, highlighting a path forward for enhanced operational efficiency in hospitals and similar environments.

Acknowledgments. This research was funded by the Natural Science and Engineering Research Council (Canada) [ALLRP-580958-22 and CREATE 510284-2018] and H. H. Angus and Associates Consulting Engineers Ltd.

References

- Aiki, H., Saito, K., Domoto, K., Hirahara, H., Obara, K., Sahara, S.: Boiler digital twin applying machine learning. *Mitsubishi Heavy Industries Technical Review* **55**(4), 1 (2018)
- Bawaneh, K., Farnaz, G.N., Rasheduzzaman, M., Dekem, B.: Energy Consumption Analysis and Characterization of Healthcare Facilities in the United States. *Energies* **12**, 3775 (2019)
- Coccagna, M., Cesari, S., Valdiserri, P., Romio, P., Mazzacane., S., Coccagna, M. . Energy consumption in hospital buildings: functional and morphological evaluations of six case studies. *International Journal of Environmental Science* **2** (2017)
- Fraile, J.-C., San-José, J., Ana, G.-A.: A boiler room in a 600-bed hospital complex: study, analysis, and implementation of energy efficiency improvements. *Energies* **7**, 3282–3303 (2014)
- Hsu, C.-J., Huang, K.-S., Yang, C.-B., Guo, Y.-P.: Flexible dynamic time warping for time series classification. *Procedia Computer Science* **51**, 2838–2842 (2015)

- Kamalov, F., Sulieman, H.: Time series signal recovery methods: comparative study. In: 2021 International Symposium on Networks, Computers and Communications (ISNCC)(IEEE), 1–5 (2021)
- Liang, J., Gau, H., Chen, K., Yu, K., Caitong, Y., Yunpeng, M.: A Survey on Intelligent Optimization Approaches to Boiler Combustion Optimization. *CAAI Artificial Intelligence Research* **2** (2023)
- Nemitallah, M., Nabhan, M., Alowaifeer, M., et al.: Artificial intelligence for control and optimization of boilers' performance and emissions: A review. *Journal of Cleaner Production*, 138109 (2023)
- Nikula, R.-P., Ruusunen, M., Leiviskä, K.: Data-driven framework for boiler performance monitoring. *Applied energy*, 1374–1388 (2016)
- Prabhu, V., Chaudhary, D.: Machine Learning enabled Condition Monitoring Models for Predictive Maintenance of Boilers. In: *2021 4th International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE)* (pp. 426–430). IEEE (2021)
- Saloux, E., Runge, J., Zhang, K.: Operation optimization of multi-boiler district heating systems using artificial intelligence-based model predictive control: Field demonstrations. *Energy* **285**, 129524 (2023)
- Shen, C., Kang, Z., Jian, G., Qingli, Z.: Analysis of building energy consumption in a hospital in the hot summer and cold winter area. *Energy Procedia* **158**, 3735–3740 (2019)
- Stock, M., Kandil, M., Macleod, B., Pilgrim, B., McArthur, J.: HVAC performance evaluation and optimization algorithms development for large buildings. *Build. Simul.* **2021**(17), 3080–3087 (2021)
- Tian, Z., Wei, S., Shi, X.: Developing data-driven models for energy-efficient heating design in office buildings. *Journal of Building Engineering* **32**, 101778 (2020)
- Xu, J., Cui, Z., Ma, S., Wang, X., Zhang, Z., Zhang, G.: Data-Driven Based Digital Twin for Operational Performance Optimization in the Cfb Boiler. Available at SSRN 4578796 (n.d.)
- Zor, K., Çelik, Ö., Timur, O., Teke, A.: Short-Term Building Electrical Energy Consumption Forecasting by Employing Gene Expression Programming and GMDH Networks. *Energies* **13**(5), 1102 (2020)

Automation and Robotics for Construction



Automated Inspection in Mechanical and Electrical Construction Leveraging OpenVSLAM and Instance Segmentation

Wei Wei¹, Yujie Lu^{2(✉)}, Lijian Zhong¹, Yufan Chen¹, Ruihan Bai¹, and Yijun Lin¹

¹ College of Civil Engineering, Tongji University, Shanghai 200092, China

² Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai 200092, China

Lu6@tongji.edu.cn

Abstract. Mechanical and electrical construction involves a multitude of components with extended lengths and intricate spatial arrangements, potentially exacerbating construction challenges and the likelihood of installation deviations. To optimize construction management efficiency, we proposed a framework for automated mechanical and electrical construction inspection leveraging panoramic simultaneous localization and mapping (SLAM) and instance segmentation. First, we extracted the coordinates of all panoramic images along the inspection path using the OpenVSLAM algorithm. Second, the PointRend algorithm was employed to segment the mechanical and electrical construction components. Finally, the installation discrepancies were automatically identified by comparing segmentation images of real scenes (as-built) with virtual images in Unreal Engine (as-planned). This framework was successfully deployed in a mechanical and electrical construction project in Shanghai, China, demonstrating precise image segmentation and accurate identification of installation discrepancies. The study establishes a research paradigm for vision-based intelligent construction inspection, which can be further expanded to encompass large-scale indoor construction progress tracking and quality monitoring.

Keywords: Construction Inspection · Instance Segmentation · Mechanical and Electrical Construction · OpenVSLAM · Unreal Engine

1 Introduction

The AI-based intelligent inspection method has greatly improved the efficiency of construction management. Numerous intelligent methods have been applied to indoor mobile inspection, such as visual camera inspection and VR inspection. By integrating computer vision technology into monocular or binocular cameras, key construction elements in inspection images, such as plastering, walls, holes, cracks, etc., can be automatically identified to determine construction progress or quality defects [1, 2]. Virtual reality (VR) technology provides an efficient and feasible method for indoor inspection. By aligning the as-planned 3D virtual model with the real space, VR devices can be worn to perform virtual-to-real visual comparisons, manually annotate objects, take measurements, identify installation deviations, and so on [3, 4].

However, the above methods still have certain limitations. In the mechanical and electrical construction scenario, there are long and slender pipeline components with large longitudinal spans, while monocular and binocular vision devices have limited fields of view, making it difficult to capture complete component information. In addition, methods based on virtual engines lack semantic information and still rely on manual judgment and selection fundamentally.

To achieve efficient indoor inspection over a large area, this study proposed an intelligent indoor electromechanical construction inspection method that integrates semantic segmentation algorithms into panoramic SLAM inspection to automatically identify mechanical and electrical construction components. By comparing real image segmentation results with virtual model images, installation deviations are automatically identified. This study provides a research paradigm for vision-based large-scale, wide-angle indoor construction inspection and serves as a practical reference for intelligent construction quality monitoring and progress tracking.

2 Related Studies

2.1 Simultaneous Localization and Mapping

SLAM can be divided into two types: direct method and indirect method. The direct method, such as LSD-SLAM [5] and DSO [6], estimates camera motion and scene depth information using the grayscale gradients of the entire image region. This method utilizes pixel-level information from the image without explicitly extracting any key points, making it adaptable to scenarios like outdoor inspection with repetitive textures and good lighting conditions. However, it is sensitive to variations in image lighting and noise. On the other hand, the indirect method, exemplified by OpenVSLAM [7], extracts key points and corresponding descriptors (e.g., ORB features) from the image through feature extraction and descriptor calculation. Camera motion and map point positions are then estimated based on the location information of these key points. This method exhibits strong robustness against disturbances such as lighting changes and noise in indoor inspections. Moreover, the indirect method is computationally more efficient compared to the direct method as it only needs to process a small number of feature points rather than every pixel.

2.2 Image Segmentation

Image segmentation algorithms can be divided into instance-based and pixel-based segmentation algorithms based on segmentation strategies. Instance-based segmentation algorithms can precisely locate and segment each instance target. However, this method tends to have lower accuracy in identifying targets with blurry boundaries, leading to potential issues of blurred boundaries. Mask R-CNN [8] is a classic two-stage instance segmentation algorithm, which is based on the Faster R-CNN [9] object detection framework and further extracts segmentation masks for multiple instances on top of object detection. SOLO [10] is a single-stage instance segmentation algorithm that transforms the segmentation task into a pixel classification and position encoding task, achieving

efficient instance segmentation. Instance-based segmentation algorithms are suitable for scenarios with a large number of instances, requiring fine-grained segmentation and identification at the individual level.

Pixel-based segmentation algorithms use strategies such as image rendering to perform semantic segmentation from the perspective of obtaining overall semantic information. PointRend [11] is an improved semantic segmentation model that enhances segmentation boundary accuracy and detail perception through a point-by-point segmentation refinement strategy. Pixel-based segmentation algorithms demonstrate noticeable effectiveness in handling complex semantic image segmentation tasks, suitable for tasks requiring high-level semantic understanding and high precision in boundary segmentation. In this study, we attempted to determine component installation status through mask area calculation and comparison. Therefore, using pixel-based segmentation algorithms to improve boundary segmentation accuracy is crucial.

3 Methods

3.1 Overall Framework

This study aims to identify key construction components through indoor panoramic inspection to determine the mechanical and electrical installation status. Furthermore, it contrasts with the completed virtual model to identify mechanical and electrical installation deviations. The research framework consists of three parts (Fig. 1):

Panoramic mobile inspection. We conducted indoor panoramic inspection using OpenVSLAM, to collect panoramic videos of on-site mechanical and electrical installations, and extract camera movement trajectories. (2) Real image segmentation. We utilized the PointRend algorithm for real image segmentation and component recognition to determine the on-site installation status. (3) Installation deviation detection based on the virtual-real comparison. First, we imported the BIM model into UE4 and extracted virtual image segmentation results as ground truth. Subsequently, the installation deviations of mechanical and electrical construction were identified through real and virtual image registration and comparison.

3.2 Panoramic Mobile Inspection Based on OpenVSLAM

OpenVSLAM [7] is compatible with various camera models, such as monocular cameras, panoramic cameras, and stereo-vision cameras. OpenVSLAM mainly consists of three core modules:

The tracking module is responsible for feature extraction, pose estimation, and keyframe detection. This module utilizes Visual Odometry technology to track camera motion by estimating the relative motion between consecutive frames through feature point matching. When a new keyframe is detected, this module initiates triangulation of map points and adds the feature point to the map.

The mapping module is primarily used for 3D point reconstruction and optimization of local poses and 3D point information. This module receives camera pose information from the tracking module and optimizes the geometric and topological structure of the

map using techniques such as triangulation and Bundle Adjustment. It also adds new keyframes and map points to the map.

The global optimization module performs loop detection and global optimization. This module conducts global Bundle Adjustment optimization for keyframes and map points in the map. Additionally, it handles loop detection and loop correction to correct positioning errors and map drift issues, thereby improving the stability and accuracy of the SLAM system. These three modules work together to enable OpenVSLAM to achieve real-time localization and map construction.

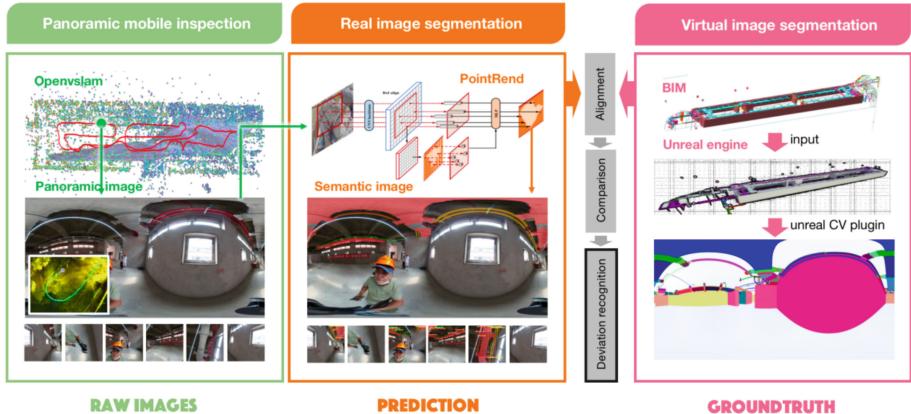


Fig. 1. Overall framework.

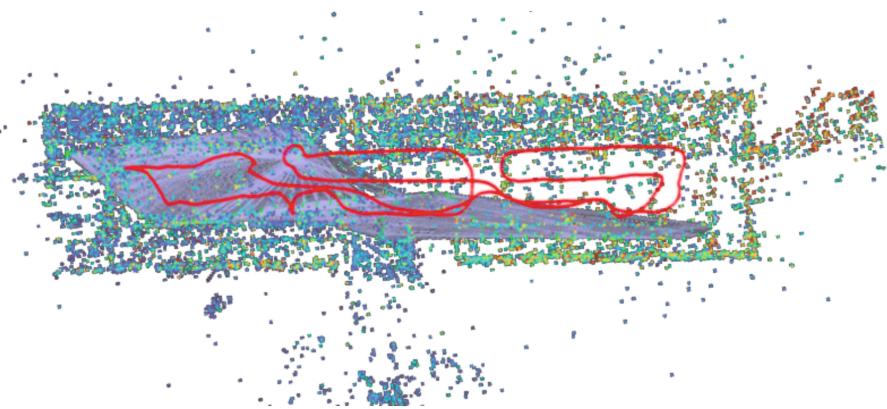


Fig. 2. OpenVSLAM.

3.3 Real Image Segmentation Based on PointRend

The PointRend [11] algorithm treats instance segmentation as a pixel-level image rendering problem. Its core idea is to first perform coarse mask prediction on features within

predicted object detection boxes and then select points with high uncertainty in boundary features for fine-grained feature prediction. By gradually constructing features on selected points, precise edge prediction is achieved. The PointRend algorithm consists of three core modules:

Point Selection Strategy: This module initially conducts coarse feature prediction on the image, selecting points with high uncertainty at the mask edges, without the need to predict all feature points.

Point-wise Feature Representation: This module performs fine-grained feature prediction on the points selected by the Point Selection Strategy. By combining low-level features (fine-grained features) and high-level features (coarse prediction), precise features are constructed at the selected mask edge points.

Point Head Module: This module aims to represent features for each selected point, which predicts segmentation labels for each point using multi-layer perceptrons.

3.4 Installation Deviation Detection Based on Virtual-Real Comparison

3.4.1 Scale Restoration of Inspection Path and Coordinate Initialization

Since panoramic cameras lack depth information, the OpenVSLAM path does not possess real spatial information. We calculated the ratio of lateral and longitudinal distance spans between the actual inspection path and the OpenVSLAM path to determine the scaling factor, thereby restoring the SLAM inspection path to its true scale.

By default, the starting position and direction of the inspection path serve as the origin and coordinate system of OpenVSLAM. We set the inspection starting point at the origin of the BIM model (referred to as the starting point in Fig. 6) and ensured alignment between the OpenVSLAM coordinate direction and the BIM coordinate system. Simultaneously, we designed the inspection path to pass through two calibration points with known real coordinates (as shown in Fig. 6) and return to the starting point to form a closed loop after inspection completion. The deviation between the predicted coordinates of OpenVSLAM and the real coordinates can be computed to evaluate the positioning accuracy of OpenVSLAM.

3.4.2 Generation of Virtual Image Segmentation Based on Unreal Engine

We converted the Revit model and imported it into Unreal Engine 4 (UE4). By using the FusionCamera Actor, we can capture panoramic roaming views of the virtual model. By comparing real panoramic images with virtual panoramic images taken from the same location and viewpoint, we can identify mechanical and electrical installation deviations.

The position information of the panoramic camera is controlled by three coordinates: x, y, and z. We ensure that the initial coordinate system and origin of UE4 are consistent with those of OpenVSLAM, allowing us to locate the corresponding positions in UE4 based on the inspection path coordinates of OpenVSLAM. The orientation of the panoramic camera is determined by the pitch, roll, and yaw angles. Due to the gyroscope, the pitch and roll remain at 0, with only the yaw changing. To obtain the observation position and viewpoint in UE4 corresponding to a certain point of the panoramic inspection, we utilized the functionality of the FusionCamera Actor to navigate the virtual

camera in UE4 to an observation point with the same position on the OpenVSLAM inspection path. Then, by adjusting the yaw angle, we ensured that the virtual camera maintained the same observation viewpoint as the inspection image, thereby obtaining real and virtual panoramic images with identical positions and observation viewpoints.

We utilized the Unreal CV plugin to automatically extract semantic information of panoramic images in the virtual engine (Fig. 3 (a)). At this point, each component has a unique RGB value, resembling the form of instance segmentation of the ground truth. Furthermore, through operations such as component RGB extraction, component mask boundary extraction, and categorization of components of the same type, we encoded components of the same type with the same colour, obtaining the form of semantic segmentation of the ground truth (Fig. 3 (b)).

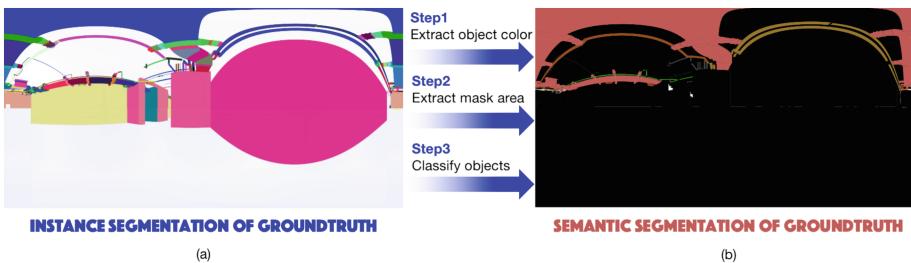


Fig. 3. Conversion from instance segmentation to semantic segmentation of ground truth. (a) Instance segmentation of ground truth; (b) semantic segmentation of ground truth.

3.4.3 Installation Deviation Detection Based on Virtual-Real Comparison

We divided the real and virtual panoramas into six single viewpoints: top, bottom, left, right, front, and back (Fig. 4). Then, we applied the PointRend algorithm to identify key construction components in the six single viewpoints of the real panorama: cable tray, fire hydrant pipeline, exhaust ventilation duct, HVAC duct, water pipe, and distribution box.

We conducted mask analysis on the six single perspectives of both real and virtual panoramas: First, we identified the mask categories in each perspective image and calculated the pixel percentage for each mask category relative to the entire image. Then, we performed a pairwise comparison between the six single perspectives of the real and virtual panoramas to determine if there were any installation deviations present in each perspective.

Taking the top perspective as an example, if there is a discrepancy in component or mask categories between the real image (prediction) and the virtual image (ground truth), then there is an installation deviation. If the component categories match, but the pixel percentage difference of the components exceeds 30%, then an installation deviation is considered. Components with installation leakage are encoded as white in the ground truth. After the comparison, the final output will display the components with installation omissions in both the panorama and single perspective images, along with the positions of the images in the inspection path.

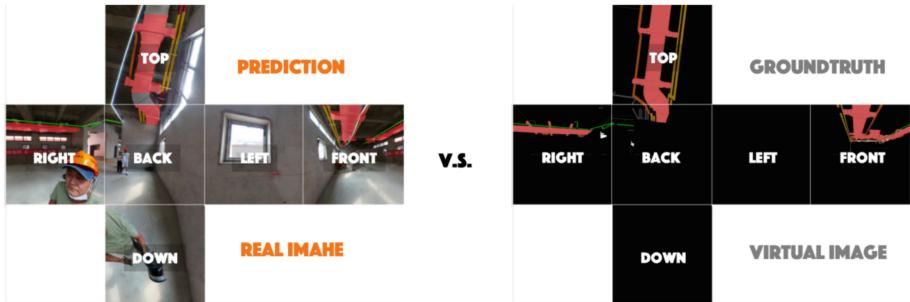


Fig. 4. Virtual-Real Comparison.

4 Data Collection and Case Study

This study conducted indoor inspections using panoramic cameras at two mechanical and electrical construction projects in Shanghai. The inspection videos were divided into six single-view videos: top, bottom, left, right, front, and back, resulting in a total of 1,170 collected images after frame extraction and filtering. We utilized Labelme for polygon annotation of construction images and divided the annotated image data into a train set (910 images) and a validation set (260 images).

As a case study, we selected another construction project in Shanghai (Fig. 5) where mechanical and electrical construction was completed but not yet capped (the mechanical and electrical pipes were visible). The panoramic inspection videos served as input data for OpenVSLAM to obtain the inspection paths. We extracted and filtered images from the single-view inspection videos, obtaining 130 images as a test set. Ultimately, we formed a mechanical and electrical component dataset of 1,300 images (train set: val set: test set = 7:2:1). The BIM model of the case study is depicted in Fig. 5 (c), serving as the ground truth for mechanical and electrical installations.

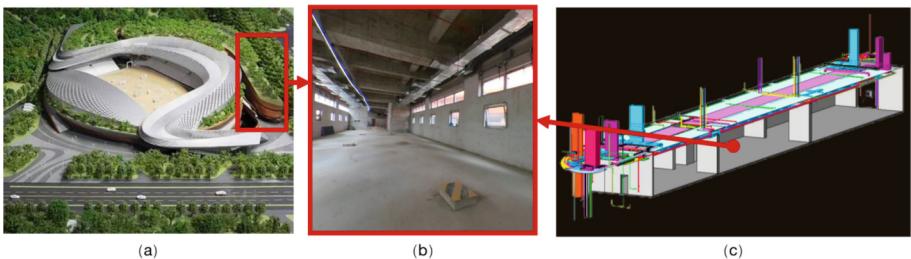


Fig. 5. Case study. (a) Architectural rendering; (b) construction site image; (c) BIM model.

5 Results

5.1 Results of OpenVSLAM

As described in Sect. 3.4, we evaluated the positioning deviation of OpenVSLAM by comparing the distance deviation between the real coordinates with the corresponding predicted coordinates of OpenVSLAM in three calibration points. The experimental results are shown in Fig. 6. The starting and end points coincide to form a closed loop. The positioning error accumulates as the inspection distance increases, reaching a maximum of 135mm at the endpoint. Since the panoramic camera can cover a complete 360-degree field of view, this positioning error has a minor impact on the coverage of the inspection field of view, thus meeting the requirements of engineering applications.

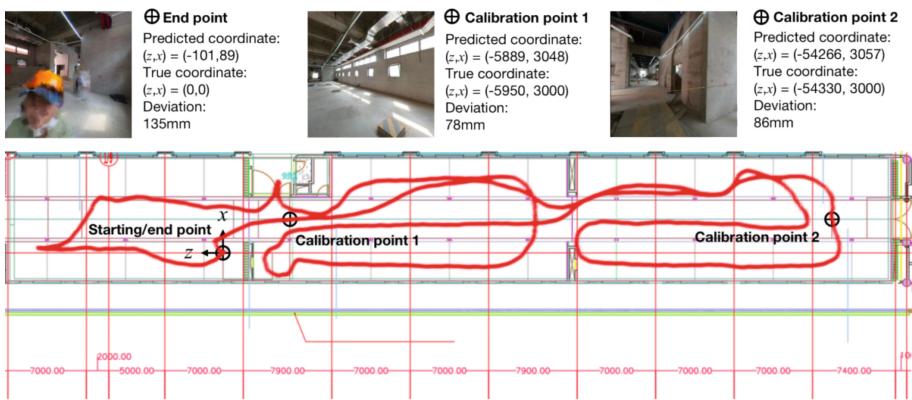


Fig. 6. OpenVSLAM path and positioning accuracy validation.

5.2 Results of PointRend

The PointRend algorithm runs on the mmdetection framework in the Windows 10 system. The hardware system utilizes a single NVIDIA GeForce RTX 3090 GPU. The network virtual environment is configured with Python 3.7, CUDA 11.1, Torch 1.8.0, mmdet 2.25.1, mmcv 1.4.2, and OpenCV 4.6.0. The algorithm training parameters are set as follows: Adam optimizer with a learning rate of 2.0×10^{-4} , 400 epochs, weight decay of 1.0×10^{-5} , batch size of 4, and image size of 896.

The visual segmentation results of PointRend detection are shown in Fig. 7. The model achieves relatively accurate recognition for mechanical and electrical components with irregular shapes and large distance spans. The training accuracy and loss of the model are depicted in Fig. 8. The loss converges around 365 epochs, with mean average precision (mAP50) for segmentation and bounding box detection reaching 78% and 85%, respectively.

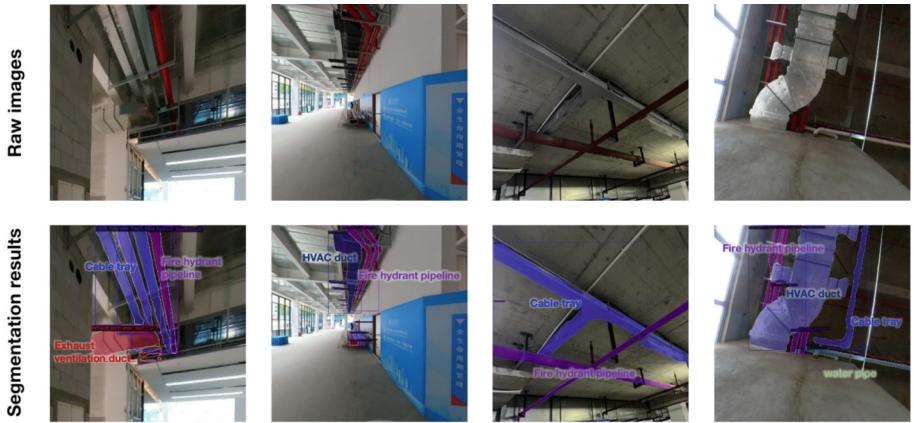


Fig. 7. Visualization results of PointRend.

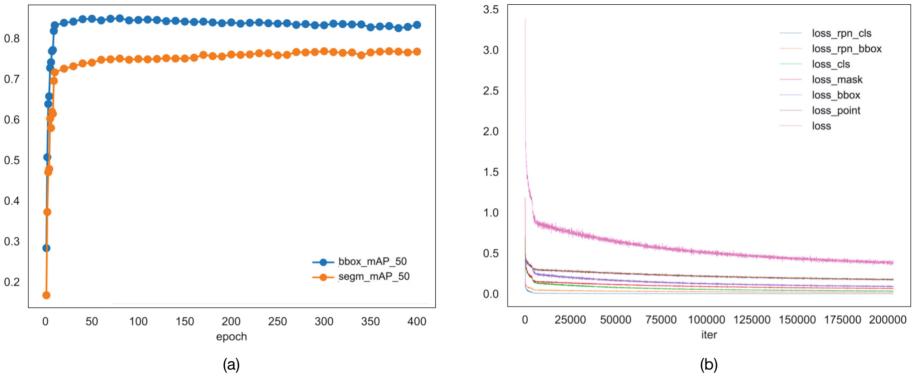


Fig. 8. Training results of PointRend. (a) Epoch-accuracy curve; (b) iteration-loss curve.

5.3 Results of Deviation Detection

This section presents an example of installation deviation detection results (as shown in Fig. 9 (a) within the red box). By comparing the segmentation results of real images from six perspectives with virtual images (ground truth), it is observed that there is a pipe installation leakage in the left and rear views (as indicated by the white masks in the ground truth image). Additionally, Fig. 9 (c) illustrates the indoor mobile inspection path, with the position of installation deviations annotated along the inspection path (Fig. 9 (c)). Through this process, the types of components with installation leakage, their positions, and the specific directions of deviations can be identified, facilitating deviation correction by managers or modification of completion models.

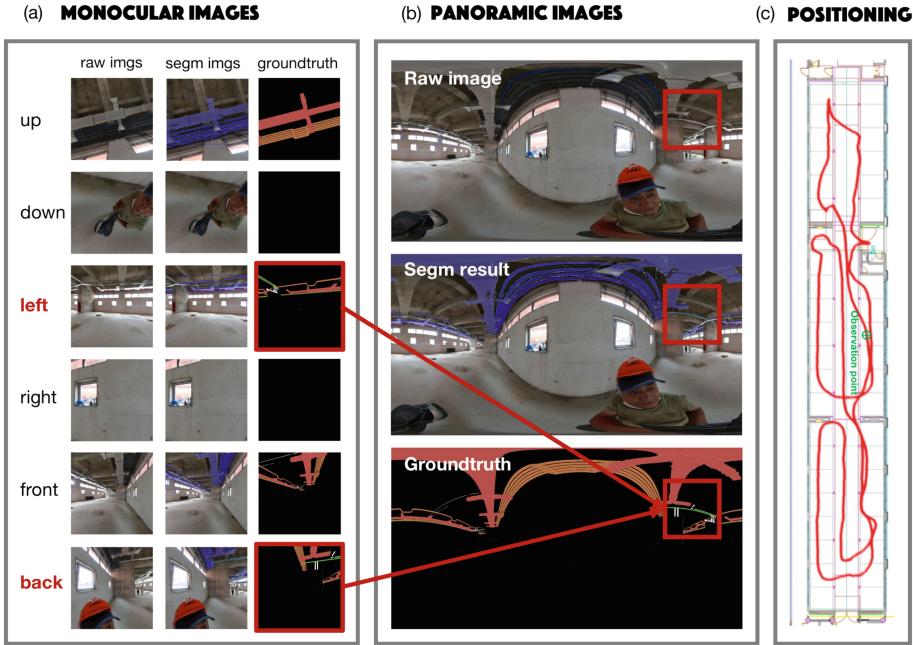


Fig. 9. Results of installation deviation recognition.

6 Discussion

6.1 Impact Analysis of SLAM Localization Accuracy

During inspections, the positioning errors of gyroscopes and yaw angle correction errors result in deviations between the observed viewpoints of real and virtual panoramic images, making it difficult to achieve complete alignment of component positions. Therefore, this study compares the pixel proportion differences of component masks in real and virtual single-view images to determine the presence of installation deviations. In the future, real-time adjustments to the camera angular parameters of UE4 models can be made by extracting spatial angular information from each frame of the SLAM algorithm. This allows for precise matching of component positions between real and virtual images, facilitating further identification of the accuracy of installation positions beyond determining the presence of component omissions.

6.2 Impact Analysis of PointRend Segmentation

The identification of installation errors is also influenced by the accuracy of segmentation in real images. Due to the lower detection accuracy of components in distant views, there is a higher likelihood of missed detections. This can easily lead to misidentifying segmented missed components as component installation omissions during the comparison of real and virtual images. In the future, we will make targeted improvements to segmentation algorithms to enhance the recognition accuracy of small targets in distant views.

6.3 Future Application

For completed projects, this study can be used to identify indoor mechanical and electrical construction installation deviation. By comparing the segmentation results of real and virtual images (ground truth), it identifies the type of omitted components, the spatial location of the omissions on the floor plan, and the specific spatial orientation (up, down, left, right, front, back), facilitating project managers in completing project acceptance and correcting installation errors.

For projects under construction, this study can also calculate the completion progress of construction by comparing the mask percentages of real and virtual images. This helps project managers quickly grasp the progress of construction for timely progress correction.

7 Conclusion

AI-based automated inspection methods have been efficiently applied in construction progress and quality management scenarios such as plastering and cavity detection. However, due to the irregular shapes and excessive lengths of pipeline components involved in mechanical and electrical construction, existing inspection methods are limited in their application for large-scale automated inspections. To achieve efficient mechanical and electrical construction inspection, this study integrates panoramic SLAM, semantic segmentation, and virtual engine technologies to propose an intelligent indoor inspection method. By comparing real and virtual panoramic images, it automatically identifies mechanical and electrical installation deviations.

The research contributes to the existing body of knowledge. In contrast to traditional inspection methods lacking sufficient semantic information and a broad field of view, we integrate semantic segmentation into panoramic SLAM inspection methods, enabling wide-angle, large-scale automated inspections. Installation deviations can be identified by comparing real inspection images with virtual images, facilitating the management of completion acceptance. This study provides a theoretical reference for vision-based indoor mobile inspection. In addition to mechanical and electrical construction scenarios, this research can also be extended to various other settings such as interior structural construction, interior finishing construction, and other scenarios for progress, quality, and safety inspection.

Despite the significant contributions, this research is subject to one limitation. It can only identify mechanical and electrical components within the visible range of the field of view. Future research will integrate laser radar and visible light technology to explore the identification and inspection of occluded components.

8 Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement. The study was supported by the National Key Research and Development Program of China (2022YFC3801700), the National Natural Science Foundation of China (52078374), the Science Foundation for the Science and Technology Commission of Shanghai Municipality (22dz1207100 and 22dz1207800), Top Discipline Plan of Shanghai Universities-Class I, the Chinese Academy of Engineering (2022-XZ-21).

References

1. Deng, H., Hong, H., Luo, D., Deng, Y., Su, C.: Automatic Indoor Construction Process Monitoring for Tiles Based on BIM and Computer Vision. *J. Constr. Eng. Manag.* **146**, 4019095 (2020). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001744](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001744)
2. Wei, W., Lu, Y., Zhong, T., Li, P., Liu, B.: Integrated vision-based automated progress monitoring of indoor construction using mask region-based convolutional neural networks and BIM. *Autom Constr* **140**, 104327 (2022). <https://doi.org/10.1016/j.autcon.2022.104327>
3. Wang, S.-K., Chen, H.-M.: A Construction Progress On-site Monitoring and Presentation System Based on The Integration of Augmented Reality and BIM. In: “Furuya Hiroshi” “Tateyama Kazuyoshi” “Osumi Hisashi” (ed.) Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC), International Association for Automation and Robotics in Construction (IAARC), Kitakyushu, Japan, pp. 147–154 (2020). <https://doi.org/10.22260/ISARC2020/0023>
4. Khairadeen Ali, A., Lee, O.J., Lee, D., Park, C.: Remote Indoor Construction Progress Monitoring Using Extended Reality, *Sustainability* **13** (2021). <https://doi.org/10.3390/su13042290>
5. Engel, J., Schöps, T., Cremers, D., LSD-SLAM.: Large-Scale Direct Monocular SLAM. In: T. and S.B. and T.T. Fleet David and Pajdla (eds) Computer Vision – ECCV 2014, Springer International Publishing, Cham, pp. 834–849 (2014). https://doi.org/10.1007/978-3-319-10605-2_54
6. Engel, J., Koltun, V., Cremers, D.: Direct Sparse Odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 611–625 (2018). <https://doi.org/10.1109/TPAMI.2017.2658577>
7. Sumikura, S., Shibuya, M., Sakurada, K., OpenVSLAM: A Versatile Visual SLAM Framework. In: Proceedings of the 27th ACM International Conference on Multimedia, Association for Computing Machinery, New York, NY, USA, pp. 2292–2295 (2019). <https://doi.org/10.1145/3343031.3350539>
8. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017). <https://doi.org/10.48550/arXiv.1703.06870>
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks, *Adv Neural Inf Process Syst* **28** (2015). <https://doi.org/10.48550/arXiv.1506.01497>
10. Wang, X., Zhang, R., Kong, T., Li, L., Shen, C.: Solov2: Dynamic and fast instance segmentation, *Adv Neural Inf Process Syst* **33**, 17721–17732 (2020). <https://doi.org/10.48550/arXiv.2003.10152>
11. Kirillov, A., Wu, Y., He, K., Girshick, R., Pointrend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9799–9808 (2020). <https://doi.org/10.48550/arXiv.1912.08193>



Multi-Agent-Based Swarm Gas Source Localization Using Nano Aerial Robots

Jan Stührenberg¹(✉), Felix S. Häusler¹, Patrick P. Neumann², Kosmas Dragos¹, and Kay Smarsly¹

¹ Institute of Digital and Autonomous Construction, Hamburg University of Technology, Hamburg, Germany

jan.stuehrenberg@tuuhh.de

² Bundesanstalt Für Materialforschung Und -Prüfung (BAM), Berlin, Germany

Abstract. Gas source localization (GSL) is crucial for mitigating the impact of industrial accidents and natural disasters, for example finding leaks in oil and gas facilities or survivors in collapsed environments. Traditional GSL methods involving human intervention may be hazardous and time-consuming. Utilizing swarms of agile and cost-effective nano aerial robots holds the potential to enhance the safety and efficiency of GSL operations. This study draws inspiration from biological swarms, particularly colonies of social insects, to coordinate and optimize the performance of nano aerial robotic swarms. While most existing swarm GSL strategies assume gas concentration maxima to be in close proximity to actual gas sources, recent research has highlighted the importance of “bouts” as a more precise indicator of gas source proximity, considering the intermittency of gas distributions. In this paper, a swarm GSL strategy is introduced that incorporates bouts as indicators of source proximity, complemented by a bio-inspired pheromone communication system. Specifically, nano aerial robots are deployed as autonomous agents. Upon detecting bouts, the agents emit pheromone markers in an artificial environment, mimicking social insects. Using the concept of artificial potential fields, the agents either exploit the search space by following pheromone gradients or explore the search space. The proposed swarm GSL strategy is implemented and validated in a real-world experiment, conducted in an indoor environment with a single gas source. The experimental results demonstrate the capability of the swarm GSL strategy to perform effectively in indoor environments and that the intermittency of gas distributions is a better source proximity indicator than the mean concentration. It is concluded that this research may provide a methodological basis for improving gas source localization techniques and enhancing disaster response capabilities.

Keywords: Gas source localization · nano aerial robots · mobile robotic olfaction · bouts · swarm robotics

1 Introduction

Gases may pose hazards to organisms, due to the toxicity or flammability. Therefore, gas source localization (GSL) becomes imperative to avoid danger and is the first step towards closing gas leaks in a range of industrial, environmental, and humanitarian applications.

Robots may be used to perform GSL, which is exceptionally useful in environments that are dangerous and/or difficult to access. Employing multiple robots may enhance search efficiency, but requires dependable coordination. Taking inspiration from biological swarms, such as communities of social insects, is a widely explored avenue for achieving robust coordination among numerous yet simple agents. In the realm of mobile robot olfaction (MRO), much research has been inspired by swarm behavior approaches that interpret GSL as a concentration optimization problem, assuming the gas concentration maximum closely approximates a source location [1]. Given this premise, prevailing swarm intelligence algorithms may be applied to address GSL.

Particle swarm optimization (PSO) and ant colony optimization (ACO) are among the most widely used algorithms for GSL performed by multiple robots, as shown in a recent review [2]. The review included 46 studies, of which 30 studies used PSO and 6 studies used ACO. However, PSO and ACO simplify the tangible and dynamic elements of the environments for computational optimization by assuming no physical interaction between the agents. Deriving direct inspiration from biological swarms may yield algorithms that are simpler yet more effective, given that biological swarms function under comparable conditions to robotic swarms. In addition, numerous investigations have illustrated that the concentration maximum may not always serve as an effective indicator of source proximity in complex environments. Alternatively, the intermittency of gas plumes, assessed by so-called “bouts”, may offer a more viable option [3].

In an attempt to counteract the disadvantages of PSO and ACO, this study introduces a swarm GSL strategy using bouts as proximity indicators, coupled with pheromone communication inspired by biological swarms (Sect. 2). Whenever agents detect bouts, the agents emit artificial pheromones at the respective position into an artificial environment. Based on the artificial environment, virtual forces are computed for navigation and collision avoidance with the artificial potential field (APF) concept. To maintain a balance between exploitation and exploration, agents switch between (i) following pheromone gradients and (ii) investigating uncovered areas in the search space. Source estimation is determined by the highest concentration of pheromones in the artificial environment. The swarm GSL strategy is implemented and validated in simulations and real-world experiments (Sect. 3). The paper closes with conclusions and future work (Sect. 4).

2 Bout-Based Swarm GSL Strategy

The swarm GSL strategy presented in this study comprises two integral components. In Sect. 2.1, the artificial (i.e., computational) environment is described, which is responsible for coordinating the swarm. Section 2.2 covers the agents, which are tasked with collecting measurements and emitting pheromones into the artificial environment upon detecting bouts. Furthermore, the interactions of the agents and the artificial environment are described.

2.1 Artificial Environment

The artificial environment serves the purpose of facilitating efficient collaboration among the agents and is implemented through the APF concept [5]. Two different types of

requests can be sent by the agents to the artificial environment, (i) the deposition of artificial pheromones in the artificial pheromone map (APM), which is part of the artificial environment, and (ii) the attractive and repulsive forces applying at the current position of the agents. The attractive forces are calculated from the APM and the repulsive forces are calculated in the so called “anti-collision component” based on the distances of the agents to walls and other agents.

Artificial pheromone map. The APM mimics the role of the environment in biological pheromone communication. The purpose of the environment in pheromone communication is characterized by two facets, (i) an arbitrary number of signals is integrated into a spatial function, and (ii) the spatial function is diffused over time. The first facet allows an arbitrary number of simple agents to cooperate without mutual awareness, while the second facet introduces a probabilistic functionality, as the effect of individually deposited pheromones diminishes over time unless the pheromones are consistently reinforced. The representation of the APM involves a discrete map covering the search space, where agents deposit pheromone markers of predetermined size and intensity at their respective locations. Following the deposition of a new pheromone marker, a diffusion kernel is applied to the APM. The event-based diffusion of pheromone markers, compared to the natural continuous diffusion in a real environment, avoids the loss of information when no new pheromone markers are deposited for a while. Still, the event-based diffusion gives newer entries more weight than earlier ones. For computing the attractive force, a copy of the APM is again smoothed with a diffusing kernel and then transformed into an APF. In this process, the kernel size governs the range of influence exerted by the pheromone markers. Subsequently, the gradient of the APF is calculated, resulting in a force field, and vortexed based on the following expression:

$$F_{vortex,x} = F_x + F_y \cdot \gamma, \quad (1a)$$

$$F_{vortex,y} = F_y - F_x \cdot \gamma, \quad (1b)$$

where $F_{vortex,x/y}$ represents the resultant vortexed force, $F_{x/y}$ denotes the initial force in the x/y direction, and γ is a factor that determines spin direction and vorticity. The vortexation results in agents not being attracted in a straight line to the current local pheromone maximum, but instead circling around the “promising” area. The circling behavior avoids repeated visits and thus the reinforcement of local optima. Finally, the vortexed force field is normalized to the range $[0, 1]$. The normalization ensures that the force is in a predefined range, which ensures that within a certain minimum distance from other agents and obstacles, the repulsive force always dominates to avoid collisions. The normalized vortex force field is the attractive force field of the APM.

Anti-collision. The purpose of the anti-collision component is to avoid collisions among agents and between agents and obstacles. The anti-collision component is made up of the static repulsive potential of the search space boundaries and potential obstacles, as well as the dynamic repulsive potentials of the agents. Following [5], the repulsive force F_{rep} should approach infinity as the distance d between the agent and obstacle goes to zero. Additionally, a minimum distance d_{min} is introduced within which the repulsive force is always greater than the maximum attractive force. Due to normalization of the vortexed force field, the maximum attractive force is one. As agents should not be

affected by far away objects, the repulsive force should be zero if the distance is larger than some maximum distance d_{max} . Finally, the function should be continuously and monotonically differentiable. In the swarm GSL strategy, an adaption of the FIRAS function presented in [5] is employed that satisfies the mentioned conditions. To avoid collisions, a repulsive force F_{rep} is calculated for each agent as follows:

$$F_{rep} = \begin{cases} \left(\frac{d^2_{min}}{\frac{1}{d_{min}} - \frac{1}{d_{max}}} \right) \left(\frac{1}{d} - \frac{1}{d_{max}} \right) \frac{1}{d^2}, & \text{if } d \leq d_{max} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The selection of parameters d_{min} and d_{max} is subject to the positioning accuracy and the safe distance at which agents can navigate past each other. The anti-collision component sums up all individual repulsive forces of an agent on each other agent and obstacle into a resultant repulsive force. The resultant repulsive forces are also vortexized according to Eq. (1). To achieve optimal navigation, it is essential that the APM and anti-collision forces share the same spin direction, which necessitates a reverse rotation for the repulsion forces generated by the enclosing walls. The same spin direction aids in preventing deadlocks during obstacle avoidance.

2.2 Agents

The agents are conceptualized as reactive particles comprised of a bout detector and a motion controller. The bout detector, inspired from [6], operates as follows: Agents calculate the second derivative of the measured concentration signal. Agents observe the second derivative to identify a positive zero crossing, termed as a bout. A bout signifies that a sensor has encountered a new plume filament. Since the sensor readings and therefore the bout detection is exposed to noise, noise-induced bouts are to be filtered out. First, the raw signal and both derivatives undergo smoothing through an exponentially weighted moving average (EWMA) filter. The EWMA filter is employed by the following function:

$$x_{s,i} = \alpha x_i + (1 - \alpha)x_{s,i-1}, \text{with} \quad (3a)$$

$$\alpha = 1 - e^{-\ln(2)/\tau} \quad (3b)$$

where x_i denotes the raw value of the current iteration, $x_{s,i}$ and $x_{s,i-1}$ denote the smoothed value at the current and previous iterations respectively, and τ represents the half-life of the raw values in terms of the iteration count. Second, a threshold is introduced to discard bouts not exceeding the threshold. The positions of accepted bouts are transmitted to the artificial environment. The artificial environment integrates the bouts into the APM as pheromone markers.

The motion controller steers the agents and switches between two attraction modes. The first mode guides the agents to exploit the swarm knowledge by pursuing the attractive force emanating from the APM. In the second mode, agents follow the attractive force

from a randomly generated setpoint for exploration of the search space. The attractive force of the random setpoint, denoted as \vec{F}_{set} , is calculated as:

$$\vec{F}_{set} = \begin{cases} \frac{\vec{X}_{set} - \vec{X}_{pos}}{\|\vec{X}_{set} - \vec{X}_{pos}\|} & \text{if } \vec{X}_{set} - \vec{X}_{pos} > \varepsilon \\ \frac{\vec{X}_{set} - \vec{X}_{pos}}{\varepsilon}, & \text{otherwise} \end{cases} \quad (4)$$

where \vec{X}_{pos} and \vec{X}_{set} represent the position vectors of the agent and setpoint, respectively, and ε is the distance in which the agent linearly reduces speed. Subsequently, the force \vec{F}_{set} is normalized. The motion controller will switch modes if the active attractive force, either towards the pheromone gradient or to the random setpoint, falls below some threshold force, indicating a lack of pheromone or reaching the vicinity of the random setpoint, respectively. Also, the motion controller switches modes after a maximum duration of the active mode. The calculated force is adjusted by the maximum linear velocity to obtain the velocity vector, which is passed to the low-level controller of the aerial robot.

3 Implementation and Validation

In the first Subsect. 3.1, the aerial robot swarm and its components are described. In Subsect. 3.2, the implementation of a simulation is specified. In the simulation, the potential of bouts as a source proximity indicator is validated. In Subsect. 3.3, the swarm GSL strategy is validated in real-world experiments.

3.1 Aerial Robot Swarm

To validate the swarm GSL strategy, a swarm of palm-sized Crazyflie 2.0 quadrocopters (Bitcraze AB, [7]) is used. Each quadrocopter of the swarm is equipped with a 3D local positioning system [8] and a customized sensor deck that includes two SGP30 metal-oxide semiconductor gas sensors (Sensirion, [9]) and an optical motion detection system [10]. A quadcopter is shown in Fig. 1.

The 3D local positioning system estimates the absolute indoor 3D position of the copter. The aerial robot measures its distances to a set of distributed Ultra-wideband-based anchors based on continuously transmitted synchronization packets from these anchors. The distances are used to calculate the absolute position of the robot, which can be used for autonomous flight and collision avoidance. With the local positioning system, an accuracy of about 0.1 m is realized. The gas sensor consists of two pixels, one sensitive to hydrogen (H_2) and the other sensitive to ethanol (C_2H_5OH). From the readings of the two pixels, the sensor is able to calculate the total volatile organic compound and CO_2 equivalent onboard. As the real-world experiments, described in Subsect. 3.3, are performed with ethanol, only the transient signals of the ethanol pixel are used for detecting bouts. The ethanol pixel is specified for a range of 0.3 to 30 ppm, with an accuracy of 15% of the measured value. With the sensors and battery mounted

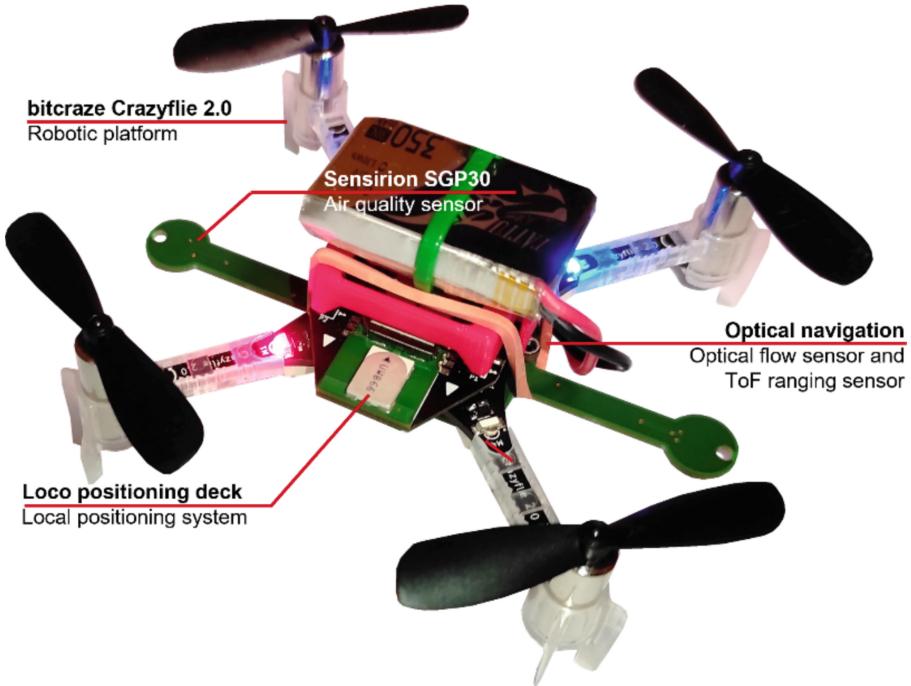


Fig. 1. The Crazyflie 2.0 quadrocopter platform with mounted sensors.

on the quadrocopter, a take-off weight of ≈ 39.6 g and flight times of up to 5 min are achieved. The swarm GSL strategy proposed in this study is implemented using the Robot Operating System (ROS) and the Crazyswarm Python API [11].

3.2 Simulation

The potential of bouts to indicate the gas source proximity was validated through simulations using GADEN [12]. GADEN is a simulation framework developed in ROS, specifically designed for mobile robotic systems and gas sensing algorithms in the field of MRO. In the simulation, 12 agents were deployed in a rectangular room measuring (6×10) m 2 . The room has two openings, one serving as the inlet and the other as the outlet. A single gas source was placed at coordinates (3.0, 1.5, 075) m. The simulation incorporated a pre-calculated wind field from GADEN, featuring strong, fluctuating air-flow. The agents traversed the room by approaching random setpoints at 0.5 m height. A screenshot of the simulation of the visualization tool RViz in ROS is shown in Fig. 2.

The agents logged the sensor measurements and positions at 10 Hz. The simulation, accelerated 10 times, ran for 30 min, resulting in approximately 180,000 sensor readings. Figure 3(a) illustrates mean sensor readings, where lower resistances correspond to higher concentrations. Figure 3(b) depicts the total bout count per cell. The figures visualize the readings in discrete maps of the search space. Both maps exhibit clusters of cells with low resistances and high bout counts near the source, although offset by 1.5 m

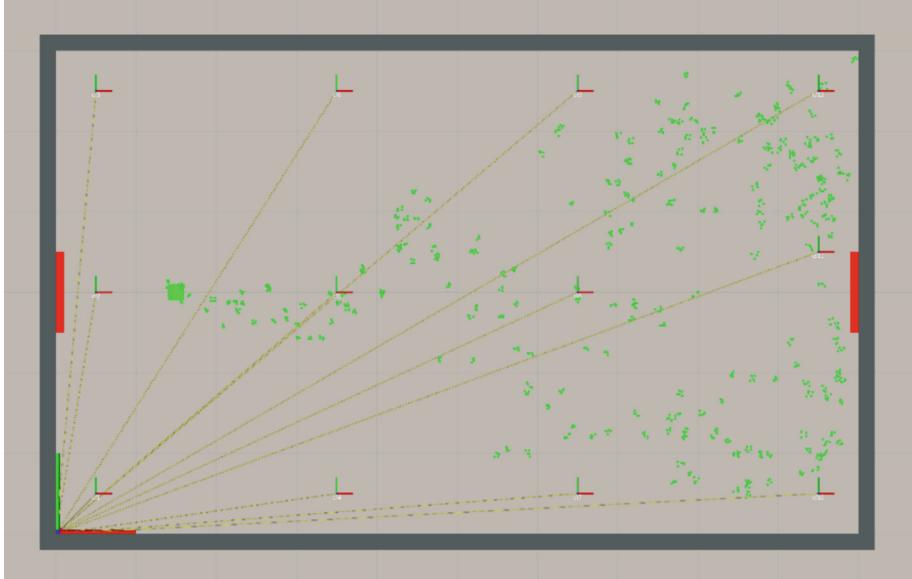


Fig. 2. Screenshot of the simulated environment in RViz, a visualization tool for ROS.

in the airflow direction. While the gradients in Fig. 3(a) appear relatively smooth with no visible boundaries, the clusters in Fig. 3(b) display a denser V-shaped distribution including steep gradients in the downwind direction. The simulation results suggest that the bout count may serve as a more accurate and less noisy indicator of source proximity compared to the mean concentration.

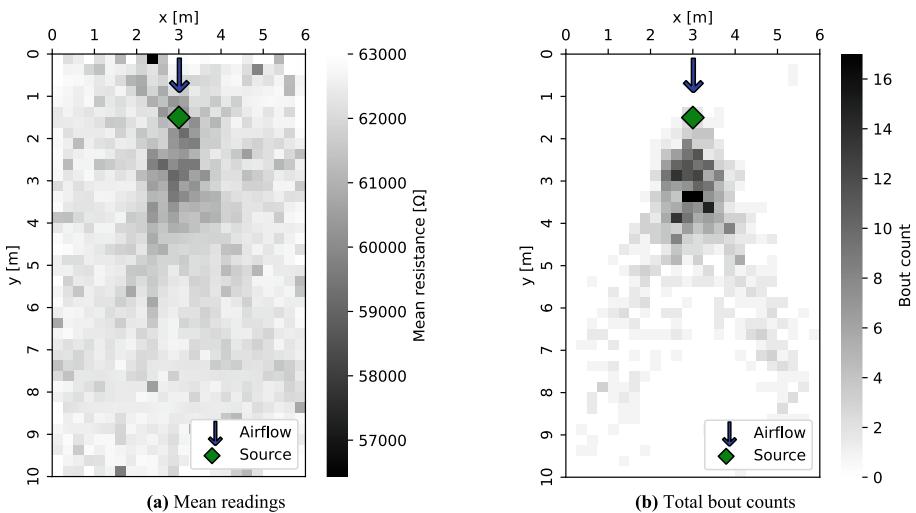


Fig. 3. Maps of the simulated sensor data.

3.3 Real-World Experiments

The validation of the swarm GSL strategy is conducted through real-world experiments. A trio of aerial robots is deployed in a cuboid indoor environment, featuring a 2D search space of approximately (3×3) m² with a single gas source, as depicted in Fig. 4. The gas source, comprising a bottle of liquid ethanol connected to a tube and a 1 W fan, is positioned at (0.80, 0.75) m, elevated 0.1 m above the ground. To stimulate evaporation, pressurized air is introduced, and the emission rate is indirectly controlled by maintaining the airflow at 2 l/min. The source faces the center of the search space and opens roughly 1 min before each run. Between runs, the source is closed, and the room is ventilated to ensure comparable starting conditions. Runs conclude either with the robots landing, crashing or running out of battery. Algorithm parameters were chosen heuristically. Sensor readings were logged at 10 Hz. Over 16 runs, around 20,000 measurements were recorded, with individual flight times ranging from 0 s (crash at takeoff) to 329 s. The combined flight times of all three quadrocopters range from 303 s to 903 s, with a mean of $613.4 \text{ s} \pm 165.2 \text{ s}$.

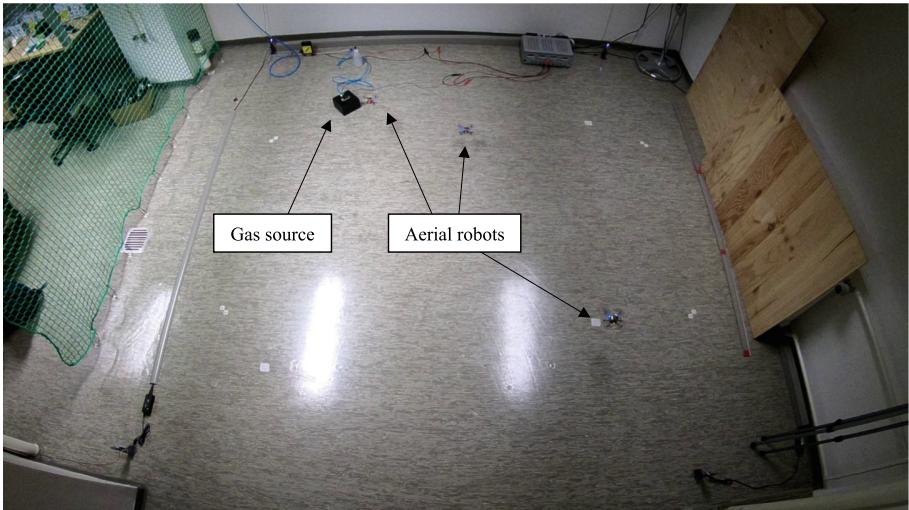


Fig. 4. Real-world experiments.

The experimental results are presented in Figs. 5 and 6. Figure 5(a) displays a mean resistance plot with smooth gradients, akin to Fig. 3(a). However, in contrast to the simulation results, the cluster of low mean resistance is shifted in the upwind direction. In Fig. 5(b), bout distributions again showcase dense clusters with steep gradients, but unlike the simulation, no V-shape is evident. Figure 6(a) illustrates the final source location estimates, with all estimation errors in the airflow direction. Notably, the mean error aligns closely with the center line of the airflow of the fan. Figure 6(b) displays the average estimation error of the combined flight time scaled by the number of active agents, encompassing results from each individual run. The source estimation, on average, improves rapidly during the initial 90 s of combined flight time. After 330 s, the

final source estimation error averages $0.7 \text{ m} \pm 0.29 \text{ m}$, ranging from 0.25 m (run #10) to 1.28 m (run #8). Improvement continues after 540 s, reducing the error to $0.6 \text{ m} \pm 0.21 \text{ m}$.

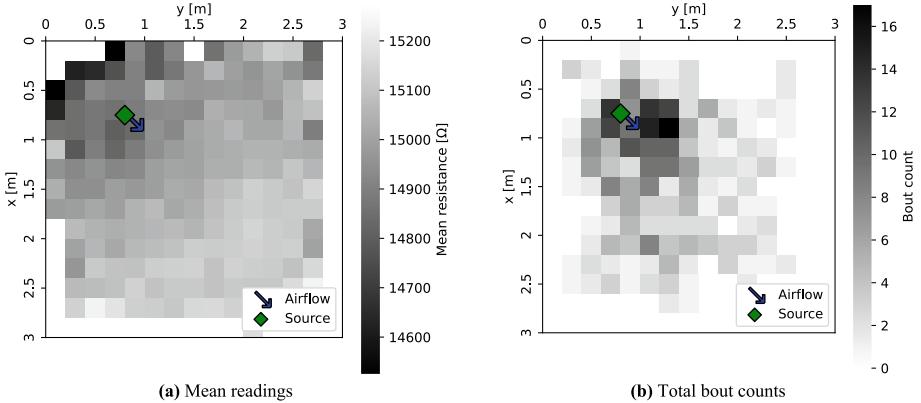


Fig. 5. Maps of the real-world sensor data.

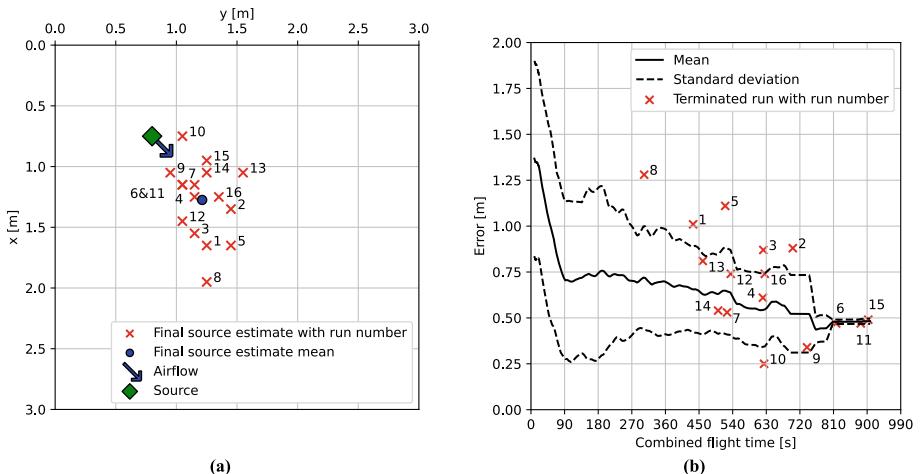


Fig. 6. Experimental results: (a) Final estimations in space and (b) smoothed mean error of combined flight time of the experiments with standard deviation over time. Run time individually scaled by the number of active agents, and estimations excluded after termination.

4 Conclusions and Future Work

This paper has presented a swarm GSL strategy utilizing bouts as indicators of source proximity and virtual pheromone communication for swarm cooperation. Unlike previous studies conducted in this field, the swarm GSL strategy proposed herein is characterized by the integration, storage, and propagation of pheromone signals, rendering this

study scalable, robust to noise and errors, responsive to changes in source position, and streamlining agents to a purely reactive nature. Both simulations and real-world experiments underscore the effectiveness of bouts as robust indicators in indoor environments, outperforming the use of mean concentrations. Despite being in its early stages, the proposed strategy exhibits promising outcomes in real-world experiments. Future research could concentrate on refining the parameter set of the swarm GSL strategy, which was chosen heuristically in this study. Enhancements might be attained by incorporating wind information and expanding the strategy into three dimensions. Additionally, a comparative analysis of the swarm GSL strategy against existing algorithms under identical environmental conditions could offer valuable insights.

References

1. Jing, T., Meng, Q.-H., Ishida, H.: Recent progress and trend of robot odor source localization. *IEEJ Trans. Electr. Electron. Eng.* **16**(7), 938–953 (2021)
2. Wang, J., Lin, Y., Liu, R., Fu, J.: Odor source localization of multi-robots with swarm intelligence algorithms: A review. *Front. Neurorob.* **16** (2021)
3. Schmuker, M., Bahr, V., Huerta, R.: Exploiting plume structure to decode gas source distance using metal-oxide gas sensors. *Sens. Actuators, B Chem.* **235**, 636–646 (2016)
4. Meng, Q.-H., Yang, W.-X., Wang, Y., Zeng, M.: Multi-robot odoplume tracing in indoor natural airflow environments using an improved aco algorithm. In: Proceedings of the 2010 IEEE International Conference on Robotics and Biomimetics. Tianjin, China, December 14 (2010)
5. Khatib, O.: Real-time obstacle avoidance for manipulators and mobile robots. In: Proceedings of the 1985 IEEE International Conference on Robotics and Automation. St. Louis, MO, USA, March 25 (1985)
6. Burgués, J., Hernández, V., Lilienthal, A.J., Marco, S.: Smelling nano aerial vehicle for gas source localization and mapping. *Sensors* **19**(3), 478 (2019)
7. Bitcraze AB: Crazyflie 2.0. Accessed November 27 (2023). <https://www.bitcraze.io/products/old-products/crazyflie-2-0/>
8. Bitcraze AB: Loco Positioning system. Accessed November 27 (2023). <https://www.bitcraze.io/documentation/system/positioning/loco-positioning-system/>
9. Sensirion AG: Datasheet SGP30 – Sensirion Gas Platform. Accessed November 27 (2023). https://sensirion.com/media/documents/984E0DD5/61644B8B/Sensirion_Gas_Sensors_Datasheet_SGP30.pdf
10. Neumann, P.P., et al.: Indoor air quality monitoring using flying nanobots: Design and experimental study. In: Proceedings of the 2019 IEEE International Symposium on Olfaction and Electronic Nose. Fukuoka, Japan, May 26 (2019)
11. Preiss, J.A., Honig, W., Sukhatme, G.S., Ayanian, N.: Crazyswarm: A large nano-quadcopter swarm. In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation. Singapore, Singapore. May 29 (2017)
12. Monroy, J., Hernandez-Bennets, V., Fan, H., Lilienthal, A.J., Gonzalez-Jimenez, J.: Gaden: A 3d gas dispersion simulator for mobile robot olfaction in realistic environments. *Sensors* **17**(7), 1479 (2017)



Autonomous Navigation of Quadruped Robots for Monitoring and Inspection of Civil Infrastructure

Aditya Tandon^(✉), Jan Stührenberg, Kosmas Dragos, and Kay Smarsly

Institute of Digital and Autonomous Construction, Hamburg University of Technology,
Hamburg, Germany
aditya.tandon@tuhh.de

Abstract. Mobile robots have increasingly been gaining recognition in monitoring and inspection of civil infrastructure, owing to their potential to improve efficiency and accuracy. Specifically, quadruped robots offer enhanced stability and adaptability, rendering the robots ideal candidates for automated monitoring and inspection. To enable quadruped robots to conduct monitoring and inspection of civil infrastructure, the robots must be capable of autonomous navigation. Existing approaches towards autonomous navigation usually incorporate joint-state information to plan motions with whole-body controllers, representing a non-linear, high-dimensional problem that is computationally expensive to solve and is a particular burden when implemented into the robots for real-time navigation. This paper presents an integrated architecture for automated monitoring and inspection that consists (i) of a mission planning framework devised for planning monitoring and inspection tasks and (ii) a robust motion planning framework to enable real-time navigation. Specifically, the motion planning problem is decoupled into high-level task-space and low-level joint-space components, and the motion planning framework, building upon the “Cartographer” simultaneous localization and mapping algorithm, combines the “batch-informed trees” and A* algorithm for global planning and the “timed elastic band” for local planning and obstacle avoidance. For validation, the integrated architecture for automated monitoring and inspection is implemented into quadruped robots, and validation tests are conducted in indoor office environments to be inspected. As a result, it is demonstrated that the quadruped robots navigate safely and collision-free in real time, accommodating both static and dynamic obstacles. Enhancing the efficiency and accuracy of autonomous navigation of quadruped robots in complex and dynamic environments, the architecture is expected to pave the way for future research in robust controller development and 3D-state space planning.

Keywords: Quadruped robots · motion planning · mission planning · batch-informed trees · timed elastic band

1 Introduction

Degradation of civil infrastructure poses significant risks to public safety. Civil infrastructure therefore requires continuous monitoring and regular inspections to ensure safe operation [1]. Existing manual monitoring and inspection approaches may be time-consuming, labor-intensive, expensive, and dangerous. With labor shortages, rising labor costs, and low productivity in the construction industry [2], monitoring and inspection may become more expensive. Adopting robot-based, automated monitoring and inspection has the potential to enhance productivity, reduce costs, and increase safety [3].

Mobile robots, equipped with sensors, have already been used for monitoring and inspections [4]. However, previous approaches have been dependent on personnel with special training, such as building inspectors and robot operators. Recent advancements in robotics and computing have fueled approaches using mobile robots, including unmanned aerial vehicles and unmanned ground vehicles, in autonomous configurations to perform visual inspections [5]. Unmanned aerial vehicles may, however, be restricted in range and operational flight time because of limited payload and battery capacity. Wheeled or tracked unmanned ground vehicles may be less constrained than unmanned aerial vehicles regarding operation time or payload capacity, but are limited by the terrain. The traversability of construction environments is challenging, as wheeled or tracked robots may be unable to climb staircases, access multi-floor workspaces, or climb over obstacles of varying sizes, such as concrete blocks, pipes, or stacks of construction materials. In contrast to unmanned ground vehicles, legged robots, particularly quadruped robots, can perform complex maneuvers, overcome obstacles, and climb staircases. Quadruped robots outperform both unmanned ground vehicles and unmanned aerial vehicles due to their enhanced mobility, allowing the quadruped robots to navigate difficult terrain and carry heavy payloads. However, to achieve robot-based automated monitoring and inspection, quadruped robots must be capable of reliable autonomous operation.

To enable reliable autonomous operation of quadruped robots, motion planning frameworks capable of real-time environment perception, localization, path planning, obstacle avoidance, and motion control are typically utilized. The goal of motion planning frameworks is to generate robot motion control commands that move the robots on collision-free trajectories to target positions. Motion planning for mobile robots has been a subject of comprehensive research [6]. The choice of motion planning frameworks primarily depends on the complexity of tasks to be accomplished and the complexity of the operational environment. The motion planning framework developed in [7] has used a combination of a global planning module that operates on a 3D volumetric map, and a local planning module for navigation and robot traversal. The motion planning framework NeBula, developed in [8], operates across several robot types, including unmanned ground vehicles, unmanned aerial vehicles, and quadruped robots. The NeBula framework includes components for state estimation, mapping, traversability, planning, and communication, and is inclined towards providing platform-agnostic autonomous navigation capabilities. In [9], an autonomous navigation framework, consisting of a locomotion control module, a state estimation module, and a planning module, has been deployed on the MIT Cheetah Vision quadruped robot in outdoor environments. The

locomotion module uses a regularized predictive control scheme with a whole-body controller (WBC), which is supplemented by information on robot joint angles and environmental characteristics perceived by the robots. The planning module utilizes an A*-based global planning module and a gait scheduler for high-level motion planning. However, motion planning is performed in the joint space. In [10], a coupled motion planning method that plans WBC motions consisting of foothold locations and horizontal motions has been presented and deployed on a quadruped robot. The WBC optimizes motion across the quadruped robot while taking into consideration center-of-mass (CoM) locations, foothold locations, robot attitude, and kinodynamic constraints to generate joint torques for robot control.

Existing motion planning frameworks for quadruped robots incorporate information on robot joint states and terrain that may be used by advanced planning components within the framework, such as foothold contact planning [11] and reachability checking [7], enabling comprehensive motion planning in the joint space. However, the motion problem caused by using advanced motion planning components may be high-dimensional and non-linear, rendering the problem complex and computationally expensive to solve in real-time, particularly on resource-constrained computing devices deployed on quadruped robots. Although WBC approaches have demonstrated promising results [12] on quadruped robots, WBC approaches are computationally complex, with this complexity being hardly justified in standard indoor environments. While outdoor environments may be uncertain, uneven, and complex, standard indoor environments are typically characterized by continuous environments that do not contain large discontinuities. In standard indoor environments without large discontinuities, computation of precise foothold locations may be less crucial, and instead, a gaited strategy that maintains the balance of the robot may be sufficient.

The motion planning frameworks mentioned above are usually either simplistic, providing limited scaling with the size and complexity of the environment, or complex, involving motion planning in the joint space. On the one hand, simplistic motion planning frameworks may fail to model the complexity of the robot and the environment; on the other hand, complex motion planning frameworks may attempt to incorporate most aspects of the motion problem, becoming computationally expensive to solve in a reasonable time and inhibiting real-time performance on resource-constrained computing devices deployed on quadruped robots. This paper proposes an integrated architecture for automated monitoring and inspection that includes a motion planning framework suitable for automated monitoring and inspection. The motion planning framework decouples the motion planning problem into two distinct components, (i) a high-level task-space problem, and (ii) a low-level joint-space problem, to ensure effective real-time operation of quadruped robots. The motion planning framework operates through high-frequency communication between the two components and utilizes the high-level interface for robot control, thereby abstracting low-level control of the robots. In addition to the motion planning framework, a mission planning framework, the second integral part of the integrated architecture, is introduced to enable planning and execution of monitoring and inspection tasks. The mission planning framework provides an integrated task planning system, used to schedule various types of monitoring, inspection, and navigation tasks.

The remainder of the paper is outlined as follows. Section 2 describes the integrated architecture for automated monitoring and inspection. First, the design and implementation of the motion planning framework is covered, followed by the design and implementation of the mission planning framework. Section 3 presents the tests conducted to validate the integrated architecture as well as the results of the validation tests. Finally, Sect. 4 concludes with a discussion of the results, a summary of the work presented herein, and an outlook on potential future work.

2 An Integrated Architecture for Automated Monitoring and Inspection

The integrated architecture for automated monitoring and inspection, including the motion planning framework and mission planning framework, is shown in Fig. 1. The software setup builds upon a distributed Robot Operating System (ROS) architecture that allows modules to be distributed across robots and computers. Offloading modules to computers reduces the computational load of the robots and facilitates real-time operation. Therefore, in this study, the motion planning framework is implemented on a robot and the mission planning framework, which is not critical for real-time operation, is implemented on a computer. In the following subsections, the motion planning framework and the mission planning framework of the integrated architecture are presented.

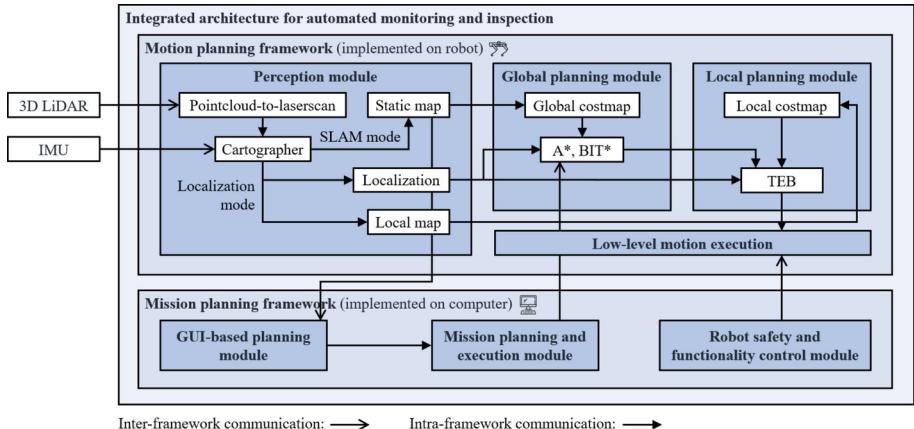


Fig. 1. Integrated architecture for automated monitoring and inspection.

2.1 Motion Planning Framework

Since robots operate in a dynamic world, reliable real-time performance of the modules of the motion planning framework must be ensured to avoid collisions. To reduce the computational complexity, the motion planning framework is decoupled into a high-level

task-space problem and a low-level joint-space problem. Decoupling the problems results in two problems with reduced complexity, which can be solved independently. The high-level task-space problem is concerned with planning collision-free paths for the robot and to generate target velocities in the task space. Using high-frequency communication, the low-level motion execution employs a gaited strategy to fulfill the target velocities while maintaining robot balance and stability. Since high-level motion planning using full 3D representations in real-time on embedded platforms remains computationally prohibitive [7], a hybrid strategy that approximates the 3D environment in a 2D format is adopted. High-level motion planning in the task space is therefore achieved by representing the quadruped robot as a holonomic robot in the SE(2) state space with a rectangular robot footprint, representing the 2D shape of the physical space occupied by the robot in the environment. Thus, the development of a generalized and modular motion planning framework is achieved. By decoupling the motion planning framework, the low-level controller can operate at high frequencies, ensuring robust and stable locomotion in uneven terrain and in the presence of external disturbances. The ROS navigation stack, used as a basis for the implementation, processes navigation goals and utilizes static and dynamic occupancy grid maps to facilitate autonomous navigation, where a global planning module plans paths globally in the static occupancy grid map and a local planning module plans paths locally in the dynamic occupancy grid to avoid obstacles. In the following paragraphs, the core modules of the framework are presented.

Perception module. Motion planning requires perception of the environment to generate maps of the environment, to localize within the maps, and to detect dynamic obstacles. In this study, perception is enabled by a light detection and ranging (LiDAR) sensor, which is the primary sensor employed within the perception module of the motion planning framework, complemented by an inertial measurement unit (IMU). Since motion planning and navigation is performed in the SE(2) state space, 2D laser scan data is employed for perception. However, 2D laser scan data offers only a 2D planar snapshot of the environment at the operational height of the LiDAR and does not effectively capture spatial information in 3D. To leverage 3D spatial information about the environment, 3D point cloud data is converted to 2D laser scan data by projecting the 3D geometry of the environment onto the horizontal 2D plane using the ROS point-cloud_to_laserscan package [13]. The IMU senses fine changes in the attitude and state of the robot to align LiDAR scan data with greater accuracy, aiding the localization and mapping processes, detailed below.

To generate a map of the environment and localize within the map, the simultaneous localization and mapping (SLAM) system “Cartographer” is used [14]. The Cartographer system leverages the 2D laser scan data and the IMU data collected by the perception module. The Cartographer system can be operated in two modes, (i) a SLAM mode, and (ii) a localization mode. The SLAM mode allows generating a static map of the environment, with the robot being controlled manually by an operator to map areas of interest. The map is stored as an occupancy grid map. Occupancy grid maps represent the environment with a map consisting of an evenly-spaced discrete grid, where each cell represents either occupied or free space for navigation. Once a static map has been generated using the SLAM mode, the localization mode of the Cartographer system is used to localize the robot in the static map.

Global planning module. The global planning module is a critical subset of the motion planning framework for enabling robots to determine optimal trajectories to navigate environments. Specifically, the global planning module identifies sequences of intermediate states that guide the robot from the initial state to a goal state. By searching through the space of all possible states, the so-called “configuration space”, the global planning module plans feasible, safe, and optimal paths, by checking against potential collisions with obstacles in the static map and minimizing distance, time, and/or energy. For searching, graph-based or sampling-based path planning algorithms are employed.

The global planning module implements the `nav_core::BaseGlobalPlanner` interface prescribed in the ROS navigation stack. Various path planning algorithms are implemented within the global planning module to enable optimum path planning performance across different operating environments. The `NavfnROS` implementation of the A* graph-based planning algorithm is utilized. A* operates on the discretized and static occupancy grid map. Given an admissible heuristic, A* guarantees an optimal solution. ROS-compatible C++ implementations of sampling-based planning algorithms are utilized from the `OmplPlanner` global planning module [15], based on the Open Motion Planning Library (OMPL) [16]. Implementations of the rapidly-exploring random trees (RRT*) and Informed-RRT* planning algorithms are used as-is, while the `OmplPlanner` is extended to include the batch-informed trees (BIT*) planning algorithm. Sampling-based path planning algorithms operate by randomly sampling the configuration space and constructing a graph by connecting points with edges if they are collision-free. The sampling-based path planning algorithms implemented in this study exhibit asymptotic optimality, i.e. the probability of converging to a solution increases to one as the search time increases. The algorithms are particularly effective for high-dimensional configurations, demonstrating runtime-independence from the dimension of the configuration space.

Local planning module. The local planning module of the motion planning framework operates on the coarse path planned by the global planning module. To account for obstacles, particularly dynamic obstacles and obstacles that may have been untracked during mapping, the local planning module adapts the path plans for a short time window in the immediate local environment of the robot at high frequencies to facilitate collision-free navigation, while ensuring plans that are generated by the module are executable by the robot controller. The local planning module utilizes sensor data received from the perception module to construct a local map centered around the robot. Local maps enable perception of the immediate environment, track dynamic and static obstacles, and allow the local planning module to generate motion commands and trajectories for collision-free navigation. To ensure minimum distances between robots and obstacles, costmaps are employed. Costmaps supplement the occupancy grid maps by associating travel costs to neighboring occupied cells.

The local planning module uses the timed elastic bands (TEB) approach for local planning [17]. The TEB planning approach modifies existing global plans by counter-balancing virtual attraction forces of target poses and repulsive forces of obstacles. The TEB planning approach generates optimized trajectories with respect to execution time while complying with kinodynamic constraints and maintaining minimum distances

when navigating around obstacles. The TEB planning approach involves the minimization of an objective function for online optimization of time-optimal trajectories, while ensuring controller actions do not violate kinodynamic bounds such as maximum velocity and acceleration. Since quadruped robots are holonomic robots and are able to move among translational axes, the TEB local planning module is configured to weaken the penalty for satisfaction of non-holonomic constraints. Thereby, trajectory generation with velocities and accelerations in all directions in the SE(2) state space is promoted.

2.2 Mission Planning Framework

To enable automated monitoring and inspection, a mission planning framework to schedule various types of monitoring, inspection, and navigation tasks is employed. The mission planning framework integrates a user interface to sequentially plan, execute, and administer sets of monitoring and inspection tasks, hereinafter referred to as “monitoring and inspection missions”, and administers robot-related safety and functionality. The mission planning framework is classified into three modules. First, the mission planning and execution module, followed by the GUI-based planning module, and lastly, the robot safety and functionality control module.

The **mission planning and execution module** models monitoring and inspection missions as hierarchical finite state machines. In particular, mission tasks are modelled as individual states, and transitions between states are defined based on the execution status of the respective tasks. The mission planning and execution module is implemented using the SMACH library in the ROS framework [18]. One waypoint is added for each state to the SMACH state-machine object, and transitions between states are created based on the successful execution status of the states, indicated by successful transitions. Goals are sent to the navigation stack from within each state by implementing an action. The state-machine object waits for completion of the goal before transitioning to the next waypoint. The state machine completes execution when the last waypoint is reached by transitioning to the “final” state.

The **GUI-based planning module** provides a panel in the ROS visualization tool RViz to intuitively create monitoring and inspection missions. Waypoints are added to a monitoring and inspection mission by marking arrows, corresponding to robot poses, on the static occupancy grid map created in the SLAM mode of Cartographer. The origin of the arrows corresponds to the target position of the waypoints and the direction of the arrows corresponds to the target heading of the waypoints. Furthermore, the GUI-based planning module allows saving, loading, and executing monitoring and inspection missions.

Last, but not least, the **robot safety and functionality control module** is devised to provide the operator with an integrated interface to view and administer control over the operation and motion status of the robot. The operator may pause the execution of a mission by pausing the operation of the robot, ensuring appropriate manual intervention during mission execution.

3 Validation Tests and Results

The integrated architecture for automated monitoring and inspection proposed in this study is validated in an indoor environment to evaluate functionality and real-time performance. The validation tests are threefold. First, the perception module of the motion planning framework is validated by creating a static occupancy grid map of the environment using the SLAM mode of Cartographer and testing the localization in the static occupancy grid map using the localization mode of Cartographer. Second, the mission planning framework is validated by planning and executing monitoring and inspection missions in the static map. The monitoring and inspection missions are designed as multi-waypoint missions to lastly validate the global planning module and the local planning module of the motion planning framework.

The motion planning framework is implemented into a quadruped robot of type “Intelligent Documentation Gadget” (IDOG), as described in [4]. The validation tests are conducted at the Institute of Digital and Autonomous Construction (IDAC) at Hamburg University of Technology, Germany. The indoor environment features a range of furniture, consisting of large tables, chairs, workstations, small cabinets, and shelves. Notably, movable furniture, such as cabinets and chairs, may alter the spatial layout to a minor extent, thereby constituting static obstacles in the environment. People walking in the office space, not captured during mapping, are considered dynamic obstacles in the environment.

The following subsections provide an overview of the validation tests performed and the results obtained, starting with the perception module, followed by the global planning module, and finally, the local planning module for autonomous navigation capabilities.

3.1 Perception Module

To validate the perception module, Cartographer is used to create a static map of the environment and subsequently to localize the robot in the static map. The LiDAR is operated at a frequency of 10 Hz and the IMU at 100 Hz. The IDOG is manually controlled, while the SLAM mode of Cartographer creates the occupancy grid map at a resolution of 0.05 m/pixel. The created occupancy grid map, shown in Fig. 2, is approximately 780 pixels in length and 400 pixels in width, corresponding to a map size of 39 m × 20 m. After starting the localization mode of Cartographer, the initially unknown location of the IDOG in the map is determined within seconds.

3.2 Global Planning

To validate the global path planning module, the mission planning framework is used to create monitoring and inspection missions set up as multi-waypoint missions. The global planning module of the motion planning framework plans paths between the waypoints. The paths are compared based on path length and path planning time. Three challenging waypoints are selected as shown in Fig. 3. The first waypoint simulates the challenges encountered when planning long – yet feasible – paths in short time intervals between waypoints that are far apart. The second waypoint simulates the complexities in navigating through intricate environments with large obstacles, narrow corridors, and

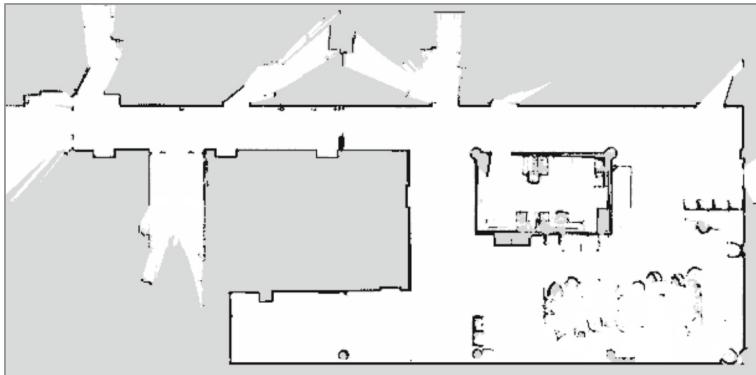


Fig. 2. Occupancy grid map of an indoor environment created by the SLAM mode of Cartographer.

around corners. The third waypoint involves traversing a narrow, cluttered space with numerous static obstacles. Validation is performed for the BIT*, Informed-RRT*, and RRT* algorithms of the OMPL planner and for the A* algorithm of the NavfnROS planner.



Fig. 3. Multi-waypoint mission, created using the mission planning framework to validate the global planning module.

The validation tests demonstrate that the global planning module is capable of planning valid paths in the indoor environment. All global planning algorithms find a path between the start position and the first waypoint. However, the path produced by RRT* is considered invalid due to its coarse trajectory through cluttered environments, resulting in collisions with furniture. This observation, supported by a comparison of optimal path costs, results in abandoning further path planning tests with RRT*. The paths planned by A*, BIT*, and Informed-RRT* between waypoints 1, 2, and 3 have comparable

lengths. However, the key distinction lies in planning time, as shown in Table 1. The A*-based planner demonstrates a planning time an order of magnitude lower than the other sampling-based planning algorithms. Despite a 1-s planning budget for the sampling-based planning algorithms, path lengths generated by the sampling-based planning algorithms exceed the length of the path generated by the A* planning algorithm. The results from the validation tests for the global planning module are tabulated in Table 1.

Table 1. Path-length and planning-time of the global planning algorithms

| Waypoint | A* | | BIT* | | Informed-RRT* | | RRT* | |
|----------|---------------|-------------|---------------|-------------|---------------|----------|---------------|-------------|
| | Length
(m) | Time
(s) | Length
(m) | Time
(s) | Length
(m) | Time (s) | Length
(m) | Time
(s) |
| 1 | 18.878 | 0.082 | 18.795 | 1.037 | 19.732 | 1.035 | 26.747 | 1.034 |
| 2 | 11.544 | 0.288 | 11.621 | 1.051 | 11.535 | 1.085 | - | - |
| 3 | 08.815 | 0.085 | 08.816 | 1.032 | 08.811 | 1.039 | - | - |

The results obtained in this study indicate that, although sampling-based planning algorithms represent the state of the art, sampling-based planning algorithms may outperform traditional graph-based planning algorithms, such as A*, only in high-dimensional problems. It may also be inferred that the existing representation of the motion planning problem may be low-dimensional, since planning is performed only in 2D. In addition, it must be noted that the size of the validation environment is small, as compared to larger environments in which monitoring and inspection are performed, such as construction or industrial sites. To facilitate future path planning in 3D and to enable operation in large environments, the BIT* planner is integrated into the motion planning framework in addition to the A* algorithm. The A* algorithm will be used for planning in 2D and small environments.

3.3 Local Planning

Validation of the local planning module of the motion planning framework is conducted by performing multiple tests. First, the path-tracking capability of the local planning module in the absence of obstacles is ascertained. The tests involve tracking a straight-line path generated by the global planning module and tracking a long global plan to navigate between waypoints while traversing long narrow passageways, maintaining a minimum distance from environment boundaries at all times. The validation tests for the local planning module in the absence of obstacles indicate that the actual trajectory traversed by the robot closely aligns with the global plan. Deviations from the global plan are noted, however, deviations from the global plan do not necessarily reflect poor performance of the local planning module, particularly in regions where minimal distance from environment boundaries must be maintained. However, in some segments of the plan traversed by the robot, slight deviations and a constant offset from the global plan is observed. Overall, the performance of the local planning module in the absence of obstacles is deemed satisfactory.

The functionality and real-time performance of the local planning module is validated against static and dynamic obstacles of variable shapes and sizes, which are likely to appear in indoor environments, despite the mapped environment in which robots operate. In fact, maps may be incomplete or change over time, thereby failing to accurately represent the updated furniture layout. Furniture that has been moved is representative of static obstacles in the environment. Moreover, the performance of the local planning module is also validated against dynamic obstacles, such as persons. The validation tests involve planning global paths using outdated maps of the environment. As the robot traverses the global path, static obstacles are added to the local costmap of the robot, based on sensor data received from the perception module, as illustrated in Fig. 4a. It is also observed that the local planning module plans around the obstacle (blue trajectory), although the global plan is in collision with the static obstacle (purple trajectory). As can be seen from Fig. 4b, the IDOG is able to successfully avoid the obstacle and complete the global plan.

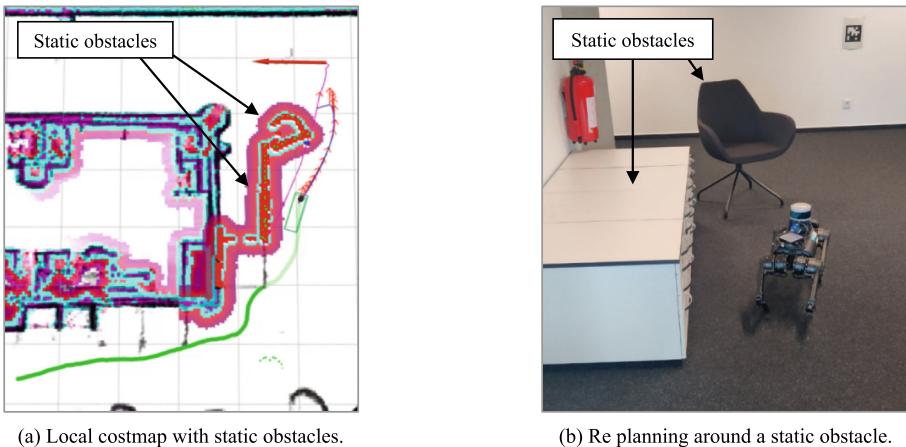


Fig. 4. Demonstration of obstacle-avoidance behaviors against static obstacles.

Furthermore, Fig. 5 shows the trajectories re-planned by the motion planning framework to avoid dynamic obstacles. As can be seen from Fig. 5, the purple trajectory signifies the local plan, while the red trajectory signifies the global plan. Although the IDOG exhibits reliable obstacle-avoidance behavior, the initial parts of the green trajectory indicate that an offset continues to exist between the real trajectory and the global plan.

To validate the autonomous navigation capability of the motion planning framework deployed on the IDOG, specific waypoints are selected for navigation through restricted passages. In the tests, waypoints are placed within an office room, requiring traversal through a door passage with a width of 1.05 m, and inside a laboratory room in the office environment, requiring navigation through a door that is 0.8 m wide. The tests aim to ascertain behaviors when navigating around obstacles and through narrow confined

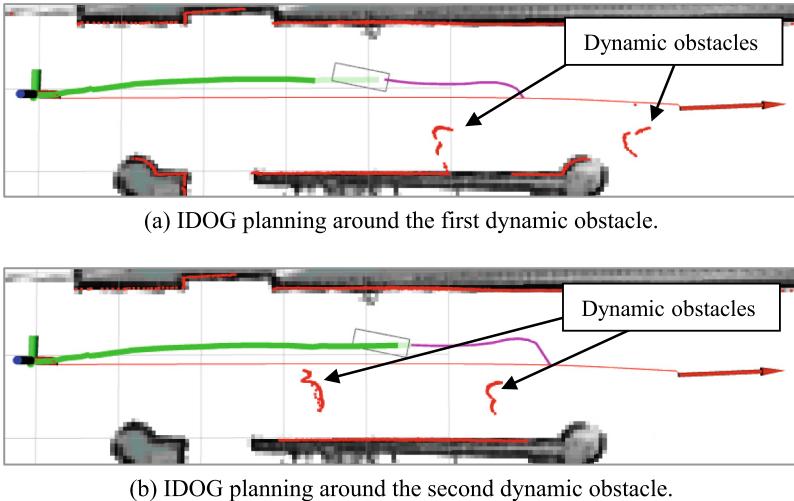


Fig. 5. Dynamic obstacle-avoidance behavior to avoid multiple persons (dynamic obstacles).

spaces simultaneously and necessitate effective global planning as well as robust local planning to adhere to obstacle clearances and to prevent getting stuck in a local minimum.

The validation tests for autonomous navigation demonstrate that the IDOG is able to successfully plan a global path, fine-tune the path locally and navigate in real time, while maintaining enough distance from the boundaries of the door passages of width 1.05 m and 0.8 m, respectively. While the autonomous navigation capabilities of the motion planning framework are satisfactory around obstacles and through narrow pathways, in confined spaces, an oscillatory behavior is observed between the path followed by the IDOG and the global plan (Fig. 6); nonetheless, the IDOG recovers and completes the global plan. The origin of the oscillatory behavior requires further investigation to improve stability of autonomous navigation.

4 Summary and Conclusions

The degradation of civil infrastructure requires continuous monitoring and regular inspections to ensure safe operation, which can be performed by quadruped robots. However, to achieve robot-based automated monitoring and inspection, it is essential that quadruped robots possess the ability to operate autonomously, which requires motion planning frameworks capable of real-time environment perception, localization, path planning, obstacle avoidance, and motion control. The motion planning framework, presented in this study, decouples the motion planning problem into two distinct components, a high-level task-space problem and a low-level joint-space problem for the real-time operation of quadruped robots. The tree-based global planning algorithm BIT*, the A* algorithm, and the TEB local planning approach are proposed for high-level robot control and obstacle avoidance.

The *integrated architecture for automated monitoring and inspection* proposed in this study has been deployed on the IDOG quadruped robot and a computer, using a

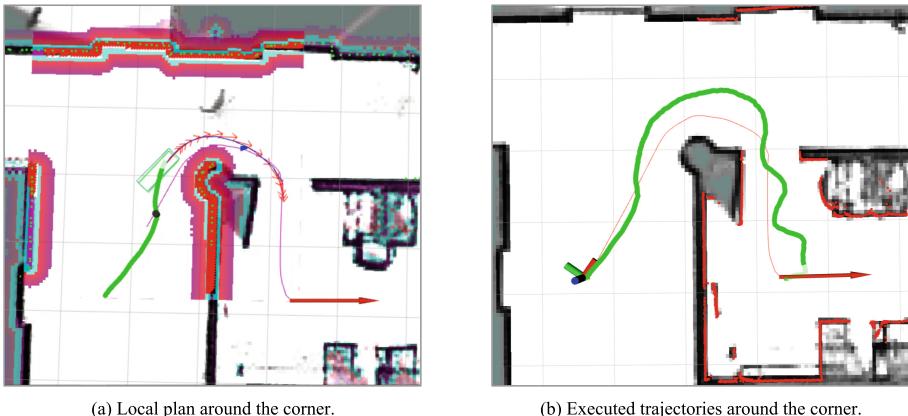


Fig. 6. Execution of a local plan in confined spaces and around corners.

distributed software architecture to enable high-level mission planning on the computer and provide autonomous navigation for the IDOG. The *mission planning framework* is deployed on the computer and uses a state machine to plan monitoring and inspection missions in an indoor office environment with static and dynamic obstacles to be completed by the IDOG. The *motion planning framework* deployed on the IDOG is validated by executing the planned monitoring and inspection missions set up as multi-waypoint missions. The results show that the motion planning framework is capable of real-time operation and successfully avoids static and dynamic obstacles. In addition, narrow doors can be traversed, although some undesirable oscillatory behavior of the robot motion is observed. In conclusion, the integrated architecture for automated monitoring and inspection provides a useful foundation for automated monitoring and inspection of civil infrastructure using quadruped robots. The mission planning framework provides a GUI to plan monitoring and inspection missions intuitively, while the motion planning framework reliably plans and controls the motion of the robot in real-time. Future work will investigate and mitigate the oscillatory behavior of the motion planning framework when traversing narrow spaces. In addition, specific monitoring and inspection tasks may be implemented and added to the mission planning framework, to further advance automated monitoring and inspection of civil infrastructure.

Acknowledgments. The authors gratefully acknowledge the support offered by the German Research Foundation (DFG) through grant SM 281/20-1. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of DFG.

References

1. Smarsly, K., Law, K.H., König, M.: Autonomous structural condition monitoring based on dynamic code migration and cooperative information processing in wireless sensor networks. In: Proceedings of the 8th International Workshop on Structural Health Monitoring 2011. Stanford, CA, USA (2011)

2. Woetzel, J. et al.: Reinventing construction through a productivity revolution, Accessed 03/05/2023 (2017). <https://www.mckinsey.com/capabilities/operations/our-insights/reinventing-construction-through-a-productivity-revolution>
3. Smarsly, K., Dragos, K.: Advancing civil infrastructure assessment through robotic fleets. *Internet Things Cyber-Phys. Syst.* **4**(2024), 138–140 (2024)
4. Smarsly, K., Dragos, K., Stührenberg, J., Worm, M.: Mobile structural health monitoring based on legged robots. *Infrastructures* **8**(9), 136 (2023)
5. Halder, S., Afsari, K.: Robots in inspection and monitoring of buildings and infrastructure: A systematic review. *Appl. Sci.* **13**(4), 2304 (2023)
6. Sánchez-Ibáñez, J.R., Pérez-del Pulgar, C.J., García-Cerezo, A.: Path planning for autonomous mobile robots: A review. *Sensors* **21**(23), 7898 (2021)
7. Wellhausen, L., Hutter, M.: Artplanner: Robust legged robot navigation in the field. *Field Robot.* **3**(1), 413–434 (2023)
8. Morrell, B., et al.: Nebula: Team costar’s robotic autonomy solution that won phase II of Darpa Subterranean challenge. *Field Rob.* **2**, 1432–1506 (2022)
9. Dudzik, T., et al.: Robust autonomous navigation of a small-scale quadruped robot in real-world environments. In: Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, NV, USA (2020)
10. Mastalli, C., Havoutis, I., Focchi, M., Caldwell, D.G., Semini, C.: Motion planning for quadrupedal locomotion: Coupled planning, terrain mapping, and whole-body control. *IEEE Trans. Rob.* **36**(6), 1635–1648 (2020)
11. Stumpf, A., von Stryk, O.: A universal footstep planning methodology for continuous walking in challenging terrain applicable to different types of legged robots. In: Proceedings of the 2022 International Conference on Robotics and Automation. Philadelphia, PA, USA (2022)
12. Liu, M., Qu, D., Xu, F., Zou, F., Di, P., Tang, C.: Quadrupedal robots whole-body motion control based on centroidal momentum dynamics. *Appl. Sci.* **9**(7), 1335 (2019)
13. Ros-perception: pointcloud_to_laserscan (2023). Accessed 12/11/2023. https://github.com/rosp-perception/pointcloud_to_laserscan
14. Hess, W., Kohler, D., Rapp, H., Andor, D.: Real-time loop closure in 2D lidar slam. In: Proceedings of the 2016 International Conference on Robotics and Automation. Stockholm, Sweden (2016)
15. ETHZ-RobotX: smb_ompl_planner (2023). Accessed 12/8/2023. https://github.com/ETHZ-RobotX/smb_path_planner/tree/master/smb_ompl_planner
16. Sucan, I.A., Moll, M., Kavraki, L.E.: The open motion planning library. *IEEE Robot. Autom. Mag.* **19**(4), 72–82 (2012)
17. Rösmann, C., Feiten, W., Wösch, T., Hoffmann, F., Bertram, T.: Trajectory modification considering dynamic constraints of autonomous robots. In: Proceedings of ROBOTIK 2012; 7th German Conference on Robotics. Munich, Germany (2012)
18. Bohren, J., Cousins, S.: The smach high-level executive. *IEEE Robot. Autom. Mag.* **17**(4), 18–20 (2010)



Investigating the Barriers to the Adoption of 3D Printing Technology in the Turkish Construction Industry

Leva Latifilkhechi¹(✉), Saman Aminbakhsh², and Emre Caner Akcay²

¹ Graduate School of Natural and Applied Sciences, Atilim University, Ankara 06830, Turkey
latifilkhechi-leva@student.atilim.edu.tr

² Department of Civil Engineering, Atilim University, Ankara 06830, Turkey

Abstract. In recent years, 3D printing has emerged as a transformative force in construction technology. This innovative technology has found applications in many construction and bridge-building projects, starting a new era of automation and efficiency. Notably, the impact of 3D printing on the construction industry has been profound, yielding benefits such as reduced labor requirements, minimized material waste, accelerated project timelines, and a significant reduction in hazardous tasks for human workers. In contrast to conventional construction approaches, 3D printing stands out for its environmentally friendly characteristics, challenging traditional notions of geometric complexities and constraints in construction processes. While 3D printing technology undeniably offers a multitude of advantages over traditional methodologies, it is crucial to acknowledge that it also introduces its own set of unique challenges and risks. The main aim of this study is to investigate the barriers that hinder the adoption of 3D printing technology in the Turkish construction industry. Towards this, first, the list of potential barriers was extracted from multiple reputable sources, including Scopus, ScienceDirect and Web of Science databases. Based on this list, an online questionnaire was prepared to assess the impact of 19 potential barriers on the implementation of 3D printing technology in the Turkish construction industry. The findings and research directions articulated in this study create fresh pathways for further inquiry and substantial contributions to the evolving field of 3D printing in the construction industry.

Keywords: 3D Printing · Barriers · Relative Importance Index

1 Introduction

The rapid growth of the global population, coupled with escalating construction costs, has presented an ever-pressing challenge of developing novel, cost-effective, and time-efficient construction methods. In this respect, in 1983, Charles W. Hull introduced a ground-breaking concept by proposing the use of UV light to solidify tabletop coatings [1], and in 1998, Behrokh Khoshnevis revolutionized the construction industry with his pioneering contributions to additive manufacturing, widely recognized as 3D printing

[2]. This pioneering endeavor marked the commencement of a new era characterized by extensive research and experimentation aimed at creating innovative construction materials, optimizing 3D printing configurations, investigating novel application methods, and designing structural systems with unparalleled complexity [2].

Although 3D printing has made remarkable steps since its emergence, it was not until 2009 that the construction sector began to witness a significant wave in interest, leading to a substantial uptick in published studies [3]. The world's first 3D-printed office building emerged in early 2015 [4]. Later, Russian civil engineer Andrey Rudenko, pioneered utilization of 3D printing technology in construction of a prototype concrete castle in Minnesota [5]. The aforesaid projects involved the assembly of separately printed components. It was not until 2016 that a new phase in 3D printing emerged, allowing for printing of an entire house directly on the construction site [2].

The construction industry is one of the most important economic pillars of the world [6]. Unfortunately, the engineering and construction sector has different challenges, including higher rates of fatalities, injuries, and accidents compared to other industries [7, 8]. Consequently, construction companies are continually in pursuit of innovative solutions [9].

3D printing in construction represents a digital technology with the capacity to automate conventional building processes [10, 11]. The foremost advantage of 3D printing is its potential to revolutionize an industry by reducing labor-intensive practices and promoting environmental sustainability [12]. Furthermore, it offers a wide range of advantages over traditional production methods, including improved efficiency, heightened design flexibility, reduced material consumption, and opportunities for prestressing, repair, and reinforcement [13]. Notably, 3D printing holds the promise of significantly reducing on-site construction timelines [14]. Despite the many promising advantages, the construction industry encounters various barriers and challenges in adopting the 3D printing method [3]. Lack of knowledge about the sustainability of 3D printing materials [15], significant startup costs associated with this technology [16], and cybersecurity risks associated with 3D printing [17] are among the barriers that can potentially hinder implementation of this revolutionizing method in the construction industry.

Examining studies from various regions enables the identification of the barriers to 3D printing adoption on a global scale. As can be followed from the work of [18], in developed countries, such as Singapore and the United Arab Emirates, a considerable volume of studies is being conducted on 3D printing. Similarly, there is considerable attention given to this area in the US and Australia [19]. While these studies systematically identify barriers in particular countries, a noticeable gap was identified as there is a dearth of such studies pertinent to the Turkish construction industry. Therefore, this study aims to fill this gap by examining the barriers specific to Türkiye. Recognizing and addressing such challenges is imperative to facilitate the seamless integration of 3D printing technology into mainstream construction processes.

Accordingly, this study aims to specifically identify and address the barriers that can potentially pose significant hurdles to the adoption of 3D printing technology within the Turkish construction industry. Since exploring the barriers to the adoption of a technology can contribute to its broader acceptance, in this study, a total of 19 distinct barriers impeding the implementation of 3D printing technology in the Turkish construction

sector were investigated and identified through a comprehensive literature review. By recognizing and understanding these barriers, this study aims to pave the way for tailored and effective strategies to facilitate the practical implementation of 3D printing technology in Turkish construction projects.

The remainder of this paper is structured as follows. In Sect. 2, the Research Methodology is detailed, explaining how the barriers to 3D concrete printing were identified. Section 3 comprises the results of the research which are further discussed and elaborated in Sect. 4. Finally, Sect. 5 concludes this paper by summarizing key findings and suggesting avenues for future research.

2 Research Methodology

As can be followed from Fig. 1, the research methodology employed in this study encompasses four key steps; namely, questionnaire design, data collection, data analysis, and the presentation of results which will be discussed below.

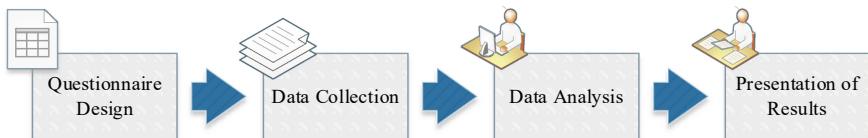


Fig. 1. Research methodology

2.1 Questionnaire Design

According to [20], the successful design of a survey relies significantly on the questionnaire. In this study, the questionnaire design involved a literature review followed by a pilot study. The research commenced with an extensive literature review to identify barriers to the adoption of 3D printing technology in the Turkish construction industry, resulting in 19 barriers. These identified barriers are presented in Table 1, accompanied by the relevant references for further exploration. Exploiting the factors tabulated in Table 1, a questionnaire survey was formulated and subsequently distributed among professionals in the field to gather insights and perspectives. The study adopted a quantitative approach, using closed-ended questions in a two-part questionnaire. The first part gathered demographic and general information, while the second part asked respondents to rate the severity of the 19 barriers on a 5-point Likert scale (1 = Not at all important, 2 = Slightly important, 3 = Moderately important, 4 = Very important, and 5 = Extremely important). In a pilot study, administered to six experts, including academics and experienced professionals, the questionnaire's appropriateness was assessed. The experts provided suggestions, and their responses were analyzed to finalize the questionnaire, incorporating minor modifications for clarity and precision.

Table 1. Barriers to adoption of 3D printing in construction industry.

| Barrier | Description | Reference(s) |
|---------|---|----------------------|
| B1 | Lack of knowledge about sustainability of 3D printing materials | [21, 22] |
| B2 | Limited availability of insulating materials | [23] |
| B3 | Possibility of negative impact on human health due to potentially harmful new substances | [24] |
| B4 | Significant startup costs associated with this technology | [25–28] |
| B5 | Transportation of large 3D printers can be challenging and costly | [2, 27–31] |
| B6 | High costs associated with the concrete used in 3D printing | [16, 32] |
| B7 | Layered and rough surface of the printed structures | [33–35] |
| B8 | Lack of knowledge about mechanical properties of the printed structures | [36–38] |
| B9 | Reinforcement of concrete can be challenging | [25, 30, 39] |
| B10 | Current technology is not ready for printing large-scale buildings | [31, 32] |
| B11 | Lack of government regulations, standardizations, and building codes on 3D Concrete Printing (3DCP) | [13, 24, 31, 40, 41] |
| B12 | Complexity and risks associated with adoption of 3D printing technology in construction industry | [2, 13] |
| B13 | Cybersecurity risks associated with 3D printing | [17, 42] |
| B14 | Lack of knowledge about the usage of 3D printers for large-scale buildings | [43] |
| B15 | Errors in design and/or mistakes in printer settings can easily lead to losses, waste, and rework | [12] |
| B16 | Leads to job losses and difficulties securing employment | [3, 34] |
| B17 | Requires specialized workforce for designing 3D objects and setting up the equipment | [2] |
| B18 | Resistance towards adoption of technological innovation by the industry | [26] |
| B19 | Lack of knowledge about buildability and/or workability of 3D printing materials | [40] |

2.2 Data Collection

The selection of appropriate target respondents plays a crucial role in surveys that involve data collection. This study focused on construction professionals from both public and private sectors. The questionnaires were distributed via email containing a web-based questionnaire link to selected professionals. Out of 107 received responses, only a single

one was incomplete and excluded. Ultimately, 106 valid responses were collected for further analysis.

2.3 Data Analysis

This study aims to quantitatively model the perspectives of the experts, with participants assessing the importance of each factor on a scale of 1 to 5. The study utilizes the Relative Importance Index (RII) approach, a common and effective ranking method, to assess the significance of barriers impacting the implementation of 3D printing technology in construction. The RII incorporates the importance derived from response scores, as per Eq. (1).

$$RII(\%) = \frac{\sum S}{H \times N} \quad (1)$$

where,

$\sum S$ = the summation of each importance score multiplied by its number of responses.

H = highest possible number (five in this study).

N = total number of respondents (106 in this study).

For example, the RII for the factor “B1: Lack of knowledge about the sustainability of 3D printing materials” is calculated as shown in Eq. (2).

$$RII(\%) = \frac{(40 \times 5) + (36 \times 4) + (22 \times 3) + (6 \times 2) + (2 \times 1)}{5 \times 106} = 0.8 \quad (2)$$

The analysis of the RII is calculated in the same manner for each of the 19 factors, as detailed in Table 2. The factors are then ranked based on their respective RII values in descending order of importance.

3 Results

3.1 Demographic Characteristics of Respondents

This section summarizes the demographic characteristics of the respondents. As shown in Fig. 2, the majority of the respondents were academicians, project managers, and design engineers constituting 23.81%, 15.24%, and 14.29%, respectively. The remaining percentages were distributed among owners, technical managers, and individuals with other job designations. The responses from individuals with other job designations amounted to 20.00%, highlighting the broad representation of professional perspectives in our study. Categorization based on participant characteristics revealed that 41.90% comprised those with 6–15 years of experience, while 32.38% accounted for participants with 1–5 years of experience. In terms of company size, 60.95% of responses were associated with large companies employing more than 500 individuals, followed by 18.10% associated with small-sized companies, having below 100 employees. This demographic diversity provides valuable insights into the industry landscape and sets the stage for comprehending the relative significance of the identified barriers.

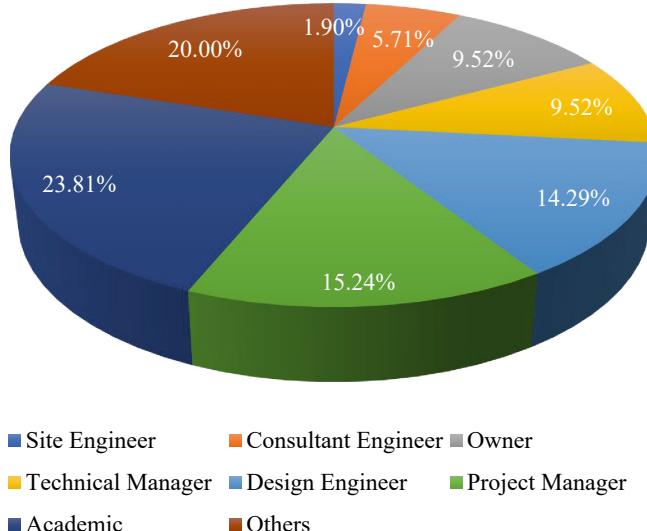


Fig. 2. Respondents according to their job description

3.2 Barriers to Adoption of 3D Printing Technology

Table 2 shows RII values in percentages and the ranking of the 19 factors, derived from completed responses from all 106 questionnaire participants in the context of this study. Based on the results presented in Table 2, the three most and the three least important barriers in the Turkish construction industry were determined as follows. “B11: Lack of government regulations, standardizations, and building codes on 3D Concrete Printing (3DCP)” ranked first according to the RII score, and the factor “B10: Current technology is not ready for printing large-scale buildings” ranked second which was followed by “B14: Lack of knowledge about the usage of 3D printers for large-scale buildings” taking up the third spot. On the other hand, the factor “B13: Cybersecurity risks associated with 3D printing” ranked the least important, followed by “B16: Leads to job losses and difficulties in securing employment” and “B7: Layered and rough surface of the printed structures” barriers.

3.3 Perceptions of Barrier Significance Within Private and Public Sectors

The scatter plot shown in Fig. 3 portrays the perceived importance of various barriers, collected based on responses from both private and public sectors. In this scatter plot the RII scores are plotted against barriers B1 to B19 which visually outlines the contrasting perceptions regarding the significance of barriers to 3D printing in construction, relative to the sector type. As can be followed from this figure, significantly higher RII scores have been assigned to the barriers by the respondents from public organizations. This can potentially be interpreted as public organizations tending to favor maintaining the status-quo, displaying a resistance to change or the adoption of new technologies, thereby exhibiting a lack of dynamism. Notably, the most significant difference in viewpoints is observed in “B6: High costs associated with the concrete used in 3D printing”.

Table 2. RII scores and ranks of the barriers.

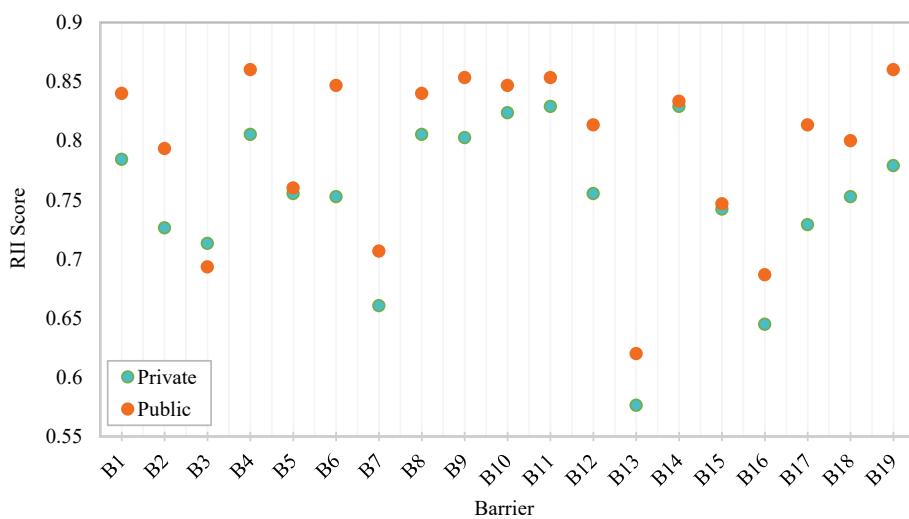
| Rank | Barrier | Description | RII (%) |
|------|---------|---|------------|
| 1 | B11 | Lack of government regulations, standardizations, and building codes on 3D Concrete Printing (3DCP) | 0.83584906 |
| 2 | B10 | Current technology is not ready for printing large-scale buildings | 0.83018868 |
| 3 | B14 | Lack of knowledge about the usage of 3D printers for large-scale buildings | 0.83018868 |
| 4 | B4 | Significant startup costs associated with this technology | 0.82075472 |
| 5 | B9 | Reinforcement of concrete can be challenging | 0.81698113 |
| 6 | B8 | Lack of knowledge about mechanical properties of the printed structures | 0.81509434 |
| 7 | B19 | Lack of knowledge about buildability and/or workability of 3D printing materials | 0.80188679 |
| 8 | B1 | Lack of knowledge about the sustainability of 3D printing materials | 0.80000000 |
| 9 | B6 | High costs associated with the concrete used in 3D printing | 0.77924528 |
| 10 | B12 | Complexity and risks associated with adoption of 3D printing technology in construction industry | 0.77169811 |
| 11 | B18 | Resistance towards adoption of technological innovation by the industry | 0.76603774 |
| 12 | B5 | Transportation of large 3D printers can be challenging and costly | 0.75660377 |
| 13 | B17 | Requires specialized workforce for designing 3D objects and setting up the equipment | 0.75283019 |
| 14 | B2 | Limited availability of insulating materials | 0.74528302 |
| 15 | B15 | Errors in design and/or mistakes in printer settings can easily lead to losses, waste, and rework | 0.74339623 |
| 16 | B3 | Possibility of negative impact on human health due to potentially harmful new substances | 0.70754717 |

(continued)

Table 2. (*continued*)

| Rank | Barrier | Description | RII (%) |
|------|---------|---|------------|
| 17 | B7 | Layered and rough surface of the printed structures | 0.67358491 |
| 18 | B16 | Leads to job losses and difficulties in securing employment | 0.65660377 |
| 19 | B13 | Cybersecurity risks associated with 3D printing | 0.58867925 |

Following this, the second-largest difference in RII scores between the two sectors is present in “B19: Lack of knowledge about buildability and/or workability of 3D printing materials”. “B13: Cybersecurity risks associated with 3D printing” is considered to have the lowest RII score by the public organizations, while B19 is regarded as the most crucial barrier. On the other hand, respondents from the private organizations appear to perceive the barriers as less stringent as they have assigned a lower level of significance to the barriers compared to those in the public sector. They perceive “B13: Cybersecurity risks associated with 3D printing” to be of the least significance with the lowest RII score; whilst, “B11: Lack of government regulations, standardizations, and building codes on 3D Concrete Printing (3DCP)” is identified to be the most significant barrier.

**Fig. 3.** Significance of the barriers according to sector type

4 Discussion

In the following section, a discussion will be provided by pinpointing some strategic recommendations for addressing the first three barriers of paramount significance.

According to the results of the survey, there is a consensus regarding “B11: Lack of government regulations, standardizations, and building codes on 3D Concrete Printing (3DCP)” being the most significant barrier to the adoption of 3D printing. The absence of existing guidelines on the utilization of 3D printing in construction in Türkiye is currently evident. In this regard, there is an undeniable need for drafting codes on the implementation of 3D printers in Turkish construction projects that will address the safety and sustainability aspects [44]. Codes devoid of any environmental impact are deemed necessary. For this, the compilation of data from around the world is deemed necessary to identify all the potential risks and challenges for the development of such codes.

Another important barrier that experts agree on is “B10: Current technology is not ready for printing large-scale buildings”. The emphasis of the absence of large-scale 3D printing machines is influenced by the inherent limitations of the nozzles employed [31], which are deemed unsuitable for utilization in extensive construction projects. The recognition of a need for technological and engineering advancements to address this interval is apparent, as the requirements of large-scale 3D construction are inadequately met by the existing machinery [30]. Uncertainty persists in aspects such as material utilization and implementation speed, along with associated costs, printer setup, and transportation in the context of large-scale 3D constructions [2].

The third most important barrier is found to be “B14: Lack of knowledge about the usage of 3D printers for large-scale buildings”. Due to a deficiency in information regarding the materials employed in large-scale 3D printing for construction, the clarity of material utilization, implementation, and associated costs is impeded. Furthermore, knowledge is absent regarding the environmental impact in the domain of large-scale constructions [24]. Conducting comprehensive laboratory tests and experiments is imperative for a thorough evaluation and analysis of the 3D printing materials, unveiling their behavior, durability, performance, and other scientific attributes [2, 45].

Finally, it is worth noting that the public sector generally displays a tendency towards maintaining the status quo, and seems not to rush to embrace new technologies unlike the private sector. This interpretation is made in reference to the scores assigned to all the barriers with the exception of “B3: Possibility of negative impact on human health due to potentially harmful new substances”. In contrast to the rest of the barriers, respondents from private organizations have assigned a higher score to B3 which can mainly be attributed to their unique experiences, values, and risk perceptions.

5 Conclusions

With the advent of 3D printing, a new era of automation and efficiency has commenced in numerous industries. Thanks to its benefits such as reduced labor requirements, minimized material waste, accelerated project timelines, and reduced hazardous tasks, 3D printing technology has found its place in the construction industry as well. While

3D printing technology undeniably offers a multitude of advantages over traditional methodologies, it is crucial to acknowledge that similar to any emerging technology, it comes with its inherent challenges, and it will require a considerable amount of time for the industry to effectively address these barriers and fully integrate it into widespread practice. This study, therefore, aimed to investigate the barriers that could hinder the adoption of 3D printing technology with a particular focus on the Turkish construction industry. Towards this, a questionnaire was designed enlisting a total of 19 barriers which were identified through a comprehensive literature review. Expert judgments were collected using the questionnaire survey. Results of the survey were analyzed employing the widely practiced Relative Importance Index (RII) method whereby the significance and priority of the barriers were identified. Guidelines were provided to overcome the foremost barriers, facilitating broader adoption of 3D printing technology within the Turkish construction industry. While the study was conducted in Türkiye, the recommendations presented in this research may have relevance to other developing nations sharing similar socioeconomic conditions. Since, the RII approach lacks consideration for the relationships between the factors and the ranking relies merely on the raw data from the questionnaire, it can be regarded as the limitation of the current study. Therefore, future research could focus on methodologies taking into account the relationships between the factors (e.g., analytic network process (ANP)) within the specified categories. This study focused solely on exploring barriers to 3D printing technology. However, future research endeavors may delve into investigating the drivers associated with its implementation as well.

References

1. CNN Business. <https://edition.cnn.com/2014/02/13/tech/innovation/the-night-i-invented-3d-printing-chuck-hall/>, last accessed 2024/01/20
2. Besklubova, S., Skibniewski, M.J., Zhang, X.: Factors affecting 3D printing technology adaptation in construction. *J. Constr. Eng. Manag.* **147**(5), (2021) [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002034](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002034)
3. Tay, Y.W.D., Panda, B., Paul, S.C., Noor Mohamed, N.A., Tan, M.J., Leong, K.F.: 3D printing trends in building and construction industry: a review. *Virt. Phys. Prototyp.* **12**(3), (2017). <https://doi.org/10.1080/17452759.2017.1326724>
4. Starr, M.: Dubai unveils world's first 3D-printed office building. CNET (2016). <https://www.cnet.com/news/dubai-unveils-worlds-first-3d-printed-office-building>
5. Baradi, K.: 3D printing as a construction process for structural members (2016). [www.academia.edu.](http://www.academia.edu/) https://www.academia.edu/en/84593993/3D_Printing_as_a_Construction_Process_for_Structural_Members
6. Horvath, J.: Mastering 3D printing. Berkeley, CA: Apress (2014). <https://doi.org/10.1007/978-1-4842-0025-4>
7. Ahmadisheykhsarmast, S., Aminbakhsh, S., Sonmez, R., Uysal, F.: A transformative solution for construction safety: Blockchain-based system for accident information management. *J. Ind. Inf. Integr.* **35**, 100491 (2023). <https://doi.org/10.1016/j.jii.2023.100491>
8. Famakin, I.O., Aigbavboa, C., Molusiwa, R.: Exploring challenges to implementing health and safety regulations in a developing economy. *Int. J. Constr. Manag.* **23**(1), 89–97 (2020). <https://doi.org/10.1080/15623599.2020.1850201>

9. Maskuriy, R., Selamat, A., Maresova, P., Krejcar, O., David, O.O.: Industry 4.0 for the construction industry: Review of management perspective. *Economies* **7**(3), 68 (2019). <https://doi.org/10.3390/economies7030068>
10. Buswell, R.A., Thorpe, A., Soar, R.C., Gibb, A.G.F.: Design, data and process issues for mega-scale rapid manufacturing machines used for construction. *Autom. Constr.* **17**(8), 923–929 (2008). <https://doi.org/10.1016/j.autcon.2008.03.001>
11. Lim, S., Buswell, R.A., Le, T.T., Austin, S.A., Gibb, A.G.F., Thorpe, T.: Developments in construction-scale additive manufacturing processes. *Autom. Constr.* **21**, 262–268 (2012). <https://doi.org/10.1016/j.autcon.2011.06.010>
12. Zuo, Z., Gong, J., Huang, Y., Zhan, Y., Gong, M., Zhang, L.: Experimental research on transition from scale 3D printing to full-size printing in construction. *Constr. Build. Mater.* **208**, 350–360 (2019). <https://doi.org/10.1016/j.conbuildmat.2019.02.171>
13. Buchanan, C., Gardner, L.: Metal 3D printing in construction: A review of methods, research, applications, opportunities and challenges. *Eng. Struct.* **180**, 332–348 (2019). <https://doi.org/10.1016/j.engstruct.2018.11.045>
14. Hou, L., Tan, Y., Luo, W., Xu, S., Mao, C., Moon, S.: Towards a more extensive application of off-site construction: a technological review. *Int. J. Constr. Manag.* **22**(11), 2154–2165 (2022). <https://doi.org/10.1080/15623599.2020.1768463>
15. Labonne, N., Rønquist, A., Manum, B., Rüther, P.: Additive construction: State-of-the-art, challenges and opportunities. *Autom. Constr.* **72**, 347–366 (2016). <https://doi.org/10.1016/j.autcon.2016.08.026>
16. Han, Y., Yang, Z., Ding, T., Xiao, J.: Environmental and economic assessment on 3D printed buildings with recycled concrete. *J. Clean. Prod.* **278**, 123884 (2021). <https://doi.org/10.1016/j.jclepro.2020.123884>
17. Berman, B.: 3-D printing: The new industrial revolution. *Bus. Horiz.* **55**(2), 155–162 (2012). <https://doi.org/10.1016/j.bushor.2011.11.003>
18. Rajan, A., Akre, V., Nassiri, N., AlAli, A.H., Sabt, Z.B.: 3D printing of buildings in UAE: Success and failure factors. In: 2018 Fifth HCT information technology trends (ITT), pp. 368–372, IEEE (2018). <https://doi.org/10.1109/CTIT.2018.8649541>
19. Alzarrad, M.A., Elhouar, S.: 3D printing applications in construction from the past and into the future. In: Proceedings of the Creative Construction Conference 2019, Budapest University of Technology and Economics, pp. 754–760 (2019). <https://doi.org/10.3311/CCC2019-103>
20. Arantes, A., Ferreira, L.M.D.F.: Underlying causes and mitigation measures of delays in construction projects: An empirical study. *J. Financ. Manage. Prop. Constr.* **25**(2), 165–181 (2020). <https://doi.org/10.1108/JFMP-03-2019-0029>
21. Liu, Z., Jiang, Q., Zhang, Y., Li, T., Zhang, H.: Sustainability of 3D printing: A critical review and recommendations. In: Proceedings of the ASME 2016 11th International Manufacturing Science and Engineering Conference. Volume 2: Materials; Biomanufacturing; Properties, Applications and Systems; Sustainable Manufacturing. Blacksburg, Virginia, USA. June 27–July 1, 2016 (2016). <https://doi.org/10.1115/MSEC2016-8618>
22. Ayegba, B.O., Egbe, K.-J.I., Nazar, A., Huang, M., Hariri-Ardebili, M.A.: Resource efficiency and thermal comfort of 3D printable concrete building envelopes optimized by performance enhancing insulation: A numerical study. *Energies* **15**(3), 1069 (2022). <https://doi.org/10.3390/en15031069>
23. Yang, S., Wi, S., Park, J.H., Cho, H.M., Kim, S.: Novel proposal to overcome insulation limitations due to nonlinear structures using 3D printing: Hybrid heat-storage system. *Energy Build.* **197**, 177–187 (2019). <https://doi.org/10.1016/j.enbuild.2019.05.048>
24. Ning, X., Liu, T., Wu, C., Wang, C.: 3D printing in construction: Current status, implementation hindrances, and development agenda. *Adv. Civil Eng.* **2021**, 1–12 (2021). <https://doi.org/10.1155/2021/6665333>

25. Teizer, J., Bickle, A., King, T., Leitzbach, O., Guenther, D.: Large scale 3D printing of complex geometric shapes in construction. In: 2016 Proceedings of the 33rd ISARC (2016). <https://doi.org/10.22260/ISARC2016/0114>
26. Olsson, N.O.E., Shafqat, A., Arica, E., Økland, A.: 3D-printing technology in construction: Results from a survey. In: 10th Nordic Conference on Construction Economics and Organization, pp. 349–356 (2019). <https://doi.org/10.1108/S2516-285320190000002044>
27. Camacho, D.D., et al.: Applications of additive manufacturing in the construction industry – A forward-looking review. *Autom. Constr.* **89**, 110–119 (2018). <https://doi.org/10.1016/j.autcon.2017.12.031>
28. Bazli, M., Ashrafi, H., Rajabipour, A., Kutay, C.: 3D printing for remote housing: Benefits and challenges. *Autom. Constr.* **148**, 104772 (2023). <https://doi.org/10.1016/j.autcon.2023.104772>
29. Anton, A., et al.: Concrete choreography: Prefabrication of 3D-printed columns. In *Fabricate 2020*, UCL Press, pp. 286–293 (2020). <https://doi.org/10.2307/j.ctv13xpsvw.41>
30. Xiao, J., et al.: Large-scale 3D printing concrete technology: Current status and future opportunities. *Cem. Concr. Compos.* **122**, 104115 (2021). <https://doi.org/10.1016/j.cemconcomp.2021.104115>
31. El-Sayegh, S., Romdhane, L., Manjikian, S.: A critical review of 3D printing in construction: Benefits, challenges, and risks. *Arch. Civil Mech. Eng.* **20**, 1–25 (2020). <https://doi.org/10.1007/s43452-020-00038-w>
32. Puzatova, A., Shakor, P., Laghi, V., Dmitrieva, M.: Large-scale 3D printing for construction application by means of robotic arm and gantry 3D printer: A review. *Buildings* **12**(11), 2022 (2023). <https://doi.org/10.3390/buildings12112023>
33. Ghaffar, S.H., Corker, J., Fan, M.: Additive manufacturing technology and its implementation in construction as an eco-innovative solution. *Autom. Constr.* **93**, 1–11 (2018). <https://doi.org/10.1016/j.autcon.2018.05.005>
34. Hager, I., Golonka, A., Putanowicz, R.: 3D printing of buildings and building components as the future of sustainable construction? *Proc. Eng.* **151**, 292–299 (2016). <https://doi.org/10.1016/j.proeng.2016.07.357>
35. Scott, J., Gupta, N., Weber, C., Newsome, S., Wohlers, T., Caffrey, T.: Additive manufacturing: Status and opportunities. Science and Technology Policy Institute (2012)
36. Liu, K., Takasu, K., Jiang, J., Zu, K., Gao, W.: Mechanical properties of 3D printed concrete components: A review. *Develop. Built Environ.* **16**, 100292 (2023). <https://doi.org/10.1016/j.dibe.2023.100292>
37. Rouf, S., Raina, A., Irfan Ul Haq, M., Naveed, N., Jeganmohan, S., Kichloo, A.F.: 3D printed parts and mechanical properties: Influencing parameters, sustainability aspects, global market scenario, challenges and applications. *Adv. Ind. Eng. Polym. Res.* **5**(3), 143–158 (2022). <https://doi.org/10.1016/j.aiepr.2022.02.001>
38. Zhu, B., Pan, J., Zhou, Z., Cai, J.: Mechanical properties of engineered cementitious composites beams fabricated by extrusion-based 3D printing. *Eng. Struct.* **238**, 112201 (2021). <https://doi.org/10.1016/j.engstruct.2021.112201>
39. Kothman, I., Faber, N.: How 3D printing technology changes the rules of the game: Insights from the construction sector. *J. Manuf. Technol. Manag.* **27**(7), 932–943 (2016). <https://doi.org/10.1108/JMTM-01-2016-0010>
40. Camacho, D.D., et al.: Applications of additive manufacturing in the construction industry – A prospective review. In: 2017 Proceedings of the 34rd ISARC, Taipei, Taiwan (2017). <https://doi.org/10.22260/ISARC2017/0033>
41. Ali, M.H., Issayev, G., Shehab, E., Sarfraz, S.: A critical review of 3D printing and digital manufacturing in construction engineering. *Rapid Prototyp. J.* **28**(7), 1312–1324 (2022). <https://doi.org/10.1108/RPJ-07-2021-0160>

42. Fey, M.: 3D printing and international security: Risks and challenges of an emerging technology. PRIF Report No. 144, Frankfurt/M (2017)
43. Wu, P., Wang, J., Wang, X.: A critical review of the use of 3-D printing in the construction industry. *Autom. Constr.* **68**, 21–31 (2016). <https://doi.org/10.1016/j.autcon.2016.04.005>
44. Mogaji, I.J., Mewomo, M.C., Toyin, J.O.: Key barriers to the adoption of 3D printing innovation in construction: A review of empirical studies. In: The Thirteenth International Conference on Construction in the 21st Century (CITC-13) Arnhem, The Netherlands (2023)
45. Ur Rehman, A., Sglavo, V.M.: 3D printing of geopolymers-based concrete for building applications. *Rapid Prototyp. J.* **26**(10), 1783–1788 (2020). <https://doi.org/10.1108/RPJ-09-2019-0244>



Electrodermal Activity as an Indicator of Cognitive Load for Human-Robot Collaboration in Construction

Seulbi Lee¹ and Bogyeong Lee^{2(✉)}

¹ Division of Architecture & Urban Design, Incheon National University, Incheon, South Korea

² Department of ICT Integrated Ocean Smart City Engineering, Dong-A University, Busan,
South Korea

blee830@dau.ac.kr

Abstract. Human-Robot Collaboration (HRC), where humans and robots work together, is revolutionizing construction site operations. While HRC offers significant productivity gains by reducing the physical demands on workers, it also presents a new challenge: cognitive overload. This study explores Electrodermal Activity (EDA) as a measurable indicator of the cognitive load experienced by workers during their interaction with robots. Utilizing a robotic arm and wearable wristbands for experiments, EDA signals were recorded from two groups: the HRC assembly group and the solely assembly group. The results revealed a noticeable difference in EDA levels between the participants in the HRC setup compared to those working independently. Specifically, the Skin Conductance (SC) feature, normalized with baseline data, outperformed other EDA signals such as SCL, PhasicMax, ISCR, and SCR, achieving an F-score of 79%. By contributing to the broader discussion on adopting HRC in the construction industry, this research highlights the importance of developing strategies to alleviate cognitive load on workers, thereby maximizing the benefits of HRC.

Keywords: Human-Robot Collaboration · Cognitive Load · Electrodermal Activity

1 Introduction

The construction industry faces several labor-related challenges. According to the Survey of Occupational Injuries and Illnesses (SOII) conducted by the United States Bureau of Labor Statistics, there were 19,380 cases of work-related musculoskeletal disorders (WMSDs) in the construction sector in 2018, yielding approximately 29 cases per 10,000 construction workers [1]. As one potential method to alleviate the physical strain on construction workers, the adoption of robotic technology in construction tasks has gained significant attention [2, 3]. Robots are not subject to fatigue or injury in the same manner as humans, rendering them well-suited for tasks like lifting heavy materials and executing high-risk activities such as demolition and excavation [4]. In addition, robots can significantly enhance efficiency, especially in tasks that are simple and repetitive [5].

For example, commercial bricklaying robots named SAM100 operate using a robotic arm capable of lifting bricks, applying mortar, and placing each brick precisely. However, the operation of robots and the oversight of overall project management tasks still fall within the purview of construction workers. This suggests that a fundamental characteristic of construction robots is their ability to collaborate with humans, serving to complement physically demanding tasks [3]. Therefore, the most preferable scenario involves humans working alongside robots as partners and managers. This concept is commonly referred to as Human-Robot Collaboration (HRC).

HRC has brought substantial changes to the construction work environment, particularly in terms of reducing the physical burden attributed to construction workers. However, concurrently, construction workers who collaborate with robots have found themselves facing unexpected challenges. For example, when sharing physical workspace with robots, workers experienced fear of collisions [6]. Poorly designed user interfaces contributed to workers' discomfort in operating robots [7]. Given that HRC is not yet fully matured in practice, certain workers experienced mental stress due to errors and malfunctions [8]. Recognizing such adverse effects, previous researchers have attempted to measure the fear, discomfort, and stress experienced by construction workers engaged in HRC. Various physiological signals, such as electroencephalogram (EEG), Photoplethysmography (PPG), and electrodermal activity (EDA), have been employed as reliable and objective methods to measure human emotion [9]. Among these physiological signals, EEG signals have garnered the most attention due to their strong correlation with the cognitive load required to complete tasks. For instance, in [2], the authors presented a system to decode human cognitive load from EEG signals and adjust robot speed when the cognitive load exceeds certain levels. However, although EEG has been widely adopted in research and applications, there remains a lack of consensus on how EDA signals should be interpreted and utilized for assessing cognitive load.

To bridge this gap, the study aims to investigate the potential of Electrodermal Activity (EDA) as an indicator for representing cognitive load experienced by workers during collaboration with robots. The experiment is conducted in an environment simulating a bricklaying robot used on construction sites, with participants performing a masonry task while interacting with a robot arm. The remainder of this paper is organized as follows: The second section reviews the literature related to EDA and cognitive load. The third section explains the experiment setup and the data collection process for EDA. The fourth section presents the results in terms of the effectiveness of EDA as an indicator of cognitive load. Finally, the last section summarizes the main findings of the study.

2 Literature Review

2.1 Utilization of Electrodermal Activity in Human-Robot Collaboration

EDA refers to changes in the electrical properties of the skin in response to sweat gland activity [10]. When humans encounter new and unfamiliar working conditions, such as collaborating with robots, their sympathetic nervous system becomes activated, leading to increased activity of sweat glands. Therefore, EDA is considered a valid physiological response of humans to stress. For instances, Dehais et al. [6] found that rapid movements of robot arms towards subjects' faces elicited strong EDA signals.

Several researchers deconvolved EDA signals into two components, tonic and phasic, for more reliable analysis. In general, the phasic component of EDA, known as the skin conductance response (SCR), has been considered a more suitable index for measuring reactions to external stimuli. Gervasi et al. [5] observed a higher average SCR during manual assembly compared to HRC assembly for the same task. They also noted that this difference was considerably reduced by the end of the task due to the learning effect. Hanajima et al. [11] observed a decrease in the average SCR when subjects blocked either the sound of the robot or the visual perception of robot motion, suggesting the potential for user comfort. Conversely, Podevijn et al. [12] utilized the tonic component, known as skin conductance level (SCL), to investigate whether human cognitive load is affected by the number of robots. The results of their study showed that a stronger SCL responses occur when confronted with 24 robots than 3 robots. They also explained that the peak of SCL appears within the first 10 s and subsequently decreases, regardless of the number of robots.

On the other hand, some researchers did not find meaningful differences in human physiological responses during HRC. Yang and Dorneich [13] reported a slight increase in EDA signals when controlling a robot vehicle in a complex maze compared to an easy maze, but no significant effect was observed. Szczurek et al. [7] compared a 2D interface on a laptop screen with a 3D interface using Mixed-reality for controlling robots; however, unfortunately, no significant patterns were observed in either the average SCR or SCL. These inconsistencies in the research findings highlight the insufficiently explored domain of physiological responses, particularly EDA signals, in assessing human cognitive load in HRC.

2.2 Features and Normalization Methods for Electrodermal Activity

The complexity of EDA signals necessitates a series of preprocessing steps for feature extraction, yet consensus on the optimal methodologies for these tasks remains elusive. Moreover, although there is widespread agreement on the need for normalization to adjust for individual differences in responses to the same stimuli, there is also no standard process for normalization. Accordingly, various researchers have adopted different normalization techniques for EDA signals to account for individual differences. Some have utilized range correction, which factors in the minimum and maximum amplitude levels. Log-transformation has been employed by some to address the skewness inherent in EDA data. Others have preferred the use of z-scores, which are calculated using the mean and standard deviation, either across the whole population or within individual participants. Additionally, normalization often involved the use of baseline data, collected in no-stimulus conditions. Table 1 lists the different features extracted from EDA signals, and Table 2 outlines the normalization methods utilized to address individual differences.

Table 1. Features extracted from EDA signals.

| Measure | Description | Unit | References |
|---------------|--|---------------------------|----------------|
| SC (EDA, GSR) | Skin conductance, Electrodermal activity, Galvanic skin resistance | μS | [3, 6, 13–15] |
| SCL (Tonic) | Tonic component of EDA, Skin conductance level | μS | [7, 12] |
| SCR | Phasic component of EDA, Skin conductance response | μS | [5, 7, 11, 15] |
| ISCR | The area of SCR peaks within the response window | $\mu\text{S}^*\text{sec}$ | [15, 16] |
| PhasicMax | The maximum value of SCR peaks within the response window | μS | [17] |

Table 2. Normalization methods to address individual differences.

| Method | Equation | References |
|---------------------|--|-------------|
| Z-score | $(X - \mu)/\sigma$ | [15, 17] |
| Log transformation | $(\log_{10}X - q_{0.05})/(q_{0.95} - q_{0.05})$ | [14] |
| Baseline adjustment | $(X - \mu_{\text{baseline}})/\sigma_{\text{baseline}}$ | [7, 12, 13] |

* X: The value of the EDA feature at a specific time, μ and σ : the mean and standard deviation of EDA feature values, $q_{0.95}$ and $q_{0.05}$: 95% and 5% quantile values of log-transformed EDA feature, μ_{baseline} and σ_{baseline} : the mean and standard deviation of EDA feature values in the no stimulus condition

3 Experiment

3.1 Participants

A total of 12 participants were involved in the experiment. To minimize individual differences, the participants consisted only of male aged between 23 and 26 years. None of the participants had prior experience with operating robot arms. The participants were randomly divided into two groups: the HRC assembly group and the solely assembly group. Each tasked with assembling blocks according to instructions, with the former group receiving assistance from a robot pre-programmed for this research and the latter group working without such aid. All participants were informed about the experiment procedures before its commencement. The experiment took place from September 15 to September 21, 2023. Participation was voluntary, and no identifiable information was collected throughout the experiment. The experiment protocol received approval from the Institutional Review Board of Incheon National University (Approval No.: 7007971-202305-011A).

3.2 Tasks

The experiment was conducted in a laboratory. As illustrated in Fig. 1(b), the experimental setup was designed to separate the robot and human participants using a shelf on

the desktop, thus minimizing the risk of collision. The procedure for the Human-Robot Collaboration (HRC) assembly group was involved multiple steps. To begin, participants were given instructions to identify and assemble blocks, each being one of seven different colors, using a total of 56 blocks for the task. Participants interacted with a tablet screen, pressing buttons that corresponded to the colors of the blocks they needed. Once a button was pressed, the robot received the command, located the specified colored block, picked it up, and placed it on the shelf. This robotic arm movement took approximately 5 s per cycle. After the robot placed the block on the shelf, participants then took the block from the shelf to proceed with the assembly. This process was repeated until the assembly task was complete. This sequence was repeated until the assembly of the blocks was completed. In contrast, the control group undertook the same task entirely manually, without any robotic assistance. Before the experiment began, participants from both groups were given a rest period of about 6 min to collect baseline data. The time required to complete the assembly was also recorded as a measure of productivity.

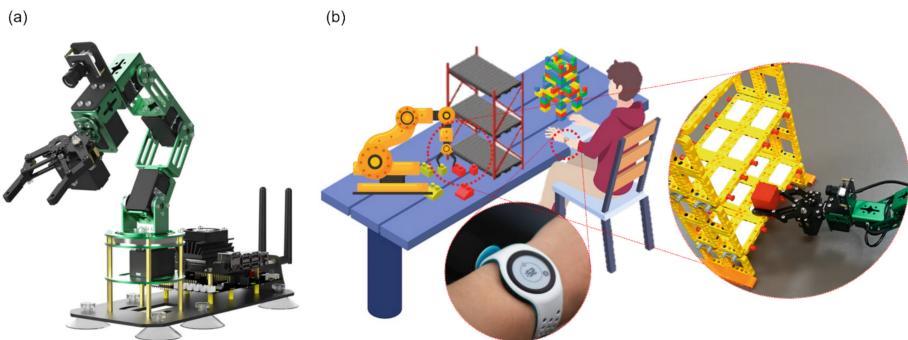


Fig. 1. (a) Robot used for this study, (b) Laboratory experimental settings.

3.3 Data Collection and Analysis

Throughout the experiment, participants in both groups wore a wristband (named Empatica Embrace Plus) to record their EDA signals. To improve the quality of the raw EDA data, which was recorded at a 4 Hz sampling rate, two preprocessing steps were applied. The first preprocessing stage involved the removal of artifacts in the EDA signals that were caused by poor electrode contact with the skin (i.e., skin temperatures below 30 °C) or by excessive movement of the participants (i.e., EDA changes faster than $\pm 10 \mu\text{S/sec}$). The second preprocessing step was the deconvolution of the EDA signals into tonic and phasic components. For the time-domain analysis, LEDALAB (<http://www.ledalab.de/>), a MATLAB-based tool commonly used by researchers to decompose skin conductance data, was utilized [5, 10, 15]. Before performing continuous deconvolution analysis with LEDALAB, the EDA data were filtered through a 4th-order Butterworth low-pass filter with a 0.5 Hz cutoff frequency. The data were then down sampled to 2Hz and smoothed using an adaptive Gaussian window. All other parameters for the continuous deconvolution analysis were maintained at their default values. The study set the time window

at 5 s, excluding the first and last 5 s of data to account for the physiological activity's time delay. All preprocessed data, as previously described in Table 1, were normalized to accommodate individual differences using four methods, as detailed in Table 2.

To determine the most effective features and normalization methods for evaluating cognitive load in Human-Robot Collaboration (HRC), this study utilized machine learning models, specifically Random Forest classifier [18]. The primary objective of this model was to classify participants into two distinct groups based on their physiological responses, as measured by electrodermal activity (EDA) signals. Subsequent analysis focused on comparing the EDA signals of the two groups to ascertain whether collaboration with robots significantly impacts human physiology.

4 Results

4.1 Task Completion Efficiency

The HRC assembly group and the solely assembly group exhibited significant differences in completion times. On average, the solely assembly group completed the assembly task in 203 ± 47 s, while the HRC assembly group required 415 ± 55 s to finish the same task. This outcome seems counterintuitive, as one might anticipate robotic assistance to enhance efficiency. Several factors could explain this discrepancy. The experiment's relatively short duration and the participants' status as first-time users of the robot likely contributed significantly. The time participants needed to adapt to effectively collaborating with the robot could have negated any potential time savings, thereby affecting the efficiency observed in the HRC group. Furthermore, the procedure involving the selection of a block on the tablet, awaiting the robot's action to pick and place the block, and subsequently assembling it, might have introduced extra steps or delays absent in the manual task, further contributing to the longer completion times.

4.2 Performance of EDA Features

Table 3 summarizes the performance metrics—Accuracy, F-score, and Area Under the ROC Curve (AUC)—of a Random Forest model applied to various EDA features across three normalization methods. The results consistently highlight SC as the most reliable indicator of human cognitive load in HRC, followed in effectiveness by SCL, PhasicMax, ISCR, and SCR, respectively. Specifically, SC has an average accuracy of 76.7% (SCL: 70.0%, PhasicMax: 66.7%, ISCR: 66.7%, and SCR: 63.0%). The average F-score for SC is 76.0% (SCL: 70.7%, PhasicMax: 66.0%, ISCR: 65.7%, and SCR: 62.0%). Furthermore, the AUC for SC stands at 80.0% (SCL: 77.0%, PhasicMax: 68.0%, ISCR: 67.3%, SCR: 57.0%). Regarding normalization methods, baseline adjustment demonstrates superior performance across all metrics when compared to z-score and log transformation. Specifically, it achieves an F-score of 71.8% for baseline, versus 66.6% for z-score and 65.8% for log transformation. While this study focused on utilizing individual features, combining these features could potentially enhance model performance further.

Table 3. Summary of performance of EDA features.

| Features | SC | | | SCL | | | SCR | | | ISCR | | | PhasicMax | | |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|-----------|------|------|
| | z | log | base | z | log | base |
| Accuracy | 0.77 | 0.73 | 0.80 | 0.73 | 0.66 | 0.71 | 0.59 | 0.64 | 0.66 | 0.65 | 0.63 | 0.72 | 0.64 | 0.63 | 0.73 |
| F-score | 0.76 | 0.73 | 0.79 | 0.73 | 0.68 | 0.71 | 0.58 | 0.63 | 0.65 | 0.64 | 0.62 | 0.71 | 0.62 | 0.63 | 0.73 |
| AUC | 0.80 | 0.74 | 0.86 | 0.80 | 0.69 | 0.82 | 0.48 | 0.57 | 0.66 | 0.68 | 0.59 | 0.75 | 0.64 | 0.63 | 0.77 |

4.3 EDA Differences Between Two Groups

Using the SC feature that achieved the highest performance, a Wilcoxon signed-rank test was conducted to assess differences between the two assembly groups. The results revealed a significant difference (p -value < 0.001) in the average SC levels across all normalization methods between the HRC assembly group and the solely assembly group. Figure 2 displays a box plot that visually represents the physiological differences experienced by humans when working alongside robots. This figure indicates that tasks involving HRC generally result in lower EDA signals, suggesting that robots have a positive impact on reducing human cognitive load. This effect can be attributed to the participants in the HRC group receiving more specific and step-by-step instructions for their tasks, in contrast to participants in the solely assembly group, who had to devise their own strategies. Moreover, considering the initial findings that the solely assembly group completed the task more quickly, it suggests that they may have experienced higher cognitive loads due to the faster pace of work.

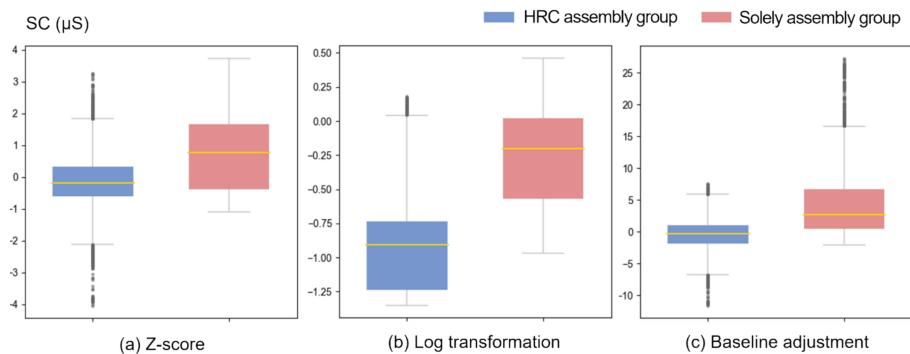


Fig. 2. Comparative analysis of the SC in two assembly groups

5 Conclusion

This study investigated the potential of EDA signals as indicators of cognitive load in tasks involving HRC. A notable finding was the significant difference in EDA signals between participants in the HRC assembly group and those working solely, suggesting that robotic assistance may reduce the cognitive load on construction workers. Furthermore, the results indicated that utilizing the SC feature, normalized using the baseline adjustment, provided superior performance in classifying these physiological changes. Thus, the study not only validates the use of SC as a reliable metric for cognitive load but also underscores the beneficial role of HRC in the construction industry. This analysis is essential for understanding the impact of robot interaction in a shared task environment on human stress levels, focus, and overall cognitive load. While these findings are promising, it's important to approach them with a degree of caution due to the study's small sample size of only 12 participants. Future research will aim to determine if these results hold across a larger population, further validating the study's conclusions.

References

1. Survey of Occupational Injuries and Illnesses Data, <https://www.bls.gov/iif/factsheets/msds.htm>, last accessed 2023/11/16
2. Liu, Y., Habibnezhad, M., Jebelli, H., Monga, V.: Worker-in-the-loop cyber-physical system for safe human–robot collaboration in construction. *Comput. Civil Eng.* 1075–1083 (2021)
3. Shayesteh, S., Ojha, A., Liu, Y., Jebelli, H.: Human–robot teaming in construction: Evaluative safety training through the integration of immersive technologies and wearable physiological sensing. *Saf. Sci.* **159**, 106019 (2023)
4. Zhang, M., Xu, R., Wu, H., Pan, J., Luo, X.: Human–robot collaboration for on-site construction. *Autom. Constr.* **150**, 104812 (2023)
5. Gervasi, R., Capponi, M., Mastrogiacomo, L., Franceschini, F.: Manual assembly and Human–Robot Collaboration in repetitive assembly processes: a structured comparison based on human-centered performances. *Int. J. Adv. Manuf. Technol.* **126**(3), 1213–1231 (2023)
6. Dehais, F., Sisbot, E.A., Alami, R., Causse, M.: Physiological and subjective evaluation of a human–robot object hand-over task. *Appl. Ergon.* **42**(6), 785–791 (2011)
7. Szczurek, K.A., Cittadini, R., Prades, R.M., Matheson, E., Di Castro, M.: Enhanced human–robot interface with operator physiological parameters monitoring and 3D mixed reality. *IEEE Access* (2023)
8. Sun, Y., Jeelani, I., Gheisari, M.: Safe human–robot collaboration in construction: a conceptual perspective. *J. Safety Res.* **86**, 39–51 (2023)
9. Seo, S., Choi, Y., Koo, C.: An interpretable machine learning approach for forecasting personal heat strain considering the cumulative effect of heat exposure. *Korean J. Constr. Eng. Manage.* **24**(6), 81–90 (2023)
10. Benedek, M., Kaernbach, C.: A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**(1), 80–91 (2010)
11. Hanajima, N., Goto, T., Ohta, Y., Hikita, H., Yamashita, M.: A motion rule for human-friendly robots based on electrodermal activity investigations and its application to mobile robot. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3791–3797. IEEE (2005)
12. Podevijn, G., O’grady, R., Mathews, N., Gilles, A., Fantini-Hauwel, C., Dorigo, M.: Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human–swarm interaction. *Swarm Intellig.* **10**, 193–210 (2016)
13. Yang, E., Dorneich, M.C.: The emotional, cognitive, physiological, and performance effects of variable time delay in robotic teleoperation. *Int. J. Soc. Robot.* **9**, 491–508 (2017)
14. Cvijanović, N., Kechichian, P., Janse, K., Kohlrausch, A.: Effects of noise on arousal in a speech communication setting. *Speech Commun.* **88**, 127–136 (2017)
15. Romine, W., Schroeder, N., Banerjee, T., Graft, J.: Toward mental effort measurement using electrodermal activity features. *Sensors* **22**(19), 7363 (2022)
16. Saitis, C., Parvez, M.Z., Kalimeri, K.: Cognitive load assessment from EEG and peripheral biosignals for the design of visually impaired mobility aids. *Wirel. Commun. Mob. Comput.* **2018**, 1–9 (2018)
17. Blanc, C., Buisson, J.-C., Kruck, J., Kostrubiec, V.: Using a haptic dynamic clamp to reduce arousal: preference, arousal, and coordination stability are related. *Exp. Brain Res.* **241**(8), 2033–2044 (2023)
18. Du, N., et al.: Predicting driver takeover performance in conditionally automated driving. *Accid. Anal. Prev.* **148**, 105748 (2020)



Development of Automatic Steel Deck Cutting System Using a Multi-Axis Robot for Power Plant Construction Work

Tsuyoshi Yamada^(✉), Naoya Shinoda, Kazuyuki Takeuchi, and Munechika Masaki

Hitachi Plant Construction, Ltd., 1-3, Higashi-Ikebukuro 3-Chome, Toshima-Ku, Japan
tsuyoshi.yamada.va@hitachi.com

Abstract. An automatic steel deck cutting system using a multi-axis robot for power plant construction work has been developed. This system uses a laser distance meter and a camera to recognize the shape of the steel deck and moves a plasma cutting torch attached on the tip of the robot to follow the unevenness of the deck. Many steel plates are embedded in the ceiling of the deck slab structure of power plants to fix equipment and piping. In order to expose these embedded steel plates during equipment installation work, it is necessary to cut the steel deck into square to match the outline of the steel plates. Conventionally, workers cut the steel deck using a gas cutting machine in an upward position, which was hard work and high risk of industrial accidents. Therefore, we decided to automate this work by combining a plasma cutting machine and a robot arm. However, steel decks are made of galvanized steel sheets several millimeters thick, and their dimensions vary from site to site. If the torch is not moved at the appropriate speed, position, and posture according to the shape of the deck, the plasma arc will break off and the cutting process will fail. Therefore, the system uses a laser distance meter and a camera to recognize the shape of the deck by image recognition. This paper describes the issues in automatic cutting of steel decks, the system configuration to solve the issues, the method of recognizing the deck shape by sensing, the results of parameter test to maintain the plasma arc during the cutting process, and the results of a steel deck cutting test using a prototype system combines these elements.

Keywords: Robot arm · Automatic system · Steel deck cutting · Line tracing · Path planning · Image recognition

1 Introduction

The equipment and pipes required for operation at electric power plants must be securely fixed in preparation for disasters such as earthquakes. Therefore, support parts are welded to embedded hardware, which is installed a surface of a reinforced concrete building to fix the equipment. Embedded hardware is exposed on a floor or wall, but concealed on a ceiling as the hardware is located between the deck plate and concrete. To safely expose this, the deck plate must be cut thermally to fix facilities on the ceiling surface.

However, conventional cutting methods pose a risk of labor accidents, as workers may be covered in melted hardware or sparks. Therefore, an automated system needs to be developed for safe deck plate cutting.

Research into automation in the construction industry aims to streamline processes and reduce labor accidents. For example, automated driving systems like SLAM [1] have been developed to transport materials to construction sites. Shimizu Corp. Has even created a robot for this purpose [2].

Meanwhile, research has been conducted on how to process construction materials to achieve characteristic designs in the processing industry. Additionally, research is currently being conducted on processing construction materials in plants for prefab construction. Braumann and colleagues have developed an automatic control platform that integrates a drawing with CAM features, allowing for the processing of complicated construction materials with individually designed using a robot arm [3]. Kerber and others conducted research on the automated welding of construction materials for prehab construction using a robot arm [4].

Additionally, research is being conducted on the automation of holding, transporting, and attaching construction materials during assembly operations. In-depth research has been carried out on bricklaying, which involves holding and accumulating block-shaped objects [5, 6]. In addition, due to the large size of the construction site and the variety of assembly operations, the system may include traveling function [7, 8], which are also promoted [9]. In addition, research is being conducted to automate construction work by stacking construction materials using robot arms [10, 11].

Meanwhile, automation research has not yet focused on cutting work for parts that have already been mounted on the construction site. Cutting work is a part of the processing operation, and it is primarily carried out in the factory which is a stable working environment to maintain processing accuracy. In addition, it is difficult to install cutting equipment at construction sites because automatic cutting of construction materials generally uses large size of the processing equipment. Furthermore, to automate the cutting work of the deck plate at the construction site, cutting equipment of suitable size must be transported and installed on the ceiling surface. Additionally, shape measurement and cutting control must be performed per cutting point to ensure accurate cutting work as the deck slab is transformed on site.

As previously mentioned, this research aims to develop an automatic cutting system for the deck plate on the construction plate. Chapter 2 analyses the existing deck plate cutting methods. Chapter 3 covers the system design for the cutting method, actuator, and control method. Chapter 4 presents the configuration of the cutting system developed based on the design and the operation processes. Chapter 5 confirms the effectiveness of the system through cutting tests. Finally, Chapter 6 summarizes the results of the system development.

2 Cutting Work of Conventional Deck Plate

Figure 1 displays a deck plate that is to be cut and embedded hardware that needs to be exposed (the view from an upper floor looking down onto the floor). A worker positions the embedded hardware on the deck plate according to the design drawing and then welds

the inflow prevention plate to the deck plate to prevent the slab from running underneath the hardware. This process allows for the exposed embedded hardware when cutting the deck plate on the ceiling. Furthermore, weld marks can be confirmed by inspecting the deck plate from below after working on the ceiling (refer to Fig. 2). The operator should draw a guideline based on these weld marks before cutting the deck plate along the guideline.

Figure 3 displays the cutting process of the top plate. The size of the embedded hardware is typically 250×250 mm or 300×300 mm. The worker cuts the deck plate on the temporary scaffold, which has been erected up to the ceiling, using gas cutting. It takes approximately ten minutes to cut a 300×300 mm plate, during which time the worker is exposed to molten hardware and sparks. Despite wearing protective equipment, workers are susceptible to burns or scratches caused by iron powder, which can lead to work-related accidents. Moreover, due to the elevated ambient temperature and the thick protective equipment worn by workers, there is a significant risk of heat stroke. As a result, it is imperative to urgently implement automation to safeguard the workers.

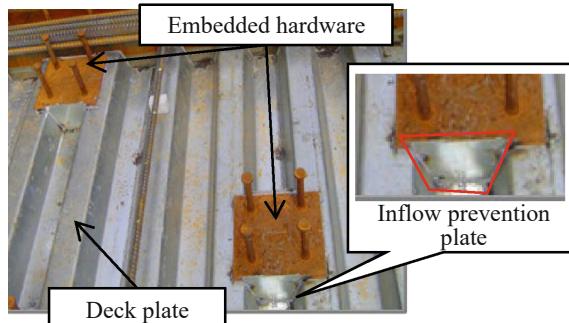


Fig. 1. Embedded hardware on the deck plate.

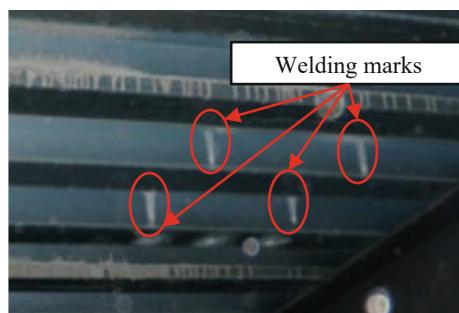


Fig. 2. Welding marks on the inflow prevention plate.

Table 1 shows the shape of the deck plate used at a power plant construction site in Japan. In Japan, according to the standard of the deck plate, JIS (Japanese Industrial



Fig. 3. Cutting work state by hand.

Standards), several types of the shape are specified. However, the deck plate is deformed on the construction site by the loads from welding and the weight of the slab, consequently the shape of the deck plate differs from the standard. Accordingly, an operator performs the cutting work according to the type and distortion of the deck plate.

Table 1. Deck plate standard used in a power plant construction [12].

| Product name | Cross-sectional shape, dimensions |
|--------------|--|
| V50A | <p>Plate thickness is 1.6mm, zinc plating processing, JIS name is ALF16.</p> |
| UKA/ UKA-N | <p>Plate thickness is 2.3mm, Zinc Plating processing, JIS name is ALN23(UKA-N), BLD23(UKA)</p> |

3 System Design

3.1 Cutting Method

The initial stage in developing an automatic deck plate cutting system involves selecting a cutting method. This decision is based on the characteristics of the cutting method, which determine the required structure and size of the cutting equipment. The chosen cutting method should be easy to carry on the temporary scaffold and suitable for automation. Two methods are available for cutting steel materials: mechanical cutting and thermal cutting.

There are two methods for cutting steel using mechanical cutting: shearing and cutting. Shearing involves cutting the iron plate with an edge tool, but it cannot be used if one side of the deck plate is covered with concrete. Cutting, on the other hand, is pushing a rotating cutter or grindstone against the steel plate to cut the object. The deck plate can be cut using this method, but a cutting device receives a large cutting reaction force. In order to withstand the cutting reaction force, the structure of the cutting equipment must be sturdy and larger. However, to carry the cutting equipment on the temporary scaffold, the equipment should be made as light as possible. Therefore, this cutting method may not be suitable for cutting the deck plate.

Thermal cutting on construction sites involves gas cutting and plasma cutting. Gas cutting is used for conventional cutting of deck plates and can be applied or carried out manually. Gas supply is visually adjusted for thermal power during gas cutting. Although gas adjustment can be automated, the equipment required is large and complicated. In plasma cutting, fire adjustment can be easily done by setting parameters such as air pressure, current, distance between cutting target and torch in the software. This makes it easier to automate compared to gas cutting, although the torch must be kept at a certain distance from the cutting target. Additionally, thermal cutting does not have a cutting reaction force, making the cutting equipment lighter and easier to handle. As previously stated, plasma cutting is the chosen cutting method for this paper due to the cutting environment of the deck plate.

3.2 Actuator

When using plasma cutting, it is important to maintain a distance of only a few millimeters between the deck plate and the cutting torch. Additionally, the actuator that moves the torch should have a high degree of freedom to accommodate for standard and distortion of the deck plate. To achieve this, a six-axis robot arm was used in this research due to its high degree of accuracy and freedom of movement.

Figure 4 displays the appearance and specifications of the robot arm. Assuming a deck plate size of 300 mm × 300 mm, we utilize an arm model with a 500 mm movable range. The robot itself weighs only 11 kg, with an additional 12 kg for the control box. This lightweight design allows for easy transportation on temporary scaffolds. Furthermore, the robot is a versatile platform capable of handling various applications and executing control programs developed in Python or C++ through open-source software.

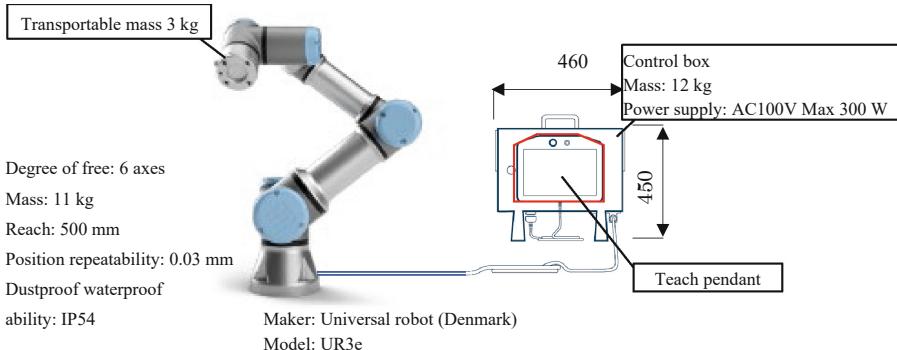


Fig. 4. Appearance of robot and specifications.

3.3 Control Method

Moving along the deck plate to the cutting torch requires a path plan for each joint of the robot arm and for tip of the cutting torch. On the other hand, the robot arm shown in Fig. 4 can calculate the path plan for each joint by the path plan for the cutting torch. Therefore, this research focus on a path planning for the tip of the cutting torch.

Path plan can be repeated if the relative relationship between the cutting place and the robot arm remains constant. For instance, the robot arm used in the line process at the plant utilizes the once-determined path planning repeatedly. However, when cutting the deck plate, several factors such as the location of the cutting target, the position of the robot arm, and the standard of the deck plate, including its distortion, may have changed. Therefore, the system is necessary to plan the path of the torch tip for each cutting point.

To plan the path, the system must recognize the cutting position. Therefore, an operator of the system must provide standard for the cutting point to the system. In this study, the cutting point is identified by recognizing the guideline created on the deck plate surface as the standard. The guideline is created by estimating the installation position of the embedded hardware based on the welding marks shown in Fig. 2 and writing the rectangle necessary to expose the embedded hardware with a pen on the deck plate. On the other hand, we attach a camera and a laser distance meter to the tip of a robot arm to enable the system to recognize the guideline through image recognition. Furthermore, we created the system to move the laser distance meter along the recognized line for measurement the shape of the deck plate. This allows for path planning of the cutting torch to be tailored to the shape of the deck plate. The process of measuring the shape of the deck plate is referred to as “line tracing” in this paper.

4 Deck Plate Cutting System

4.1 System Configuration

Figure 5 shows the appearance of the robot arm in the system. The plasma cutting torch, the camera, the laser distance meter and the cutting torch are installed at the tip of the robot arm. Incidentally, the collimation direction of the camera is inclined by 30° to the

torch axis depending on the camera specifications. Accordingly, we installed the laser distance meter at same angle to measure the center of the image captured by the camera. Figure 5(c) shows the method of fixing the robot arm.

The robot stand was made of an 80 mm square aluminum frame. This stand was suspended from the deck plate using three neodymium magnets (adsorption force of 1.3 kN each). By suspending the stand from the deck plate, we can fix the equipment regardless of the conditions of the temporary scaffold. In addition, we attach the special fireproof cover to the robot body to protect it from sparks.

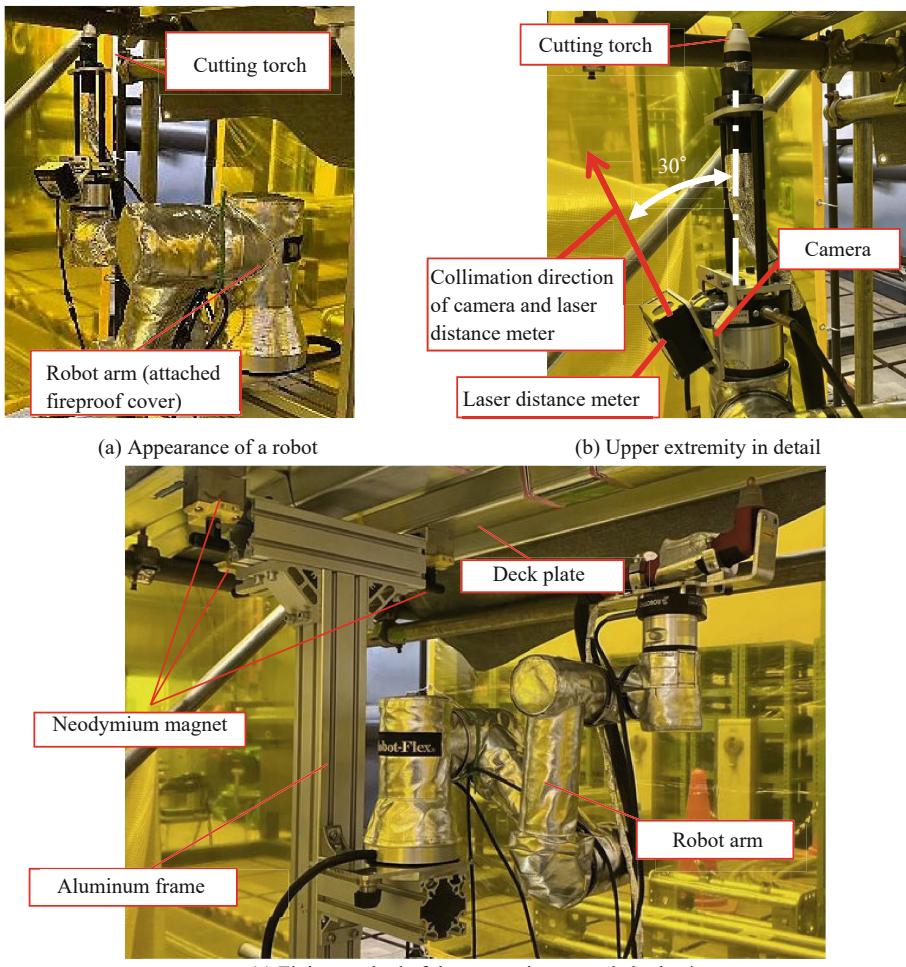


Fig. 5. The appearance of the system (Robot arm part).

Figure 6 displays the system configuration, which comprises the robot arm, control box, plasma cutting machine, camera, laser distance meter, and the system control PC with the automatic cutting program installed. The robot arm, camera, and laser distance meter are connected to the control box. In this configuration, each member communicates through the control box, which is also connected to the control PC.

The control PC program acquires measured values from the camera image and laser distance meter. Subsequently, the program proceeds to the path planning of the cutting torch tip and sends the information to the control box. In addition, the control box calculates the velocity of each joint and the rotation amount based on the path planning and activates the robot. After the above, the system cuts the deck plate at the command of the operator. Incidentally, in this research, the switch operation of the plasma cutting machine is performed manually by the operator.

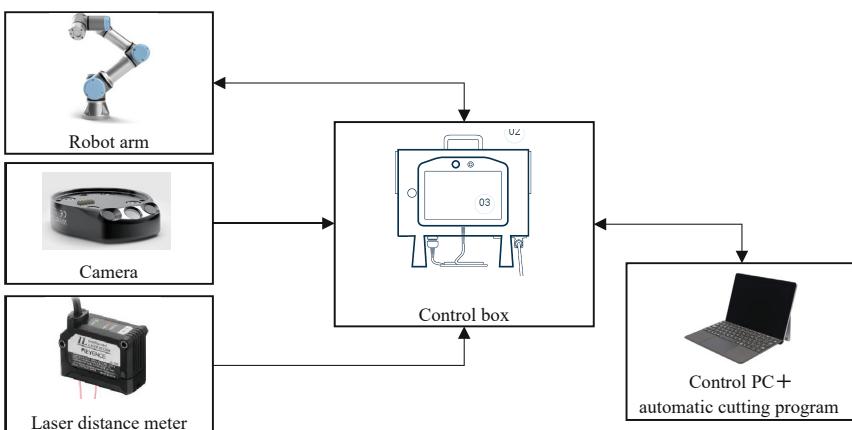


Fig. 6. System configuration.

4.2 Operation Process

Figure 7 illustrates the system's operation process. The broken line boxes indicate the movements of the robot arm, while the other boxes represent calculations or the operator's input. A simple explanation of each process's activities is as follow.

To begin, set up the system devices at process 1 and fix the stand for the robot at the cutting site. Connect each piece of equipment shown in Fig. 6. Next, the operator should draw a line at the cutting point with a thickness of at least 1mm. In process 2, make the robot arm moved automatically to measure the flat surface of the deck plate. The system recognizes the relative position between the robot arm and the deck plate by measuring the flat surface of the deck plate. After this process, the system adjusts the position of laser distance meter and performs measurements based on this flat surface.

During process 3, the operator prepares line tracing such as moving the robot arm to the appropriate position and adjusting parameters for recognizing the camera image. At first, the operator moves the robot arm to positioning the line in the center of the

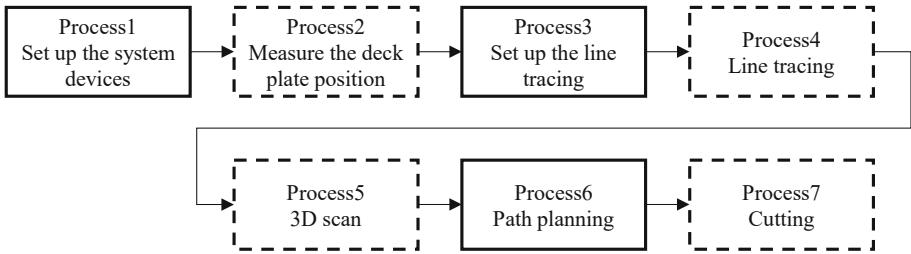


Fig. 7. Operation process.

camera image. Then, the operator adjusts the distance between the robot arm and the deck plate based on the measuring range of the laser distance meter. Additionally, the operator adjusts the color parameters for image recognition of the guideline and confirms the result. If the result is not satisfactory, the operator repeats adjusting the color parameters until the system recognizes the guideline appropriately. This setting enables measurement of the guideline during line tracing.

During process 4, the system performs line tracing, as shown in Fig. 8. The robot arm is moved to position the line in the center of the camera image and in the terminal direction of the line. Simultaneously, the system uses a laser distance meter to measure the shape of the deck plate. This measurement corresponds to the shape of the deck plate where the line is drawn, as the laser distance meter is placed to measure the center of the camera image. The system stores the shape of the deck plate of the cutting place as a coordinate value within the robot.

During process 5, the system scans the shape of the deck plate around the cutting area. Figure 9 illustrates the scanning process. If only the measurement information from process 4 is used, the shape of the deck plate around the guideline is not taken into account during path planning, which can result in the cutting torch and the laser distance meter colliding with the deck plate. To prevent this collision, it is needed to measure the shape of the deck plate around the cutting area. In this paper, considering the size of the cutting point, the system automatically scans a square of 250 mm from the center of the guideline. Implementing interference checks using acquired data during path planning prevents the deck plate from colliding with the robot arm during cutting.

During process 6, the system calculates the path planning for the torch tip using measured data. In this process, the system assumes the x-z plane shown in Fig. 8. The path conditions are as follows: At first, the torch tip is on the x-z plane and passes a position far from the cutting position by an offset distance in the vertical direction. Subsequently, the system calculates the vertical direction from the deck surface and the position of the torch tip along the deck shape in 1mm increments. Currently, the system checks for interference by using the shape data around the cutting position measured in process 5 and the robot model equipped with the torch and laser distance meter. If interference is detected, the system adjusts the torch attitude by 1° from its current value and confirm again. This process is repeated until the system no longer detects interference, allowing for path planning that ensures the robot arm does not collide with the deck plate in process 7.

Finally, in process 7, the operator uses the robot arm and plasma cutting machine to cut the deck plate. The system enters the path planning of the cutting torch to the control box, which performs feedback control of each joint, setting the path planning as the control target. Simultaneously, the operator starts the plasma cutting, and the system cuts the deck plate according to the path planning.

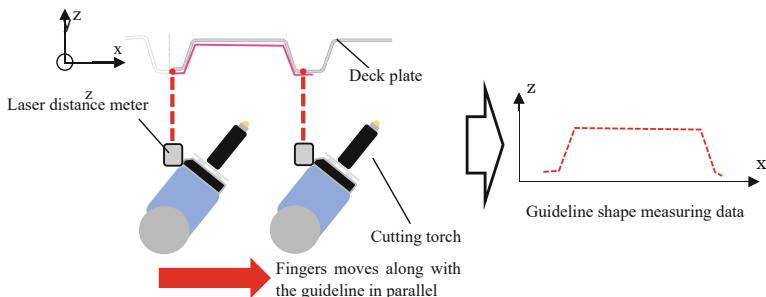


Fig. 8. Guideline shape measuring image by line tracing.

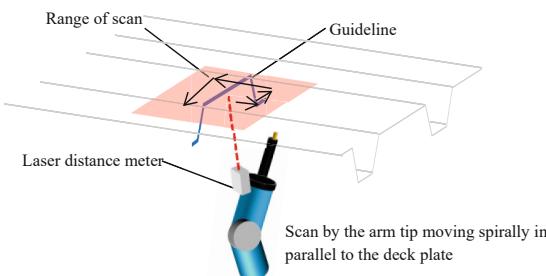


Fig. 9. Scan image around the guideline.

5 Cutting Test

5.1 Contents of the Test

The system's effectiveness was confirmed through a cutting test of the deck using the V50A deck plate standard (Table 1), which was commonly used at a specific construction site. Prior to the test, we verified the parameters, including the output of the plasma cutting machine, velocity of the cutting torch, and offset distance between the cutting torch and the cutting surface. We have confirmed that the parameters of output 5 kW, velocity 400 mm/min, and offset distance 1 mm are suitable for cutting.

This paper confirms the test results of process 4, 6, and 7 for automated deck plate cutting in the operation processes outlined in Sect. 4.2. The following items were confirmed in the test:

- (1) The line tracing operation (process 4) can measure the shape of the deck plate along the guideline.
- (2) It is possible to plan a path for the robot arm that avoids collisions with the deck plate (process 6).
- (3) The robot arm can be moved to cut the deck plate (process 7).

5.2 The Results of the Test

We confirm the results of the test according to each check item described in the previous section.

- (1) It is possible to measure the shape of the deck plate along the guideline (process 4)

Figure 10 displays the three-dimensional plot of the deck plate's shape measured in process 4 using a laser distance meter. The arm was moved along the guideline drawn on an even surface of the deck plate, and the measured data confirms the shape of the deck. The y-axis represents the direction of the guideline, and the z-axis represents the height direction. To ensure the guideline is straight and at least 1mm thick, it is acceptable to use colorful tape instead of a pen. Additionally, adjusting the color parameters in process 3 is crucial for accurately tracing the guideline.

When we used the measured data to conduct process 6 and activated the robot arm to move the cutting torch, we visually confirmed that the tip of the torch moved along the center of the guideline at a certain distance. Therefore, we concluded that the measured data is reasonable.

- (2) It is possible to make path planning in which the robot arm does not collide with the deck plate (process 6).

Figure 11 displays the interference check screen. The system created a path that avoids colliding robot model and deck model by using the 3D scan data of the deck plate acquired in process 5 and the simple model of the robot tip, which includes the cutting torch and the laser distance meter. Figure 12 displays the robot arm when activated along this path. The images at II, IV, VII, and X demonstrate that the robot does not collide with the deck plate even when the cut section switches, and the arm angle changes significantly. Additionally, we verified that the torch tip maintains 1mm from the cutting point as per the setting value. Furthermore, we have confirmed that by adjusting the position or dimensions of the cutting torch or laser distance meter, we can still achieve suitable path planning using the modified model. As previously stated, the path-making method, which takes into account interfere avoidance in process 6, is a reasonable approach.

- (3) It is possible to cut the deck plate by activating the robot arm (process 7).

Using the path planning from process 6 and cutting parameters outlined in Sect. 5.1, we were able to cut the deck plate into a square without breaking plasma arc during cutting, as shown in Fig. 13. This confirms the effectiveness of the system was also confirmed by the fact that we were able to cut the deck plate into a square using this system, as shown in Fig. 14.

* Unit of each axis is [m]

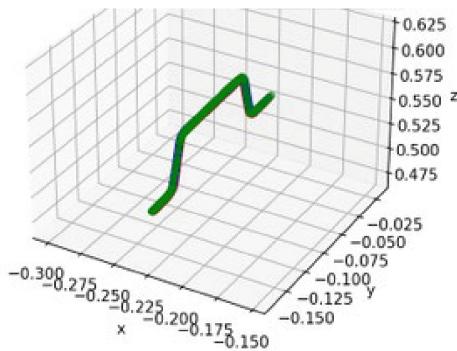


Fig. 10. The result of measurement of the shape of the deck plate.

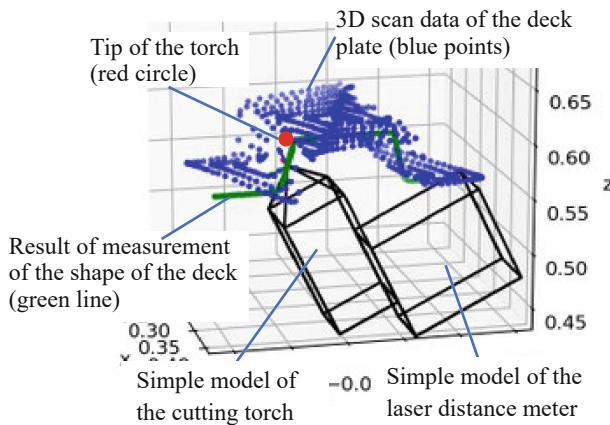


Fig. 11. Example of the display of the interference check.

6 Conclusion

We have developed an automatic cutting system for the deck plate to reduce the risk of labor accidents.

- (1) Our research has led us to conclude that the plasma cutting method is the most suitable for automatic cutting in construction sites. Additionally, a multi-axis robot arm, which allows a cutting torch to be operated high accuracy and a high degree of free operation, is the most suitable actuator in the system. The robot arm will be controlled by planning the cutting torch's path using data acquired through line tracing.

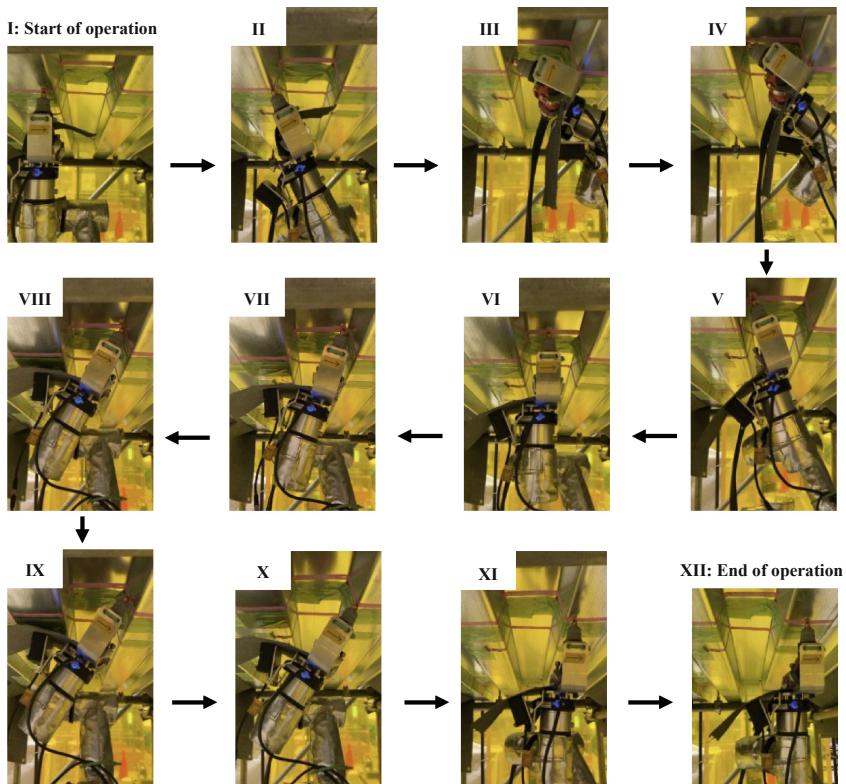


Fig. 12. Confirmation of the copying motion of the guideline by the cutting torch.



Fig. 13. The state of the time of cutting using the system.

- (2) The cutting torch, camera, and laser distance meter were installed on the tip of the robot arm to enable automatic cutting. A suspended stand was attached to the deck plate on the ceiling using a magnet to avoid any influence from the temporary scaffold. The robot was then fixed onto the stand.



Fig. 14. The state after cutting into a square.

- (3) The automatic cutting program was developed with a feature that measures the shape of the deck using a camera and laser distance meter to trace lines. It then calculates the position and attitude of the torch tip to create a movement plan. The program was successfully implemented into the system.
- (4) The cutting system of the deck plate was tested and confirmed to be capable of measuring the shape of the deck plate. To avoid interference between the deck plate and the arm, the deck plate should be cut reasonably. As mentioned Sect. 5.2, our developed system allows for automatic cutting and provides its effectiveness.

As a future challenge, we aim to improve the line tracing settings and path creation method. Currently, the operator must adjust the color parameters or camera-acquired images in the line tracing settings, which requires knowledge of image processing or the robot arm's actions and is very difficult. Therefore, we plan to develop a system that does not require the operator's expertise. Additionally, we currently use a straight line as the guideline to create the path. Although not the focus of this research, there is a circular cutting position on the deck plate. Therefore, we will improve the path-making method to automatically cut along a curved guideline. We will evaluate the effectiveness and challenges and consider further system improvements by applying it to the site.

References

1. Grisetti, G., Kümmerle, R., Stachniss, C., Burgard, W.: A tutorial on graph-based SLAM. *IEEE Intell. Transp. Syst. Mag.* **2**(4), 31–43 (2010)
2. Shimizu Corp. Homepage, <https://www.shimz.co.jp/en/topics/construction/item12/>, last accessed 2023/12/6
3. Braumann, J., Brell-Cokcan, S.: Parametric robot control integrated CAD/CAM for architectural design. In: Proceedings of the 31st Annual Conference of the ACADIA, pp.242–251. ACADIA 2011, Alberta, Canada (2011)
4. Kerber, E., Heimig, T., Stumm S.: Towards robotic fabrication in joining of steel. In: 35th ISARC, pp. 444–452. ISARC 2018, Berlin, Germany (2018)
5. Wos, P., Dindorf, R.: Develop and Implement a masonry algorithm control in a bricklaying robot. In: The 11th International Conference on Applied Mechanics, pp. 020027-1-020027-10. AIP Publishing, Bydgoszcz, Poland (2023)
6. Wos, P., Dindorf, R., Takosoglu, J.: Bricklaying robot lifting and levelling system. *Commun. Sci. Lett. Univ. Zilina* **23**(4), B257–B264 (2021)

7. Helm, V., Ercan, S., Gramazio, F., Kohler, M.: Mobile robotic fabrication on construction sites: DimRob. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4335–4341. IEEE, Vilamoura, Portugal (2012)
8. Liu, Y., Jebelli, H.: Human-robot co-adaptation in construction: Bio-signal based control of bricklaying robots. In: Raymond, R., Issa, A. (eds.) The ASCE International Conference on Computing in Civil Engineering 2021, pp. 304–312. ASCE, Orlando, USA (2021)
9. Fastbrick Robotics Homepage, <https://www.fbr.com.au/view/hadrian-x>, last accessed 2023/12/6
10. Dirrenberger, J.: From architectured materials to large-scale additive manufacturing. In: Bier, H. (ed.) Robotic Building, Springer Series in Adaptive Environments, pp. 79–96. Springer International Publishing, Berlin (2018)
11. Watson, N.D., Meisel, N.A., Bilén, S.G., Duarte, J., Nazarian, S.: Large-scale additive manufacturing of concrete using a 6-axis robotic arm for autonomous habitat construction. In: 2019 International Solid Freeform Fabrication Symposium, pp. 1583–1595. University of Texas at Austin, Austin, USA (2019)
12. Nippon Steel Metal Products Corp. Homepage, <https://www.ns-kenzai.co.jp/a2uka.html>, last accessed 2023/12/6



Toward Future-Proof Asset Information Management

Marlena Block¹(✉), Alexander Buttgereit², and Markus König¹

¹ Computing in Engineering, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany

marlena.block@ruhr-uni-bochum.de

² Jade University of Applied Sciences Oldenburg, Ofener Str. 16/19, 26121 Oldenburg, Germany

Abstract. Municipalities are responsible for planning, constructing, and maintaining resilient infrastructure in the contemporary urban landscape. Recognizing the pivotal role of information management in meeting this challenge, this study focuses on municipal road preservation to identify opportunities where Asset Information Management (AIM) can be effectively leveraged. Based on a case study and an analysis of macroscopic trends, this research provides valuable insights to guide municipalities through the complexities of modern infrastructure management. In a previous article, the authors have formulated consequences and requirements for information management in municipal building administrations resulting from macroscopic changes. The present contribution aims to highlight areas where efficient AIM can be utilized in the field of municipal road preservation. The results are divided into two main sections. Firstly, the restoration process in municipal road preservation is analyzed as a sample. Secondly, the effects of macroscopic trends on the various phases of municipal road preservation processes are analyzed. The results are presented using a matrix that analyses the influence of these trends on the various process steps and underlines the support potential of AIM.

Keywords: Asset Information Management · municipalities · infrastructure management

1 Introduction

Nowadays, municipalities play a key role in shaping our urban environments. They face the challenge of planning, constructing, and maintaining sustainable infrastructure. Sustainable infrastructure in a municipality refers to infrastructure systems designed to meet the needs of the present without compromising the ability of future generations to meet their own needs [1]. This involves integrating environmental, social, and economic considerations into infrastructure projects.

To fulfill these requirements, comprehensive decision-making support in public administration is essential. However, decision-making in municipal administration is

characterized by many different aspects [2]. Political framework conditions such as digital administration, obligations at the federal and state levels to use Building Information Modelling (BIM), the need for citizen involvement and participation, the mobility transition, and the consolidation of finances create a complex decision-making context. Legal factors also have a significant influence on planning and construction processes. Compliance with environmental protection laws, applying the EU taxonomy, implementing the EU Construction Products Directive, and changes in the awarding and procurement process are essential. Economic factors such as the shortage of building materials, inflation, energy prices, developments in outsourcing or insourcing, and the refurbishment backlog are putting pressure on local authorities' financial resources. Socio-cultural aspects such as ensuring mobility participation and accessibility, demographic change, changes in urban planning, and increased traffic create further challenges. Despite these complex framework conditions, some technical concepts and possibilities enable municipalities to tackle these challenges.

Therefore, different infrastructure systems should be integrated in a structured way to acknowledge the complexity and interconnection of these systems. It is also helpful to prioritize utilizing information and communication technologies and data more effectively. Additionally, there's a need to reconsider the current assessment of infrastructure [1]. Building upon these general requirements for technical solution concepts, numerous specific approaches exist. For more transparency towards the citizens and as a basis for political decision-making, for instance various forms of Urban Data Platforms are used to visualize underlying data, enabling performance monitoring, or offering publicly accessible, machine-readable data bar [3]. Different concepts for real-time generation of condition data, such as condition monitoring of bridges, continuously provide data that helps in the agile planning of measures and thus reduces the increase of refurbishment backlogs [4]. To counter the shortage of building materials, municipalities strive to assess their anthropogenic stocks and implement appropriate material flow management solutions for an efficient circular economy [5]. Common to all these approaches is the need for data and information about existing assets and the generation of data that must be managed long-term within urban administration.

Integrating such approaches often involves municipal employees from different departments serving as the interface, primarily utilizing communication forms such as email or telephone. They then provide data and information asynchronously, resulting in partial gaps and outdatedness. In consideration of these circumstances, the implementation of efficient Asset Information Management (AIM) should focus on successfully tackling the challenges of urban development information to enable those responsible to make effective decisions.

In a previous article, the authors addressed the question of which macroscopic trends currently exist and are likely to impact municipal infrastructure management in the future [2]. Employees of a municipal building administration reflected the consequences of the identified trends, and effects and requirements for the building administration were derived. The influence analysis will continue in the next step based on municipal infrastructure management processes. The authors pursue two key questions: (1) Where is strategic intervention needed to make a municipal building administration future-proof? (2) Where is a concrete need for AIM?

To achieve this, the basics of municipal Asset Management (AM), asset information management, and the tasks of municipal road maintenance are briefly laid out in Sect. 2. Section 3 sets out the overarching research concept and describes the underlying approach in this article. Section 4 describes a selected road maintenance process and finally presents the results of the article. Section 5 concludes with specific answers to the fundamental questions.

2 Operation and Maintenance of Municipal Infrastructure

The following section delves into two key areas: municipal AM and AIM. In Sect. 2.1, we explore the challenges and strategies of municipal AM, emphasizing the need for its integration into German regulations for effective infrastructure maintenance. Section 2.2 focuses on AIM as a crucial tool for handling asset information across municipal responsibilities. Lastly, Sect. 2.3 outlines the tasks involved in municipal road preservation, emphasizing the importance of effective management for extending asset lifespan and ensuring proper usage. The content of the last subsection is required to apply the methodological approach.

2.1 Municipal Asset Management (Infrastructure)

Asset management represents a crucial instrument for effectively managing resources such as personnel, equipment, and finances across the entire lifecycle of tangible assets. Its objective is to achieve defined performance and cost goals while taking into account relevant legal, geopolitical, economic, social, technological, and environmental conditions [6]. Unlike quality management, which describes process flow for task execution, AM encompasses coordinated, multidisciplinary organizational processes. Particularly in infrastructure maintenance, various assets like sewers, roads, and transportation systems are consolidated and managed within an asset portfolio [7]. Figure 1 shows the context of AM within the managing organization.

Despite being well-researched and practical internationally, the adoption of AM, especially at the municipal level, remains insufficient in Germany [8]. Hence, there is a need to develop a long-term strategy for implementing AM in municipalities to expedite integration.

International standards delineate the fundamentals and requirements of AM. The structuring of AM can be done through the continuous sequence of goal definition, planning, decision-making, implementation, and monitoring [9]. Although AM is described as an international standard, a specific implementation for German road maintenance management is lacking. The concept has not been fully integrated into the German Association for Road and Transportation's regulations, particularly for road maintenance management [10]. Nevertheless, the systematic approach of asset management can serve as guidance in establishing an AM system. Short-term control aims to achieve a balanced annual outcome, while long-term control seeks sustainable resource utilization to ensure the long-term performance of municipalities [7].

To control AM efficiently, goals, metrics, and key indicators must be defined. Long-term control aims to ensure intergenerational equity and sustainability. Short-term control focuses on achieving a balanced annual result. AM emphasizes preserving the value

of assets over their life cycle to minimize environmental impacts and use financial resources economically.

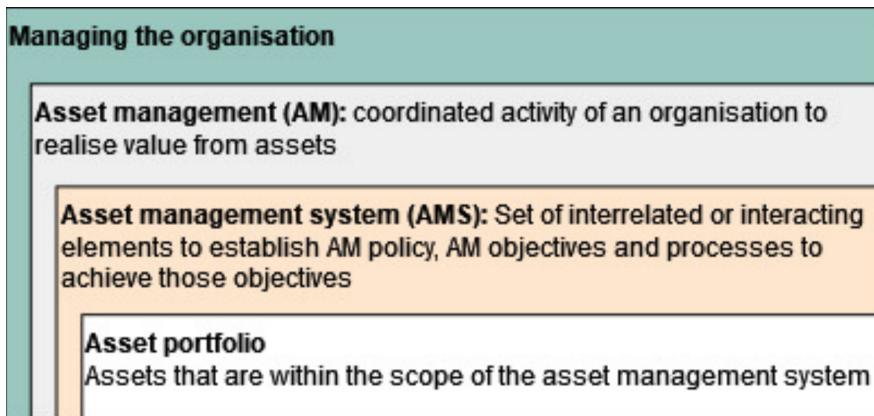


Fig. 1. Context of the terminology based on ISO 55000 [9].

2.2 Asset Information Management

The asset information model is the result of the ordered asset information requirements, which show which information is required by the customer regarding administrative, commercial, and technical aspects of the creation of asset information [11]. The standard lists equipment inventories, cumulative maintenance costs, installation and maintenance schedules, ownership, and other relevant information as examples of information in an AIM.

AIM is a technical system that controls the task-orientated handling of asset information from various municipal areas of responsibility at a strategic, tactical, and operational level. In this regard, it is a tool of the AM (see Fig. 2). This enables municipalities to manage multiple service areas strategically, considering information from other technical systems and assets such as sewers, bridges, or roads. In contrast, many cities still utilize Pavement Management Systems (PMS). These systems are technical data storage systems for road preservation information focusing primarily on economic evaluations and monetary asset valuations rather than encompassing the entire lifecycle [6]. Such AIM are managed with the help of the data collected in the AMS and the information generated from it, for example, by using a balanced scorecard [12]. This contains the city's objectives and the associated vital figures and metrics and presents the current target achievement very simply and clearly. The realization of many goals in this strategic city management takes the form of projects, i.e., measures that are generally clearly limited in time. During project processing, information from the AIM is required on the one hand, and project-specific information is generated on the other.

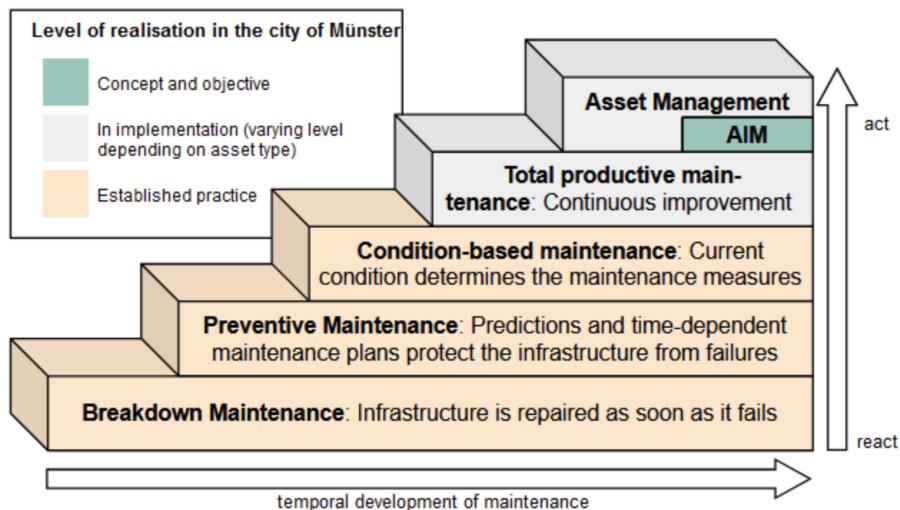


Fig. 2. Classification of AIM and development of maintenance systems over time using the example of the city of Münster.

2.3 Tasks of Municipal Road Preservation

The relevant tasks of municipal road preservation include constructing, maintaining, restoring, and renewing traffic areas, roads, and their components [13]. Road preservation measures preserve traffic surface pavements' substance and utility value, including adjacent areas and environmental compatibility. Construction maintenance includes smaller-scale construction measures to maintain the substance of road pavements, carried out by hand or machine with little effort immediately after localized damage. This includes construction methods such as filling potholes or individual cracks, surface treatment of individual damaged areas, filling open joints, and grinding off deformations in small areas. This type of maintenance must be distinguished from operational maintenance. These are non-structural measures to ensure the proper and safe use of roads, such as regular street cleaning and the maintenance of green spaces. Restoration involves construction measures to maintain the substance or improve the surface properties of traffic surface pavements, which are usually carried out on continuous areas in lane widths up to a thickness of 4 cm. This includes construction methods for surface treatment, application of thin asphalt surface courses, but also back molding of the surface course. These measures are carried out to safeguard or extend the asset's life. In the case of a fundamental renewal, a road renewal is usually carried out up to the subgrade, including the associated equipment and drainage system. Renewal includes partial renewal from the bound structure to the entire superstructure. Measures in this area are aimed at the complete reconstruction of a traffic surface pavement or parts if more than just the asphalt surface course is affected. These measures often lead to a restart of the technical service life of the asset.

3 Positioning in the Overarching Objective and Methodological Approach

As already described, this contribution continues a previously made consideration. It will not ultimately fulfill the overarching goal of a strategic introduction of asset information management for the planning, construction, and maintenance of sustainable and generation-appropriate municipal infrastructure. The previous contribution focused on identifying the consequences of changes in the political, economic, socio-cultural, technological, economic, and legal framework conditions of municipal infrastructure management, which are attributable to macro trends and formulated specific requirements. The focus of this contribution is to identify fields of action for further progress towards an efficient municipal AIM and to differentiate this from measures that are not or only marginally in AIM implementation. To modularize this project and make it tangible, the entire municipal AM will be narrowed down, and a combination of process analysis and expert discussions will be carried out as a method for further action. This contribution focuses on the area of municipal road preservation. As an example, the restoration process is developed in cooperation with employees of the City of Münster; the resulting process description is presented in Sect. 4.1. Since a single measure was used as the subject of the discussion with the employees to make the analysis process as tangible as possible, processes at a higher level, such as creating the construction program, where several measures are weighed against each other, are only roughly outlined. These are not the focus of the analysis but must be put into context.

In a second session, an open expert discussion identifies the influences of external factors, their consequences, and requirements on the respective process steps. This contribution's underlying guiding questions were also the expert discussion's guiding questions.

4 Results

The results consist of two main areas. Firstly, the exemplary process of restoration was analyzed and described. Efforts were made to streamline the process as much as possible, although it was noted that there was more implicit knowledge among the employees than was depicted in the subsequent section. The main objective - to be able to name and describe consecutive processes clearly - was achieved. The second part of the results focuses on the actual research questions.

4.1 Process of Restoration in Municipal Road Preservation

Restoration begins with identifying the demand for measures and prioritizing the damages to road infrastructure. This is done by assessing damages by district engineers and inspectors and possibly using measurement methods such as radar surveys or visual-sensitive inspections. These damages are further identified through citizen reports. Additionally, statistical damage frequencies may be used to assess the need for repairs. The recorded requirements include recording attributes such as location and condition value, linking images or measurements, and documentation of a repair proposal based on the

inspectors' experiences. This results in a list of measures considered when creating the construction program. The construction program, in turn, is prepared with additional measurement data, such as bore cores and excavations, to specify the previously recorded repair proposals. The measures are prioritized for one to five years, with condition value, road relevance, and available budget playing a crucial role. An essential distinction is made in budget management between consumptive and capital expenditures. Consumptive expenditures (operating costs) relate to the maintenance and repair of the facility during its life and generally do not affect the asset value. On the other hand, investments are intended to create long-term benefits and primarily involve fixed assets such as comprehensive renewal and reconstruction.

External and internal consultations are carried out to specify the time and resource planning. External consultations include coordination with operators of telecommunications infrastructure, electricity providers, and waterways operators, as well as inter-municipal transportation networks such as state or federal roads. Internal consultations take place within the municipality and involve all departments involved in planning, construction, and operation. Often, consultations are bilaterally using simple lists.

Prioritized measures are transferred to the tendering phase following a political decision. This is followed by the awarding of contracts and implementation of the measures. Depending on the type and size of the measure, various committees such as the Environment Committee or Transport Committee - if available - are consulted in addition to the City Council. For implementation, execution planning information is provided and adjusted if necessary due to situational dependencies (time, location, weather conditions, etc.). The services provided are accepted and documented by the municipal contractor. Defect-free services are invoiced. Upon completion of the activities, traffic is released; strictly speaking, all relevant information regarding measures must have been transferred to the asset management system at this time to ensure operation. However, in practice, this often occurs with significant delay. Before the end of the warranty period, a warranty acceptance is carried out; this process step is typically deferred 1–4 years after the actual measure. Any deviations from the information in the asset management system must be adequately and promptly implemented again.

4.2 Consideration of Macro Trends in the Various Process Steps of Municipal Road Maintenance

Together with experts, the list of external factors was iterated process by process. In some cases, the meaning of the factors had to be clarified, for which the discussion was interrupted, and a shared understanding of the specific subjects of the questions was reached. The questions are the same for each entry in the matrix.

For example: Does the Digital Administration trend impact the process of identifying the demand for a maintenance measure? Yes, it does, as digital administration must sometimes provide opportunities for citizens to participate, such as digital access to the damage reporting system. Conversely, digital administration also requires that municipal work is presented transparently, so it would be feasible at this point to present identified needs transparently and publicly in a map application. At the same time, this argument also justifies the influence of citizen involvement and participation and identifies demand.

Although both areas naturally contain even more potential for action. A place for strategic action has thus been identified.

The follow-up question to the identification was: Is there potential that AIM can address at this point? In this case, the answer is yes, as the map application, for example, can be fed from the AIM data, thus fulfilling the requirement for transparency in digital administration. A technical realization would be feasible.

All cells of the matrix (Table 1) were analyzed similarly. The existence of an effect was documented with plus +. The grey background of the cell showed the positive answers to the second question.

Notably, none of the participants saw any potential for AIM in tendering and procurement processes. This is because there is already an existing and satisfactory technical solution that fully meets the legal and administrative requirements.

The potential benefits of integrating processes into AIM are more frequently perceived in the political and sociocultural domains. This can be explained by aligning the goals within these factors (digital administration, federal and state levels committed to using BIM, and citizen involvement and participation). The underlying political strategies aim to increase the transparency of administrative processes, improve efficiency, and reduce long-term costs through error prevention and faster turnaround times. These are also key concepts in AIM.

The area of socio-cultural trends and climate change is characterized by a high degree of complexity and issues at the transport network level. Effective AIM is assumed to enable the technical integration of data sources and the IT capture of external parameters to map trends and support decision-making. Artificial intelligence appears to be useful for handling this complexity.

5 Conclusion and Outlook

This paper continues a previous consideration and aims to make a valuable contribution to the efficient management of municipal infrastructure. It focuses on identifying areas for efficient Asset Information Management (AIM) in the municipal road maintenance sector. The results are divided into two parts. Firstly, planning and recording measures in municipal road administrations are examined. Secondly, macroscopic trends in the various steps of municipal road preservation are considered. The study aimed to identify the points at which strategic action is required to achieve a future-proof, i.e., resilient municipal construction administration fields of action for AIM were also identified. The results are presented in a matrix that investigates the effects of trends on process steps and highlights the potential for support by AIM.

The limitations of this contribution lie in the study's uniqueness and the small sample. The results were elaborated in close cooperation with employees of a single municipality (City of Münster) and, therefore, represent motivation and by no means a contradiction to the overarching approach but are difficult to generalize. Examining municipal challenges at the international level is undoubtedly complex, given significant variations in national politics and laws. Nevertheless, there are areas where uniformly similar development potentials exist beyond national borders. These need to be identified and subjected to scientific research. It is essential to differentiate between issues in this context, addressing

Table 1. Overview of results: Influence of macro trends on the renovation and renewal process steps.

| | | renovation and renewal process | | | | | | | | |
|-----------------|---|--------------------------------|---|----------------|-------------------|--------------------|---------------------------|--------------------------|----------------------|---|
| | | Identifying demand | Preparation of the construction program | | | Political decision | Tendering and procurement | Construction realization | Information feedback | |
| external factor | | | Creation of a list of measures | Prioritization | Budget management | | | | | |
| political | digital administration | + | + | + | + | + | + | + | + | + |
| | federal and state levels committed to the use of BIM | + | + | + | + | + | + | + | + | + |
| | citizen involvement and participation | + | (+) | | | | + | | + | |
| | transformation of mobility | | | + | | | | | | |
| | consolidation of finances | | | | + | | + | | | |
| economical | shortage of building materials | | | + | | | | + | + | |
| | inflation | | | | + | | + | | | |
| | energy prices | | | | + | | + | + | | |
| | out- or insourcing | | | | | | | | | |
| | refurbishment backlog | | + | + | + | + | + | | | |
| socio-cultural | mobility participation and accessibility | + | + | + | | + | + | | | |
| | demographic change | + | + | + | | + | + | | | |
| | changes in urban planning | + | + | + | | + | + | | | |
| | increase in traffic | + | + | + | | + | + | | | |
| technical | technological development of building materials and processes | | | | | | | + | + | |
| | automation of machines and processes | | | | | | | | + | + |
| | increasing amounts of data | | | | | | | | | + |
| | artificial intelligence | + | + | + | | + | | | | + |
| ecological | climate change | + | + | + | (+) | | + | + | | + |
| | resource-efficient construction | | | | | | | + | + | |
| | concept of the "Sponge City" | | | + | | + | | + | + | |
| legal | environmental protection laws | | | + | | + | + | + | | |
| | EU taxonomy | | | | + | | | | | |
| | EU Construction Products Directive | | | | | | | + | + | + |
| | changes in procurement and contracting | | | | | | | + | | |
| | EU-MAK values | | | | | | | | + | |

the scientific community, and those that can be dealt with solely through targeted and structured consulting services or change and implementation projects. The relevance of municipal work has increased significantly due to its crucial role in shaping society by providing a future-proof built environment, especially according to today's fast-paced macroscopic changes. The pressure to invest public funds efficiently and with minimal economic loss in this task persists and increases. Additionally, the shortage of personnel in public administration has become a bottleneck, driven by a lack of skilled workers and demographic changes, at least in Germany.

Efficient management of municipal tasks requires comprehensive data for well-founded decisions, presenting opportunities for research and technological developments. These can contribute to deploying resources more efficiently and developing innovative solutions. Therefore, involving municipalities more actively as partners in research is crucial to gaining practical insights and addressing evolving needs effectively and sustainably in changing times.

References

1. Roelich, K., Knoeri, C., Steinberger, J.K., et al.: Towards resource-efficient and service-oriented integrated infrastructure operation. *Technol. Forecast. Soc. Chang.* **92**, 40–52 (2015). <https://doi.org/10.1016/J.TECHFORE.2014.11.008>
2. Buttgereit, A., et al.: Potential of holistic asset information management. In: Biondini, F., Frangopol, D.M. (eds.) Life-cycle of structures and infrastructure systems: Proceedings of the Eighth International Symposium On Life-Cycle Civil Engineering (IALCCE 2023), 2–6 July, 2023, Politecnico di Milano, Milan, Taylor & Francis, Erscheinungsort nicht ermittelbar, pp. 2313–2320 (2023)
3. Barns, S.: Smart cities and urban data platforms: Designing interfaces for smart governance. *City Cult. Soc.* **12**, 5–12 (2018). <https://doi.org/10.1016/j.ccs.2017.09.006>
4. Sonbul, O.S., Rashid, M.: Towards the structural health monitoring of bridges using wireless sensor networks: A systematic study. *Sensors (Basel)* **23**, (2023). <https://doi.org/10.3390/s23208468>
5. Flamme, S., et al.: Lösungsansätze für einen ressourceneffizienten Tiefbau - Ressourcenplan kommunaler Tiefbau: RekoTi. Deutsches Ingenieurblatt (2022)
6. Ruitenburg, R.J., Braaksma, A., van Dongen, L.: A multidisciplinary, expert-based approach for the identification of lifetime impacts in asset life cycle management. *Procedia CIRP* **22**, 204–212 (2014). <https://doi.org/10.1016/j.procir.2014.07.007>
7. Buttgereit, A., Gomolluch, S., Wesselmann, M.: Erhaltungsmanagement - Ein Baustein im Asset Management in Münster (Teil 2). *Straße und Autobahn*: 33–39 (2021)
8. Abdirad, H., Dossick, C.S.: Rebaselining asset data for existing facilities and infrastructure. *J. Comput. Civ. Eng.* **34**, (2020). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000868](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000868)
9. ISO (2014) Asset management—Overview, principles and terminology (55000)
10. M FinStrKom: Merkblatt über den Finanzbedarf der Straßenerhaltung in den Kommunen, FGSV, Köln (2019)
11. ISO: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling: Part 1: Concepts and principles 35.240.67(ISO 19650) (2018)
12. de-Almeida-e-Pais, J.E., et al.: Measuring the performance of a strategic asset management plan through a balanced scorecard. *Sustainability* **15**, 15697 (2023). <https://doi.org/10.3390/su152215697>
13. FGSV: Begriffsbestimmungen für das Straßen- und Verkehrswesen: BBSV, Stand Juni 2020. FGSV Verlag GmbH, Köln (2020)



Evaluating Automated Floorplan Generation: Benchmark on Residential Buildings

Abdullah Elsafty^(✉) and Timo Hartmann

Civil Systems Engineering, Technische Universität Berlin, Berlin, Germany
a.elsafty@tu-berlin.de

Abstract. In the rapidly advancing field of automated building assessment, the accurate generation of floorplans from point cloud data is crucial, particularly for residential buildings which form a significant part of the urban environment. This study presents detailed evaluation of the latest state-of-the-art methods in automated floorplan generation, focusing exclusively on their application in indoor residential buildings. Our analysis assesses these methods using a diverse range of metrics, including accuracy, efficiency, and scalability, to understand their performance in interpreting complex residential environments. We evaluate two approaches, uncovering their strengths and shortcomings in various scenarios. The results of this comparative study are critical; they not only highlight the current capabilities and limitations of these methods but also pave the way for future enhancements. Our findings provide valuable insights for both academic researchers and industry professionals, emphasizing the need for further innovation and precision in the field of automated residential floorplan generation. This work contributes to the ongoing development of automated building assessment methodologies, aiming to optimize the process of transforming point cloud data into accurate and functional floorplans for residential buildings.

Keywords: Floorplan Generation · Benchmark · Point Cloud · Automated Building Assessment · Residential Building

1 Introduction

In architectural design and urban planning, the generation of accurate and efficient floorplans from point cloud data is a major challenge, particularly in residential environments [1, 2]. Automated floorplan generation holds significant potential for both existing building maintenance and new construction planning, especially given the complexity and diversity of residential spaces [3, 4]. While substantial research has been conducted in the realm of automated floorplan generation, the primary focus has mainly been on commercial, public buildings, and infrastructure [3, 5, 6]. These studies have provided valuable insights into the application of automated methods in these environments but have left a significant gap in our understanding of their effectiveness in residential buildings. Residential structures, with their unique architectural features and varied layouts, present a distinct set of challenges that are not typically encountered in commercial or

public buildings [7]. This discrepancy in research focus has resulted in a lack of targeted solutions and insights for one of the most common and essential types of buildings in urban environments. This study addresses the gap in automated floorplan generation for residential buildings, a crucial yet often overlooked area in architectural technology. Our focus is on critically evaluating the accuracy and applicability of contemporary methods in generating floorplans for indoor residential buildings. We aim to compare the efficacy of various state-of-the-art methods, identifying their strengths and limitations in dealing with the complexities of residential structures [8]. This research is pivotal for advancing automated building assessment, enhancing the precision and utility of these methods in residential settings. The central problem this study seeks to address is the current inadequacy of automated floorplan generation methods when applied to residential buildings. Many existing approaches struggle to accurately interpret the unique characteristics of residential interiors, such as varied layouts and architectural nuances. This often results in floorplans that are imprecise or not fully functional for practical applications [8]. By focusing on the specific challenges associated with residential buildings, this research endeavors to uncover the limitations of existing methods and identify potential areas for technological advancement. The goal is to refine these methods to better accommodate the distinct needs of residential floorplan generation, thereby supporting more effective and efficient urban and residential development. Our study contributes to narrowing the knowledge gap in this field, offering valuable insights for both academic research and practical applications in architectural technology. The goal is to improve the efficiency and reliability of floorplan generation processes, ultimately leading to better planning, resource allocation, and quality in residential building projects.

2 Background

In the field of architectural planning and maintenance, automated floorplan generation from point cloud data has gained significant momentum. Despite the advancements in 3D modeling (Scan to BIM), 2D floorplans remain essential for quick decision-making and efficient maintenance processes [3]. Historically, the creation of floorplans transitioned from manual drawings, which were time-consuming and prone to inaccuracies, to semi-automated methods that blended manual inputs with computer-aided design [9]. The current situation, however, is increasingly turning to fully automated approaches. Driven by technological advancements in computer vision and machine learning, these methods aim to streamline the process of transforming point cloud data into detailed and accurate 2D floorplans, thereby reducing the need for manual intervention, and enhancing overall efficiency and precision [10]. Reviewing recent literature on automated floorplan generation from point clouds, several notable approaches and their inherent challenges have been identified. A study proposed a method utilizing RGB images alongside dense 3D meshes for multilevel plane detection [11], combining image semantics with point cloud geometry. However, this method lacked detailed semantics for features like doors and windows and struggled in areas devoid of the initial mesh. Another approach [7], focused on generating engineering drawings through complex plane projection and boundary extraction. While innovative, this technique encountered difficulties with complex geometries and shorter wall structures. Similarly, a geometric priors-based method

outlined a five-step floorplan generation process from 3D point clouds [3]. Despite its ingenuity, it relied on another method [12] -FloorSP- for internal room boundary generation, potentially limiting its global optimality. Advancements in Modeling generation were observed in [13], where a deep learning-based approach using RGB-D or LiDAR point clouds was followed by energy minimization for space boundary optimization. Despite its advanced approach, it suffered from geometric accuracy loss. In a different vein, another method introduced a technique detecting wall planes from LiDAR or MVS point clouds [14], employing Integer Linear Programming (ILP) for ground plane slicing. While versatile, it faced challenges in capturing smaller planes in deficient point clouds. A strategy focusing on space decomposition into polygonal partitions was presented in [1], using energy minimization and wall-structure edge selection. Relying on room labels from Mask R-CNN and detected wall planes, it was prone to geometric errors due to inaccurate inputs. Addressing cluttered environments, [15] projected 3D point cloud data onto 2D planes using voxel space, extracting walls through density histograms and the Hough transform. However, it was vulnerable to occlusions common in cluttered spaces. Similarly, [16] extracted vertical structures by refining planar surface extraction with door detection algorithms but faced limitations in occlusion-heavy environments. Recent models have aimed to increase resilience to clutter. Methods in [17, 18] extracted linear structures from voxelized spaces, using RANSAC and k-means clustering, respectively. However, their reliance on scene-specific knowledge or trajectory information limited their broader applicability. Addressing indoor challenges, [19] detected walls within 2D grids but depended on panoramic imaging, a cost-prohibitive approach for many. Conversely, [20] used LiDAR-based point clouds for planar surface classification yet faced issues with noise sensitivity and point density inconsistencies. In the construction industry, capturing complex geometry for as-is conditions has led to multi-dimensional watertight models, as seen in [21, 22]. These methods, adept at modeling wall structures, required external scene capture and multi-sensor integration, facing limitations indoors due to occlusions. Exploring line structure extraction in building facades, [23, 24] assumed unobstructed targets, an assumption not realistic for indoor environments. Addressing this, [25] utilized semantic labeling and mobile laser scanning, optimized with deep learning and machine learning-based edge line estimation algorithms, presenting a forward-thinking approach to indoor scanning challenges. The complexity of residential buildings adds another layer of difficulty to automated floorplan generation. Unlike commercial or public buildings, which often follow standard design patterns, residential buildings can vary greatly in terms of size, layout, and architectural features. To summarize, many current automated methods for generating floorplans from point cloud data rely on supplementary data, such as images or meshes, to enhance the accuracy and detail of the generated plans. However, a few cutting-edge approaches focus exclusively on utilizing the raw point cloud data in its simplest form (x , y , z coordinates) for floorplan generation. These methods, which represent some of the latest state-of-the-art techniques in the field, have been selected for evaluation in our study [26, 27], by relying solely on point cloud data, these approaches present a more simplified and potentially scalable solution for automated floorplan generation, making them particularly suited for widespread application in residential settings.

3 Methodology

To achieve the objectives of this study, we employ a structured methodology centered around the evaluation of two latest automated floorplan generation methods. This evaluation is conducted using a new benchmark dataset, comprising a diverse range of point cloud data from various residential buildings. The dataset is carefully collected to represent a wide range of residential structures, ensuring a thorough and unbiased assessment of the methods. The dataset consists of point cloud data from two distinct residential environments (see Fig. 1): a first floor from residential house (D1) and a high-rise apartment (D2). These selections offer a varied representation of residential structures, from the intricate layouts of a multi-story house to the unique challenges posed by high-rise apartments. The data, acquired using Terrestrial Laser Scanning (TLS) and Mobile Laser Scanning (MLS), captures detailed indoor environments, including elements of noise and everyday clutter, to provide a realistic setting for evaluation. Reference models for these datasets were created using Autodesk Revit™ (see Fig. 1). These models, derived solely from the point cloud data, feature comprehensive floorplans, and serve as benchmarks for assessing the accuracy and effectiveness of the automated floorplan generation methods.

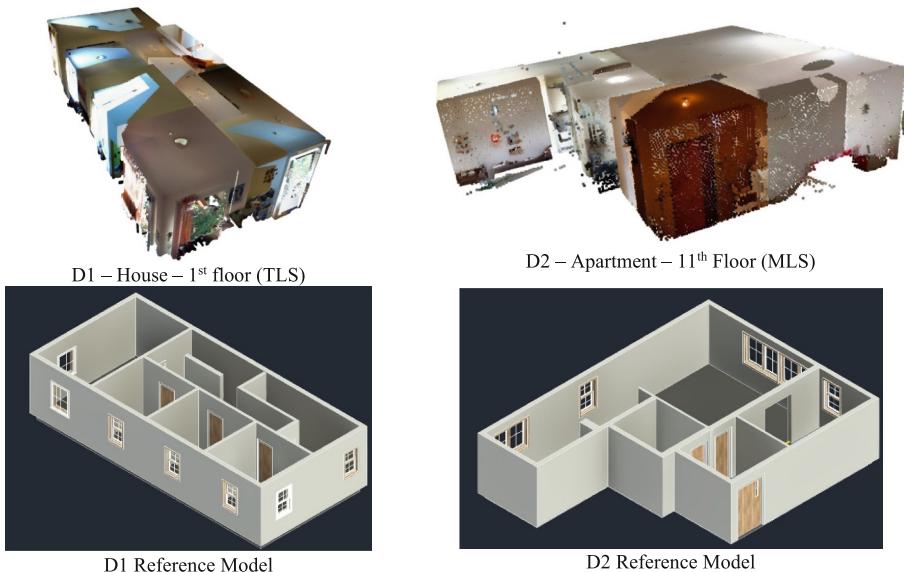


Fig. 1. Benchmark datasets with 3D reference models.

Our approach begins with an in-depth study of each selected automated method, focusing on understanding their algorithms, techniques, and underlying assumptions [28]. This theoretical understanding forms the basis for their practical implementation on datasets. The implementation phase adapts each method to our specific datasets, ensuring fidelity to their original design while accommodating the unique characteristics

of D1 and D2. The aim is to recreate each method in a simplified yet effective form, adhering as closely as possible to the original design but also making adaptations as necessary for our specific datasets. This step is critical for gaining insights into the feasibility, adaptability, and resource demands of each method. The effectiveness of the floorplan generation methods is quantitatively assessed using established metrics: recall, precision, and the F1-score. These metrics provide a comprehensive evaluation of the methods' precision and accuracy. Additionally, scalability is assessed, reflecting the diverse sizes and complexities of residential floorplans. This methodology aims to deliver a nuanced understanding of each method's performance, particularly in the context of the varying layouts and challenges presented by datasets. The selected methods for this evaluation are among the most recent advancements in automated floorplan generation from point cloud data. The criteria for their selection were twofold: firstly, their reliance solely on point cloud data as input, eschewing the need for supplementary data types like images; secondly, their demonstrated performance, with one method exhibiting an accuracy exceeding 90%, and the other having been successfully implemented in real-world residential settings. These criteria align perfectly with the main scope of our paper, making these methods ideal candidates for evaluation.

Method 1 employs advanced segmentation techniques to transform 3D point cloud data into 2D geometric models, boasting over 90% accuracy. This method is notable for its precision, essential in residential settings where exact spatial representation is critical [26].

Method 2 utilizes machine learning algorithms to detect key feature points in point clouds, converting these into geometric models. This method is distinguished by its efficiency in extracting geometric primitives from complex point clouds, demonstrating high accuracy in digital model generation [27].

4 Experimental Setup

4.1 Overview

Our study's experimental setup is tailored to assess the accuracy and scalability of two state-of-the-art automated floorplan generation methods using point cloud data from residential buildings. Our experimental structure is both systematic and iterative. It begins with a comprehensive understanding of each selected method, followed by a practical implementation on D1 dataset. This initial implementation serves as a baseline to gauge the method's performance and adaptability. The implementation process involves adjusting parameters within the scope of the original method's design, ensuring fidelity to its core principles while also making necessary adaptations for our specific dataset. Each method is then sequentially tested across the second dataset in our benchmark, with parameter adjustments and considerations made for each unique point cloud dataset. This iterative process allows us to fine-tune the methods for diverse residential environments, ensuring a robust evaluation. The experimentation unfolds in a structured manner. After the initial implementation and testing, each method is applied to the second dataset. This is followed by a rigorous evaluation of the results against our constructed reference

models. The final stage involves a comparative analysis, comparing the performance of the methods across all datasets, thereby offering a detailed view of their potential and limitations in automated floorplan generation for residential buildings.

4.2 Implementation

For the implementation of both methods, Python was used alongside the Open3D library. This choice of software and tools reflects a balance between accessibility and computational power. In our implementations, no modifications were made to the core algorithms of the methods. Instead, we focused on adjusting parameters to optimize performance for each dataset, this approach ensured that while the integrity of the methods remained intact, they were fine-tuned to suit the unique characteristics of each residential dataset, such as point density, noise and furniture level, and spatial complexity. The parameter adjustments were key in adapting these methods to different types of residential environments, ensuring that they could effectively process and analyze the varying point cloud data. This strategy highlights our focus on assessing the methods' adaptability and performance in a range of real-world settings, an essential aspect of their evaluation.

4.3 Challenges and Considerations

A key challenge was the fine-tuning of parameters, requiring precise adjustments to optimize method performance across different datasets. This process was complicated by the sensitivity of the outcomes to parameter changes and the lack of standardized tuning guidelines. The incomplete or ambiguous descriptions in the original papers added to the complexity of replication. In some cases, this led to developing our implementation strategies to align with the conceptual frameworks provided. The complexity and variability in residential buildings necessitated a flexible and adaptable approach in method implementation. Finding a general effective set of parameters was challenging due to the diverse architectural styles and layouts in residential structures.

5 Results and Discussion

5.1 Visualized Results of Method 1

Method-1 employs a structured four-step process (see Fig. 2) to convert point cloud data into 2D floorplans: slicing, voxelization, coarse segmentation, and fine segmentation, where each step contributes to the method's overall accuracy.

In our study, Method-1 was applied to datasets, with the generated floorplans compared against reference models. This comparison (see Fig. 3), highlights the method's ability to capture the unique characteristics and complexities of different residential environments.

Quantitative performance of Method-1, shown in Table 1, provides insights into its effectiveness across these datasets, evaluated through recall, precision, and F1-score metrics. The precision peaks in D1 and D2 suggest a strong ability of Method-1 to accurately identify actual floorplan elements. However, variations in recall highlight

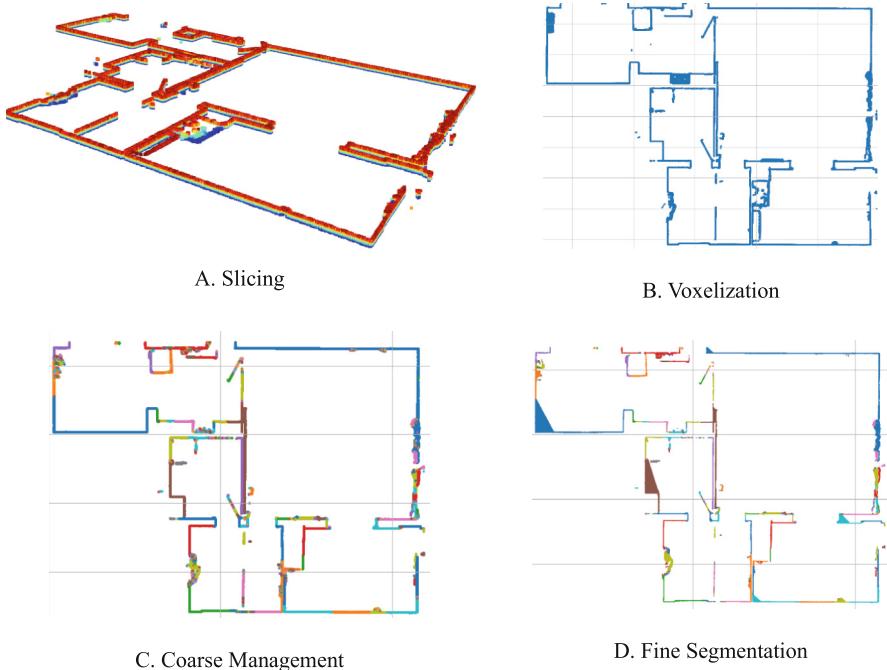


Fig. 2. Method-1 implementation approach

Table 1. Evaluation metrics for method-1 implementation

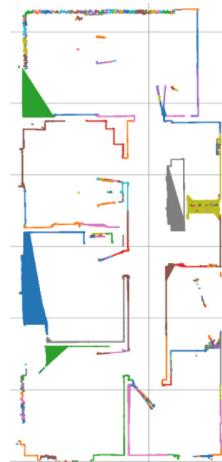
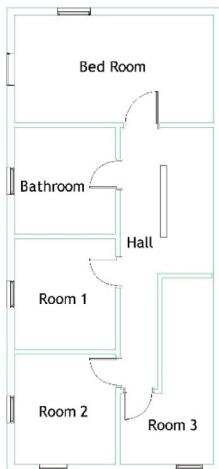
| | Recall | Precision | F1-score |
|----|--------|-----------|----------|
| D1 | 0.84 | 0.89 | 0.86 |
| D2 | 0.70 | 0.80 | 0.74 |

the influence of specific dataset characteristics, such as complexity and noise, on the method's performance.

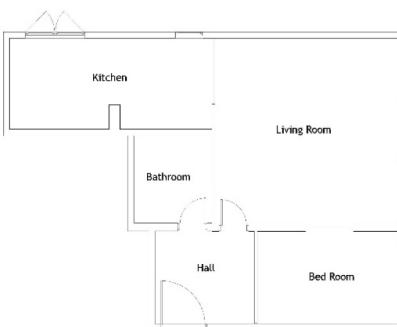
The quantitative analysis reveals potential areas for enhancement in Method-1, suggesting the need for further optimization or algorithmic adjustments to better handle the diverse challenges present in residential buildings. This highlights the importance of adaptability and dataset-specific considerations in the application of automated floorplan generation methods.

5.2 Visualized Results of Method 2

Method-2 follows a four-step process for extracting floorplans from point cloud data, involving edge detection, edge thinning, non-convex hull algorithm for feature point conversion, and line refinement using RANSAC (see Fig. 4). Despite its structured



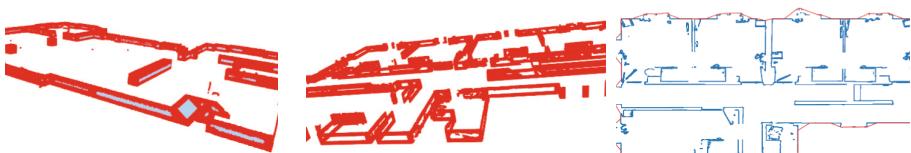
D1: Reference model alongside with the generated floorplan by method-I.



D2: Reference model alongside with the generated floorplan by method-I.

Fig. 3. Method-1 implementation results along with the reference models.

approach, all datasets presented significant challenges, particularly at the non-convex hull algorithm stage, affecting the accuracy of the generated floorplans.

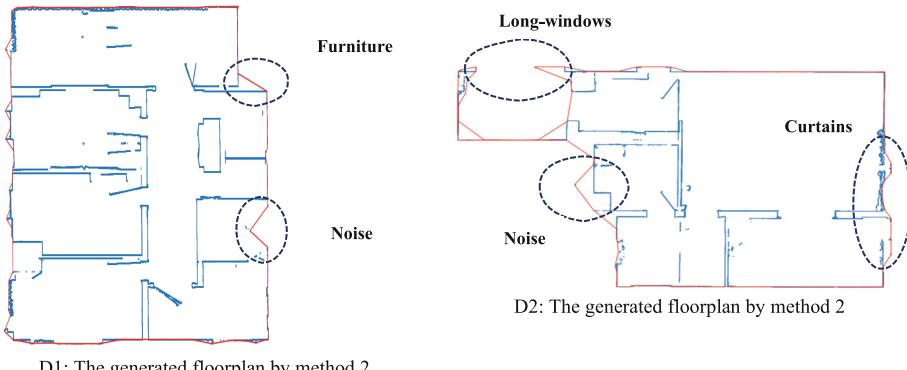


- Edge Detection
- Edge thinning
- Non-Convex Hull

Fig. 4. Method-2 implementation approach.

When applied to datasets, Method-2 showed initial effectiveness in edge detection and refinement. However, its performance was consistently hindered during the non-convex hull algorithm phase (see Fig. 5), failing to accurately convert feature points

into solid lines for each dataset. The challenges varied across the datasets: D1, faced issues with noise, complex features like furniture. D5's challenges included noise, long windows, and curtains that the algorithm struggled to process.



D1: The generated floorplan by method 2.

Fig. 5. Method-2 non-convex hull results along with the challenges per dataset.

Despite its successful application in a controlled residential environment in the original study, our real-world datasets introduced complexities that impeded Method 2's performance. The presence of noise, intricate designs, and various household elements, which were prevalent across our datasets, posed significant challenges. These included structural elements and reflective surfaces, which may not have been as prominent in the dataset used in the original study.

These findings indicate that while Method 2 demonstrates potential, its adaptability to diverse and unfiltered real-world environments needs enhancement. The variation in performance between the controlled environment of the original study and our real-world datasets underscores the necessity of testing automated floorplan generation methods across various settings to ensure their broad applicability and resilience.

5.3 Discussion

In the implementation of both methods, we faced the intricate task of parameter adjustment, a process that proved to be both critical and challenging. Each method came with a multitude of parameters, each playing a pivotal role in the outcome of the floorplan generation process.

For instance, in Method I, parameters like slicing height, voxel size, and pixel threshold were carefully adjusted. While these adjustments were crucial for tailoring the method to our specific datasets, they also introduced a level of complexity. Altering a single parameter could significantly affect the results, sometimes leading to less-than-optimal outcomes. This necessitated a process of meticulous fine-tuning, where each parameter was systematically varied, and its impact observed. Similarly, in Method II, parameters such as neighbors num, kernel bandwidth, and non-convex hull threshold required precise tuning. The challenge here was not just in adjusting these parameters

but in understanding their interplay and collective impact on the final model. The fine-tuning process was not straightforward; it lacked a systematic approach, often requiring a trial-and-error method to find the optimal settings. This aspect of the implementation highlighted the delicate balance needed between adhering to the method's theoretical underpinnings and adapting it to real-world, varied datasets.

These experiences underscore the nuanced nature of working with advanced floorplan generation methods. Fine-tuning parameters is not merely a technical task but an art that requires a deep understanding of both the methods and the data. It involves striking a balance between fidelity to the original algorithm and the practical realities of diverse and complex point cloud datasets.

6 Conclusion

The objective of our paper was to critically assess the accuracy and applicability of advanced automated floorplan generation methods using point cloud data, with a special emphasis on residential buildings. Our approach was rooted in a literature review, identifying a gap in the evaluation of such methods specifically for indoor environments. By implementing and adjusting two state-of-the-art methods across two datasets, we sought to provide a detailed comparison of their performance and reveal significant insights for future technological developments in architectural assessment.

Our exploration into automated floorplan generation methods concluded with significant findings, the evaluation of Method I across residential datasets showcased its robustness, with high accuracy levels that, while commendable, did not reach the 90% threshold achieved within the original study's datasets. This discrepancy highlights the challenges of replicating controlled experimental results within a broader range of real-world settings. Method II, despite its solid theoretical foundation, struggled when applied to the practical complexities found within our residential datasets. This method's practical challenges underline the necessity for automated floorplan generation techniques that can adapt to the intricate and varied features of real-world residential spaces. The study revealed that residential buildings exhibit unique characteristics that demand a tailored approach to automated floorplan generation. The necessity of fine-tuning parameters for each dataset serves as evidence of this diversity. Notably, even though Method II was previously successful in other residential implementations, it did not effectively adapt to the residential datasets in our study, suggesting that successful application in one residential environment does not guarantee efficacy across all. These findings collectively emphasize the importance of developing flexible and robust floorplan generation methods that can address the nuances of different residential environments and maintain high accuracy to be truly effective in real-world applications.

Our study is constrained by the availability and variety of datasets representing residential environments. The need for more benchmarks capturing the full spectrum of these settings is apparent. A richer collection of datasets would enhance the robustness of our evaluations and provide a clearer picture of the capabilities and limitations of automated floorplan generation methods. Also, there is a need for more extensive evaluations of the proposed methods across a wider range of datasets. Such an endeavor would not only validate the effectiveness of the methods in diverse scenarios but also help chart a path

for future developments in the field, informing us of our current standing and the next steps to take. While we implemented two advanced methods, there remains an element of uncertainty in fully replicating the original authors' work. The lack of detailed implementation guidance and code availability from the authors means we cannot be certain our implementation matches theirs at a one-to-one level, potentially affecting the comparability of results. The study's methodological approach, while thorough, suggests room for improvement. Future research might explore more systematic ways of parameter tuning and algorithm adaptation to better handle the intricacies of different residential datasets, leading to advancements in the field of automated floorplan generation.

Acknowledgment. We acknowledge the support of the ReconTOP project, and the funding provided by the Deutsche Forschungsgemeinschaft (DFG), which were crucial in the realization of this research.

References

1. Fang, H., Lafarge, F., Pan, C., Huang, H.: Floorplan generation from 3D point clouds: A space partitioning approach. *ISPRS J. Photogramm. Remote Sens.* **175**, 44–55 (2021). <https://doi.org/10.1016/j.isprsjprs.2021.02.012>
2. Pintore, G., Mura, C., Ganovelli, F., Fuentes-Perez, L., Pajarola, R., Gobbetti, E.: State-of-the-art in automatic 3D reconstruction of structured indoor environments. *Comput. Graph. Forum* **39**, 667–699 (2020). <https://doi.org/10.1111/cgf.14021>
3. Cai, R., Li, H., Xie, J., Jin, X.: Accurate floorplan reconstruction using geometric priors. *Comput. Graph. (Pergamon)* **102**, 360–369 (2022). <https://doi.org/10.1016/j.cag.2021.10.011>
4. Spanedda, F., Fusaro, M.C.: New approaches to housing complexity: designing dwellings in the age of cognitive economy. *City, Territ. Arch.* **7** (2020). <https://doi.org/10.1186/s40410-020-00128-5>
5. Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y., Behnood, A.: 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. *Adv. Eng. Inf.* **51** (2022). <https://doi.org/10.1016/j.aei.2021.101501>
6. Spencer, B.F., Hoskere, V., Narazaki, Y.: Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering* **5**, 199–222 (2019). <https://doi.org/10.1016/j.eng.2018.11.030>
7. Kong, Q., Liao, L., Yuan, C.: Rapid generation of editable engineering drawings from 3D point cloud reconstruction for large-scale buildings. *J. Build. Eng.* **63** (2023). <https://doi.org/10.1016/j.jobe.2022.105486>
8. Dey, E.K., Awrangjeb, M., Kurdi, F.T., Stantic, B.: Building boundary extraction from LiDAR point cloud data. In: 2021 digital image computing: techniques and applications (DICTA), IEEE, 2021: pp. 1–6. <https://doi.org/10.1109/DICTA52665.2021.9647371>.
9. Houshiar, H.: Documentation and mapping with 3D point cloud processing, Dissertation. Universität Würzburg (2017). <https://doi.org/10.1515/9783112339923>

10. Chmelar, P., Rejsek, L., Nguyen, T.N., Ha, D.H.: Advanced methods for point cloud processing and simplification. *Appl. Sci. (Switzerland)* **10** (2020). <https://doi.org/10.3390/APP10103340>
11. Han, J., Liu, Y., Rong, M., Zheng, X., Shen, S.: FloorUSG: Indoor floorplan reconstruction by unifying 2D semantics and 3D geometry. *ISPRS J. Photogramm. Remote Sens.* **196**, 490–501 (2023). <https://doi.org/10.1016/j.isprsjprs.2023.01.020>
12. Chen, J., Liu, C., Wu, J., Furukawa, Y.: Floor-SP: Inverse CAD for floorplans by sequential room-wise shortest path, In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2661–2670 (2019). <https://doi.org/10.1109/ICCV.2019.00275>
13. Tang, S., Li, X., Zheng, X., Wu, B., Wang, W., Zhang, Y.: BIM generation from 3D point clouds by combining 3D deep learning and improved morphological approach. *Autom. Constr.* **141** (2022). <https://doi.org/10.1016/j.autcon.2022.104422>
14. Han, J., Rong, M., Jiang, H., Liu, H., Shen, S.: Vectorized indoor surface reconstruction from 3D point cloud with multistep 2D optimization. *ISPRS J. Photogramm. Remote Sens.* **177**, 57–74 (2021). <https://doi.org/10.1016/j.isprsjprs.2021.04.019>
15. Okorn, B., Xiong, X., Akinci, B.: Toward automated modeling of floor plans. In: Proceedings of the Symposium on 3D Data Processing, Visualization and Transmission, Paris, pp. 1–8 (2010). <https://api.semanticscholar.org/CorpusID:3334144>
16. Babacan, K., et al.: Towards object driven floor plan extraction from laser point cloud. *Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **41**, 3 (2016). <https://doi.org/10.5194/isprarchives-XLI-B3-3-2016>
17. Hossein Pouraghdam, M., Saadatseresht, M., Rastiveis, H., Abzal, A., Hasanlou, M.: Building floor plan reconstruction from slam-based point cloud using ransac algorithm. *Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci. ISPRS Archives* **42**, 483–488 (2019). <https://doi.org/10.5194/isprarchives-XLII-4-W18-483-2019>
18. Stojanovic, V., Trapp, M., Richter, R., Döllner, J.: Generation of approximate 2D and 3D floor plans from 3D point clouds. In: VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications 1, pp. 177–184 (2019). <https://doi.org/10.5220/0007247601770184>
19. Fotsing, C., Hahn, P., Cunningham, D., Bobda, C.: Volumetric wall detection in unorganized indoor point clouds using continuous segments in 2D grids. *Autom. Constr.* **141**, 104462 (2022). <https://doi.org/10.1016/j.autcon.2022.104462>
20. Gankhuyag, U., Han, J.H.: Automatic 2D floorplan CAD generation from 3D point clouds. *Appl. Sci. (Switzerland)* **10** (2020). <https://doi.org/10.3390/APP10082817>
21. Vo, A.V., Truong-Hong, L., Laefer, D.F., Bertolotto, M.: Octree-based region growing for point cloud segmentation. *ISPRS J. Photogramm. Remote Sens.* **104**, 88–100 (2015). <https://doi.org/10.1016/j.isprsjprs.2015.01.011>
22. Heo, J., et al.: Productive high-complexity 3D city modeling with point clouds collected from terrestrial LiDAR. *Comput. Environ. Urban Syst.* **41**, 26–38 (2013). <https://doi.org/10.1016/j.compenvurbsys.2013.04.002>
23. Pu, S., Vosselman, G.: Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS J. Photogramm. Remote Sens.* **64**, 575–584 (2009). <https://doi.org/10.1016/j.isprsjprs.2009.04.001>
24. Ripperda, N., Brenner, C.: Application of a formal grammar to facade reconstruction in semiautomatic and automatic environments. In: 12th AGILE International Conference on Geographic Information Science, pp. 1–12 (2009). <http://www.ikg.uni-hannover.de/agile/fileadmin/agile/paper/44.pdf>
25. Wang, C., et al.: Semantic line framework-based indoor building modeling using backpacked laser scanning point cloud. *ISPRS J. Photogramm. Remote Sens.* **143**, 150–166 (2018). <https://doi.org/10.1016/j.isprsjprs.2018.03.025>

26. Kim, M., Lee, D.: Automated two-dimensional geometric model reconstruction from point cloud data for construction quality inspection and maintenance. *Autom. Constr.* **154**, 105024 (2023). <https://doi.org/10.1016/j.autcon.2023.105024>
27. Kim, M., Lee, D., Kim, T., Oh, S., Cho, H.: Automated extraction of geometric primitives with solid lines from unstructured point clouds for creating digital buildings models. *Autom. Constr.* **145**, 104642 (2023). <https://doi.org/10.1016/j.autcon.2022.104642>
28. Weber, R.E., Mueller, C., Reinhart, C.: Automated floorplan generation in architectural design: A review of methods and applications. *Autom. Constr.* **140**, 104385 (2022). <https://doi.org/10.1016/j.autcon.2022.104385>



Multi-Objective Optimization of a Sustainable LC3 Mortar for 3D Printing

Willy Jin^{1,2}, Claudiane Ouellet-Plamondon², and Jean-François Caron¹(✉)

¹ Navier Laboratory, École des Ponts ParisTech, Univ. Gustave Eiffel, CNRS, Marne-La-Vallée, France

jean-francois.caron@enpc.fr

² Department of Construction Engineering, École de Technologie Supérieure, Montreal, QC, Canada

Abstract. The construction industry is compelled to evolve toward industrialized and sustainable means and processes. Indeed, the Portland cement production alone is responsible for 4 to 8% of global greenhouse gas emissions. In this context, concrete 3D printing demonstrates a significant potential for the reduction of material use (formwork-free, structural optimization), but the majority of 3D printing materials still hold a high carbon intensity mainly due to the rheological constraints associated with pumpability, extrudability and buildability. The present study proposes to integrate a parametric life cycle assessment model in the multi-objective optimization of a 3D printing mortar in order to minimize the environmental impact with a significantly reduced mixture design workload. Applied to a limestone calcined clay cement-based mortar with high cement substitution, a non-dominated sorting genetic algorithm is used to decrease the climate change score while maintaining a set of rheological properties which are predicted by artificial neural networks and define the printability. In this way, the non-linear behavior of cementitious materials is represented in the design and the multiplicity of independent and dependent variables is handled adequately. As a result, starting from a reference mixture with already low cement content (6.8 wt%) and a dataset of 20 to 30 mixtures, this methodology allows the identification of more sustainable printable mixtures in as low as 7 additional formulations. Besides, as the process advances, its efficiency is enhanced. This methodology is reproducible with locally sourced materials and for the majority of 3D printing materials, which are usually designed through empirical trial and error. Therefore, it could also be applied with different objective functions to bicomponent materials, which offer added printing flexibility. This study introduces a systematic optimization process which establishes the sustainability at the core of its objectives and includes new tools for the formulation of cementitious materials.

Keywords: Sustainability · 3D printing · LC3 · Optimization · Machine learning

1 Introduction

This study is part of a necessary move towards automation in the construction industry, which must occur in association with a reduction of its environmental impacts. The development of the concrete 3D printing technology is seen as a way to increase productivity, reduce manual labor and add functionality to concrete structures. In addition to that, it offers an opportunity to cut down on the use of raw materials by eliminating the need for formwork and allowing structural optimization [1]. However, this saving is often overshadowed by the use of materials displaying high cement contents, which represents the main contributor to environmental impacts in concrete. Indeed, the rheological constraints associated with 3D printing, especially without accelerating admixtures, as well as the limitation in aggregate size, lead to a higher cement content in the binder along with a higher binder content in the concrete.

The nature of concrete mix design makes it demanding in time and resources. The non-linear behavior of cementitious materials and the variability in raw materials result in a complex optimization of material properties. This leads to design methods which do not explore all possibilities of the search space and stop at satisfactory but unoptimized concretes [2]. As a consequence, a solution involving artificial neural networks has been developed for the optimization of high-performance 3D printing concrete with discrete factors and fresh to hardened state properties [3]. This active learning process is implemented in this study with continuous parameters and with sustainability as its foremost objective. As such, the material of interest is a limestone calcined clay cement-based mortar (LC3) with high cement substitution. The goal is to propose a reproducible, automated methodology for tuning the rheological and environmental properties of this LC3.

In that regard, the materials and methods are presented in the Sect. 2, the optimization process is detailed in the same part and the results of the study are depicted in the Sect. 3.

2 Materials and Methods

This section presents the materials used as well as the printability characterization procedure at fresh state. In addition to that, the framework for the calculation of climate change (ClCh) score is quickly described and a focus is put on the global optimization methodology.

The LC3 is composed of general use cement blended with 8% silica fume (GUbSF), calcined clay (CC), limestone filler (LF), sand (Sa) and polycarboxylate ether superplasticizer (SP).

2.1 Methods

A range of rheological tests are performed at fresh state in order to characterize the printability of initial and optimized mortar mixtures. The results of these tests also serve as a training dataset for artificial neural networks (ANNs). Along with the calculation of ClCh score, the predictions of these ANNs constitute the objective functions employed in a multi-objective Pareto optimization. The resulting LC3 is suited for a continuous

mixing system which involves a *MAI 4MULTIMIX* mortar mixing pump and an *ABB IRB 6620* 6-axis industrial robot.

Rheological tests

The characterization protocol represents the three main parts of a printing process, namely **pumping, deposition and structuration**. Thus, the flow table test [4] is performed at 2 min after mixing in order to evaluate the **flowability** in the pumping time-frame. A truncated cone is filled with two layers, both tampered 20 times with a rubber rod. The cone is then lifted and the table is dropped 25 times in 15 s, with a steady pace. The **shape retention** is measured through the slump test with a miniature Abrams cone with a base diameter of 100 mm, a top diameter of 50 mm and a height of 150 mm according to the norm ASTM C143/C143M [5]. The mold is filled with three layers of mortar, each tampered 25 times with a steel rod. The mold is then removed vertically in 5 s at the 8 min mark, letting the mortar slump under its own weight. The resulting static yield stress is estimated through Eq. 1 [6], with τ the static yield stress, ρ the material density, g the gravity, H_0 the initial height of the sample and s the slump value:

$$\tau = \frac{\rho g (H_0 - s)}{\sqrt{3}} \quad (1)$$

The structuration rate (Athix) is considered linear in the dormant period [7] with the form:

$$\tau(t) = \tau_0 + Athix * t \quad (2)$$

With τ_0 the initial static yield stress after deposition and *Athix* the structuration rate in Pa/s. In order to measure this *Athix* value, the direct shear test is carried out at 30 min and 60 min after mixing which allows us to obtain static yield stress measurements in the tens of kPa range. The mortar sample is placed in a circular shear box that restricts the horizontal strain but allows shearing on a horizontal plane where the plates of the box touch. The methodology is similar to [3], adapted from [8]. The box is filled in two layers and compacted on a vibrating table for 5 s. The samples are then loaded on a *Humboldt* automatic shear testing machine equipped with a displacement sensor, a load sensor, a linear displacement motor and a 100 kN load cell. The test was carried out on three samples at 30 and 60 min with a displacement rate of 12.7 mm/min. A linear regression between these values and the slump test yield stress gives a structuration rate in Pa/s.

A literature review along with large-scale 3D printing experiments allowed us to identify several rheological thresholds that define printability. The flow must be at least 120% at 2 min after mixing, then the slump must be between 60 and 70 mm 6 min after mixing. Finally, the *Athix* must satisfy an arbitrary criteria of 3.5 Pa/s in the first hour after mixing. These constitute our **experimental goals**.

Artificial Neural Networks (ANNs)

In order to train the ANNs for the prediction of these properties, we conduct a design of experiments which maximizes the representation of the design space while setting a feasible number of runs (18). The python package *Tensorflow* then allows us to generate an

ANN for each characterization test using optimal hyperparameters. Indeed, the training of these ANNs is subject to a single-objective optimization in order to determine the best activation function, learning rate, number of hidden layers and number of neurons in each layer. The assessed activation functions are *exponential linear unit* (elu), *scaled exponential linear unit* (selu), *rectified linear unit* (relu), *hyperbolic tangent activation function* (tanh), *sigmoid activation function* (sigmoid), *softmax activation function* (softmax), *softplus activation function* (softplus), *softsign activation function* (softsign) and *exponential activation function* (exponential).

Life cycle assessment

The ClCh score is calculated with a life cycle assessment (LCA) model using the python package Brightway2. We conduct a cradle-to-gate analysis, excluding the printing phase, the use phase and the end of life. The functional unit is the **production of 1m³ of 3D printable mortar**. The focus is placed on the global warming potential (GWP), which is estimated with the life cycle impact assessment method Environmental Footprint 3.0 baseline model of 100 years. The reference LC3 mixture used as a starting point for the optimization process has a ClCh score of **251 kgCO₂eq**.

2.2 Optimization Process

In order to obtain suggestions of efficient mixtures, the concept of multi-objective optimization is used through a genetic algorithm named *NSGA-II* [9], implemented in the python package *Pymoo*. This algorithm iteratively evaluates sets of mixture populations with specific cost functions in order to generate a set of non-dominated solutions, which means no objective can be improved without penalizing another. When the individual is a global optimum in the search space, it is called a Pareto-optimal solution. The mechanisms of this non-dominated sorting, which includes elitism and a crowding distance fitness value is detailed on the Table 1 [9]:

In our case, the objective functions correspond to the ClCh score and the prediction of each of the trained ANN for flow, slump and Athix. The objectives are to **minimize the ClCh score while attaining the printability goals** defined above.

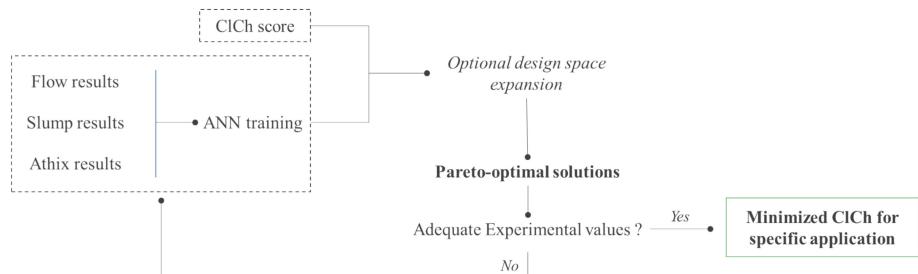
The general optimization process consists of 5 steps, depicted on the Fig. 1.:

- Step 1: Training of artificial neural networks for the prediction of flow, slump and Athix values
- Step 2: Optimization run with specified parameter boundaries
- Step 3: Experimental validation of three **selected** suggested mixtures
- Step 4: Optional expansion of design space
- Step 5: Repetition of steps 1 to 4 until a satisfactory mixture is found

The progressive expansion of the design space is necessary since the generalization ability of the ANNs is limited considering the dataset size. The parameter boundaries are widened if desired and when predictions improve. In the step 3, the selection of mixtures to validate is the responsibility of the user and must aim at checking different design regions in order to provide the largest amount of information for the subsequent training of ANNs (step 5).

Table 1. Non-dominated sorting genetic algorithm II (NSGA-II) process [9]

| | |
|--|--|
| $R_t = P_t \cup Q_t$ | Combine parent and offspring population |
| $F = \text{fast-non-dominated-sort}(R_t)$ | $F = (F_1, F_2, \dots)$, all nondominated fronts of R_t |
| $P_{t+1} = \emptyset$ and $i = 1$ | |
| until $ P_{t+1} + F_i \leq N$ | until the parent population is filled |
| crowding-distance-assignment(F_i) | Calculate crowding-distance in F_i |
| $P_{t+1} = P_{t+1} \cup F_i$ | Include i th nondominated front in the parent pop |
| $i = i + 1$ | Check the next front for inclusion |
| Sort(F_i, \prec_n) | Sort in descending order using \prec_n |
| $P_{t+1} = P_{t+1} \cup F_i [1:(N - P_{t+1})]$ | Choose the first $(N - P_{t+1})$ elements of F_i |
| $Q_{t+1} = \text{make-new-pop}(P_{t+1})$ | Use selection, crossover and mutation to create a new population Q_{t+1} |
| $t = t + 1$ | Increment the generation counter |

**Fig. 1.** Optimization process

3 Results and Discussion

Including the design of experiments and the results from the trial-and-error campaign carried out to identify an appropriate design region, the total number of runs before applying the optimization process is 30. The latter resulted in **5 repetitions** of the steps 1 to 5 with **one expansion** of the design space. The latter was adapted for the second round of optimization when the suggested results tended towards the lower GUbSF bound (Table 2).

The results for all suggested mixtures are presented in the Table 3, from iteration 1 (i1) to iteration 5 (i5). For each iteration, 3 mixtures were formed (i-1,3), this table presents the GWP reduction compared to the reference mixture (251 kgCO₂eq) and the experimental errors compared to the printability objectives.

In the first round of optimization, the genetic algorithm explores unseen design regions in iterations i1 and i2, thus generating substantial errors. This is expected as the ANN models struggle to generalize, which is a known issue with machine learning methods, especially when considering the initial dataset size. Nevertheless, the goal

Table 2. Parameter boundaries for the 2 rounds of optimization

| Parameter | Round 1 (kg/m ³) | Round 2 (kg/m ³) |
|-----------|------------------------------|------------------------------|
| GUbSF | 140–180 | 120–180 |
| CC | 110–150 | 90–150 |
| LF | 340–380 | 340–400 |
| W | 210–250 | 200–260 |
| Sa | 1200–1400 | 1200–1450 |
| SP | 5.5–7.5 | 5.5–8 |

of this study is to overcome this matter by repeating the training process in order to progressively explore the design space with a very limited number of runs. For the third iteration, the ANNs exhibit an increase in prediction quality, especially for mixtures i3-2 and i3-3. Indeed, the mixture i3-3 satisfies all the printability constraints while reducing the ClCh score by 6% compared to the reference mixture. Although this constitutes a local optimum, the expansion of the design region allows for an exploration of lower ClCh regions.

In the second round of optimization, the suggested mixtures also tend towards the lower GUbSF bound, which is an unseen experimental region. However, in as low as 2 iterations (6 mixtures), a suggested mixture (i5-3) satisfies the printability thresholds while generating a 10% reduction in ClCh score. Although we can observe that this very low cement content does not affect the rheological properties, it decreases the compressive strength. As such, the decision to stop the design space expansion is made.

The mean coefficients of determination across all iterations between ANN predictions and all experimental data are respectively 0.96, 0.90 and 0.76 for the flow, the slump and the Athix with optimized hyperparameters. From these five repetitions of hyperparameter optimization, the best activation functions for each rheological test can be pointed out. For the flow table test and the slump test, the *scaled exponential linear unit* function gives the best results, meaning the minimal error between test data and predictions. On the other hand, for the Athix, the *hyperbolic tangent* and *softsign* activation functions alternatively generate the best performances.

In terms of global warming potential, the reference mixture (251 kgCO₂eq) already corresponds to a low carbon 3D printing mortar compared to the literature [10]. Given that under a certain ClCh score, the formulation of more sustainable materials is complexified, we can see that this methodology is very efficient in identifying such mixtures. Indeed, this active learning methodology identified two printable mixtures in different search spaces after respectively 9 and 15 additional mixtures. From all the experimental results, it is observed that the structuration rate can be tuned by the proportion of CC, which enhances it with additional hydrates and higher specific surface [11].

From the Table 3, we can see a global decrease in errors along the iterations. As the optimization process advances, design regions are explored and those depicting the best potential are further investigated by the genetic algorithm. Although the user is responsible for the selection of non-dominated solutions to validate, this choice is

Table 3. Results of optimization: GWP reduction and errors from printability objectives

| Mix | GWP reduction (%) | Flow error (%) | Slump error (mm) | Athix error (Pa/s) | Printable |
|-------------|-------------------|----------------|------------------|--------------------|------------|
| i1-1 | 5 | 28 | 49 | 0,1 | No |
| i1-2 | 4 | 8 | 30 | 0,3 | No |
| i1-3 | 3 | 24 | 37 | 1,1 | No |
| i2-1 | 10 | 16 | 18 | 0,7 | No |
| i2-2 | 13 | 32 | 50 | 1,3 | No |
| i2-3 | 4 | 32 | 60 | 0,5 | No |
| i3-1 | 12 | 28 | 40 | 1,1 | No |
| i3-2 | 7 | 9 | 8 | 0,6 | No |
| i3-3 | 6 | 0 | 1 | 0,0 | Yes |
| i4-1 | 21 | 32 | 54 | 1,7 | No |
| i4-2 | 13 | 24 | 26 | 0,9 | No |
| i4-3 | 8 | 21 | 24 | 0,2 | No |
| i5-1 | 19 | 33 | 31 | 0,4 | No |
| i5-2 | 16 | 19 | 45 | 0,1 | No |
| i5-3 | 10 | 1 | 1 | 0,7 | Yes |

considerably easier to make than progressing through trial-and-error. In this process, the recommendation for first iterations is to choose solutions depicting the widest region in the bounded search space, prioritizing desired objectives. Then, as the predictions improve, the choice must be directed to the most favorable region.

4 Conclusion

In only 15 additional mixtures, this optimization process allowed us to adequately predict three rheological contradictory properties of 6 parameter mixtures in a defined design space. As a consequence, a 3D printing mortar with 10% ClCh score reduction was identified with significant savings in time and material resources. As this methodology is independent of both input parameters and output responses, it can be applied to multiple types of materials for countless applications. More importantly, the use of locally sourced materials is possible and recommended.

For instance, this process could be applied to accelerated concretes with different rheological objective functions, such as open time and reactivity with admixtures. In a next step and along with the 3D printing process requirements, application-related objectives will also be integrated in this optimization in order to consider mechanical and durability performance. With such hardened state considerations, the relevance of this multi-material multi-application optimization tool could be affirmed.

Acknowledgement. The authors would like to thank the Canada research chair program and the LABEX MMCD for funding this research.

References

1. Alhumayani, H., Gomaa, M., Soebarto, V., Jabi, W.: Environmental assessment of large-scale 3D printing in construction: A comparative study between cob and concrete. *J. Clean. Prod.* **270**, 122463 (2020)
2. Zhang, C., Nerella, V.N., Krishna, A., et al.: Mix design concepts for 3D printable concrete: A review. *Cement Concr. Compos.* **122**, 104155 (2021)
3. Sergis, V., Ouellet-Plamondon, C. M.: Automating mix design for 3D concrete printing using optimization methods. *Digit. Discov.* 10.1039/D2DD00040G (2022)
4. ASTM, C.: Standard test method for flow of hydraulic cement mortar, C1437 (2007)
5. ASTM, C.: 143/C143M: Standard test method for slump of hydraulic-cement concrete. Annual Book of ASTM Standards, V. 4, pp. 89–91 (2010)
6. Roussel, N., Coussot, P.: ‘Fifty-cent rheometer’ for yield stress measurements: From slump to spreading flow. *J. Rheol.* **49**(3), 705–718 (2005)
7. Roussel, N.: A thixotropy model for fresh fluid concretes: Theory, validation and applications. *Cem. Concr. Res.* **36**(10), 1797–1806 (2006)
8. Wolfs, R.J.M., Bos, F.P., Salet, T.A.M.: Early age mechanical behaviour of 3D printed concrete: Numerical modelling and experimental testing. *Cem. Concr. Res.* **106**, 103–116 (2018)
9. Deb, K., Pratap, A., Agarwal, S., et al.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**(2), 182–197 (2002)
10. Flatt, R.J., Wangler, T.: On sustainability and digital fabrication with concrete. *Cem. Concr. Res.es.* **158**, 106837 (2022)
11. Beigh, M., et al.: Structural build-up behavior of limestone calcined clay cement (LC3) pastes in the context of digital concrete construction (2019)



An Assessment of the Usage of Collaborative Robotics in the South African Construction Industry

Ayanda Boya¹(✉), Clinton Aigbavboa¹, and Opeoluwa Akinradewo²

¹ Cidb Centre for Excellence & Sustainable Human Settlement and Construction Research Centre, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

ayandaboya@gmail.com, caigbavboa@uj.ac.za

² Department of Quantity Surveying and Construction Management, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa

Abstract. In recent years, Human-Robot Collaboration (HRC) has garnered considerable attention. The emergence of Collaborative Robots (Cobots) as the second generation of robots offers new possibilities for the development of single-task construction robots that collaborate with human workers. Although Cobots are becoming more recognised as work partners, new methods will be needed to manage an African workforce that still has to build trust, only collaborates with one another, and fears losing their jobs. This study discusses the level of awareness of Cobots in the South African construction industry. Corresponding literature was reviewed, and a quantitative research methodology was adopted to extract information about the research objectives. To achieve this, a simple random sampling technique was used, and a survey questionnaire was distributed to 100 construction professionals, who were situated in the Gauteng province of South Africa. According to the rankings extracted from the calculated mean item scores and standard deviation, it was determined that the respondents are most aware of the Automated Product Assistant (APAS) which is used to replace labor-intensive or repetitive lifting operations performed by people in material handling applications. The results also showed that the respondents are least aware of the Baxter Robot which is used for handling lightweight components during material assembly. The level of awareness of these Cobots could be linked to the availability of these Cobots in the South African construction industry. Based on the findings of the study, it can be recommended that more research, exposure, and strategies need to be applied in order to increase awareness so as to improve the usage of Cobots in the industry.

Keywords: Cobots · Construction Industry · 4IR · HRC · Robotics · Technology

1 Introduction

In the 1970s, robots were primarily used in the construction sector in Japan to enhance the prefabrication of modular houses and industrialized structural components. Subsequently, plans for the use of robotics in the construction industry progressed, and the first

construction robots were utilized on building sites in the 1980s. ‘Integrated automated’ construction sites were further developed in the 1990s [1]. The latest advancements in hardware, software, and machine learning techniques have accelerated the overall development of robotics, therefore enhancing the functionality and effectiveness of construction robots [2]. Adversely, the construction environment is unpredictable and constantly changing. Due to this, robots are dependent on human intervention in order to execute a variety of tasks, particularly those that require high precision and sophisticated human assessment [3]. The introduction of Collaborative Robots (Cobots) thus signified a profound transformation in robotic implementation, although a comprehensive study of the collaborative industrial robotic use in the construction industry has not yet been thoroughly investigated [4].

In recent years, Human-Robot Collaboration (HRC) has garnered considerable attention. It is an integrative topic that focuses on how people and robots work together to achieve mutual objectives [5]. Collaborative robots (Cobots) are designed to work closely with people to safely execute various jobs which include assembling, manufacturing etc. Previously, it was sought to be too dangerous for human beings and robots to function in close contact [4] however, Cobots are becoming more common particularly in the automotive and industrial sectors [6]. There are currently over 100 Cobot manufacturers throughout the world. Investors have shown significant interest in Cobots, which are regarded as a viable business niche. Between 2017 to 2019, global sales of Cobots increased significantly, reaching 30,000 units in 2019, indicating a 36.4% annual increase [7]. Some international regions are exceeding when it comes to HRC implementation. An example is European manufacturers who have a strong presence in the global robotics market and are actively seeking to maintain this position in the emerging collaborative robotic sector. Europe has a very strong robotics research base and a major centre of expertise is the Fraunhofer Institute for Factory Operation and Automation (IFF). Here, several programs, often with industrial partners, involve the development of critical Cobot technologies, most notably the safety aspects [8]. In the German industry, reasons for choosing collaborative robot applications included the increase in operation efficiency, innovation, improvements in physical and cognitive ergonomics, new assembly processes, reduced assembly time, reduced monotony, improved quality and flexibility, and demographic change. Incorporating Cobots in their workplaces enables human workers to minimize tedious and mundane jobs, allowing them to concentrate on elevated or more creative work. Furthermore, Cobots can function as a smart tool, assisting in the locomotion of difficult loads, hence reducing ergonomically associated health problems [9]. Complementary to its unique benefits, Cobots also have some notable drawbacks [10] conducted three exploratory industrial case studies to investigate the design of human-Cobot teams and the application of Cobots in different firms for the purpose of machining and manufacturing. The tasks performed included picking and placing, welding, and assembling. Although it was not particularly difficult to install these Cobots, the case study revealed a few challenges. These included: (1) Lack of reflectiveness of Cobots; (2) Lack of feasibility to employ Cobots; and (3) Worker and Cobot collision concerns. Challenges can occur when a worker is unable to perform consistently because of disorganized and unpredictable human-robot collaborations and

studies have specifically shown that it might be challenging for humans to collaborate with robots on their mechanical aspects [11].

Despite the necessity for extensive research in the HRC field, the work of [12] shows the remarkable growth of publications between the years 1996 and 2015. While the majority of research on HRC concentrates on the technical aspects of Cobots, novel approaches are still required to integrate Cobots into human-centered workstations and evaluate the feasibility of implementing HRC in construction workspaces [13, 14]. As the construction sector becomes more intricate and encounters new obstacles, interest in robotic technology has grown [15]. Thus far, the usage of robots in building has mostly concentrated on automating specific industrial operations, with the goal of lowering operating costs by eliminating the need for a human operator or enhancing efficiency through machine control [16]. Presently, the application of construction robotics is limited, with most autonomous or semi-autonomous robots being used either in research studies [16] or in pre-fabrication phases – e.g., precast concrete components [17]. Building sites are set up in a challenging way for robot implementation since these settings are typically unstructured, crowded, congested, and are constantly changing depending on the construction activity [16, 18]. Furthermore, the area of a construction jobsite is usually restricted and varies from project to project, making the usage of robotics difficult [17]. Cobots have marked a significant shift in robotics applications, however there has been minimal utilization of collaborative robotics in the construction industry [4]. This shows that there is an overall gap in existing literature, knowledge and application of Cobots.

The overall construction process poses health risks to labourers and consists of tedious and harmful tasks [19]. Through the application of Cobots, some of these risks may be alleviated by substituting humans with robots, to perform the tasks that are deemed too risky or dangerous, and to increase productivity and efficiency in the workplace. Although advanced collaborative technology was developed primarily to address the flexibility needs of the automotive, electrical, and manufacturing sectors, it also gives prospects for robot solutions in the building industry. The application of Cobots in construction is slowly growing, therefore it is important to thoroughly understand the functions of Cobots and how they will be beneficial in the near future. Since Cobots have not been fully implemented in the South African construction industry, the purpose of the research is to evaluate the level of awareness of Cobots and to bridge the gap in existing literature. Identifying the level of awareness of Cobots can help determine whether or not the South African construction industry is ready to implement this new technology.

2 Level of Awareness of Collaborative Robotics in the South African Construction Industry

It has been determined that automation and robots are the primary and most effective ways to shift the construction industry toward the simplification and computerization of essentially all construction operations and procedures [20]. The revolutionary aspect of Industry 4.0 causes significant changes in work operations, necessitating a new strategy towards how work is conducted [21]. According to [22], the impact of Industry 4.0 on

job descriptions will be profound. As a result, technology advancements, both new and existing will demand new skill sets [23]. Collaborative robots capable of interacting with workers will make up a large portion of the workplace in the future. This suggests that even while the level of automation may vary among industries and job kinds, its effects will be experienced by everyone [24]. Research by [25] shows that South African companies are aware of the global production advancements surrounding the adoption of human-robot collaboration and its potential impact on the African workforce. For South African and African firms, human-robot collaboration is a feasible option. As South Africa is a developing country, all innovation advancements must be advantageous to promote economic growth, enhance task precision, empower labourers to operate more skilfully, and enhance human work execution [26].

South Africa's automotive industry now employs the majority of manufacturing robots in Africa [27]. In many circumstances, robot assistance in the automotive industry is beneficial, and as new functional capacities emerge, robots will increasingly be required to assist workers in the labour market. Accuracy, productivity, and safety in the automotive industry are all enhanced through using robots for assembly line tasks. As the automotive industry develops, the human-robot interaction becomes one of the most important aspects of effective implementation in the industry. Given that the industry is characterized by a variety of automobile parts and erratic market demands, the assembly line is the appropriate workspace where robotic assistance should be applied [26]. According to [28], this application and assistance will reduce anxiety in human employees brought on by ergonomic factors whilst also improving quality control across the production line. As a result, the South African Department of Trade and Sector has acknowledged the significance of human-robot interface in the automotive industry.

2.1 Theoretical Research from South African Institutions

2.1.1 Stellenbosch University

At Stellenbosch University, the Department of Mechanical and Mechatronic Engineering houses the Mechatronics, Automation and Design (MAD) Research Group. The MAD group has been researching the control and design of contemporary manufacturing systems for the past ten years. From 2017, the group's research has been centred on achieving the Industry 4.0 objective, which involves a widely interconnected world of smart applications and people, that use the internet and real-world data to gather insight and improve business operations [29].

The adoption of robots and automation in the construction industry has been gradual, and this is a challenge that many countries encounter throughout the world, including the South Africa [30]. Although human safety in robot-occupied environments has improved significantly (due to Cobot technology), there remains a need to improve human-robot collaboration. Cobots, for example, permit the "safe" interaction of a robot and a human - although this safety is essential, smart collision avoidance and subsequent modification of robot motion would provide more effective collaboration. To enhance smart human-robot collaboration, the MAD Research Group is researching the implementation of a collaborative robot digital twin [29]. The MAD Research group proposed a Digital Twin (DT) architecture for HRC. In order to predict future trends and adjust robot control

instructions to maximize factors such as safety, throughput, and the consumption of energy, the DT will be utilized to record the status of crucial elements in a collaborative work-cell [31].

2.1.2 University of the Witwatersrand

The University of the Witwatersrand's robotics research program is primarily focused on issues related to how robots learn and behave under unexpected circumstances. Their primary objective in their robotics research focuses on how robot systems learn [29]. Their work in this area has mostly included concepts from reinforcement learning with emphasis on knowledge acquisition that can transcend across different situations and tasks. Understanding how to respond in circumstances where the robot lacks access to all of the information about the environment, particularly the thoughts of a collaborating human, is a crucial component of learning [32]. The Wits research team has created methods based on partially observable Markov decision processes for this situation in order to infer a proxy for a human's thought processes in the same workspace as the robot and utilize this to enhance how the robot interacts [33].

2.1.3 Nelson Mandela Metropolitan University

The Nelson Mandela Metropolitan University has been researching the existing and future use of Cobots internationally, as well as its influence on the African workforce. Their primary objectives are to investigate the discrepancy between domestic and foreign adoption of Cobots, as well as how Cobots possibly influence the manufacturing and assembly industries in Africa, particularly on the African Workforce. According to their research, the key hurdles South African and African enterprises face when integrating Cobots at the organizational level include union resistance, technology installation costs, change management, training, and managing workers' heightened fear of job loss [25].

2.1.4 University of KwaZulu-Natal

Research at the University of KwaZulu-Natal proposed a new system that aims to replace existing safety methods and uses a new approach that includes reducing risks while additionally enabling human-robot interaction. The suggested safety system can be directly installed onto a robot and is considered to be a "third-party" system. The primary goal is to establish an adaptable virtual zone around every piece of equipment that could potentially be dangerous [34].

2.2 Types of Collaborative Robotics Implemented in the South African Construction Industry

2.2.1 SAM 100 (the Bricklaying Robot)

Bricklaying robots are single task construction robots (STCRs) that build brick walls with minimal human involvement. They almost entirely substitute the physical work required by human labourers to lay bricks. They were created in order to increase productivity whilst consuming fewer resources [35]. However, while the semi-automated mason

would be a worthy investment and therefore would be advantageous towards construction businesses that implement them, the South African construction sector is not yet ready to adopt such advances. The expense of procuring the semi-automated mason, the lack of essential skills to operate the robot, and the unpredictability of the operating costs, considering regular maintenance costs and the operator's wage, are among the challenges that construction businesses face when considering using SAM [30].

3 Research Methodology

The research approach that was conducted for this study was a quantitative data analysis. Using a deductive logic method, the data was collated using surveys/questionnaires. To compare the relationship between the variables, the explorative technique was implemented [36] states that the quantitative research approach collects numerical data that can be ranked, categorised, or assessed using measurement units [37] elaborated that quantitative technique involves conducting a systematic survey of questionnaires. This study was carried out in Gauteng, South Africa. Targeted respondents included professionals and organizations within the construction and engineering industry (Mechanical Engineers, Civil Engineers, Industrial Engineers, Construction Engineers, Construction Project Managers, Construction Managers, Quantity Surveyors, and Architects). A simple random sampling technique was used for this study. Surveys/questionnaires were randomly distributed to the above-mentioned professionals in South Africa. They represent all professionals that encounter robotics in a construction related workplace. This sampling technique was used because it will not be feasible to take samples from the whole African continent. The Cobots selected in the questionnaire included: APAS robot, KUKA robot, SAM100 robot, AUBO, Baxter robot, and ABB YuMi. A total of 100 applicable responses were obtained, and the Statistical Package for Social Science (SPSS) and Microsoft Excel were used to analyse the data. The data was analysed using two metrics: "Standard Deviation" (SD) and "Mean Item Score" (MIS). The Kruskal-Wallis test was used to assess the variances among the sampled groups, and the reliability of the results was assessed using the Cronbach's Alpha coefficient, the value of 0.948 indicated that the study's conclusions could be relied on. The limitations encountered in this study were due to many professionals not completing the survey which could have improved the results and findings.

4 Findings and Discussion

4.1 Findings

According to the rankings shown on Table 1, extracted from the calculated mean item scores, standard deviation, and P-values of the level of awareness of collaborative robotics in the South African construction industry, it was determined that out of all the Cobots mentioned, the respondents were most aware of the Automated Product Assistant (APAS), which was ranked first with a mean (MIS) of 2.27 and a standard deviation (SD) of 1.162. The Semi-Automated Mason (SAM100) was ranked second, with a mean (MIS) of 2.18 and a standard deviation (SD) of 1.067. The KUKA LBR

iiwa Robot was ranked third with a mean (MIS) of 2.12 and a standard deviation (SD) of 1.076. The ABB YuMi was ranked fourth with a mean (MIS) of 2.09 and a standard deviation (SD) of 1.006. The AUBO Robot was ranked fifth with a mean (MIS) of 2.08 and a standard deviation (SD) of 0.981. The Baxter Robot was ranked sixth (last) with a mean (MIS) of 2.02 and a standard deviation (SD) of 0.974. Considering all the Cobots and responses, approximately 62.67% were either Not Aware or Fully Not Aware of any Cobot. Approximately 14.17% respondents were either Aware or Fully Aware of all the Cobots. The Kruskal-Wallis H-Test was applied in this study to determine the significance of variations in the perspectives of the various construction professionals on the level of awareness of the collaborative robots that were identified. The table shows that every Cobot's data set has a p-value greater than 0.05. The result that can be drawn from this finding is that, with a 95% degree of confidence, there is a statistically significant correlation between the professionals' awareness of collaborative robotics in the South African construction industry.

Table 1. Level of awareness of Cobots in South Africa

| | | | | | Kruskal-Wallis |
|--|------|----------------|---------|--------|----------------|
| | Mean | Std. Deviation | Ranking | Chi-Sq | Sig |
| Level of awareness of Automated Product Assistant (APAS) | 2.27 | 1.162 | 1 | 9.128 | 0.332 |
| Level of awareness of Semi-Automated Mason (SAM100) | 2.18 | 1.067 | 2 | 6.338 | 0.609 |
| Level of awareness of KUKA LBR iiwa Robot | 2.12 | 1.076 | 3 | 7.510 | 0.483 |
| Level of awareness of ABB YuMi | 2.09 | 1.006 | 4 | 7.197 | 0.516 |
| Level of awareness of AUBO Robot | 2.08 | 0.981 | 5 | 5.582 | 0.694 |
| Level of awareness of Baxter Robot | 2.02 | 0.974 | 6 | 8.113 | 0.422 |

4.2 Discussion

According to the rankings extracted from the calculated mean item scores and standard deviation of the level of awareness of collaborative robotics in the South African construction industry, approximately 62.67% respondents were either Not Aware or Fully Not Aware of any of the aforementioned Cobots, and only 14.17% respondents were either Aware or Fully Aware of all the Cobots. This asserts literature by [30] that the adoption of robots and automation in the South African construction industry has been

gradual. The high percentage in respondents lacking awareness in the Cobots could be as a result of ongoing research that is still taking place across South African institutions, and not enough publications on the different types of Cobots, that could be implemented in the industry. The Automated Product Assistant (APAS) was ranked highest as the Cobot which had the highest level of awareness. This implies that there is some knowledge of the Cobot although there has not been practical application in the construction field. The development of these new, generally smart Cobots presents significant safety, legal and ethical concerns. Uncertainty about technology can cause opposition to new technology adoption, which can impede an organization's or company's development towards advanced and efficient practices [38]. If workers lack confidence in or are sceptical about new technology, it will be challenging to get them to embrace or utilize them. Humans are less inclined to work hand in hand with or alongside robots when they fear it will be hazardous, despite the actual degree of safety [39, 40] suggested that safety related concerns such as errors, reliability, and faulty behavior affect trust and acceptance in HRC work environments. Previous study also found that employees frequently associated the adoption of robotic technology with the fear of being replaced by robots [41], which is a typical impediment in the adoption of and implementation of robotic technology since the first industrial revolution [42]. The likelihood of being replaced by a Cobot undermines the workers' sense of self-worth and elevates concern about their long-term economic circumstances [43]. This underlines the psychological importance and rationalizes the possible detrimental influence on employee behaviour. The principles of robot design and operation, as well as accountability for the consequences of Cobot actions, must be determined. The underlying principles of civil law will have to be radically altered. It is entirely plausible that a new category of law will be required to address "autonomous robots" or merely "robots". Currently, preliminary work is being done to establish legal standards particularly for the devices which will control robot behaviour and allow for the evaluation of circumstances caused by them or in conjunction with their activities [44].

4.3 Implication of Findings

The literature review and data analysis demonstrate that the level of awareness of Cobots in the South African construction industry is still minimal. The data analysis revealed that the construction industry professionals are more aware of the Automated Product Assistant (APAS), the Semi-Automated Mason (SAM100), and the KUKA LBR iiwa Robot. It also revealed that the professionals were least aware of the ABB YuMi, the AUBO Robot, and the Baxter Robot. Although professionals are more aware of the APAS robot, they seem to be unaware of its purpose and application. The literature review showed that, apart from the SAM100 robot, there is a lack of knowledge, information, and practical references on the application of Cobots in the South African construction industry.

5 Conclusion and Further Research

The research objective of this study was to assess the awareness level of collaborative robotics in the South African construction industry which was defined and achieved by the study. The demography of the survey questionnaire was limited to professionals working in the construction industry in South Africa. As a result, the study's findings supported methodological techniques identified in the level of awareness of Cobots in the construction industry. According to the literature review, there is not much HRC implementation in the construction industry globally. However, a number of institutions in South Africa are currently investigating HRC and how to incorporate them in different construction settings and assembly lines. The SAM 100 Cobot (Brick-Laying Robot) is the only evident Cobot that has been introduced and used in the construction industry, although it has not been fully implemented by construction companies in South Africa. Results generated from the questionnaire based in Gauteng, South Africa demonstrated that the respondents were most aware of the Automated Product Assistant (APAS), The Semi-Automated Mason (SAM100), and the KUKA LBR iiwa Robot. This indicates that there is a low level of awareness of Cobots in the South African construction industry. This low level of awareness can be attributed to the unavailability of policies and regulations guiding the adoption of Cobots in the construction industry despite the long history of the incorporation of robots to the construction industry processes. Based on the findings of the study and the identified limitations, the following recommendations are made: organisations should focus on engaging employees in trainings that will help develop skills needed to properly operate, interact with, and monitor robotic systems, encourage employees to give early input and test prototypes in order to facilitate iterative changes and promote technology adoption, consideration for safety first in the development and application of Cobot systems and guaranteeing that these innovations are meant to improve worker safety rather than compromise them, and allaying concerns about job displacement by emphasising the possibility of shifting to more valuable responsibilities, improving productivity, and enhancing safety conditions at work. For future research, a study on the virtual replications of construction projects through digital twins can be explored, so as to determine suitable environments for Cobot applications and how they can be implemented. As a result, project planning, monitoring, and decision-making are enhanced through improved visualisation, simulation, and analysis. Additionally, building trust in robotic technologies among construction workers requires a multi-faceted approach. Future research projects can examine other socio-technical aspects that impacts the adoption of emerging technologies in the construction sector, such as levels of competitive advantage, reluctance towards change, and support from top management.

References

1. Bock, T.: Construction robotics. *Auton. Robot.* **22**(3), 201–209 (2007)
2. Balaguer, C., Abderrahim, M.: Trends in robotics and automation in construction. Chapter 1. In: *Robotics and Automation in Construction*. IntechOpen, London, pp. 1–20 (2008)
3. Liu, Y., Habibnezhad, M., Jebelli, H.: Brainwave-driven human-robot collaboration in construction. *Autom. Constr.* **124**, 103556 (2021)

4. Bogue, R.: Robots that interact with humans: a review of safety technologies and standards. *Ind. Robot Int. J.* [Post-print] **44**(4), 395–400 (2017). Available from <https://doi.org/10.1108/ir-04-2017-0070>
5. Villani, V., Pini, F., Leali, F., Secchi, C.: Survey on human–robot collaboration in industrial settings: safety, intuitive interfaces and applications. *Mechatronics* **55**, 248–266 (2018)
6. Muller, R., Vette, M., Scholer, M.: Inspector Robot—A new collaborative testing system designed for the automotive final assembly line. *CIRP Conf. Assem. Technol. Syst.* **23**, 59–64 (2014)
7. Liu, L., Guo, F., Zou, Z., Duffy, V.G.: Application, development and future opportunities of collaborative robots (Cobots) in manufacturing: a literature review. *Int. J. Hum. Comput. Interact.* **40**(4), 915–932 (2024)
8. Bogue, R.: Europe continues to lead the way in the collaborative robot business. *Ind. Robot Int. J.* **43**(1), 6–11 (2016)
9. Bauer, W., Bender, M., Braun, M., Rally, P.E.T.E.R., Scholtz, O.: Lightweight Robots in Manual Assembly—Best to Start Simply, p. 1. Frauenhofer-Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart (2016)
10. Kadir, B.A., Broberg, O., Souza da Conceição, C.: Designing human-robot collaborations in industry 4.0: explorative case studies. In: DS 92: Proceedings of the DESIGN 2018 15th International Design Conference, pp. 601–610 (2018)
11. Dragan, A.D., Bauman, S., Forlizzi, J., Srinivasa, S.S.: Effects of robot motion on human–robot collaboration. In: 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 51–58. IEEE (2015)
12. Ajoudani, A., Zanchettin, A.M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., Khatib, O.: Progress and prospects of the human–robot collaboration. *Auton. Robot.* **42**, 957–975 (2018)
13. Meyer, R.D., Schraft, C.: The Need for an Intuitive Teaching Method for Small and Medium Enterprises. ISR Robotik, Germany (2012)
14. Mourtzis, D., Vlachou, E., Dimitrakopoulos, G., Zogopoulos, V.: Cyber-physical systems and education 4.0—the teaching factory 4.0 concept. *Procedia Manufact.* **23**, 129–134 (2018)
15. Son, H., Kim, C., Kim, H., Han, S., Kim, M.: Trend analysis of research and development on automation and robotics technology in the construction industry. *KSCE J. Civ. Eng.* **14**(2), 131–139 (2010)
16. Saidi, K., Bock, T., Georgoulas, C.: Robotics in construction. In: *Handbook of Robotics*. Springer-Verlag, Berlin Heidelberg, pp. 1493–1520 (2016)
17. Bruckmann, M., Spengler, R., Malkwitz, K., Bruckmann, T.: Automated construction of masonry buildings using cable-driven parallel robots. In: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, 33, 1 (2016)
18. Feng, C., Xiao, Y., Willette, A., Mcgee, W., Kamat, V.: Vision guided autonomous robotic assembly and asbuilt scanning on unstructured construction sites. *Autom. Constr.* **59**(1), 128–138 (2015)
19. Kim, W., Lorenzini, M., Balatti, P., Nguyen, P., Pattacini, U., Tikhanoff, V., Peternel, L., Fantacci, C., Natale, L., Metta, G., Ajoudani, A.: A reconfigurable and adaptive human–robot collaboration framework for improving worker ergonomics and productivity. *IEEE Robotics Autom. Mag.* (2019)
20. Oke, A., Akinradewo, O., Aigbavboa, C., Akinradewo, O.: November. Benefits of construction automation and robotics in the South African construction industry. In: IOP Conference Series: Earth and Environmental Science, vol. 385, no. 1, p. 012063. IOP Publishing (2019)
21. Prifti, L., Knigge, M., Kienegger, H., Krcmar, H.: A competency model for “Industrie 4.0” employees (2017)
22. Sackey, S.M., Bester, A.: Industrial engineering curriculum in Industry 4.0 in a South African context. *South Afr. J. Ind. Eng.* **27**(4), 101–114 (2016)

23. Maisiri, W., Darwish, H., Van Dyk, L.: An investigation of industry 4.0 skills requirements. *South Afr. J. Ind. Eng.* **30**(3), 90–105 (2019)
24. Chui, M., Manyika, J., Miremadi, M.: Where machines could replace humans—and where they can't (yet), pp. 1–16. McKinsey & Company, Chicago (2016)
25. Calitz, A.P., Poisat, P., Cullen, M.: The future African workplace: the use of collaborative robots in manufacturing. *SA J. Hum. Resour. Manag.* **15**(1), 1–11 (2017)
26. Chigbu, B.I., Nekhwevha, F.H.: The collaborative work experience of robotics and human workers in the automobile industry in South Africa. *Afr. J. Sci. Technol. Innov. Dev.* **14**(1), 280–287 (2022)
27. Hendrikse, H.: Industrial robotics global market trends and its effect on South Africa. In: Siemens Future of Manufacturing Conference (2014)
28. Gleeson, B., Maclean, K., Haddadi, A., Croft, E., Alcazar, J.: Gestures for industry: intuitive human-robot communication from human observation. In: ACM/IEEE International Conference on Human-Robot Interaction, pp. 349–356 (2013). <https://doi.org/10.1109/HRI.2013.6483609>
29. Boje, E., Christopher, R.L., Fernandes, J., Hepworth, J.H., Kuriakose, R.B., Kruger, K., Lorimer, T., Luwes, N., Mouton, H.D., Patel, A. and Rosmang, B.: A review of robotics research in South Africa (2019)
30. Harinarain, N., Caluza, S., Dondolo, S.: Bricklaying robots in the South African construction industry: *The Contractors Perspective*. In: Scott, L., Neilson, C.J. (eds.) Proceedings of the 37th Annual ARCOM Conference, 6–7 September 2021, UK, Association of Researchers in Construction Management, pp. 36–45 (2021)
31. Joseph, A.J., Kruger, K., Basson, A.H.: An aggregated digital twin solution for human-robot collaboration in industry 4.0 environments. In: Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future: Proceedings of SOHOMA 2020 (pp. 135–147). Springer International Publishing (2021)
32. Görür, O.C., Rosman, B., Sivrikaya, F., Albayrak, S.: February. Social Cobots: Anticipatory decision-making for collaborative robots incorporating unexpected human behaviors. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 398–406 (2018)
33. Görür, O.C., Rosman, B., Albayrak, S.: May. Anticipatory Bayesian policy selection for online adaptation of collaborative robots to unknown human types. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, pp. 77–85 (2019)
34. Hoskins, G.O., Padayachee, J., Bright, G.: January. Human-robot interaction: the safety challenge (an integrated frame work for human safety). In: 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), pp. 74–79. IEEE (2019)
35. Bock, T., Linner, T.: Site Automation. Cambridge University Press (2016)
36. Creswell, J.W.: Research Design_ Qualitative, Quantitative, and Mixed Method Approaches, 4th edn. SAGE Publications Inc., Thousand Oaks, CA (2014)
37. Tan, W.C.K.: Practical Research Methods. Pearson Custom (2011)
38. Pan, M., Pan, W.: Understanding the determinants of construction robot adoption: perspective of building contractors. *J. Constr. Eng. Manag.* **146**(5), 04020040 (2020)
39. Atkinson, D.J., Clark, M.H.: Methodology for study of human-robot social interaction in dangerous situations. In: Proceedings of the Second International Conference on Human-Agent Interaction (pp. 371–376) (2014)
40. Weiss, A., Wortmeier, A.K., Kubicek, B.: Cobots in industry 4.0: a roadmap for future practice studies on human–robot collaboration. *IEEE Trans. Hum. Mach. Syst.* **51**(4), 335–345 (2021)

41. Richert, A., Müller, S., Schröder, S., Jeschke, S.: Anthropomorphism in social robotics: empirical results on human–robot interaction in hybrid production workplaces. *AI & Soc.* **33**(3), 413–424 (2018)
42. Maurtua, I., Ibarguren, A., Kildal, J., Susperregi, L., Sierra, B.: Human–robot collaboration in industrial applications: safety, interaction and trust. *Int. J. Adv. Rob. Syst.* **14**(4), 1729881417716010 (2017)
43. Granulo, A., Fuchs, C., Puntoni, S.: Psychological reactions to human versus robotic job replacement. *Nat. Hum. Behav.* **3**(10), 1062–1069 (2019)
44. Zbigniew, P., Klimasara, W., Pachuta, M., Słowikowski, M.: Some new robotization problems related to the introduction of collaborative robots into industrial practice. *J. Autom. Mob. Robot. Intell. Syst.* 91–97 (2019)

Analysis, Simulation and Sensing



Development of an Electronic Guidance Device “eGuiDev” Based on Wireless Communication Technology for Pedestrian Navigation

Abduaziz Juraboev^{1,2(✉)}, Abdulrahman Othman¹, Rüdiger Kern¹, Joaquín Díaz¹, and Uwe Rüppel²

¹ THM, University of Applied Sciences Mittelhessen, 35390 Giessen, Hesse, Germany
abduaziz.juraboev@bau.thm.de

² TUD, Technical University of Darmstadt, 64287 Hesse, Germany

Abstract. With the increasing complexity and digitalization of the built environment, the need for barrier-free accessibility of buildings in relation to assistive devices is one of the remaining challenges towards navigation. Orientation in complex buildings currently relies on analogue information such as signs and emergency plans. There is a lack of centimetre accurate visual and audible guidance systems. In this study, a centimetre-accurate wireless electronic Guidance Device “eGuiDev” based on Ultra High Frequency Radio Frequency Identification (UHF RFID) technology is developed to assist visually impaired and mobility-limited people with pedestrian navigation. We defined the requirements for UHF technology based on the compactness of smaller planar antennas with a high bandwidth (892–940) MHz. Thus, we investigated an electronic reader consisting of UHF RFID components such as antenna, reader module, microcontroller and, for wireless communication, Bluetooth module, integrated into a guidance device. These electronic components can be embedded in a 3D printed, human-operated white cane with a handheld device that guides the user to the desired RFID tagged targets. Passive tags are integrated into the flooring. However, the focus of this paper is on the development of the multifunctional device, the components and the implemented algorithm. A literature review presents the basis for identifying suitable components for the implementation of a passive RFID system. Serial interfaces have been used to enable communication between the microcontroller, the RFID reader and other smart devices via the Bluetooth Low Energy standard. The “eGuiDev” features low power consumption, a 2600 mAh lithium battery and USB Type-C for recharging the battery. The reader emits a vibration and audio signal when it reads passive transponder information, and communication with mobile phones can provide further information that are suitable for a navigation. This study shows that passive UHF RFID components enable navigation with centimetre accuracy.

Keywords: Accessibility in the Built Environment · Navigation · Auto-ID Technology

1 Introduction

1.1 Background

Accessibility and usability of buildings are essential to ensure equal participation in social and professional life at all stages of life. People spend about 90% of their time in buildings and urban spaces, such as infrastructure, roads, bridges, parks and other built environments [1]. It is therefore essential to adopt a sustainable approach that takes into account efficient use potential, health aspects and accessibility. This requires buildings and the built environment to meet higher functional requirements, which can be achieved through digitalisation, modernisation and the use of new technologies and methods.

Buildings are becoming smarter with the integration of sensors such as wireless communication technologies and automatic identification and data collection systems such as Radio Frequency Identification (RFID), Ultra-Wideband (UWB), QR-Code, Bluetooth and more [2]. These sensors and Internet of Things are linked to software applications that link the physical World to a Digital Twin in real time, providing dynamic and specific information. One of the key foundations for the digital transformation of the built environment is the holistic and process-oriented approach of Building Information Modelling (BIM) [3], based on a collaborative partnership between project participants (Fig. 1).

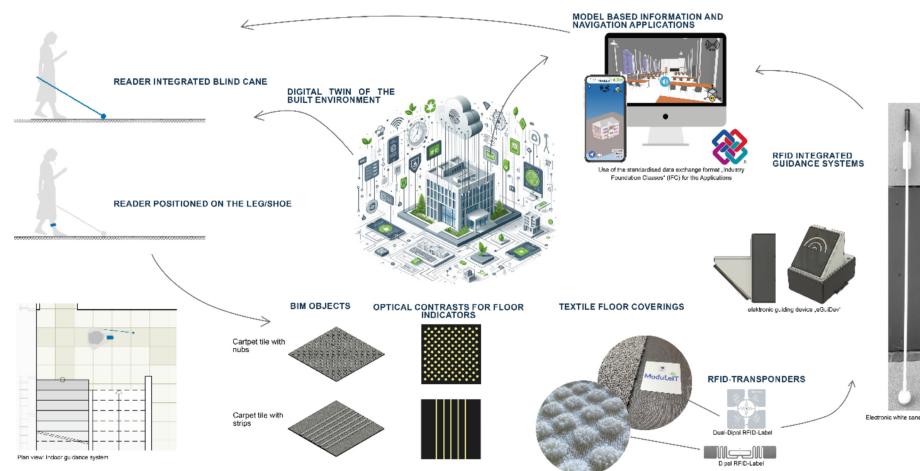


Fig. 1. Overview of RFID and BIM-based information and guidance systems for the built environment

The user-centred design and implementation of an accessible building for people with visual, hearing, cognitive or motor impairments is therefore one of the most important functional aspects of sustainability. Public buildings and workplaces need to be accessible without particular difficulty. The World Health Organization reports that over 2.2 billion people worldwide have a visual impairment, either near or distant [4]. To ensure their inclusion in society, it's crucial for these individuals to be able to move

around independently. One of the most valuable aids for this is a cane, which can be used to navigate tactile guidance systems like floor strips.

1.2 Objective

The primary objective of this research paper is the comprehensive investigation and development of an electronic guidance device, designed as “eGuiDev”. The device uses advanced wireless communication technology to create an auditory built environment that enhances spatial awareness and independent mobility. The research focuses on transforming accessible visual barrier-free contrasts into auditory information. This transformation is critical in facilitating enhanced navigation and significantly improving the spatial awareness of visually impaired users.

An essential component of this study is the detailed analysis of various hardware elements that are necessary for the construction of an electronic pedestrian navigation device. This analysis includes both the research and development of an electronic reader, which is essential to the functionality of the “eGuiDev”.

Ultimately, this project aims to contribute substantially to the broader understanding of technological advancements in the realms of the Internet of Things (IoT) and wireless technologies. It particularly focuses on their application in the improvement of pedestrian navigation systems, thereby addressing a key need within the visually impaired people.

1.3 Research Questions

To achieve this objective, three research questions are going to be investigated in this paper:

1. How can a heterogeneous system be achieved, and what role does interoperability play in an IoT system considering the interaction of numerous independent components?
2. How can wireless technologies be used to develop an electronic guidance device for pedestrian navigation?
3. What are the essential components required to implement an electronic guidance device?

To achieve this goal, we analysed various hardware components and conducted research to create an electronic reader.

2 Related Work

2.1 Integration of RFID into Materials

Research into the use of wireless communication technologies like RFID and Sensors has been going on for decades. The integration of sensors into materials has made it easier to connect physical and virtual things and environments [5]. By integrating RFID transponders into materials, different user groups can access relevant information throughout the life cycle of the materials. RFID technology plays a critical role in this integration by enabling the unique identification of physical entities within a building. By

associating additional data with these entities, users can access the information by scanning RFID-tagged building components. In addition, the information can be delivered to users through applications in both visual and auditory formats, increasing transparency and accessibility. Several academic studies have focused on this approach, particularly in the development of RFID and BIM-based electronic signage systems tailored for the visually impaired.

2.2 Use Case Digital Accessibility in the Built Environment

The field of guidance systems for individuals with visual impairments is rapidly growing area of research around the world. This encompasses a variety of approaches, including tactile guidance systems and digital navigation solutions. An important aspect of this research is the use of tactile floor indicators [6], which are standardised in public spaces according to ISO 23599:2019 – “Assistive products for blind and vision-impaired persons, Tactile walking surface indicators” [7] and DIN 32984:2023 – “Ground indicators in public spaces” [8]. While these tactile systems are well established, there is a lack of formal standards or guidelines for digital guidance systems for the visually impaired.

Despite this gap in standardisation, digital wayfinding systems have been successfully implemented in various public and indoor environments, demonstrating their practicality and effectiveness. This disparity between the well-regulated tactile systems and the emerging yet unstandardized digital solutions presents an opportunity for further exploration and development in this field.

3 Materials and Methodology

3.1 Overview

In order to find the suitable Radio-Frequency Identification (RFID) reader for mobility assistance, a thorough evaluation of different reader-antenna-transponder combinations is needed. This is because it's difficult to predict which specific reader-antenna configuration will perform better with a particular transponder.

At the beginning of this research, we focused on RFID systems that operate in the Ultra High Frequency (UHF) range, which comply with ISO 18000-6c standards [9]. These systems have extended reading ranges that are particularly useful for locating people and objects.

Next, we conducted a detailed assessment of commercially available UHF antennas and transponders to determine their suitability and performance. When selecting an appropriate reader antenna and microcontroller, we considered critical factors such as effective reading range, as well as the physical attributes of the antenna, such as size, weight, and durability. These factors are essential to ensure practical usability in mobility devices.

As a result of this research, we have developed a prototype that can be easily integrated into assistive devices, such as a foot cuff or an electronic cane, designed for visually impaired individuals.

3.2 Components

After conducting thorough research, we selected the following components for UHF RFID readers: antennas, microcontrollers, charging and vibration modules, etc. (Tables 1 and 2).

UHF RFID Reader/Antenna Module

The DwarfG2 has been specially selected for use with the ARRRN5 antenna, which has a long reading range and a robust design. The DwarfG2 is compact and suitable for mounting on a cane. The maximum transmission power of the DwarfG2 is 50 mW, which was considered sufficient and reliable for the required applications. Because DwarfG2 reader is not an embedded system, we need an extra antenna to scan the tags [10].

The ARRRN5 antenna has a size of 4×4 cm and is known for its robust construction and small cross section. These features make it a suitable choice for integration into a cane for the blind, as it offers a good read/transmit range without being too large or heavy. The ARRRN5 antenna can be selected for the project due to its performance, especially its read/transmit range [11].

The module M5Stack is characterised by low power consumption and high performance, with a remarkable transmission power of 100 mW, enabling it to effectively read UHF RFID tags over a distance of more than 1.5 m. It features serial interface communication and supports an encapsulated “Attention (AT)” command set for easy integration and user-friendly operation [12].

The choice of M5Stack for the development of this device was due to the fact that the device is embedded, has light weight, cost effective and has a longer range.

Table 1. RFID readers and antenna

| Item | RFID Reader | Operating Principle | Resonance Frequency [MHz] | Supported Protocol | Read Range [mm] | Range after tests [mm] | Costs [€] | Source |
|------|------------------|--|---------------------------|------------------------------------|-----------------|------------------------|---------------|--------|
| 1 | MetraTec DwarfG2 | only reader | 868–928 | EPC Class 1 Gen 2 v2, ISO 18000-6C | up to 1200 | 300 | approx. 400 | [10] |
| 2 | ARRRN5 antenna | antenna | 868–915 | ISO 18000-6C | | 300 | approx. 14,50 | [11] |
| 3 | M5Stack JRD-4035 | embedded RFID ceramic reader & antenna | 868–928 | EPC Class 1 Gen 2, ISO 18000-6C | up to 1500 | 800 | 73,47 | [12] |

Raspberry Pi Pico W

The Raspberry Pi Pico W is a powerful and affordable microcontroller board that offers a wide range of features, making it ideal for a variety of projects. It is based on

the Raspberry Pi RP2040 microcontroller chip, which has dual-core Arm Cortex M0+ processors, 264 KB of SRAM, and 2 MB of flash memory. It also includes several built-in peripherals, such as SPI, I²C, UART, PWM, and a temperature sensor [13].

The following Table 1 gives a summary of all important used modules.

Table 2. Microcontroller and other modules

| Item | Moduls | Clock Rate [MHz] | Serial Flash [KB] | RAM [KB] | Bluetooth | Serial Interfaces | Source |
|------|---------------------|------------------|-------------------|----------|-----------|-------------------|--------|
| 1 | Raspberry Pi Pico W | 133 MHz | 2048 | 256 | Yes | 2 UART | [13] |
| 2 | STM Nuclei 32 | 80 MHz | 256 | 69 | No | 1 UART | [14] |
| 3 | Arduino Nano Every | 20 | 48 | 6 | No | 1 UART | [15] |

Lithium Battery Charging Module

The TP4056 module, with its USB Type-C compatibility, offers efficient and convenient battery charging, ensuring faster charging times and universal compatibility across a variety of electronic devices. Its compact design and safety features make it a reliable choice for portable applications [16].

DC-to-DC converter

The batteries typically supply 3.7 V. This voltage can vary continuously. Therefore, a module is used to constantly provide the components with a stable 5 V, offering protective mechanisms [17].

3.3 “eGuide” Architecture and Flow Diagram

In the early stages of system design, a power source capable of delivering a voltage of 3.7 V is essential. To meet this requirement, a dedicated module is used for the dual purpose of charging the battery and facilitating voltage transfer. At the same time, an additional module is integrated into the system architecture. The main function of this module is to regulate the voltage, thus ensuring its compatibility with the microcontroller and the reader.

The system architecture (Fig. 2) also includes the integration of several peripheral modules that interface with the microcontroller. These include an acoustic signaling device (buzzer), a tactile feedback mechanism (vibration module) and the reader itself. This configuration is critical to the efficient operation and functionality of the overall system.

Upon powering on the device, the initialization of modules and declaration of variables take place, setting the default state to ‘Instant By’ mode within a continuous loop. Users have the capability to alternate between different operational modes, including ‘Standby’ and ‘Running’, by pressing the designated ‘Switch Button’. When activated

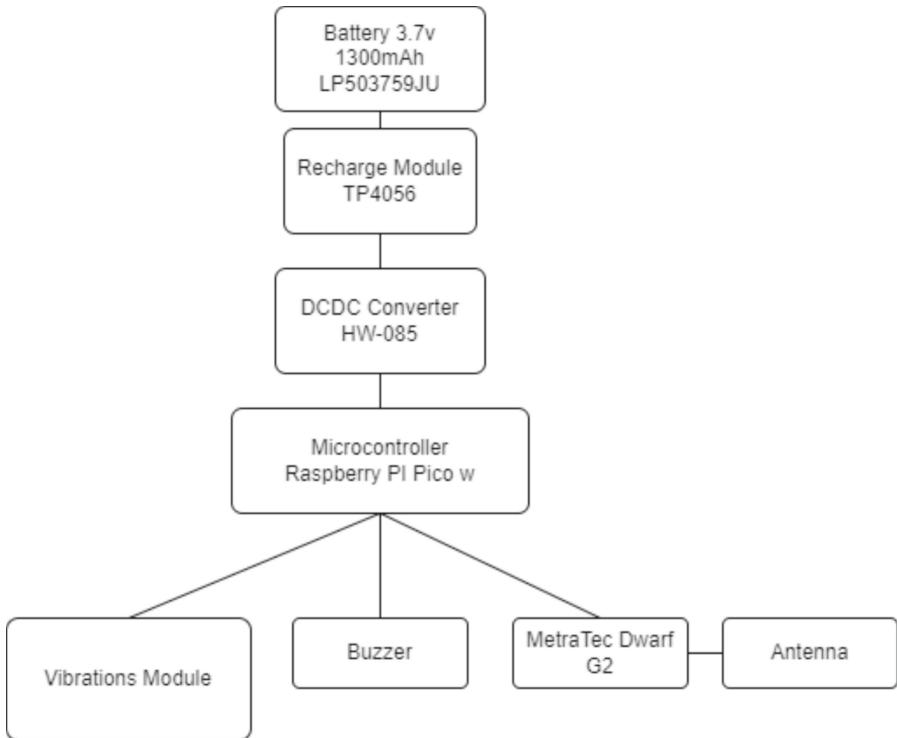


Fig. 2. Architecture of “eGuiDev”

and set to ‘Running’ mode, the device initiates object initialization and establishes a Bluetooth Low Energy (BLE) connection with a paired smartphone (see Fig. 3).

In this operational state, the device is designed to detect and read UHF RFID tags. Upon encountering a designated tag, it captures the Electronic Product Code (EPC) along with other pertinent data. This information is then transmitted wirelessly to the smartphone application via the established BLE connection. To enhance user interaction, the device incorporates an integrated buzzer and vibration feature. The buzzer serves to audibly notify the user of successful tag reading, thereby obviating the need for a visual display. Similarly, the vibration feature provides tactile feedback, further augmenting the user experience.

This multifunctional approach, combining tactile and auditory signals with advanced RFID and Bluetooth technologies, ensures efficient and user-friendly operation of the device across various modes.

3.4 Implementation

MicroPython Framework

The hardware was mainly implemented using MicroPython, a streamlined and efficient version of the Python 3 programming language. It has a limited set of the Python

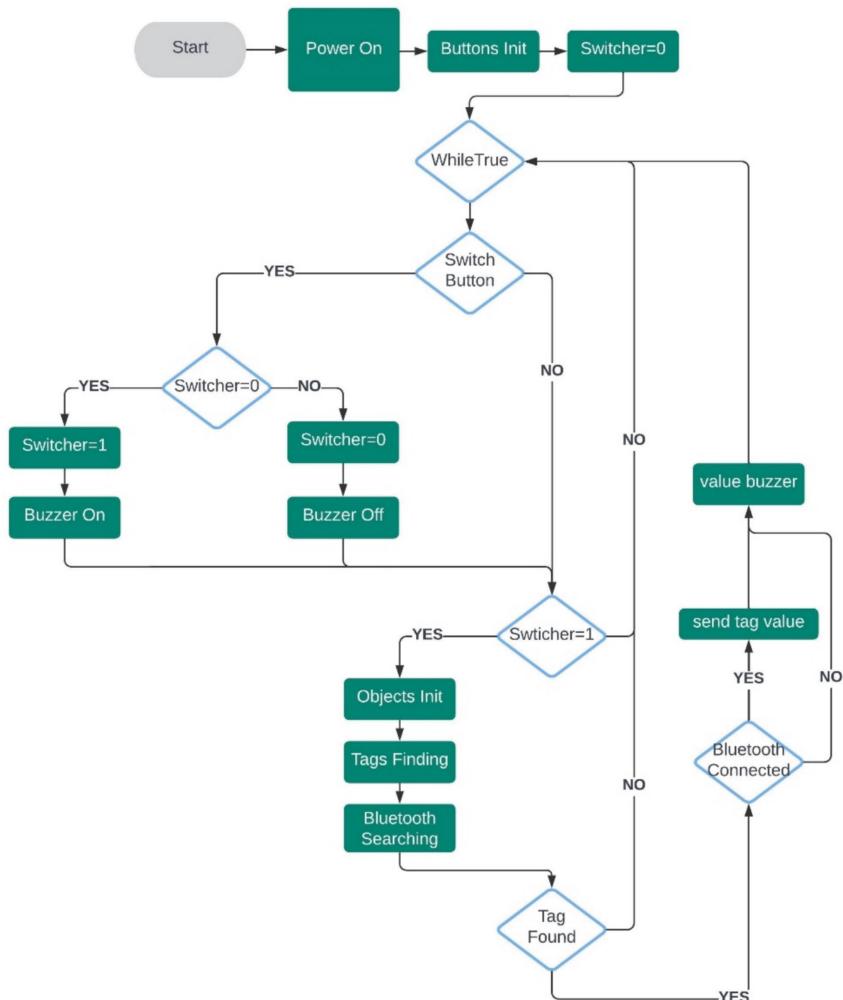


Fig. 3. Flow Diagram of the system

standard library and is specifically designed to work effectively on microcontrollers and in environments with limited resources. The MicroPython pyboard is a small electronic circuit board that runs MicroPython directly, providing a Python operating system at a low level. This makes it suitable for managing various electronic projects.

Unit Testing

Unit testing is a crucial aspect of MicroPython development, ensuring the robustness and reliability of the codebase. By employing unit testing methodologies, developers have rigorously validated individual units or components of the MicroPython code, systematically assessing their functionality in isolation. This approach allows for the identification of potential issues early in the development process, promoting a more efficient and error-free implementation. The utilization of unit testing in MicroPython

development exemplifies a commitment to producing high-quality, dependable software for microcontrollers and constrained environments. Figure 4 shows the prototype device with integrated antenna and reader.

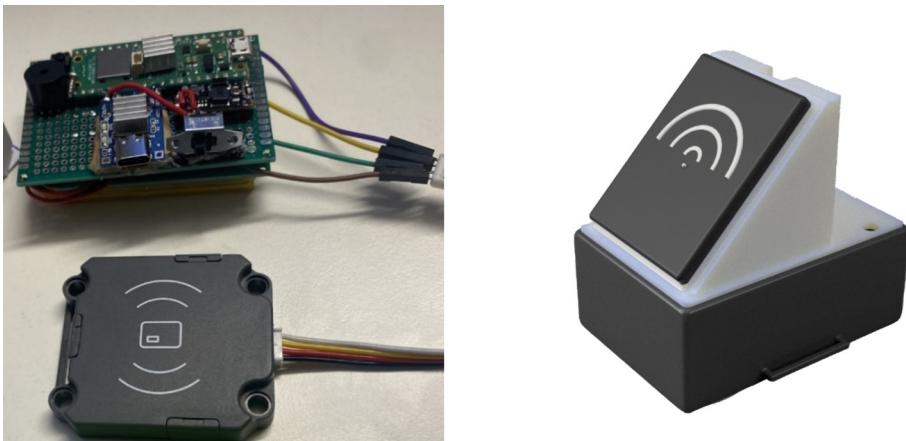


Fig. 4. The developed and 3D-printed RFID reader “electronic Guidance Device - eGuiDev”

Next, the embedded RFID system and the electronic components were integrated into the 3D printed case. The further experiments were carried out with this physical reader.

4 Experimental Results

Laboratory experiments were conducted to determine the optimal transponders for the “eGuiDev” reader, focusing on their interaction with different materials, particularly carpet tiles (Fig. 5). The experiments involved assessing the effect of directional RFID radiation on transponder performance. This was achieved by placing the transponder within the inlay of a carpet and taking measurements from four different orientations relative to the transponder.

The methodology included two primary measurement orientations: a perpendicular, vertical approach at 90° to the transponder, and an oblique approach at a 45° angle. The results were graphically represented to illustrate the variance in reading ranges contingent upon the transponder attachment type and the angle of air measurement.

In the graphical evaluation of the measured values as shown in Fig. 6, the reading ranges for different types of transponder attachment for air measurement at a 90° angle and at a 45° angle were shown. The measurement results show that the read ranges vary depending on the measurement type and transponder type. For example, the read range for the UHF RFID label “Thunder Trace 105 × 25 mm, PET with foam in the background, white permanent, antenna shape dipole” was 650 mm for the air measurement, while it was 550 mm for the label “TH24 Hamtrace Monza 4D/E/QT, 100 × 100 mm, PET, white permanent without foam in the background, antenna shape dual dipole”. The range of the

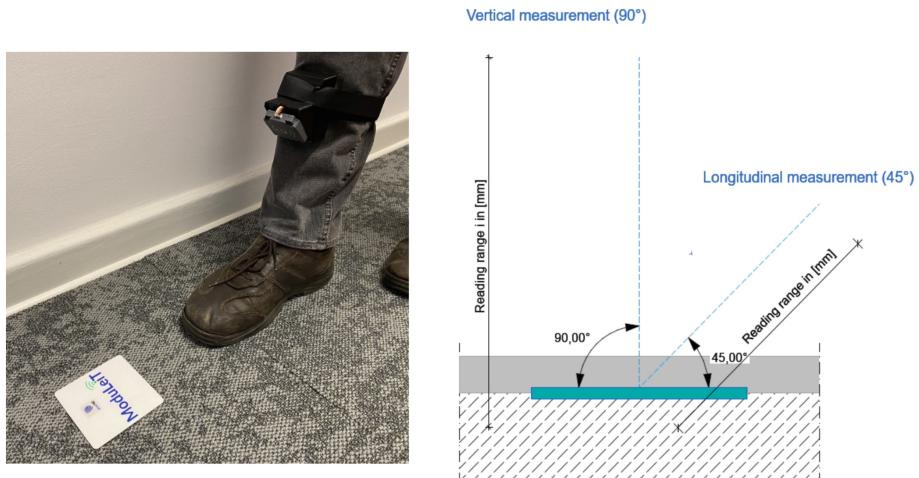


Fig. 5. Integration of RFID labels under carpets and techniques for measuring the reading range

“Thunder Trace” label deteriorated when the foam underneath the label was removed. The tests have shown that a thin layer of foam under the transponder can lead to an increase in reading range.

These individual results are important for developing a future installation pattern based on the read range in floor coverings, e.g. how many RFID labels are installed per square metre of carpet tile in order to develop a higher positioning accuracy for navigation.

Based on the static measurement results, a test was carried out at a normal walking speed (1 m/s) on a walkway with an integrated RFID carpet tile of 500 × 500 mm. The RFID reader was attached to the leg and the tags were integrated into the carpet tile with 1 to a maximum of 4 inlay tags. As the number of tags increases, so does the probability of the tags being read. The reader scanned one inlay tag per metre. With a step length of 70 cm, this equates to reading 3 tags in 4 steps.

5 Conclusion and Future Work

In conclusion, the study highlights that environmental conditions (e.g., the placement of the transponder in or above the material) and material properties (e.g., carpet over screed, carpet over wooden floors) substantially influence the reading range of the “eGUiDev”-electronic reader. The influence of foam on the reading range of RFID transponders varies depending on its placement and thickness, particularly as the transponder does not rest directly on the concrete screed. Transponders equipped with dual dipole antennas were identified as particularly effective, offering extensive read ranges from all directions and demonstrating efficient antenna design. “eGUiDev” will be initially inserted into a foot manchette and tested in a real environment.

The findings from this research provide a foundational framework for subsequent optimizations and modifications aimed at enhancing accessibility and mobility for visually impaired individuals in urban environments. This study’s insights into the effective

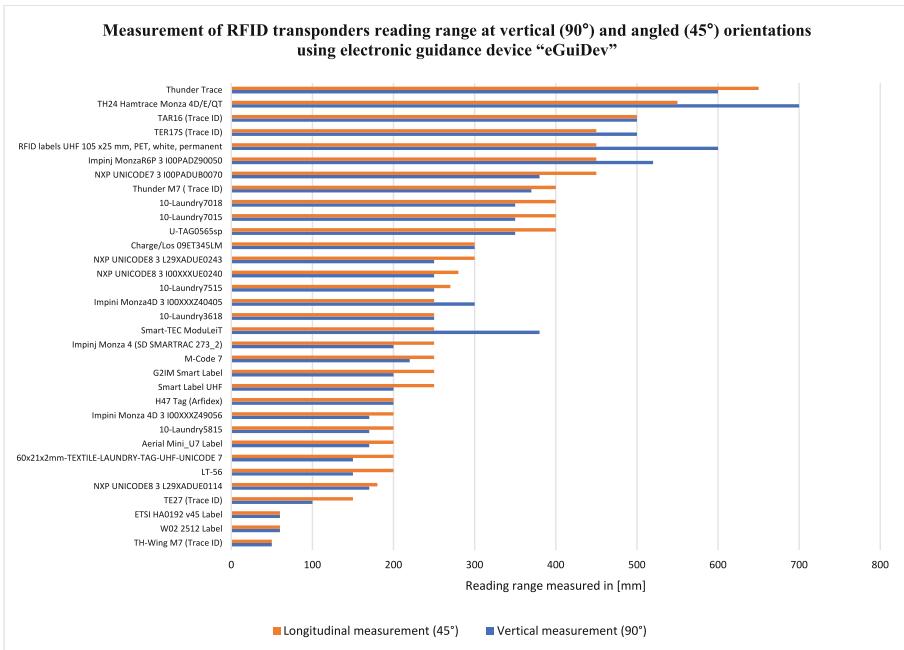


Fig. 6. Results of RFID label measurements over range with electronic reader

use of RFID technology, particularly in relation to transponder placement and material interaction, hold significant potential for developing navigational aids that are more responsive to the unique challenges faced in urban landscapes.

By understanding how environmental factors and material properties affect RFID read ranges, future innovations can be tailored to create more inclusive and navigable urban spaces for people with visual impairments.

Acknowledgements. This research is part of the “ModuLeiT” project (Nr. 22683 BG/2), which focuses on the development of a design guideline for modular, haptic and optical floor indicators for indoor guidance systems for visually impaired people and RFID-based navigation systems based on textile floor coverings. The project is funded by the Federal Ministry for Economic Affairs and Climate Action of Germany.

References

1. van Bueren, E., et al.: Sustainable Urban Environments, An Ecosystem Approach. Springer Science+Business Media B.V., Heidelberg London, New York, pp. 1–3. <https://doi.org/10.1007/978-94-007-1294-2> (2012)
2. Rüppel, U., Marcus Stübbe, K., Zwinger, U.: Indoor Navigation Integration Platform for firefighting purposes. In: 2010 International Conference on Indoor Positioning and Indoor Navigation, Zurich, Switzerland, pp. 1–6 (2010). <https://doi.org/10.1109/IPIN.2010.5647401>

3. Mazzoli, C., Iannantuono, M., Giannakopoulos, V., Fotopoulou, A., Ferrante, A., Garagnani, S.: Building information modeling as an effective process for the sustainable re-shaping of the built environment. *Sustainability* **13**, 4658 (2021). <https://doi.org/10.3390/su13094658>
4. World Health Organization.: A report about increasing eye care interventions to address vision impairment (who.int) (2021)
5. Juraboev, A., Díaz, J., Kern, R.: Embedding RFID tags into precast structural components for tracking and holistic real-time lean construction management. In: S. Skatulla, H. Beushausen (Hrsg.), *Proceedings of 19th International Conference on Computing in Civil and Building Engineering*, Cape Town, South Africa, October 26–28, 2022 (Volume 1, Bd. 49, 9 pages), A Springer book series LNCE. Book: *Advances in Information Technology in Civil and Building Engineering*, ISBN: 978-3-031-35398-7 (2022). https://doi.org/10.1007/978-3-031-35399-4_49
6. Juraboev, A., Suess, S., Kern, R., Díaz, J.: Embedding RFID tags into modular textile floor coverings and integration in BIM. In: *Proceedings of CLEaR—Construction Logistics, Equipment, and Robotics*, 09. – 11. October, Munich 2023, LNCE 390, pp. 1–10, 2024. Springer Nature (2024). https://doi.org/10.1007/978-3-031-44021-2_15
7. ISO 23599:2019-01 Assistive products for blind and vision-impaired persons. Tactile walking surface indicators, pp. 1–41.
8. DIN 32984:2023-04, Tactile walking surface indicators in public areas, pp. 1–71. <https://doi.org/10.31030/341404>
9. ISO/IEC 18000-6:2013, Information technology. Radio frequency identification for item management, Part 6: Parameters for air interface communications at 860 MHz to 960 MHz General
10. MetraTec GmbH, Magdeburg, Docu_DwarfG2_r0103.pdf (metratec.com)
11. Abracon ARRRN5-868 MHz, Abracon | Inventory
12. M5Stack Technology Co. Ltd., Shenzhen, UHF RFID Unit (JRD-4035) | m5stack-store
13. Raspberry Pi Foundation, Raspberry Pi Documentation - Raspberry Pi Pico and Pico W, 2023 <https://www.raspberrypi.com/documentation/microcontrollers/raspberry-pi-pico.html>
14. NUCLEO-32 STM32L432KC EVAL BRD, 2023 <https://www.digikey.com/en/products/detail/stmicroelectronics/NUCLEO-L432KC/6132763>
15. Arduino Nano Every, 2023 <https://store.arduino.cc/products/arduino-nano-every>
16. Entwicklerboards - Ladeplatine für 3,7V Li-Akkus, USB-C, 1A, 2024, https://www.reiche_lt.de/
17. MT3608 DC-DC power supply adapter step up module, 2024, <https://www.az-delivery.de>



Simulation-Based Predictive Analytics for Construction Safety Incidents

Abbey Dale Abellanova^(✉), Hamidreza Golabchi, Stephen Hague, Simaan AbouRizk,
and Yasser Mohamed

Construction Engineering and Management, Holve School of Engineering, University of Alberta,
Edmonton, AB, Canada
abellano@ualberta.ca

Abstract. Productivity loss is one key concern in managing a construction project. This issue is brought by many influencing project factors and is often discussed in the literature. However, isolating the magnitude of productivity loss due to health and safety incidents is yet understudied. This productivity loss is briefly defined as the construction productivity safety incidence rate (CPSIR). A deeper understanding of CPSIR is relevant because when workers are injured, there is a direct loss in productivity due to the recovery period from the injury, fluctuations in workforce allocation, training and rehiring for the vacated skilled position, deflated work morale, and work stoppage. This phenomenon also creates a feedback mechanism as the management continually updates the construction plan based on the risk events that may happen. One approach which can tackle these project dynamics and CPSIR is through simulation. However, simulating CPSIR is overwhelming in practice mainly due to the following reasons: (1) poor understanding of which construction project metrics are accurate predictors of a specific safety outcome on the level of analysis required; (2) metrics for potential precursors of incidents as indicated in literature are rarely collected and if they are, it is fragmented within various departments of an organization; and (3) lack of understanding on the complex interrelationships of different attributes affecting incident rates. This high degree of complexity for a simulation study necessitates a much more sophisticated tool such as high-level architecture (HLA). In the present study, a simulation-based approach through developing a safety federate is used to explore the relationship between construction scenarios that affect incident rates. This paper aims to contribute to understanding the dynamics between construction work package attributes and construction safety. This is demonstrated through the development of the construction safety incident federate and how it interoperates with other federates of a construction project.

Keywords: Safety Management System · High Level Architecture (HLA) · Distributed Simulation

1 Introduction

Enhancing construction productivity stands as a pivotal challenge for stakeholders overseeing costs and quality in construction projects. However, the multifaceted nature of productivity loss defies easy categorization. Lam et al. (2001) reveal that productivity

fluctuates dynamically within the same work item over the construction period. Moreover, low productivity is affected by diverse latent factors and has caused delays in construction projects (Enshassi et al. 2009). Hussain (1979) underscores the indispensable role of construction in development, emphasizing that productivity is its cornerstone.

From a management perspective, system improvements commence with employing proper metrics to measure work performance. In terms of measuring productivity, Randolph Thomas and Yiakoumis (1987) introduced a factor model for measuring construction productivity with factors such as design, management, site, and environmental factors. Moreover, the study of Enshassi et al. (2009) underscored the profound impact of safety factors on productivity, asserting that accidents impede workers, leading to absenteeism. In addition to that, safety incidents and injuries not only jeopardize worker well-being but also result in project delays and cost overruns (Ead et al. 2019). This study seeks to propose a mechanism for calculating productivity loss by focusing on safety-related incidents at construction sites, defining this loss as the construction productivity safety incidence rate (CPSIR).

Simulation emerges as a powerful tool for generating safety incidents. Simulation is capable of handling the dynamic variables of a construction operation to be encoded to prediction models in generating safety incidents and their corresponding severity levels. Ead et al. (2019) proposed a combined continuous and discrete-event simulation to model a safety-related schedule delay. Moreover, Tixier et al. (2017) built a synthetic stochastic simulator capable of generating univariate and bivariate construction safety risks at the situational level. In another study, Abellanova et al. (2022) used a discrete-event simulation to quantify the impact of safety inspections on the safety incidents and the project duration. However, these studies are yet limited in terms of incorporating other important factors in predicting safety incidents. The data that can be used to predict the occurrence of a safety incident may be sourced from a different data warehouse such as environmental, work package plan, spatial, and project progress attributes. These fragmented measures which are collected by various project agents are not easily accessible for a more sophisticated decision, thus limiting the practicality of implementing such assessment methods (Pereira et al. 2020). Building upon these insights, this study proposes utilizing distributed simulation modeling to integrate independent simulators, enabling seamless communication among them and enhancing safety incident predictions.

Despite sophisticated risk assessment models, many construction risks remain poorly understood (Shang and Hossen 2013). When it comes to safety risk assessment, injury prevention in construction projects begins with a clear understanding of construction injury factors (Gondia et al. 2022). Though there are a variety of influence factors attributed to construction injuries and fatalities as widely discussed in the literature, expert judgment is still considered critical (Rodrigues et al. 2016) in assessing the risk of the construction environment and how certain conditions of the site may cause safety incidents and further trigger a loss in productivity. According to (Fayek 2020), it is challenging to use either the extremes of expert knowledge or statistical methods in mapping relationships between variables since there is a plethora of variables that affect productivity. Numerous studies have employed fuzzy systems to assess safety risks (Kim et al. 2016; Liu and Wang 2017; Patel et al. 2016; Zhang et al. 2020), enabling the processing of expert linguistic terms into numeric values.

The present study illustrates the development of a safety prediction model as an independent simulator as a component of a larger distributed simulation model. This federate is capable of generating safety incidents, along with its severity level and the quantified CPSIR. The data inputs used in the prediction model are the updated values received from the other autonomous simulators of the federation. A Fuzzy Inference System (FIS) is the running engine used to predict values of the crisp values of CPSIR and the corresponding category of incidents.

2 High Level Architecture Simulation Framework

The High Level Architecture (HLA) is a standard that is developed by the U.S. Department of Defense to integrate independent autonomous simulators into a single distributed simulation referred to as a “federation” (Azimi et al. 2011). Within this framework, each autonomous simulator is designated as a “federate”. The safety simulation model that is presented in this study is one of the federates and is named the “Safety Federate”. All the federates running under the HLA framework follow a set of standards contained in the IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) – Federate Interface Specification (2010). These standards enable effective interfederate communication, which is essential for seamless interaction with infrastructure software. The exchange of information among federates is facilitated through the standardized services and interfaces provided by the HLA runtime infrastructure (RTI). The RTI is mainly responsible for the coordinated exchange of information when the federates are executed in the distributed simulation environment.

There are existing distributed simulation models running through the HLA implementation for construction simulation purposes. A bidding game was developed by Abourizk et al. (2009) as a training tool for students’ skills in decision-making. Their game used the HLA for a distributed bidding process in construction to explore the rules of a competitive bidding environment. Elnimr and Mohammed (2010) utilized the HLA framework to create a visualization framework for modeling a real pipe spools assembly yard and pipe modules site installations. In another study, Azimi et al. (2011) applied the HLA in automated project monitoring and control to facilitate corrective actions when deviations from the plan occur. Their implementation of the HLA was developed for the fabrication phase of the steel construction allowing for proactive mitigation of potential damage of ongoing steel fabrication. A prediction model to determine the availability of equipment for earthmoving operations was also developed by Pereira et al. (2014) using HLA. Another study by Pereira et al. (2020) employed the HLA framework to create a distributed simulation assessing safety risk levels in a construction project by simulating the impact of safety-related measures on safety performance.

In this study, the focus is on leveraging the HLA framework to develop a distributed simulation that comprehensively evaluates safety risk levels in construction projects. The safety federate, as a crucial component, utilizes HLA standards to interact with other federates and contribute to a holistic understanding of safety incidents and their impact on project performance.

3 Development of the Safety Federate

The role of the safety federate is to generate safety incidents along with their corresponding severity level and lost time or CPSIR. To achieve this objective, the safety federate creates predictions on every work package that is in progress which is dictated by the other simulators in the federation. The prediction is done daily when the safety federate receives updated information from other federates within the federation. Once the predictions are generated, the safety federate then returns the values of severity level, CPSIR, and the affected work package to the distributed simulation through the “PublishObjectClassAttributes”. This receiving of updates is handled by the HLA service called “SubscribeObjectClassAttributes”.

The safety federate joins the federation and synchronizes with other federates at the following synchronization points:

1. Ready to Declare: At this synchronization point, the safety federate declares which attributes are of interest in subscribing.
2. Ready to Initialize: During the initialization phase of the federation, the safety federate discovers all the work package instances created by the planner federate.
3. Ready to Execute: The federation advances time on a daily interval. To ensure that the safety federate is running at the same simulation time as the other federates, advancing time is controlled by the Time Advance Grant service of the HLA. During the Time Advance Grant service, the safety federate keeps track of all the work packages that are in progress.

Since the safety federate is interested in the updates of the work package attribute values of elevation, percent complete, resource type, and assigned resource quantity, daily updates of these attribute values are recorded for prediction purposes. As the attributes of each work package are updated, the values of these attributes are reflected in the safety federate. As it receives updates from the other federates through the RTI, it also notes the daily updates of the global attributes such as temperature and wind speed.

If a work package is in progress, then the safety federate feeds all the input values to the FIS prediction model and it then returns the incident category, the crisp value of lost hours due to safety, and the affected work package.

If the prediction output of the FIS model is an incident, i.e. “very minor”, “minor”, “mild”, “severe”, “very severe”, or “fatality”, the safety federate then creates an instance of a safety incident and updates the lost hours due to safety, incident category, and the affected work package based on the output values of the FIS. In the rare event of a fatality, the lost hours due to safety value is then converted to lost days due to safety since simulates a work stoppage upon the occurrence of a fatal incident. The work stoppage is assumed to be in the range of 5 to days. In most cases where the prediction is a “no incident” category, the safety federate just does nothing for that specific work package on that specific day.

4. Ready to Terminate: The safety federate terminates the simulation once all the work packages achieve 100 percent completion.

4 Safety Incidents and Lost Time Through Fuzzy Inference System

Developing a safety-incident-generating model with a very realistic performance has its complications. The first reason is that every project has its dynamic features and plenty of moving variables that are very difficult to capture in a single model. Second, gathering data for incidents in a single project is not sufficient when dealing with low-probability events. The third is mapping the relationships between the selected variables to the severity of incidents and their corresponding lost time is not straightforward. Working with low-resolution information in generating specific types of incidents and quantifying lost time due to safety requires a model capable of handling vague inputs. One of the available tools capable of handling these vague inputs is the FIS developed by Mamdani and Assilian (1975).

FIS is a variant of Fuzzy Logic which was formally introduced by Zadeh (1965). FIS delivers a means to represent human linguistic terms and vague attributes of the real world in computing so far as variables are concerned (Selase et al. 2015). The logic behind the FIS is that variables, which are defined in a continuum of human language, are mapped into a grade of membership ranging between zero and one (Zadeh 1965). The fuzzy inference engine used in the present study is the Mamdani fuzzy system, originally implemented to investigate the possibility of interaction with a controller system from experienced human operators (Mamdani and Assilian 1975). Similarly, the present study harnesses the interpretable set of rules contained within the Mamdani FIS to generate safety incidents and predict the lost time due to safety as its crisp numeric output.

The Mamdani FIS operates through several key steps:

1. Determining a set of fuzzy rules: A set of 39 fuzzy rules is established, connecting fuzzy input variables to fuzzy output variables in human-readable linguistic terms. These rules form an interpretable set of guidelines for the FIS.
2. Fuzzifying the inputs using the input membership functions: Each fuzzy input variable described in Table 1 is mapped to specific numeric values, depending on its linguistic scale. The linguistic scale of these fuzzy input variables serves as a spectrum of a human being would describe the input variable without providing a crisp numeric value.
3. Combining the fuzzified inputs according to the fuzzy rules to establish a rule strength: The combination of certain fuzzy input variables allows for a more comprehensive description of the situation. An example of this rule is: IF (Temperature IS cold) AND (Wind_Speed IS high) AND (Assigned_Resource_Quantity IS average) AND (Percent_Complete IS small) AND (Elevation IS high).
4. Finding the consequence of the rule by combining the rule strength and the output membership function: When the combination of fuzzy inputs is defined, each fuzzy rule is then mapped to a corresponding fuzzy output. For instance of an output from the established rule in the previous step, the expected output is “very minor” with a corresponding lost man-hour value within 0 and 8 h. The membership values of the output variables are described in Table 2.
5. Combining the consequences to get an output distribution: Since the decisions are inferred by testing through all the rules of the FIS, the output consequence must be combined in some form. This is done through the method of aggregation. Aggregation

is the step whereby the fuzzy sets that correspond to the output of each rule are combined into a single fuzzy set. This is only executed once for each output variable right before the defuzzification step. The final output of this step is a fuzzy set for each of the fuzzy output variables.

6. Defuzzifying the output distribution: The defuzzification step outputs the crisp numeric value, which is the lost man hours. The calculation behind this step could be done in many ways. However, in the present study, the ‘centroid’ method is used. This is done by locating the position of the centroid of the output distribution described in the previous step.

Table 1. Fuzzy Sets and Membership Functions of Inputs

| Fuzzy Input Variable | Linguistic Scale | Membership Function |
|----------------------------|------------------|---|
| Temperature | extreme cold | Trapezoidal($a = -40, b = -35, c = 0, d = 5$) |
| | cold | Trapezoidal($a = 0, b = 5, c = 10, d = 15$) |
| | warm | Trapezoidal($a = 10, b = 15, c = 20, d = 25$) |
| | hot | Trapezoidal($a = 20, b = 25, c = 30, d = 35$) |
| | extreme hot | Trapezoidal($a = 30, b = 35, c = 40, d = 45$) |
| Wind Speed | low | Triangular($a = 0, b = 0, c = 6$) |
| | medium | Triangular($a = 0, b = 6, c = 12$) |
| | high | Triangular($a = 6, b = 12, c = 12$) |
| Elevation | low | Triangular($a = 0, b = 0, c = 2$) |
| | medium | Triangular($a = 0, b = 2, c = 4$) |
| | high | Triangular($a = 2, b = 4, c = 4$) |
| Resource Type | module | Triangular($a = 1, b = 1, c = 1.1$) |
| | pipefitter | Triangular($a = 1.9, b = 2, c = 2$) |
| | ironworker | Triangular($a = 3, b = 3, c = 3.1$) |
| | electrical | Triangular($a = 3.9, b = 4, c = 4$) |
| | mechanical | Triangular($a = 5, b = 5, c = 5.1$) |
| Percent Complete | early | Triangular($a = 0, b = 0, c = 0.5$) |
| | mid-stage | Triangular($a = 0, b = 0.5, c = 1$) |
| | completing | Triangular($a = 0.5, b = 1, c = 1$) |
| Assigned Resource Quantity | few | Triangular($a = 0, b = 0, c = 5$) |
| | average | Triangular($a = 0, b = 5, c = 10$) |
| | many | Triangular($a = 5, b = 10, c = 10$) |

In such cases where the prediction of the FIS is categorized as a ‘fatality’, the defuzzified crisp output is overridden by a random number generator to convert the prediction to lost days instead of lost hours. This is to simulate an actual case where the

Table 2. Fuzzy Sets and Membership of Lost Hours Due to Safety and Scale of Incident Severity

| Fuzzy Output Variable | Linguistic Scale | Membership Function |
|--------------------------|------------------|------------------------------------|
| Lost Hours Due to Safety | no incident | Triangular(a = 0, b = 0, c = 0.1) |
| | very minor | Triangular(a = 0, b = 0, c = 8) |
| | minor | Triangular(a = 0, b = 8, c = 16) |
| | mild | Triangular(a = 8, b = 16, c = 24) |
| | severe | Triangular(a = 16, b = 24, c = 32) |
| | very severe | Triangular(a = 24, b = 32, c = 40) |
| | fatality | Triangular(a = 39, b = 40, c = 40) |

rare occurrence of a fatality leads to a work stoppage until the local regulations permit the project to resume. The lost days due to fatality incidents are assumed to last for 5 to 10 days.

5 Conclusion

The present study describes the development of a construction safety incident simulation to predict the types of incidents and productivity loss through the utilization of the HLA framework. The primary goal at this stage was to create an autonomous federate capable of receiving information from other federates within the simulation, thereby generating safety incidents and linking them to productivity loss. While the study has made significant strides in developing the framework, it is important to note that the current implementation is limited to the available input variables in the federation simulation. The Fuzzy Inference System (FIS) used for incident prediction relies on these variables, and assessing incident severity along with corresponding lost hours demands a broader consideration of influence factors within the FIS model.

Moving forward, the evolution of this predictive simulation model will involve expanding the set of influence factors, refining membership functions, and subjecting the model to real-case datasets for validation. The incorporation of additional factors will enhance the accuracy and comprehensiveness of the FIS model, allowing for a more realistic and reliable simulation of safety incidents and their impact on productivity. This continuous development and refinement process will contribute to a robust predictive tool that can offer valuable insights into construction safety risk management.

References

- Abellanosa, A.D., Wei, Y., Abbasianfar, V., Mohamed, Y., Golabchi, H., Pereira, E.: Discrete-event simulation of construction safety inspections. In: 21st International Conference on Modeling and Applied Simulation, MAS 2022 (2022). <https://doi.org/10.46354/i3m.2022.mas.011>
- Abourizk, S.M., Hague, S., Moghani, E.: Developing a Bidding Game Using High Level Architecture (2009)

- Azimi, R., Lee, S., Abourizk, S.M., Alvanchi, A.: A framework for an automated and integrated project monitoring and control system for steel fabrication projects. *Autom. Constr.* **20**(1), 88–97 (2011). <https://doi.org/10.1016/j.autcon.2010.07.001>
- Ead, R., Elkholyosy, H., Eltahan, A., Hague, S., Abourizk, S.: Using combined discrete-event/continuous modeling for simulation of safety incidents on project networks. *Winter Simulation Conference, 2019-December*, pp. 3041–3052 (2019). <https://doi.org/10.1109/WSC40007.2019.9004752>
- Elnimr, A.A., Mohammed, Y.: A simulation driven visualization framework for construction operations: development and application. *Construction Research Congress*, 257–266 (2010). [https://doi.org/10.1061/41109\(373\)26](https://doi.org/10.1061/41109(373)26)
- Enshassi, A., Choudhry, R.M., Abualqumboz, M.: Safety and productivity in the construction industry. *Arab Gulf Journal of Scientific Research*, **27**(3), 139–155 (2009). <https://www.researchgate.net/publication/299020038>
- Fayek, A.R.: Fuzzy logic and fuzzy hybrid techniques for construction engineering and management. *Journal of Construction Engineering and Management* **146**(7) (2020). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001854](https://doi.org/10.1061/(asce)co.1943-7862.0001854)
- Gondia, A., Ezzeldin, M., El-Dakhakhni, W.: Machine learning-based decision support framework for construction injury severity prediction and risk mitigation. In: ASCE-ASME J. Risk Uncertain. in Eng. Syst. Part A: Civil Eng. **8**(3) (2022). <https://doi.org/10.1061/ajrua6.0001239>
- Hussain, A.: Construction productivity factors. *Issues Eng. J. Prof. Act.* **105**(4), 189–195 (1979)
- IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)– Federate Interface Specification. (2010). IEEE Std 1516.1-2010 (Revision of IEEE Std 1516.1-2000), pp. 1–378. <https://doi.org/10.1109/IEEEESTD.2010.5557728>
- Kim, H., Kim, K., Kim, H.: Vision-based object-centric safety assessment using fuzzy inference: monitoring struck-by accidents with moving objects. *J. Comput. Civ. Eng.* **30**(4), 04015075 (2016). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000562](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000562)
- Lam, K.C., Lee, D., Hu, T.: Understanding the effect of the learning-forgetting phenomenon to duration of projects construction. *Int. J. Project Manag.* **19**(7), 411–420 (2001). [https://doi.org/10.1016/S0263-7863\(00\)00025-9](https://doi.org/10.1016/S0263-7863(00)00025-9)
- Liu, H., Wang, Y.: The assessment of construction project safety management security system based on the fuzzy comprehensive evaluation. *ICCREM* **2017**, 262–271 (2017)
- Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* **7**(1), 1–13. [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2)
- Patel, D.A., Kikani, K.D., Jha, K.N.: Hazard assessment using consistent fuzzy preference relations approach. *J. Constr. Eng. Manag.* **142**(12). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001192](https://doi.org/10.1061/(asce)co.1943-7862.0001192)
- Pereira, E., Ali, M., Wu, L., Abourizk, S.: Distributed simulation-based analytics approach for enhancing safety management systems in industrial construction. *J. Constr. Eng. Manag.* **146**(1) (2020). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001732](https://doi.org/10.1061/(asce)co.1943-7862.0001732)
- Pereira, E., Mohamed, Y., Abourizk, S.: Predicting equipment availability using a high level architecture framework simulation-based analytics for fabrication quality-associated decision support view project. In: Proceedings of the European Modeling and Simulation Symposium 2014, pp. 325–333 (2014). <https://doi.org/10.13140/2.1.4996.5761>
- Randolph Thomas, H., Yiakoumis, I.: Factor model of construction productivity. *J. Constr. Eng. Manag.* **113**(4), 623–639 (1987)
- Rodrigues, M., Leão, C., Nunes, E., Sousa, S., Arezes, P.: A fuzzy logic approach in the definition of risk acceptance boundaries in occupational safety and health. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B: Mech. Eng.* **2**(4), 044501 (2016). <https://doi.org/10.1115/1.4032923>
- Selase, A. E., Xing, C., Agbadze, O.K., Thompson, B. E.: The General Overview of the Phrase ‘Fuzzy Logic’ **5**, 68–73 (2015)

- Shang, K., Hossen, Z.: Applying fuzzy logic to risk assessment and decision-making. *Casualty Actuar. Soc.* 1–59 (2013)
- Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B.: Construction safety risk modeling and simulation. *Risk Anal.* **37**(10), 1917–1935 (2017). <https://doi.org/10.1111/risa.12772>
- Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**(3), 338–353 (1965). [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zhang, M., Cao, Z., Yang, Z., Zhao, X.: Utilizing computer vision and fuzzy inference to evaluate level of collision safety for workers and equipment in a dynamic environment. *J. Constr. Eng. Manag.* **146**(6) (2020). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001802](https://doi.org/10.1061/(asce)co.1943-7862.0001802)



Formation Process Simulation Method Based on the Real Construction Data for Vibro Stone Columns

Daoxiang Chen¹, Peng Lin^{1,2(✉)}, Zhiwei Zhang¹, and Guo Li³

¹ Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China
celinpe@tsinghua.edu.cn

² Sichuan Energy Internet Research Institute, Tsinghua University, Chengdu 610213, China
³ China Huaneng Group Co., Ltd., Beijing 100031, China

Abstract. Vibro stone columns are widely used worldwide for the improvement of soft foundations. The characteristics of underground engineering make it almost impossible to perceive the morphology of columns, and the prediction of the morphology of columns formation is also relatively preliminary. A simulation method for the formation of vibro stone columns based on real construction data is proposed. The construction process parameters are extracted from monitoring data and then transferred to simulation parameters. Based on the coupled Eulerian–Lagrangian method, parametric modeling techniques, and script-based post-processing techniques are used to implement the simulation of continuous construction processes. The formation mechanism of vibro stone columns is analyzed based on the formation process simulation. This method can be used for quality evaluation and digital twin of vibro stone columns.

Keywords: Vibro stone columns · Formation process simulation · Construction data · Formation mechanism · Foundation improvement

1 Introduction

Vibro Stone Columns (VSC) have been extensively utilized for foundation reinforcements across various construction domains, including dams, weirs, and both industrial and civil structures (Bell 2004; Dheerendra Babu et al. 2013). VSC serves multiple foundation improvement objectives such as enhancing bearing capacity, augmenting shear strength, mitigating liquefaction, facilitating drainage and decompression, and managing settlement. Furthermore, the VSC technology is characterized by its high degree of mechanization, swift construction, cost-effectiveness, and environmental sustainability.

Depth and diameter are two critical design parameters of vibro stone columns (Priebe 1995). However, depth and diameter easily fail, cases have been reported by Bell (2004) and Serridge (2016). The digital and informational monitoring system has been established to inspect the quality of vibro stone columns. However, the monitoring data has not been utilized adequately. The diameter assessment method of vibro stone columns

based on construction data is lacking, a numerical method based on construction data can be used to reflect the real pile formation process.

Many scholars have studied the deformation and load transfer mechanisms of VSC through numerical methods. In summary, multiple numerical simulation methods are available for modeling vibro stone columns, each with limitations and specific applicability. One method simplifies the foundation as a homogeneous structure with drainage wells (Stacho and Sulovska 2019), suitable for predicting bearing capacity and settlements but lacking in capturing pile-soil interaction. Modeling stone columns and soil separately is also widely used, allowing for interaction analysis and damage prediction, Ambily and Gandhi (2007) studied the behavioral characteristics of stone columns based on single pile and group pile load tests. Tan et al. (2021) and Indraratna et al. (2015) studied the pile-soil interaction mechanism through single pile load tests and discrete-continuous coupling numerical simulations. Some simulations analyze the reinforcement effect of soil between columns during construction but overlook the interaction between stone and soil, a procedure developed by Liu (2015) simulates the generating and dissipation process of excess pore pressure of soil under vibrational loading during the process of installation of VSC. However, numerical simulation methods for the process of VSC formation are still missing. Nagula et al. (2022) adopted the coupled Eulerian-Lagrangian method to simulate the construction process of the vibro-impact compacting method, revealing the feasibility of the method in the formation process simulation of the vibro stone columns.

To predict the pile diameter of vibro stone columns more accurately, this paper proposes a simulation method for the pile forming process of vibro stone columns based on real construction data. The construction process is extracted from monitoring data and served for modeling and loading. The construction parameter-simulation parameter mapping method is established. The coupled Eulerian-Lagrangian method is used for the simulation of the larger deformation of soil. The formation mechanism of vibro stone columns is analyzed based on the formation process simulation.

2 Methodology

2.1 Formation Process Simulation Method

A numerical simulation method for the vibro stone column formation process was established based on the coupled Eulerian–Lagrangian method with actual construction data. The soil and stone are modeled with the Eulerian domain while the vibrator is modeled with the Lagrangian body, and a two-dimensional surface is contained to realize the interaction between the soil domain and stone domain. The workflow for continuous simulations of the formation process is shown in Fig. 1.

Firstly, an initial soil domain with holes was created, regardless of how the holes were formed. The focus was on the process of filling stones into the holes and compacting them into columns. After establishing the soil domain, the stone domain was created and the vibrator position was initialized according to construction records of filling volumes. The initialization strategy for the vibrator position was 10cm above the upper surface of the stone domain. Based on the established model, a single vibration deposition process was simulated. The updated soil domain range was then extracted from the simulation

results. The updated soil domain model replaced the pre-vibration domain model. New construction data was used to establish another model and simulate it, repeating until the individual pile construction was complete.

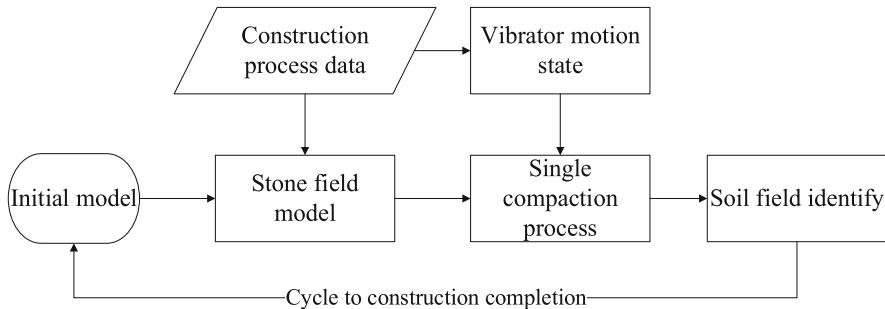


Fig. 1. Formation process simulation process

This section elucidates the method of obtaining the construction process from the construction sequencing data. Generally, in order to ensure superior construction quality, the difference in depth between two consecutive compaction processes is typically within the range of 0.2 m to 0.4 m. Therefore, for a 10m deep stone columns, there are approximately 25 to 50 compaction processes. Figure 2(a) shows a construction sequencing, which contains about 26 compaction processes within 6m. The construction process extraction method is outlined as follows.

Firstly, by analyzing the variations in depth and referring to the current characteristics, the corresponding time intervals for each compaction process can be determined. Then, the relevant data such as current, cumulative filling quantity, compaction time, and compaction position to their respective time frames are matched. This process ultimately yields compaction process data ($k, t_k, D_k, T_k, I_k, SV_k$), where k is the order of increment; t_k is the construction time; D_k is compaction depth; T_k is compaction time; I_k is compaction current; SV_k is cumulative feed volume.

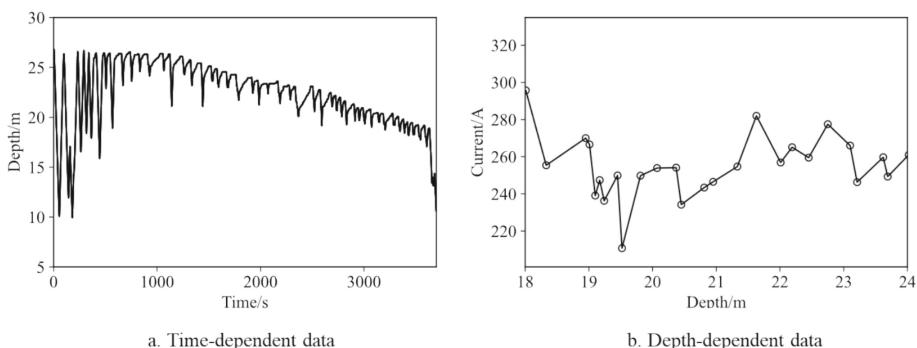


Fig. 2. Construction process detection method

2.2 Simulation Parameters Set

The main parameters set in the simulation include determining the geometric shape of stones and soil bodies, as well as controlling the motion state of the vibrator, as shown in Fig. 3. In terms of geometric shapes, it will be important to realistically model key dimensions such as the varying size distributions of individual stones, as well as properties of the surrounding soil matrix. Regarding the vibrator motion, the simulation would need to define variables like frequency, amplitude, and duration of vibration cycles. This aims to accurately represent the dynamic compaction process induced by the vibrator tool.

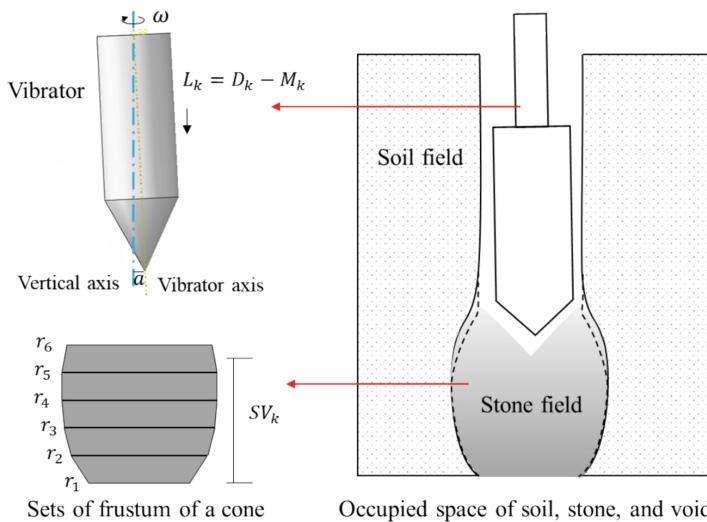


Fig. 3. Schematic of simulation parameters

Geometric parameter. The entire computational domain can be divided into three parts: soil, stone, and void. The task of geometric shape determination is to identify the soil domain and stone domain for further modeling. The geometric shape determination is divided into two steps: The first step is to determine the shape of the space occupied by the stone domain and void domain, also known as the non-soil domain. This equivalently determines the soil domain. After determining the shape of the non-soil domain space, the shape of the stone column space is determined based on the numerical accumulation of the filling amount and the law of mass conservation.

Considering the discrete nature of numerical simulation, for ease of identification of each domain, the non-soil domain is modeled as segmented cylindrical sections of equal heights. At the same time, it is assumed that the upper portion of the stone domain is flat. According to the volume formula for cylinders, the volume of each cylindrical section can be calculated and accumulated, until the sum of cylinder volumes is greater than the cumulative filling volume. Then according to Formulas (3) and (4), the height

of the stone fill can be determined.

$$V_i = \frac{1}{3}\pi H_i(r_i^2 + r_{i+1}^2 + r_i r_{i+1}) \quad (1)$$

$$\arg \max_m \left\{ \left(\sum_1^m V_i \right) > V_{stone}^k \text{ and } \left(\sum_1^{m+1} V_i \right) < SV_k \right\} \quad (2)$$

$$SV_k - \sum_1^m V_i = \frac{1}{3}\pi H_i \left(\frac{r_{i+1}^3 - (r_i + x)^3}{r_{i+1} - r_i} \right) \quad (3)$$

$$h = H_i \left(1 - \frac{x}{r_{i+1} - r_i} \right) \quad (4)$$

where: V_i , r_i , r_{i+1} refer to the volume, lower base radius, and upper base radius of the i-th frustum of a cone; SV_k refer to the cumulative filling volume after the k-th compaction.

Vibrator motion state. The purpose is to determine the spatial motion state of the vibrator through feedback based on monitored current data. This includes the angle between the vibrator axis and the vertical direction, the vertical displacement of the vibrator, and the angular velocity of vibrator rotation. The angle between the vibrator axis and the vertical direction is calculated based on the relationship between current and amplitude, as shown in Formula (5). The vertical displacement of the vibrator L_k is obtained by calculating the initial position of the vibrator P_k and the compaction position D_k .

$$\frac{F}{\omega(M+m)} \sqrt{F_{max}^2 - F^2} = \omega^3(M+m)a \sqrt{a_{max}^2 - a^2} = \lambda \sqrt{3} \cos \varphi U(I - I_0) \quad (5)$$

$$L_k = D_k - P_k \quad (6)$$

where: a refers to the amplitude of the vibrator; F refers to the vibration force of the vibrator; ω refers to the angular velocity of the vibrator; M refers to the vibrator shell mass; m refers to the eccentric block mass; F_{max} refers to the maximum vibration force; U refers to the voltage; I refers to the current; φ refers to the phase difference between voltage and current; I_0 refers to the no-load current; λ refers to the energy conversion efficiency of the motor of the vibrator; L_k refers to the vertical displacement of the vibrator; D_k refers to the compaction position of the vibrator; P_k refers to the initial position of the vibrator.

2.3 Test Condition

The validity of the method is verified by the following simulation modeling.

Geometric Properties. A cubic soil Eulerian domain with length, width, and height of 4, 4, and 6 m respectively are modeled, while a cubic stone Eulerian domain with length, width, and height of 2, 2, and 6 m respectively is modeled. The top 1 m of both the soil domain and stone domain were set as void to accommodate the material flow. The vibrator is simplified and modeled as a combination of a cylinder with a height of 1.7 m and a diameter of 0.426 m, and a right circular cone with a height of 0.3 m and a base diameter of 0.426 m. The interface is modeled based on the outline of the stone domain.

Table 1. Properties of Materials Used

| Material | ρ
(kg/m ³) | E
(kPa) | ν | C_u
(kPa) | ϕ
(°) |
|-----------|--------------------------------|--------------|-------|----------------|---------------|
| Clay | 1650 | 12,500 | 0.3 | 10 | 37 |
| Stones | 2000 | 200,000 | 0.17 | — | 43 |
| Interface | 2000 | 1,000 | 0 | — | — |

Material Properties. The Mohr-coulomb plasticity is used. Three materials are contained in the simulation, i.e. clay, stone, and interface, as shown in Table 1.

Boundaries and loads. The bottom boundary of both the soil domain and stone domain were restrained against vertical movement. Meanwhile, the external lateral boundaries of the soil domain were restrained against horizontal movement while no external lateral boundaries were restrained to the stone domain due to the stone material cannot reach the boundary during normal calculations. A cylindrical coordinate system was chosen for the contact surfaces, allowing only radius displacements.

3 Results

3.1 Formation Process Analysis

The initial hole diameter is 0.6 m and the finished pile diameter is about 1.4 m. However, the bottom pile diameter is only 0.6 m, as shown in Fig. 4.

From the perspective of pile diameter, two stages can be observed. It can be clearly seen from Compaction 1 to 6 that there is a noticeable bulging space, which is caused by the smaller diameter of the bottom and top columns. From Compaction 7 to 15, the intermediate crushed stones have been adequately expanded, and the diameter of the stone columns is relatively uniform.

From the perspective of soil deformation, it is mainly influenced by the squeezing effect caused by the expansion of the crushed stone. From Compaction 1 to 11, the squeezing effect is manifested by the phenomenon of necking in the upper hole. As the depth of vibration decreases, the position of necking also moves upward. Considering the effect of water, the protruding soil will be washed away by water flow, restoring the integrity of the hole. Until Compaction 12 and beyond, the squeezing effect mainly causes the uplift of the hole opening.

3.2 Mechanism Analysis

Based on the formation process analysis, a fine pile formation model including the process of expanding-cavity-filling-compacting is established.

The crushed stone is in a loose accumulation state when it is filled into the hole. As the vibrator descends, it will squeeze the crushed stone, and the crushed stone will be expanded horizontally and vertically, its transverse expansion process squeezes the soil

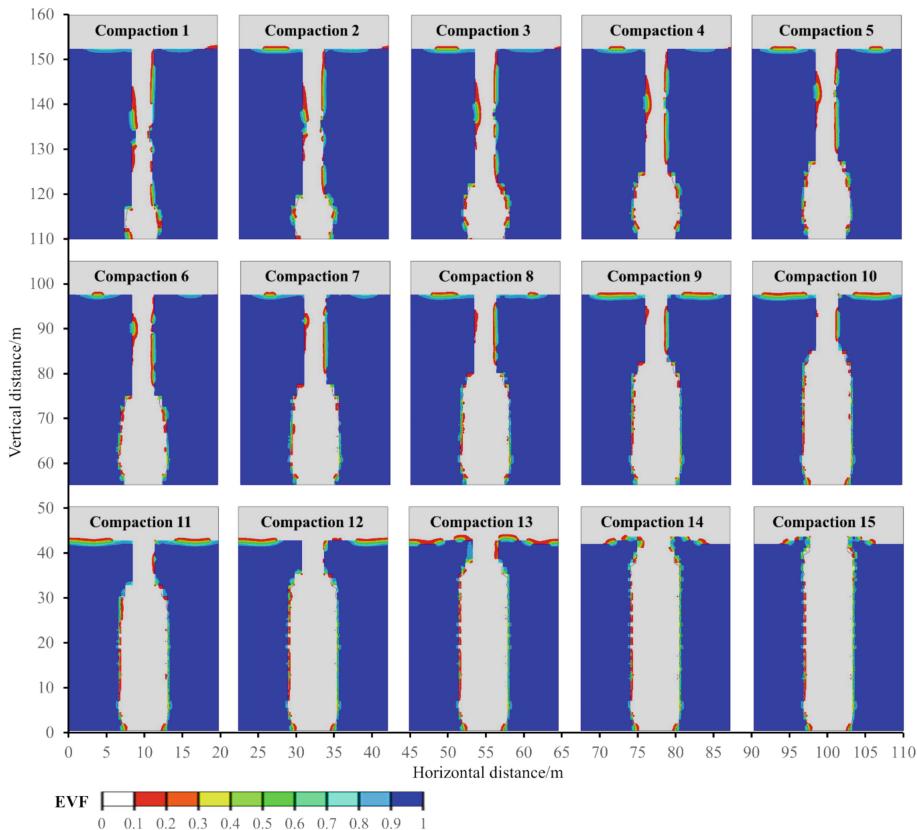


Fig. 4. Results of the formation process

to expand the hole, and at the same time, the longitudinally expanding crushed stone will fill up the gap between the vibrator and the original upper soil, and less transverse expansion will occur under the transverse vibration of the vibrator due to the small amount of crushed stone, and then a drum-shaped space is formed, which is the process of expanding. In the process of expanding, if the time of vibration is long enough, the vibrating force of the vibrator and the lateral earth pressure of the surrounding soil form a dynamic balance, and the diameter of the stone columns is no longer expanded.

After the vibrator is lifted up, the crushed stone will settle and pile up under the action of gravity, occupying the space previously occupied by the vibrator and forming a cavity in the upper part. Meanwhile, the crushed stone which has been compacted by extrusion and preliminary vibration will be loosened by unloading until the next filling and stay vibration to fill the cavity formed by the uplift of the vibrator and vibration to compact the fill, when the encryption current is large enough, the pile compactness is mainly controlled by the time of stay vibration.

4 Conclusions

This study proposes a simulation method based on real construction data to model the formation of vibro stone columns. As a result of this study, the following conclusions can be drawn:

The formation process simulation method mainly consists of construction data extraction, parametric modeling, and script-based post-processing. The construction data is extracted from the monitoring data with the process of construction interval delineation and construction parameter characterization. The coupled Eulerian–Lagrangian method is practicable to deal with the process of vibro stone column formation process with parametric modeling and script-based post-processing techniques. Eulerian domains can handle large deformations and displacements of stone and soil. Meanwhile, the two-dimensional surface can be well realized for the interaction between gravel and soil.

Through the simulation, the formation mechanism of vibro stone columns is analyzed. A fine pile formation model including the process of expanding-cavity-filling-compacting is established, which can guide the control of vibro stone column construction.

Acknowledgments. This research was supported by the China Huaneng Group., Co., Ltd. (No: HNKJ21-H35, HNKJ21-H74), and the science and technology programme of Xizang Autonomous Region (No: XZ202401ZY0078).

Declaration of Competing Interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Ambily, A.P., Gandhi, S.R.: Behavior of stone columns based on experimental and FEM analysis. *J. Geotech. Geoenviron. Eng.* **133**, 405–415 (2007)
- Bell, A.L.: The development and importance of construction technique in deep vibratory ground improvement. In: *Ground and Soil Improvement*, pp. 101–111 (2004)
- Dheerendra Babu, M.R., Nayak, S., Shivashankar, R.: A critical review of construction, analysis and behaviour of stone columns. *Geotech. Geol. Eng.* **31**, 1–22 (2013)
- Indraratna, B., Ngo, N.T., Rujikiatkamjorn, C., Sloan, S.W.: Coupled discrete element–finite difference method for analysing the load-deformation behaviour of a single stone column in soft soil. *Comput. Geotech.* **63**, 267–278 (2015)
- Liu, Y., Yan, H.: Vibro-stone column reinforcement mechanism and numerical analysis. In: Du, X., Zheng, J., Yan, W., Li, Y., Zhang, J. (eds.) *Trends in Civil Engineering*, PTS 1-4. pp. 1940–1943. Coll Architecture & Civil Engineering; Beijing University of Technology; School of Civil Engineering & Architecture; Zhejiang University of Technology (2012)
- Nagula, S.S., Katt, J., Grabe, J.: Optimisation of vibroflotation by amplitude in 1g tests and coupled Eulerian-Lagrangian simulations. *Int. J. Phys. Modell. Geotech.* **22**, 143–156 (2022)
- Priebe, H.J.: The design of vibro replacement. *Ground Eng.* **28**, 31–37 (1995)
- Serridge, C.J.: Briefing: learning from vibro stone column failures. *Proc. Inst. Civ. Eng.-Ground Improv.* **169**, 1500012 (2016)

- Stacho, J., Sulovska, M.: Numerical modelling of the subsoil zone between stone columns. *Slovak J. Civil Eng.* **27**, 32–39 (2019)
- Tan, X., Hu, Z.B., Feng, L.J., Zhao, M.H.: Three-dimensional discrete-continuous coupled numerical simulation of a single stone column in soft soils. *Chinese J. Geotech. Eng.* **43**(2), 347–355 (2021). (in Chinese)



Prototyping a Wireless Hardware System for Automated Collection and Transmission of Audio and Kinematic Data at Construction Jobsites

Seyedeh Zahra Golazad¹(✉), Farzad Ordubadi², Abbas Mohammadi¹, Armin Tajalli²,
Abbas Rashidi¹, and Sadegh Asgari³

¹ Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT,
USA

{Mina.Golazad, abbas.mohammadi, abbas.rashidi}@utah.edu

² Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT,
USA

{farzad.ordubadi, armin.tajalli}@utah.edu

³ Department of Civil Engineering, Merrimack College, North Andover, MA, USA
asgaris@merrimack.edu

Abstract. This study introduces a novel wireless sensory system for automated monitoring of construction machinery, leveraging audio and kinematic data to enhance job site operations, including scheduling, cost forecasting, safety assessments, and fleet optimization. The system features two embedded units based on ARM processor architecture: a sensor unit integrating a microphone, accelerometer, gyroscope, and radio module, capable of transmitting data up to 100 feet, and a receiver unit for data collection and processing. This design enables flexible, real-time monitoring, overcoming the limitations of traditional wired systems. The accuracy of the hardware was validated through extensive field tests conducted at three different distances: 30, 60, and 100 feet. The results obtained from these tests were promising, indicating high reliability and effectiveness. In particular, the system's multi-modal approach, which combines audio and kinematic data, demonstrated an average accuracy of 79.46% in recognizing construction machinery activities. This significantly outperforms models that rely solely on single data types and underscores the efficacy of integrating diverse data sources for precise activity recognition. Such accuracy is vital for improving operational efficiency and ensuring safety at construction sites. However, the research also underscores the need to optimize sensor placement and data transmission, particularly over longer distances, to enhance system performance further. These results offer significant insights into the development of advanced, automated monitoring systems in the construction industry, marking a step forward in the field's ongoing digitalization and innovation.

Keywords: Construction Monitoring · Wireless Sensing · Activity Recognition

1 Introduction

The construction industry faces numerous challenges in efficiency, productivity, and safety management, as highlighted by Ref. [1, 2]. The construction research area encompasses a wide range of topics to address various challenges within the industry. It has many aspects, such as sustainable construction [3], construction materials [4], construction safety and health [5], infrastructure management [6], and construction technology and automation [7]. One of the subfields of construction technology and automation is site management. A key aspect of construction site management is monitoring machinery activities, which traditionally rely on manual methods. The advent of technology in construction, particularly in automated systems, presents new opportunities for enhanced monitoring and management. In this context, the automated recognition of construction machinery activities using sensory data is gaining importance. This technological advancement aligns with the industry's increasing focus on digitalization and automation to improve efficiency, safety, and decision-making processes [8].

Current monitoring methods in construction, including manual observation and basic electronic tracking, are limited in their effectiveness. Manual monitoring is labor-intensive, subject to human error, and often lacks real-time processing capabilities. Additionally, the existing automated systems are often limited in their range of data collection, struggling to capture complex machinery movements and environmental interactions accurately [7]. These limitations hinder the timely and accurate gathering of data, which is crucial for effective project management and safety compliance.

The primary objective of this research is to develop and validate a wireless hardware system for the automated collection and transmission of audio and kinematic data from construction machinery. This system aims to overcome the limitations of current monitoring methods by providing a comprehensive, real-time prediction of machinery activities at construction sites. By integrating audio and kinematic data, the system seeks to enhance the accuracy and reliability of activity recognition, contributing to improved operational efficiency and safety management.

While modern construction machinery is often equipped with built-in sensors, these sensors are primarily designed for operational control and basic diagnostics rather than detailed activity recognition and data analysis. These inherent sensors typically focus on parameters such as engine performance, fuel efficiency, and maintenance alerts [9]. However, they fail to capture the nuanced audio and kinematic data required for comprehensive activity monitoring and analysis in construction environments. The developed sensory board addresses this gap by integrating advanced audio and motion detection capabilities. Unlike standard sensors, our system provides a multidimensional view of machinery operations, including capturing subtle vibrations, movements, and sound patterns specific to different construction activities. This level of detail is crucial for accurate activity recognition, operational efficiency, and safety compliance, which are not achievable with the limited scope of existing sensors in construction machinery. The sensory board's ability to collect and transmit a rich dataset in real-time enables a deeper understanding of machinery behavior, paving the way for advanced analytics and intelligent decision-making on construction sites.

This study has a potential impact on the field of construction technology. The successful development of an automated, wireless sensory system for construction machinery

monitoring can transform how construction sites are managed. It can improve project scheduling, cost forecasting, safety assessments, and fleet optimization. The findings of this study are expected to contribute to the growing body of knowledge in automated construction monitoring.

2 Literature Review

The advancement of technology in construction monitoring has led to significant improvements in efficiency and safety management. Recent studies have emphasized the need for automated data acquisition and real-time monitoring in construction operations. Ref. [10] highlights the limitations of manual methods and introduces the concept of automated site data acquisition using remote sensing technologies such as GPS, UWB, RFID, WSN, and digital imaging. It also suggests integrating these technologies with information models like BIM to enhance efficiency.

Recent advancements in construction monitoring technology have significantly enhanced efficiency and safety management. On-Board Instrumentation (OBI), an integral part of construction machinery, now includes various sensors for operational efficiency and safety. Externally, geospatial tools such as RFID, ultra-wideband tags, and GPS are utilized for object tracking on construction sites. Additionally, vision-based technologies and 3D laser scanning have become pivotal in progress monitoring, offering detailed 3D site information. Integrating these technologies with BIM further augments construction site management, facilitating real-time monitoring and comprehensive data analysis [11].

Similarly, Ref. [12] presents a scalable technical approach for real-time 3D visualization and construction equipment monitoring, using sensor data to update 3D models in digital twins. Furthermore, a different approach discusses using a construction sound library and a sound classification framework for autonomous audio-based site monitoring, emphasizing its role in enhancing safety and progress analysis [13].

Expanding on the theme of audio-based monitoring in construction, Ref. [14] investigates the impact of hardware and software settings on the audio-based analysis of construction operations. This study developed an audio-based system to recognize routine sounds of construction machinery, achieving an accuracy of over 85% and demonstrating its practical value in calculating equipment productivity rates. Additionally, Ref. [15] introduces a cost-effective, practical multi-label sound classification method for recognizing heavy construction equipment activities using Short-Time Fourier Transform and Convolutional Neural Network. This innovative approach, requiring only a single-channel microphone, effectively identifies activities on job sites without needing to separate sound signals.

Adding to the utilization of audio data in construction, Ref. [16] proposes an audio-based cycle time forecasting system for predicting cycle times in construction projects. This innovative system uses signal processing, machine learning, and statistical analysis to estimate productivity, leveraging sound as a primary source of information. This approach marks a significant contribution to the field, considering sound as a valuable source of information for productivity estimation and cycle time forecasting in construction activities.

In the realm of activity recognition, several studies have employed machine learning methodologies. One study focuses on classifying construction worker activities using accelerometer data, achieving high accuracy through both user-dependent and user-independent approaches [17]. Another introduces a machine learning method using Random Forest and fractional calculus-based feature augmentation for activity recognition of construction equipment, demonstrating its effectiveness through various case studies [18]. A comprehensive review of the state-of-the-art methods for activity recognition in construction discusses the challenges and limitations of existing automated methods, underscoring the need for further research and development in this area [19].

Additionally, smartphone-based tracking of construction workers' movements has been explored, with machine learning algorithms trained for activity recognition showing promising results [20]. Another study validates the use of built-in mobile sensors for construction simulation input modeling, highlighting its applicability in real-world scenarios [21].

Wearable technologies and sensor integration have also been a focus area. A notable study uses artificial neural networks to recognize complex construction activities through wearable EMG and IMU sensors, achieving a high accuracy in recognizing scaffold builder activities [22].

Collectively, these studies contribute to our understanding of how automated systems, machine learning, and advanced sensor technologies can revolutionize monitoring and management practices in the construction industry. The integration of these technologies promises not only enhanced efficiency and safety but also paves the way for innovative approaches in construction project management.

3 Methodology

3.1 Design of the Sensory System

For this project the data from the construction machinery is collected using two separate circuit boards. These ARM microcontroller-based boards are functioning as a transmitter circuit and a remote receiver circuit.

The boards are the ST Microelectronics B-L475E-IOT01A Discovery kit. These kits feature multiple on-board modules, including audio, gyroscope, accelerometer, time, radio frequency transmission and reception sensors. Moreover, USB connections are available for further processing of the data from the on-board modules.

The discovery kits are mainly meant for IoT applications and were a suitable choice for the task at hand. Therefore, the development of the hardware platforms for the goals of this project was done using the B-L475E-IOT01A discovery kits.

Transmitter Board

The purpose of the transmitter board is to be mounted on the construction machinery and collect the desired sensory data. This was done by securing the kit inside a suitable box meant for this purpose.

Using the onboard LSM6DSL module, the accelerometer within a range of $\pm 2\text{ g}$ and a sensitivity of $0.061 \frac{\text{mg}}{\text{LSB}}$ is initialized. Furthermore, using LSM6DSL, the gyroscope is also initiated with an operating range of $\pm 2000\text{dps}$ and a sensitivity of $70 \frac{\text{mdps}}{\text{LSB}}$.

The two digital omnidirectional microphones MP34DT01 available on the kit are utilized to collect the audio signals from the environment with less impact from the present noise. The audio frequency bandwidth is limited to a range from *DC* to *6kHz* as the limits of the kits dictated for this application. Moreover, this was considered sufficient for the audio signals from the contraction equipment in use on the job site [7, 23, 24].

The collected data from the previously mentioned sensors was also time indexed by the clock oscillator available on-board the kits.

Finally, the data is packeted and sent over the radio to be collected from a remote board (the receiver). This is done using the Sub-GHz (868 MHz or 915 MHz) low-power-programmable RF module (SPSGRF-868/SPSGRF-915) provided on the kits. Considering the amount of gathered sensory data each second, the radio limitations from the kit resulted in the reliable wireless communication range being limited to roughly 50/60 m.

Receiver Board

For the discovery kit operating as the receiver, only the programmable RF module (SPSGRF-868/SPSGRF-915) and the USB connections are configured. The receiver will gather the data from the transmitter that is mounted on the machinery and store the data on the computer for further processing.

Python language was used to develop the USB interface between the receiver board to store the sensory data as a file on the computer. The raw sensory data from the discover kits are stored in an encoded manner for each on-board module. Therefore, MATLAB functions were developed to decode and format these data for further processing. The overall diagram of the system is shown in Fig. 1.

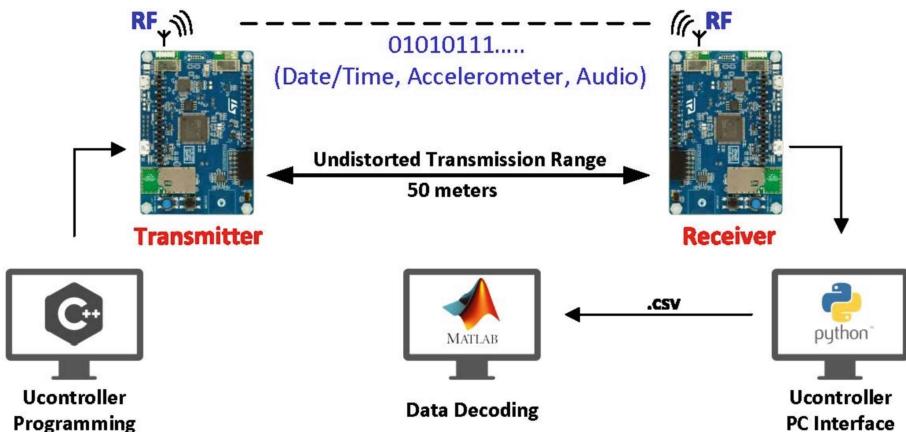


Fig. 1. The overall hardware system developed for the transmission and collection of data from the construction machinery

3.2 Data Collection Process

Site and Equipment Description

Data was collected from a road construction site located in Farmington, Utah. The primary subject of the tests was a CAT 320 EL, a common piece of construction equipment. This selection was strategic, ensuring the results would be applicable to widely used machinery in the industry.

Sensor Setup

The core of data collection involved a sensory board specifically designed for this study. This sensory board was securely attached to the CAT 320 EL. Concurrently, a comparison sensor, which in this study was a smartphone, was used. This choice was inspired by the findings of Ref. [25], who demonstrated the efficacy of smartphones in collecting reliable data for activity recognition of construction equipment. To mitigate any interference or external vibrations, dampers were employed on both the sensory board and the smartphone. Figure 2 shows the setting of the tests in this study.



Fig. 2. (a) The CAT 320 EL equipped with the sensory board attached to its body, demonstrating the practical application of the sensory system in a real-world construction environment. (b) The data receiver setup connected to a laptop for data acquisition and processing, as positioned on the construction site near the equipment.

Testing Protocol

Our research involved a series of 36 tests, evenly split between focusing on accelerometer data and gyroscope data. These tests were conducted at three distinct distances between the receiver and the transmitter: 30, 60, and 100 feet. The rationale behind selecting these specific distances was to gather data across short, medium, and long-range transmission scenarios. This approach provides insights into the sensor's performance and reliability in different operational ranges typically found in construction sites.

Operational Tasks

During each test, the CAT 320 EL was engaged in two specific operational tasks: moving forward/backward and arm raising/lowering. These tasks were selected to represent common movements in construction equipment, providing a relevant context for activity

recognition objectives. Classifying actions such as movements forward/backward and arm lowering/raising is meaningful as these are common tasks that provide a relevant context for activity recognition objectives and significantly contribute to the operational understanding of construction machinery movements. Future work will incorporate a broader range of actions and their corresponding magnitudes as classification and regression task objectives to enhance the system's applicability and accuracy.

Duration and Data Collection

Each test lasted approximately 30 s, a duration determined to be sufficient for collecting a comprehensive set of data while maintaining operational efficiency. Throughout each test, the sensory board and smartphone sensors continuously recorded data, capturing vital information about the equipment's movements and environmental conditions.

4 Analysis Steps and Machine Learning Model

This section details on the data analysis methods used in this study. Two distinct approaches were employed: 1) comparative analysis using accelerometer and gyroscope data; and 2) training of machine learning models for activity recognition with or without audio data and distance-based analysis.

4.1 Comparative Analysis

First, accelerometer and gyroscope data were preprocessed. In this regard, first, the comparison sensor data was shifted to cover the gap between their recording start times. The comparison sensor data was downsampled to match the sensor board. Then, a low-pass filter was applied to both sensors' data to remove high-frequency noise. The data was also smoothed using a moving average filter.

Key statistical measures, including mean difference (bias), standard deviation of differences, and root mean square error (RMSE), were computed to compare the performance of the introduced sensor against a standard sensor. Correlation coefficients were calculated to assess the similarity between the studied sensor data and the standard sensor data across various processing stages (denoised, smoothed, and shifted data).

4.2 Machine Learning Model Training and Distance-Based Analysis

Model Design

The study employed an XGBoost classifier for activity recognition, chosen for its effectiveness in handling large datasets and preventing overfitting through advanced regularization. This model is well-suited for scenarios requiring high precision, like sensory data-based activity recognition. The classifier was configured with a binary logistic objective for the binary classification of two activities: "arm lowering/raising" and "moving forward/backward." The architecture included a maximum depth of 3 for base learners, a learning rate of 0.1, and 100 boosted trees.

To evaluate the system's performance, three scenarios were considered: using only audio signals, only kinematic signals, and a combination of both to train an XGBoost

model. The combined approach involved aggregating features from both data types, enhancing the discriminative power of the model.

The study also involved training three models on the combined dataset of both signal types. These models were specifically trained on data collected from distances of 30, 60, and 100 feet from the transmitter. This approach was undertaken to assess how distance impacts the effectiveness of activity recognition, providing insights into the robustness of the system across various operational ranges.

These methods formed the basis of data analysis, preparing the dataset for effective machine learning model training and ensuring reliable activity recognition.

Training Process

For the analysis of data collected using the sensory board, a comprehensive approach was employed to ensure accurate activity recognition. This section outlines the techniques used for data preprocessing, feature extraction, and preparing the data for machine learning modeling.

Data Preprocessing: The data preprocessing process involves several key steps. In the accelerometer data preprocessing, gravity is removed to focus solely on dynamic accelerations. A low-pass filter, set between 0.1 to 0.5 Hz, isolates gravitational components from the sensor readings. The filter updates gravity values $g(t)$, initially zero, and subtracts these from the sensor's amplitude $s(t)$. This method, detailed by Ref. [26], effectively separates static gravity effects from the accelerometer's dynamic readings, yielding true kinematic values indicative of the equipment's relative motion.

Then a moving window method was used on all data to identify and remove outliers. This approach calculates the mean and standard deviation within a window, identifying and replacing outlier values with NaN, which are then filled using backfill methods. The results showed that removing outliers with a window size of 3 works well on the data.

The median filter, applied here with a kernel size of 21, is a strategic choice for audio data denoising. It operates by replacing each sample with the median of its neighbors, effectively smoothing the signal. This technique is particularly useful for maintaining the crucial characteristics of audio signals, such as sharp transitions or edges, while efficiently reducing background noise. Its implementation in this context ensures cleaner audio data, essential for precise analysis and interpretation.

Feature Extraction: The audio and kinematic data needed to be synchronized due to their different sampling frequencies. This study chose to up-sample the kinematic data to match the audio data's higher sampling rate, as down-sampling the audio would lead to information loss. This up-sampling was achieved using linear interpolation. Then, a Short-Time Fourier Transform (STFT) was performed on both data types to convert them into the time-frequency domain, extracting features relevant for activity recognition. The features extracted are standardized through Z-score normalization to mitigate scaling effects, ensuring that features with a broader range do not overshadow those with a smaller range.

Model Training: The training set was prepared by splitting the dataset into training and testing sets with a ratio of 80:20. The XGBoost model was trained using cross-validation with 5-folds, assessing its performance on the training set. Parameters such as max depth, learning rate, and the number of estimators were carefully selected to optimize the model's performance.

Model Evaluation

The model's performance was evaluated on both the training and testing sets. Cross-validation results provided an average accuracy score on the training set, indicating the model's consistency and reliability. The final evaluation of the testing set involved accuracy measurements and a detailed classification report. These metrics demonstrated the model's effectiveness in accurately recognizing the specified activities of "arm lowering/raising" and "moving forward/backward."

This comprehensive approach, combining robust data preprocessing, strategic feature extraction, and careful model training and evaluation, ensured accurate activity recognition from the sensory data collected.

5 System Testing and Results

5.1 Comparative Analysis

In evaluating the performance of our wireless sensory board for the automated collection and transmission of audio and kinematic data, a comparative analysis was conducted between the custom sensor mounted on construction machinery and a comparison sensor across a series of tasks and distances.

Figure 3 shows the impact of task type and transmission distance on the board's accelerometer performance. For accelerometer data, the correlation coefficients ranged from 0.1839 to 0.3437, indicating a moderate linear relationship and general agreement between the sensors' readings. Notably, the arm raising/lowering tasks tended to yield higher correlation values, which suggests that this type of activity may be more consistent in terms of movement patterns detectable by both sensors. Moreover, the lower correlation in moving forward/backward tasks could be attributed to the conditions of line-of-sight data transmission, which makes it harder to maintain the data collection in these tasks.

Also, the relative mean differences were relatively low across all tasks, with values oscillating narrowly, which speaks to the precision of the sensory board in capturing the intensity of movements. As expected, the variability in data, as indicated by Standard Deviation and RMSE, increased with the distance, reflecting the expected degradation of signal quality over longer transmission ranges. This degradation is a common phenomenon in wireless communication and is exacerbated by physical barriers and environmental factors that disrupt the line of sight between the transmitter and the receiver.

Figure 4 shows the impact of task type and transmission distance on the board's gyroscope performance. The gyroscope data, however, presented a wider range of correlation coefficients, from as low as 0.1521 to as high as 0.6537. This larger range suggests that gyroscope data may be more sensitive to the specific nature of the tasks, perhaps due to the complex rotational movements involved. The negative relative mean differences in some tests indicate a potential systematic bias or calibration issue between the two sensors, requiring further investigation. Notably, the task of arm raising/lowering at the 100-foot distance produced the highest relative mean difference and RMSE values, indicating that this task's gyroscope readings were significantly more affected by distance

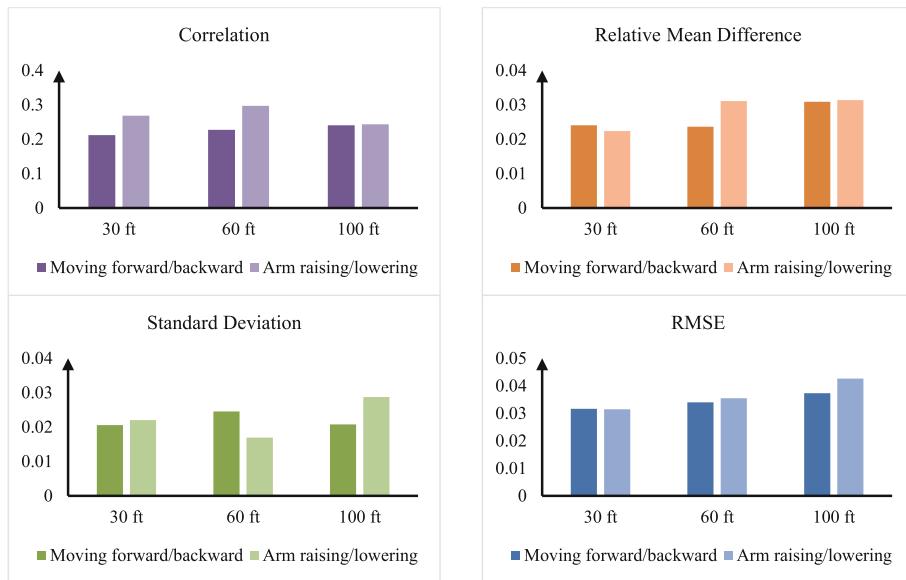


Fig. 3 Comparative Analysis of Accelerometer Data. Four bar charts depict the correlation, relative mean difference, standard deviation, and root mean square error (RMSE) for the accelerometer data collected from construction machinery during two different tasks at 30 ft, 60 ft, and 100 ft

than the accelerometer readings. This shows a significant challenge for line-of-sight transmission in capturing consistent gyroscope data.

The comparative analysis has demonstrated that while both sensors are capable of detecting construction machinery activities, there are nuances in their performance that are influenced by the nature of the task and the distance of data transmission. The accelerometer shows relatively consistent performance with a moderate degree of correlation between sensors, whereas the gyroscope's performance is more variable, with some tasks showing high correlation and others demonstrating significant differences. This variability could be attributed to several factors, including sensor sensitivity, data transmission stability, and environmental conditions at the construction site.

The implications of these findings are two-fold. Firstly, they highlight the necessity for the development of advanced data processing techniques to mitigate the effects of transmission distance and line-of-sight disruptions on data integrity. Secondly, they underscore the importance of optimizing sensor placement on the construction equipment to minimize obstructions and maximize the quality of the transmitted signal.

In the next section, the gathered data will be used to train a machine learning (ML) model to assess the system's efficacy in activity recognition. The ML model's performance will be a critical measure of the system's practical application, as it needs to reliably interpret sensor data to inform job site monitoring and decision-making processes.

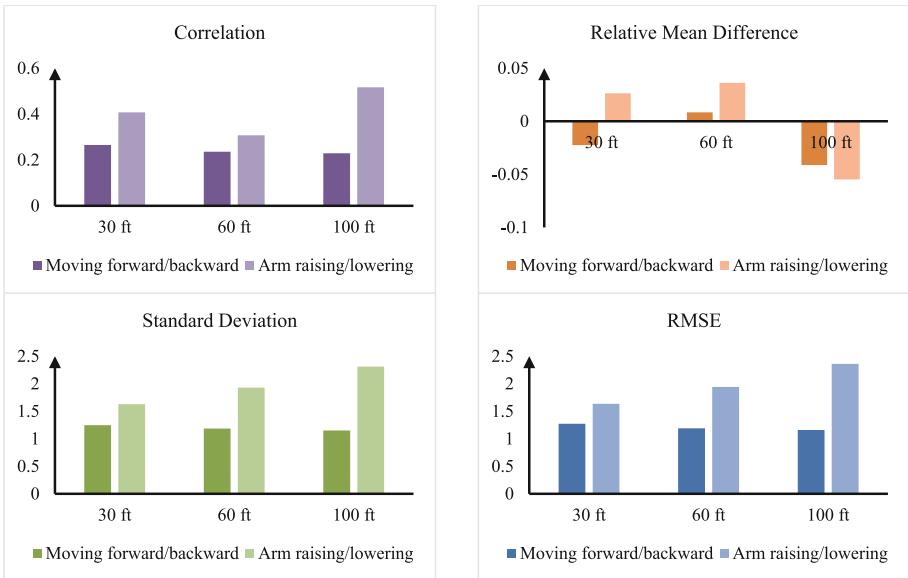


Fig. 4 Comparative Analysis of Gyroscope Data. Four bar charts depict the correlation, relative mean difference, standard deviation, and root mean square error (RMSE) for the gyroscope data collected from construction machinery during two different tasks at 30 ft, 60 ft, and 100 ft

The potential of this system, if fine-tuned, is substantial. It could offer a significant leap in automated job site monitoring, providing real-time, accurate insights into machinery operations, which are crucial for enhancing efficiency, safety, and operational oversight in construction environments.

5.2 Machine Learning on Data

This section analyzes the performance of machine learning models trained on different types of data collected by the sensory board: kinematic, audio, and a combination of both. The results of the trained models are presented in Table 1. The objective is to understand how each data type contributes to the accuracy and precision of activity recognition in construction machinery, specifically identifying movements forward/backward (Class 0) and arm lowering/raising (Class 1).

Performance of Combined Kinematic and Audio Data

The model trained on the combination of kinematic and audio data shows an average accuracy of 79.43% on the training set and 79.46% on the test set. This model demonstrates a balanced performance, with a precision of 76% for Class 0 and 84% for Class 1, and recall rates of 88% for Class 0 and 70% for Class 1. The F1-scores for the classes are 82% and 77%, respectively, indicating a strong ability to distinguish between the two activities.

The combination of kinematic and audio data harnesses the strengths of both sensor types, leading to a finer and more accurate activity recognition. This synergy is evident in the balanced precision and recall scores across both classes.

Table 1 Comparative performance metrics of machine learning models trained on combined kinetic and audio data, kinetic data only, and audio data only

| | Average Training Accuracy | Average Test Accuracy | Precision | | Recall | | F1-Score | |
|-----------------------------------|---------------------------|-----------------------|-----------|---------|---------|---------|----------|---------|
| | | | Class 0 | Class 1 | Class 0 | Class 1 | Class 0 | Class 1 |
| Combined Kinematic and Audio Data | 0.7943 | 0.7946 | 0.76 | 0.84 | 0.88 | 0.7 | 0.82 | 0.77 |
| Kinematic Data | 0.7150 | 0.7137 | 0.68 | 0.77 | 0.85 | 0.57 | 0.76 | 0.65 |
| Audio Data | 0.7398 | 0.7430 | 0.68 | 0.9 | 0.95 | 0.52 | 0.79 | 0.66 |

Performance of Kinematic Data Only

When trained only on kinematic data, the model achieved an average accuracy of 71.50% on the training set and 71.37% on the test set. This configuration yielded a precision of 68% for Class 0 and 77% for Class 1, with recall rates of 85% and 57% respectively. The lower recall rate for Class 1 suggests a challenge in accurately detecting arm raising/lowering activities with kinematic data alone.

Performance of Audio Data Only

The model trained exclusively on audio data demonstrated an average accuracy of 73.98% on the training set and 74.30% on the test set. This model showed a high precision of 90% for Class 1 but a lower precision of 68% for Class 0. The recall rates were 95% for Class 0 and 52% for Class 1, indicating a stronger capability in detecting forward/backward movements but a significant limitation in identifying arm movements.

Comparative Analysis and Implications

The combined data model outperforms the models using only kinematic or audio data, highlighting the advantage of a multi-sensor approach in capturing a more comprehensive picture of machinery activities. The combination of kinematic and audio data allows for a more robust activity recognition system. While kinematic data is crucial for understanding movement patterns, audio data provides additional context, especially useful in complex or noisy environments.

The results suggest that the introduced sensory board integrating kinematic and audio sensors is preferable for more accurate and reliable activity recognition in construction machinery. This has significant implications for enhancing operational monitoring and safety in construction sites.

5.3 Distance-Based Analysis

This section analyzes the performance of machine learning models trained on data collected from different distances—30 feet, 60 feet, and 100 feet—from the transmitter. The performance metrics are considered for two classes: Class 0, representing the movement

of construction machinery (forward and backward), and Class 1, representing the arm raising/lowering activities. The results of the trained models are presented in Table 2.

Table 2 Comparative performance metrics of the machine learning model at different transmission distances (30 Feet, 60 Feet, and 100 Feet)

| | Average Training Accuracy | Average Test Accuracy | Precision | | Recall | | F1-Score | |
|----------|---------------------------|-----------------------|-----------|---------|---------|---------|----------|---------|
| | | | Class 0 | Class 1 | Class 0 | Class 1 | Class 0 | Class 1 |
| 30 Feet | 0.8808 | 0.8810 | 0.85 | 0.94 | 0.96 | 0.79 | 0.9 | 0.86 |
| 60 Feet | 0.8713 | 0.8709 | 0.84 | 0.91 | 0.92 | 0.82 | 0.88 | 0.86 |
| 100 Feet | 0.8483 | 0.8516 | 0.82 | 0.9 | 0.91 | 0.79 | 0.86 | 0.84 |

Results of the 30-foot collected data show that the overall accuracy is considerably high, showing a strong ability of the model to distinguish between the two activities with higher precision for arm activities. Results of the 60-foot collected data present a slight decrease in performance compared to the 30-foot distance, possibly due to increased transmission challenges, but the model still maintains robust classification capabilities. Finally, the results from 100-foot data show a noticeable reduction in model performance, highlighting the impact of increased distance on data transmission quality.

The results indicate a clear trend: as the distance between the transmitter and the receiver increases, the accuracy of the model decreases. This trend is observed in both classes but is more pronounced in Class 1 (arm activities), where precision and recall rates show a significant drop. This could be due to the more complex nature of arm movements compared to straightforward movement patterns.

It's also noteworthy that the receiver's requirement to be in the line of sight with the transmitter at all distances suggests that environmental factors and physical obstructions could play a significant role in data transmission quality. The degradation in model performance at longer distances is likely a result of these factors impacting the integrity of the transmitted signal.

These findings highlight the importance of considering the transmission range and environmental factors in the deployment of wireless sensory systems on construction sites. The results suggest a need for optimization in sensor placement and data transmission methods to enhance the system's effectiveness, especially over longer distances.

6 Discussion

The study's findings provide insights into developing and implementing a wireless hardware system for monitoring construction machinery activities. This system's primary objective was to automate audio and kinematic data collection and transmission at construction sites, aiming to enhance job site monitoring for various applications.

The results achieved through this study align well with the stated objectives. The wireless sensory system, which integrated audio and kinematic data collection, showed promising results in accurately recognizing and monitoring construction machinery activities. The data collected at different distances (30, 60, and 100 feet) provided crucial insights into the performance and reliability of the sensors under varied operational ranges typically found at construction sites.

The machine learning model that combined kinematic and audio data demonstrated higher accuracy in activity recognition than models using single data types. This confirms that a multi-modal approach can more effectively capture a comprehensive picture of machinery activities, which is aligned with the results of Ref. [25]. Furthermore, the decrease in model accuracy with increasing distance highlights the challenges of maintaining data integrity over longer transmission ranges, an essential consideration for practical applications.

One of the primary strengths of this study lies in integrating audio and kinematic data for activity recognition. The use of machine learning models further enhances the system's ability to process and interpret complex data sets effectively. Additionally, the real-world testing of the system at a construction site adds practical validity to the research findings.

However, the study also has its limitations. The range limitation of the wireless system, particularly evident at 100 feet, poses a challenge for larger construction sites. Furthermore, environmental factors such as noise, physical obstructions, and varying weather conditions, which can significantly impact the quality of data transmission and sensor performance, were not fully explored in this study. The reliance on specific hardware (ARM microcontroller-based boards) also limits the generalizability of the findings.

The results of this study contribute to the growing body of literature on automated monitoring systems in construction. Previous research has primarily focused on either audio or kinematic data for activity recognition. Integrating both data types in the sensory board and using machine learning for analysis distinguishes this study from existing work.

The challenges identified in maintaining data integrity over longer transmission ranges and the impact of environmental factors agree with findings from similar studies in the field. For instance, Ref. [7, 25] have highlighted the challenges in data collection and accuracy in dynamic construction environments. This study's findings build upon these existing studies, offering new insights into the potential of multi-modal sensory systems in overcoming some of these challenges.

7 Conclusion

This study successfully developed a wireless sensory system for the automated collection and transmission of audio and kinematic data from construction machinery. The system, tested at various distances, demonstrated high accuracy and reliability in data collection and processing, offering significant improvements over traditional monitoring methods.

The practical implications of this research are substantial. The system's ability to provide real-time, accurate data about machinery activities can greatly improve the

efficiency and safety of construction sites. Theoretically, this research contributes to the field of construction technology by demonstrating the feasibility and effectiveness of sensors that integrate audio and kinematic data for automated monitoring.

Future research could focus on enhancing the range and reliability of wireless transmission, especially in large and complex construction environments. Investigating the impact of various environmental factors on data transmission quality and exploring the potential for integrating additional types of sensors could also be valuable. Moreover, adapting this technology for different types of construction machinery and exploring its application in other areas of construction management would be beneficial.

Acknowledgements. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not reflect the views of the funding agency.

References

1. Pariafsai, F., Behzadan, A.H.: Core competencies for construction project management: literature review and content analysis. *J. Civil Eng. Educ.* **147**(4), 04021010 (2021)
2. Heydari, M., et al.: A dynamic model to assess the role of site supervision systems in the safety performance of construction projects. *J. Constr. Eng. Manag.* **150**(3), 04024001 (2024)
3. Golazad, S., et al.: Integrating GIS, agent-based, and discrete event simulation to evaluate patient distribution policies for enhancing urban healthcare access network resilience. *Sustain. Cities Soc.* **111**, 105559 (2024)
4. Mohammadi, A., Ramezanianpour, A.M.: Investigating the environmental and economic impacts of using supplementary cementitious materials (SCMs) using the life cycle approach. *J. Build. Eng.* **79**, 107934 (2023)
5. Lee, Y.-C., et al.: Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents. *Autom. Constr.* **113**, 103127 (2020)
6. Golazad, S., Heravi, G., AminShokravi, A.: People's post-disaster decisions and their relations with personality traits and decision-making styles: the choice of travel mode and healthcare destination. *Int. J. Disaster Risk Reduct.* **83**, 103395 (2022)
7. Cheng, C.-F., et al.: Activity analysis of construction equipment using audio signals and support vector machines. *Autom. Constr.* **81**, 240–253 (2017)
8. Ray, S.J., Teizer, J.: Real-time construction worker posture analysis for ergonomics training. *Adv. Eng. Inform.* **26**(2), 439–455 (2012)
9. Jiang, Y., He, X.: Overview of applications of the sensor technologies for construction machinery. *IEEE Access* **8**, 110335 (2020)
10. Moselhi, O., Bardareh, H., Zhu, Z.: Automated data acquisition in construction with remote sensing technologies. *Appl. Sci.* **10**(8), 2846 (2020)
11. Nakanishi, Y., Kaneta, T., Nishino, S.: A review of monitoring construction equipment in support of construction project management. *Front. Built Environ.* **7** (2022)
12. Talmaki, S.A., Kamat, V.R.: Sensor acquisition and allocation for real-time monitoring of articulated construction equipment in digital twins. *Sensors* **22**(19), 7635 (2022)
13. Xie, Y., et al.: Construction data-driven dynamic sound data training and hardware requirements for autonomous audio-based site monitoring. In: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. IAARC Publications (2019)

14. Cheng, C.-F., et al.: Evaluation of software and hardware settings for audio-based analysis of construction operations. *Int. J. Civil Eng.* **17**, 1469–1480 (2019)
15. Sherafat, B., Rashidi, A., Asgari, S.: Sound-based multiple-equipment activity recognition using convolutional neural networks. *Autom. Constr.* **135**, 104104 (2022)
16. Sabillon, C., et al.: Audio-based bayesian model for productivity estimation of cyclic construction activities. *J. Comput. Civil Eng.* **34**(1), 04019048 (2020)
17. Sanhudo, L., et al.: Activity classification using accelerometers and machine learning for complex construction worker activities. *J. Build. Eng.* **35**, 102001 (2021)
18. Langroodi, A.K., Vahdatikhaki, F., Doree, A.: Activity recognition of construction equipment using fractional random forest. *Autom. Constr.* **122**, 103465 (2021)
19. Sherafat, B., et al.: Automated methods for activity recognition of construction workers and equipment: state-of-the-art review. *J. Constr. Eng. Manag.* **146**(6), 03120002 (2020)
20. Akhavian, R., Behzadan, A.H.: Smartphone-based construction workers' activity recognition and classification. *Autom. Constr.* **71**, 198–209 (2016)
21. Akhavian, R., Behzadan, A.H.: Construction activity recognition for simulation input modeling using machine learning classifiers. In: Proceedings of the Winter Simulation Conference 2014. IEEE (2014)
22. Bangaru, S.S., et al.: ANN-based automated scaffold builder activity recognition through wearable EMG and IMU sensors. *Autom. Constr.* **126**, 103653 (2021)
23. Campbell, B., et al.: Gear noise reduction of an automatic transmission through finite element dynamic simulation. *SAE Trans.* 2883–2893 (1997)
24. Cheng, C.-F., et al.: Audio signal processing for activity recognition of construction heavy equipment. In: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. IAARC Publications (2016)
25. Sherafat, B., et al.: A hybrid kinematic-acoustic system for automated activity detection of construction equipment. *Sensors* **19**(19), 4286 (2019)
26. Bayat, A., Pomplun, M., Tran, D.A.: A study on human activity recognition using accelerometer data from smartphones. *Procedia Comput. Sci.* **34**, 450–457 (2014)



Interactive Light and Sound-Sensitive Mortar for Architectural Environments

Karina Hwang Arcolezi¹, Claudiane Ouellet-Plamondon¹(✉), and Bora Ung²

¹ Department of Construction Engineering, École de Technologie Supérieure, Montreal,
QC H3C 1K3, Canada

cclaudiane.ouellet-plamondon@etsmtl.ca

² Department of Electrical Engineering, École de Technologie Supérieure, Montreal, QC H3C
1K3, Canada

bora.ung@etsmtl.ca

Abstract. Integrating technological advancements in construction materials, such as smart devices, presents an opportunity to enhance the physical environment. This study explores the design of mortars incorporating plastic optical fibres (POF), LEDs, and sound sensors, to enhance architectural interactivity and user experiences. The mix design of the mortars, aimed to reduce Portland cement content, includes 150 kg/m³ of Portland cement for general use, 100 kg/m³ of calcined clay, and 400 kg/m³ of limestone filler. Compressive strength tests compared reference samples without optical fibres, to samples with varying numbers (1, 4, and 9) of transverse POF. Results showed approximately 1% and 5% increase in compressive strength in samples with 1 and 4 POF compared to the reference, respectively, with a decrease observed at 9 fibres, suggesting optimal performance was obtained with 4 POF under these conditions. Additionally, the material's interactivity was showcased through POFs connected to visible LEDs, enabling dynamic colour changes based on the sound amplitude measured by acoustic sensors. This interaction enhances the aesthetic dimension of architectural elements and introduces the potential for innovative applications in architectural design; for example, adaptability to different sound environments, customizable programming, along with artistic and aesthetic possibilities. The material's contribution to reduced cement consumption and environmental awareness, also enriches the comprehensive exploration of this interactive construction element.

Keywords: Interactive Construction Materials · Sound Sensor · Material Innovation

1 Introduction

As advancements in internet technology, virtual reality, and smart devices converge, there is a growing opportunity to enhance our physical environment. This transformation involves exploring new materials and technologies that can dynamically enrich spaces and respond intelligently to evolving needs [1]. In addition to their dynamic behaviour and ability to respond to external stimuli, smart materials could enable the creation of

a sensible environment based on individual activities and physical characteristics. This external stimulation can be considered as an impulse input to the material system [1, 2]. Cementitious materials, such as concrete and mortars, have long been renowned for their durability, strength, and cost efficiency. However, traditional concrete can revolutionize the field, becoming a smart material that truly appeals to aesthetics and architecture by incorporating advanced functionalities to become capable of interacting, transmitting light, and performing functions [3, 4].

A promising field of research is the development of cementitious materials with optical fibre, which offers a combination of aesthetics and functionality. This technology can improve the conventional use of concrete in architecture and design, unlocking new possibilities in lighting design, energy efficiency, artistic expression, and user interactivity [5, 6]. While significant research has been conducted on smart concrete over the last few decades, the development of smart and translucent concrete is a relatively new area that combines the durability of traditional cementitious materials with the highly efficient light-transmitting properties of optical fibres [3, 4] resulting in enhanced capabilities and versatility in various engineering fields [5, 7]. Light-transmitting concrete, made from standard concrete components along with translucent materials like optical fibres, offers a unique combination of aesthetics and functionality. These optical fibres, constituting approximately 4–6% by volume, transmit light through the material, presenting an innovative approach to lighting design and energy efficiency [8, 9]. However, the incorporation of optical fibres in concrete typically comes with a tradeoff. Optimal mechanical performance is observed at low fibre ratio content, while higher ratios may be suitable for energy-saving purposes [10].

As our lives and homes become increasingly dependent on processing digital data, it is important to develop novel architectural environments that can “sense” temperature and mechanical changes, “hear”, “communicate” and interact with users, for example. In this context, this work aims to partly address the following broad research question: How can we make normally passive and opaque construction materials such as mortar and cement more translucent, luminous, interactive and “smart”. And more specifically: How can the integration of light-transmitting mortar with sound sensors contribute to the creation of an adaptive lighting system that dynamically adjusts based on ambient sound amplitude? The approach taken in this work to answer this question is to develop an innovative mortar that incorporates plastic optical fibres capable of transmitting light and is connected to sound sensors. Potential applications of this research includes the development of interactive concrete for responsive urban and interior design, road safety, public art, and smart lighting.

2 Materials and Methods

2.1 Materials

The materials used in this study included Portland cement (Type GU, Ciment Québec, QC, Canada) under CSA A3000-18 standards [11], with a density of 3.15 g/cm^3 . The binder materials comprised calcined clay and limestone filler (Pulverized limestone, Graymont, QC, Canada), while local natural sand (BOMIX). To improve the workability and flow of the mortar, a superplasticizer, polycarboxylate MasterGlenium, was added.

Polymethylmethacrylate (PMMA) optical fibres with a diameter of 1 mm were selected to create light-transmitting sound-based interactive mortars, owing to their excellent visible light transmittance properties. Table 1 provides an overview of the physical properties of the PMMA optical fibres.

Table 1. Specifications of PMMA optical fibre properties.

| | |
|-----------------------|---------------|
| Core refractive index | 1.49 |
| Numerical aperture | 0.5 |
| Operating temperature | -55°C to 70°C |
| Attenuation loss | ≤200 dB/km |
| Tensile strength | 65 N |

In addition to the PMMA optical fibres, the experimental setup incorporated essential electronic components, including RGB LEDs for light transmission and Arduino microcontrollers for precise control. These components formed the integral infrastructure of the experimental setup, facilitating investigations and accurate measurements of the mortar's responses to sound stimuli and environmental variations.

2.2 Experimental Procedure

The mortar mixes were prepared and tested according to ASTM C305 [12] and ASTM C109 [13] standards, respectively, with a water binder ratio of 0.384. The mix design of the mortars is presented in Table 2. After casting, the samples were demolded after 24 h and placed in a moisture room at 23 ± 2 °C with a relative humidity level of not less than 95%. Compressive strength tests were conducted on the samples at 28 days. For each batch, three samples were prepared for the compressive strength tests, and six additional samples were also produced for sound-based interactivity tests.

Table 2. Mix design.

| Materials
(kg/m ³) | Cement | Calcined Clay | Limestone Filler | Sand | Water | Superplasticizer |
|-----------------------------------|--------|---------------|------------------|------|-------|------------------|
| 150 | 100 | 400 | | 1400 | 250 | 12 |

Optical fibre distribution and compressive strength

Regarding the mechanical properties of the samples, this stage of the experimental plan is crucial for understanding how the mechanical properties of the mortars will be affected by different numbers of optical fibres placed transversally in the mortar samples. To this end, two types of samples were tested: 1) the reference samples of mortar, without POF, and 2) mortar samples with 1, 4, and 9 POF.

The distribution of the POFs in the mortar samples followed a square lattice symmetry (as shown in Fig. 1). This arrangement of the fibres ensured a consistent and repeating pattern, much like a checkerboard, with equal spacing and uniform orientation along both horizontal and vertical directions. Figure 1 illustrates the distribution of the POF pattern for varying numbers of POF.

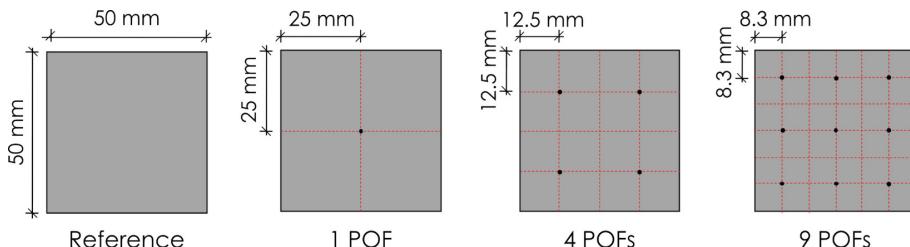


Fig. 1. Distribution of plastic (PMMA) optical fibres in mortar samples for different fibre quantities.

Flow test

The flow test was conducted according to ASTM C1437-20 standards [14], so as to assess the consistency of hydraulic cement mortar. In summary, the procedure involves demolding the fresh mortar on the standard table, which is dropped 25 times during the period of 15 s. The flow mix is determined by using a caliper in four different measurements scribed on the table. For this paper, the flowability aimed to reach at least 130% to ensure proper casting of the mortar. Achieving adequate flowability is crucial as it ensures that the mortar can effectively fill the space between the fibres, preventing voids or gaps in the interior structure of the material.

Sound sensors properties

The sound sensing properties were implemented via an Arduino microcontroller programmed to assign specific colour values to different ranges of sound amplitudes. The programming was determined to trigger the RGB LED with different colours according to the sound amplitude. Sound amplitudes between 50 and 80 would trigger the emission of the red colour, while amplitudes between 80 and 160 would prompt a green colour. Additionally, sound amplitudes above 160 would result in a blue colour, and below 50, the LED remained off. This setup allowed for dynamic colour changes in response to varying sound levels detected by the sensor, enhancing the interactive capabilities of the mortar samples.

For this experiment, a specific setup was utilized to assess the properties of the sound sensor and its integration with optical fibres. This setup involved the bundling of optical fibres and inserted into a metallic tube, which was then mated directly over the LED. The LED, along with the sound sensor, was connected to the Arduino microcontroller to enable synchronized operation and real-time response to sound stimuli (Fig. 2).

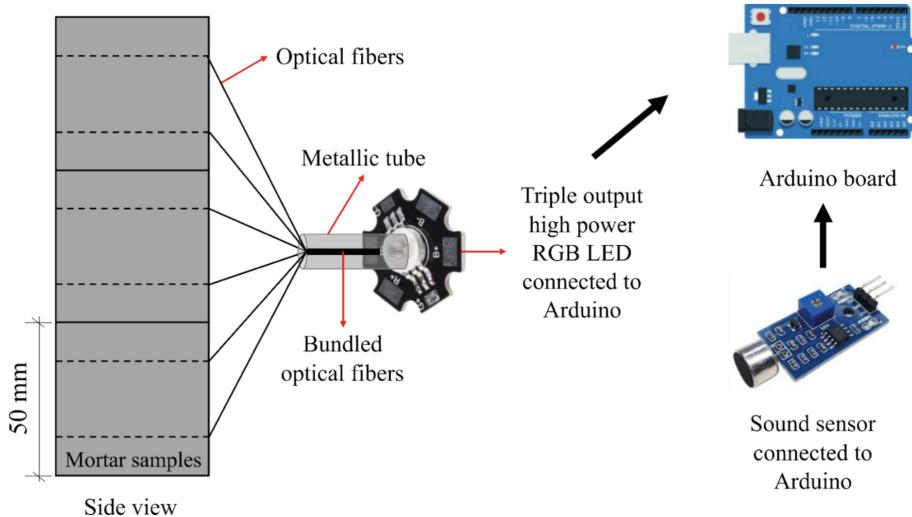


Fig. 2. Experimental setup for sound-responsive mortar.

3 Results and Discussions

3.1 Flow Test and Compressive Strength

The physical and mechanical properties of the mortar samples were evaluated to assess their suitability for application. The flow test results indicated that the mortar mixture achieved a flowability of 140%, ensuring proper casting and filling of voids within the material (Fig. 3).

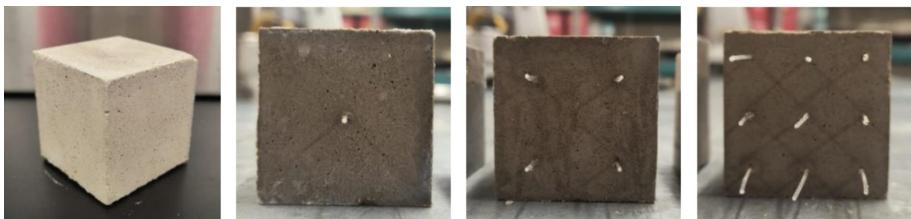


Fig. 3. Mortar samples with varying amounts of POFs.

The results of the compressive strength tests revealed that the addition of optical fibres affected the compressive strength of the mortar both positively and negatively. Figure 4 presents the compressive strength values for the samples. Samples with 1 and 4 POF exhibited approximately 1% and 5% increases, respectively, compared to the reference (22.9 MPa). On the other hand, a decrease occurred with 9 POFs, resulting in 20.1 MPa (Fig. 4). It is noticeable from the results that the compressive strength of the mortar samples can vary according to the number of POFs in its composition.

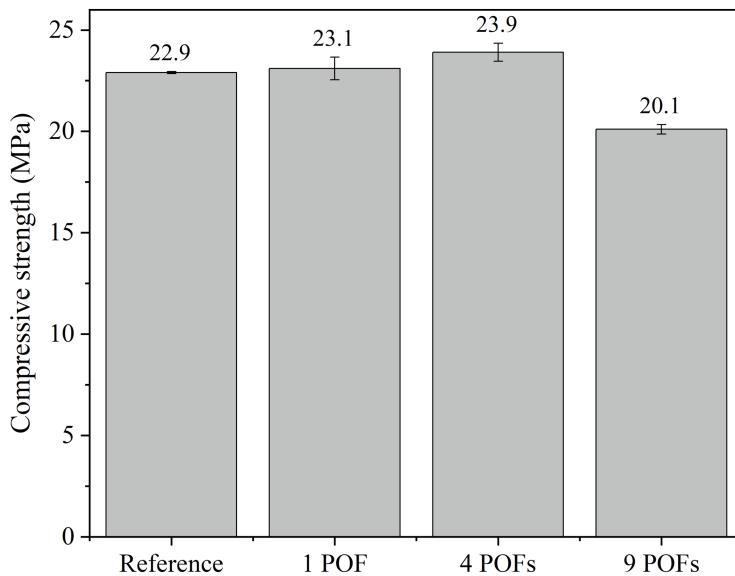


Fig. 4. 28-day compressive strength.

A particularly noticeable decrease was observed in the sample containing 9 POFs, which decreased by approximately 12% compared to the reference sample and almost 16% compared to the sample with 4 POFs, which exhibited the highest average compressive strength at 23.9 MPa. This finding highlights the importance of conducting tests with different fibre quantities to demonstrate their optimal inclusion levels. While the addition of POFs holds promise for enhancing the properties of the mortar, these results highlight the need for careful consideration and optimization to ensure that the desired mechanical performance is achieved [10]. Under test conditions, samples with 4 POF (23.9 MPa) demonstrated optimal compression strength, providing crucial practical insights for future applications.

Furthermore, it is noteworthy that the results concerning compressive strength are even more significant when considering the low cement content. Despite a reduction in cement content, the samples demonstrated strengths exceeding 20 MPa. This aspect not only underscores the technical feasibility but also the potential for reducing environmental impact associated with cement production.

3.2 Sound-Based Interactivity

The integration of sound sensors with the mortar samples allowed for real-time responsiveness to environmental stimuli. Through programming, the LED connected to the mortar's POFs changed colour based on varying sound amplitudes, showcasing its interactive capabilities. In the experiments, the sound amplitudes were monitored in real time using sound sensors. The resulting graph (Fig. 5) depicts the variation in sound amplitude during a specified period, illustrating the mortar's real-time responsiveness to environmental sound stimuli, in this case, music. Based on these sound amplitudes, the

RGB LED embedded within the mortar emitted different colors or remained off, as programmed. This visual representation demonstrates the material's ability to dynamically respond to changes in its surroundings, enhancing its interactive functionality.

Moreover, the optical fibres embedded within the mortar further amplify its interactive capabilities. These fibres, connected to the LED that transmits light, enable visual feedback based on sound amplitude. As the sound level changes, the optical fibres emit light of varying colours, providing a direct and tangible representation of the environmental stimuli. This integration of sound sensors, optical fibres, and LED technology not only enhances the aesthetic appeal of the mortar but also adds a layer of interactivity, making it a versatile and engaging material for architectural applications (Fig. 6).

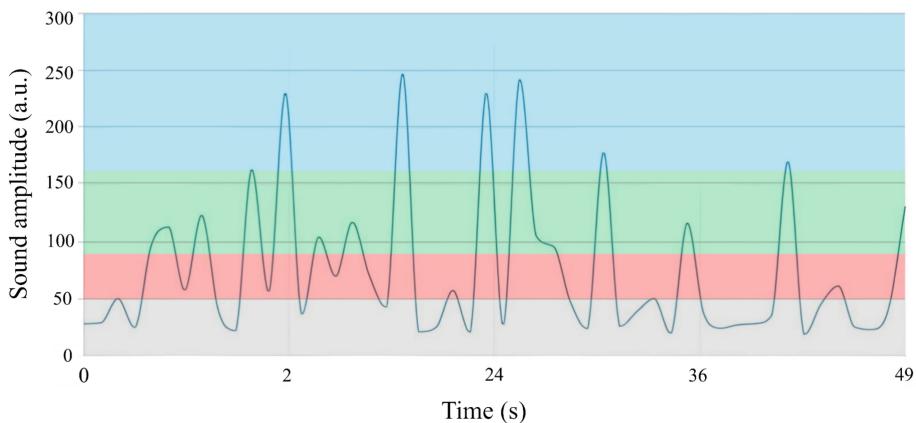


Fig. 5. LED colour variation by amplitude range of a sound amplitude curve over time.

However, despite these promising results, the practical application of this technology in architectural and design contexts remains unclear. Further research and development are needed to explore and optimize the integration of sound-based interactive mortar in real-world scenarios. Nonetheless, the findings suggest potential applications in various fields, including urban signage, architecture, interior design, road safety, urban furniture design, public art, and smart home lighting. These examples demonstrate how optical fibres, sensors, and smart materials in interactive concrete can contribute to redefining how people interact with their environment, offering dynamic and personalized experiences.

4 Conclusions

The findings presented in this study underscore the innovative potential of interactive concrete, which integrates optical fibres, LEDs, and sound sensors. Beyond influencing mechanical properties, the incorporation of optical fibres in the mortars, along with their connection to sound sensors and LEDs, enables significant interactivity. This not only enhances aesthetic dimensions but also offers practical opportunities for innovative applications in architectural design.

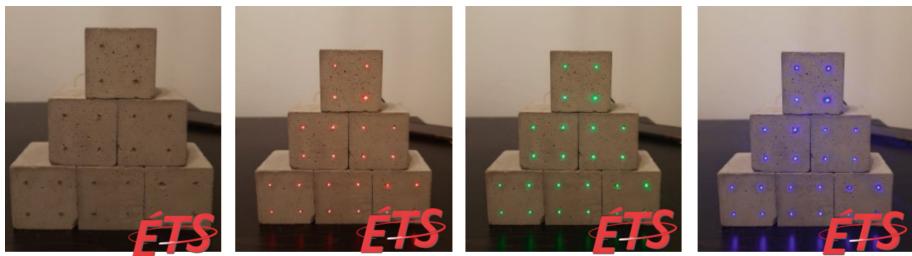


Fig. 6. Mortar samples interacting to sound with colour changes.

In addition to the examples mentioned earlier, the integration of interactive concrete in public infrastructure, such as bridges and tunnels, can enhance safety measures by providing real-time feedback on structural integrity and environmental conditions. Similarly, in urban planning, interactive concrete can be used to create responsive streetscapes that adapt to pedestrian and vehicular traffic patterns, improving both functionality and aesthetics.

Furthermore, the development of interactive concrete opens up new avenues for sustainable design practices. By incorporating energy-efficient lighting systems and responsive elements, buildings, and public spaces can reduce energy consumption and promote environmental stewardship. Additionally, the use of locally sourced materials and modular construction techniques can further enhance the sustainability of interactive concrete projects.

While compressive strength is a fundamental aspect, this paper's exploration also highlights the material's ability to interact through visible light and sound sensors, contributing to the ongoing development and applications of interactive concrete. Moreover, the interdisciplinary nature of interactive concrete research fosters collaboration between fields. This collaborative approach not only enriches the design process but also fosters a culture of innovation and experimentation within the built environment sector. Overall, the study highlights the transformative potential of interactive concrete in shaping the future of architectural design and urban development. By reimagining the relationship between materials, technology, and human interaction, interactive concrete offers a glimpse into a more dynamic, responsive, and sustainable built environment.

Acknowledgements. The authors would like to thank the support of the program FRQNT Art Prisme.

References

1. Marin, P., Philippe, L., Blanchi, Y.: Interactive concrete surface: an exploration of smart materials (2014). https://doi.org/10.5176/2301-394X_ACE14.38
2. Peiris, R. L., Nakatsu, R.: TempTouch: a novel touch sensor using temperature controllers for surface based textile displays. In: Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS 2013), pp. 105–114. ACM, New York, NY, USA (2013). <https://doi.org/10.1145/2512349.2512813>

3. Chiew, S.M., Ibrahim, I.S., Mohd Ariffin, M.A., Lee, H.S., Singh, J.K.: Development and properties of light-transmitting concrete (LTC)—a review. *J Clean Prod* **284** (2021) <https://doi.org/10.1016/j.jclepro.2020.124780>
4. Luhar, I., Luhar, S., Savva, P., Theodosiou, A., Petrou, M.F., Nicolaides, D.: Light transmitting concrete: a review. *Buildings* **11**(10), 1–18 (2021). <https://doi.org/10.3390/buildings1100480>
5. Larsen, L.K.: Feedforward and feedback modalities for embedded touch interaction with hi-fi systems. Master thesis, Aarhus University (2015)
6. Su, X., Zhang, L., Luo, Y., Liu, Z.: Energy & buildings an energy analysis of translucent concrete embedded with inclined optical fibres. *Energy Build* **273**, 112409 (2022). <https://doi.org/10.1016/j.enbuild.2022.112409>
7. D'Alessandro, A., Birgin, H.B., Cerni, G., Ubertini, F.: Smart infrastructure monitoring through self-sensing composite sensors and systems: a study on smart concrete sensors with varying carbon-based filler. In: Proceedings of the 2022 International Conference on Infrastructures (INFRA 2022), pp. 1–15. Springer, Berlin, Germany (2022). <https://doi.org/10.3390/infrastructures7040048>
8. Said, S.H.: State-of-the-art developments in light transmitting concrete. *Mater Today Proc.* **33**, 1967–1973 (2020). <https://doi.org/10.1016/j.matpr.2020.06.128>
9. Palanisamy, C., Krishnaswami, N., Velusamy, S., Kumar, Krishnamurthy, H., Velmurugan, H., Kumaar, Udhayakumar, H.: Transparent concrete by using optical fibre. *Mater Today Proc.* **65**, 1774–1778 (2022). <https://doi.org/10.1016/j.matpr.2022.04.799>.
10. Elgheznawy, D., Eltarabily, S.: A review of translucent concrete as a new innovative material in architecture. *Civil Eng. Architect.* **8**(4), 571–579 (2020). <https://doi.org/10.13189/cea.2020.080421>
11. CSA A3000-18. Cementitious materials compendium. CSA 9781488313233
12. ASTM International.: C305—14 standard practice for mechanical mixing of hydraulic cement pastes and mortars. ASTM Standard B. 14–16 (2015)
13. ASTM International.: C109/C109M—16a standard test method for compressive strength of hydraulic cement mortars using 2-in. or [50-mm] Cube Specimens). ASTM Stand. B. 1–9 (2016)
14. ASTM International.: C1437-20—Standard Test Method for Flow of Hydraulic Cement Mortar. ASTM (2020)



Automated FEM-Based Determination of Thermal Diffusivity and Temperature Profile for Rockfall Prediction

Jaka Dujc¹(✉) and Mateja Jemec Auflič²

¹ Faculty of Civil and Geodetic Engineering Ljubljana, University of Ljubljana, Ljubljana, Slovenia

jdujc@fgg.uni-lj.si

² Geological Survey of Slovenia, Ljubljana, Slovenia

Abstract. A novel method for the determination of thermal diffusivity and the generation of a continuous temperature profile of a rock formation based on temperature measurements at several depths is presented. The methodology combines finite element modeling and optimization techniques to accurately simulate the heat transfer processes in the rock. The optimization procedure is divided into three successive steps using three different finite element models. The first model uses optimization to calculate the thermal diffusivity by minimizing the difference between simulated and measured temperatures. The second model uses optimization to determine the unknown surface temperatures of the rock. The third model uses finite element analysis to create a continuous temperature profile over the entire rock depth. It uses the optimized thermal diffusivity and surface temperatures determined by the first two models. The optimization procedures use gradient-based methods and consider sensitivity analysis to determine the derivatives of the objective functions. The methodology is able to accurately determine the thermal diffusivity and create a continuous temperature profile even if the measured data has inaccuracies.

Keywords: Finite Element Method (FEM) · Optimization · Thermal Diffusivity · Temperature Profile · Rockfall

1 Introduction

Rockfall, the abrupt detachment and downward movement of rock fragments from a cliff face, poses a significant threat to infrastructure and human life. These events are often triggered by a complex interplay of geological factors and meteorological conditions, in particular freeze-thaw cycles. Freeze-thaw cycles, where water freezes and expands in cracks and pores, can place significant stress on the rock, leading to cracking and weakening. Understanding the thermomechanical processes that affect rock faces is crucial for developing strategies to mitigate rockfall hazards.

An important parameter in the study of heat transfer in rock is thermal diffusivity. Accurate knowledge of thermal diffusivity is an essential prerequisite for the development of reliable models that can predict the effects of freeze-thaw cycles on the stability of rock faces.

This paper presents a novel automated method for determining the thermal diffusivity and temperature profiles in rocks. The method uses a finite element model (FEM) of heat transfer in the rock formation, coupled with in-situ temperature measurements and an optimization algorithm. The optimization process minimizes the difference between simulated and measured temperatures to determine the optimal thermal diffusivity and temperature profile.

The proposed method has been successfully applied at several sites in Slovenia and has been shown to be able to accurately characterize the thermal diffusivity and temperature profiles in rocks in different geological environments.

The paper is organized as follows: Sect. 2 gives a brief overview of rock temperature measurements in the field. Section 3 deals with the finer details of the finite element model. Section 4 describes the optimization procedure used to obtain the thermal diffusivity and temperature profile and the corresponding results. The last section contains concluding remarks.

2 On-Site Rock Temperature Measurements

As part of the ongoing research project “Deciphering the sensitivity of rock faces to climatic changes and freeze-thaw cycles in permafrost-free regions” (J1-3024), extensive field observations are being conducted at five sites in Slovenia. Using a variety of sensors and instruments, data on precipitation, air temperature and humidity, rock temperature, and rock displacement and strain is recorded several times per hour and transmitted to a central database via cellular network.

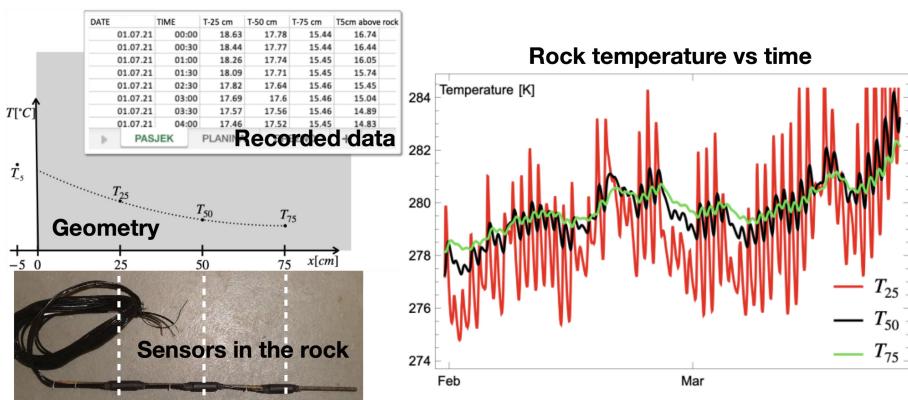


Fig. 1. A schematic of the rock temperature measurements (left) and the rock temperature versus time curves for the three depths (right).

For our study, the rock temperature measurements are of primary interest in terms of freeze-thaw cycles. Temperature measurements are taken approximately every 10 min at depths of 25, 50, 75, and 5 cm above the rock surface. Figure 1 shows a schematic of the rock temperature measurements by depth, accompanied by a small portion of the observed temperature data represented by temperature versus time curves.

3 Finite Element Model of the Heat Transfer in Rock

The collected rock temperature data is invaluable for comprehending the impact of freeze-thaw cycles. However, due to the constraints of the measuring instruments, temperature readings are only available at three discrete points (depths) within the rock and one point above the surface. Therefore, we use numerical simulations to extrapolate from the discrete temperature measurements and create a continuous temperature profile.

Assuming heat transfer occurs primarily along the depth of the rock, we can represent the rock formation (half-space) using a one-dimensional (1D) model, as shown in Fig. 2. The heat transfer throughout the rock formation is represented by the following strong form equilibrium equation:

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = 0, \quad \alpha = \frac{k}{\rho c_p} \quad (1)$$

where T stands for temperature, t indicates time, x denotes the spatial coordinate (depth), and α represents thermal diffusivity defined as the ratio of thermal conductivity k to density ρ and specific heat capacity c_p . The equilibrium Eq. (1), coupled with the relevant boundary conditions, which will be further discussed in Sect. 4, is transformed into its weak form (see e.g. [1]) and resolved using the finite element method.

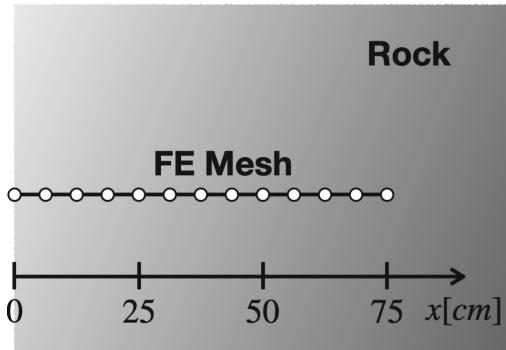


Fig. 2. The rock formation represented by the finite element mesh from the rock surface to the deepest temperature measurement point.

It is worth highlighting that we utilize Kelvin (K) throughout the simulation process instead of degrees Celsius (°C). This approach contributes to stabilizing the solution and circumventing potential issues related to temperature sign changes.

4 Optimization Procedure

The optimization procedure aims to determine (i) the thermal diffusivity and (ii) the entire temperature profile of the rock, which depends on the unknown rock surface temperature. This can be achieved either by formulating a single, comprehensive optimization problem or by breaking it down into several sequential steps. In the sequential approach, the thermal diffusivity is first estimated, followed by an independent optimization procedure to determine the rock surface temperature and the temperature profile. We opted for the sequential strategy as it exhibits superior convergence properties due to the smaller number of unknown design parameters involved in each optimization step.

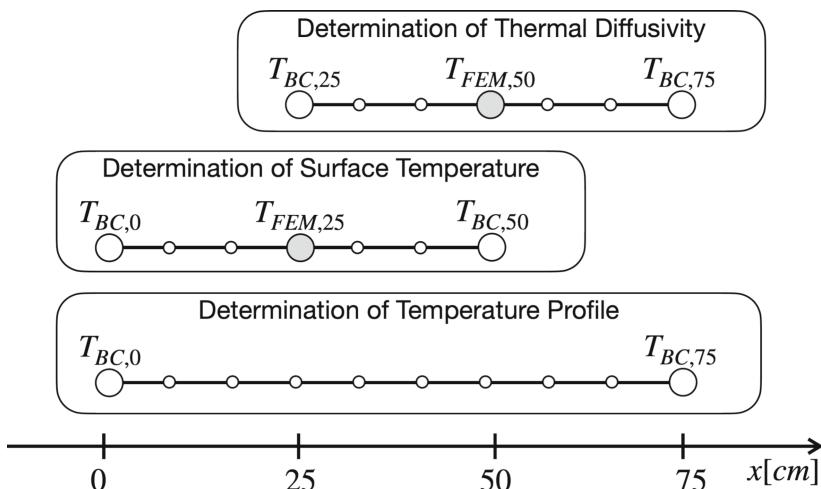


Fig. 3. The finite element geometries and boundary conditions of the models used to determine the thermal diffusivity (above), the surface temperature (middle), and the continuous temperature profile (below).

Our study employed three distinct finite element models:

1. First model (Sect. 4.2) determines the thermal diffusivity,
2. Second model (Sect. 4.3) obtains the unknown rock surface temperatures,
3. Third model (Sect. 4.4) establishes the continuous temperature profile.

Figure 3 illustrates the geometries and boundary conditions for each of these finite element models. The first two models combine optimization with finite element analysis, while the third model employs only finite element analysis.

4.1 Optimization Procedure Outline

The optimization procedures employed in Sects. 4.2 and 4.3 involve identifying a vector of design variables ϕ that minimize the specified objective function $\Phi(\phi)$. These design variables represent the unknown parameters of the system, such as the thermal diffusivity

or the rock surface temperatures. By iteratively modifying these values, the optimization algorithm attempts to find the combination that leads to the minimal value of the objective function.

The applied optimization procedures can be summarized as follows (see e.g. [2, 3]):

1. **Initialization:** set the optimization iteration counter to $l = 0$ and choose the initial values of the design parameters $\phi^{(0)}$.
2. **Finite element analysis:** employ the AceFEM finite element environment [4] to perform the finite element analysis and compute the nodal temperatures $\mathbf{T}^{(l)}$ for the chosen set of measurements.
3. **Objective function computation:** utilize the results of the finite element analysis to compute the objective function $\Phi(\phi^{(l)})$.
4. **Sensitivity analysis:** perform sensitivity analysis to compute the derivatives $\left(\frac{\partial \Phi}{\partial \phi}\right)^{(l)}$, which also includes computation of the derivatives $\left(\frac{\partial \mathbf{T}}{\partial \phi}\right)^{(l)}$. The AceGen [4] symbolic code generator is used to derive analytical expressions for these derivatives [5].
5. **Gradient-based optimization:** employ Mathematica's [6] *FindMinimum* gradient optimization method to compute an update of the design parameters $\Delta\phi^{(l)}$.
6. **Updating design parameters:** update the design parameters $\phi^{(l+1)} = \phi^{(l)} + \Delta\phi^{(l)}$.
7. **Convergence check:** evaluate the convergence criteria for the optimization loop. If the criteria are met, terminate the optimization procedure. If not, increment the optimization iteration counter $l = l + 1$ and return to step 2.

4.2 Determination of Thermal Diffusivity

The objective of the first model is to determine the thermal diffusivity of the rock material. The model (see Fig. 3 above) employed Dirichlet boundary conditions at depths of 25 cm and 75 cm to simulate heat transfer within the rock formation between these depths. Notably, the applied temperature values were derived from actual measurements

$$T_{BC,25}(t_m) = T_{m,25}(t_m), T_{BC,75}(t_m) = T_{m,75}(t_m), \quad (2)$$

with a specific emphasis on the temperature at $x = 50$ cm within the model's center to evaluate the alignment between the measurement data and simulation results.

The optimization problem in this context is defined as: determining the value of the design variable α that minimizes the difference between the measured and simulated temperatures. Mathematically, this involves finding a vector of design variables $\phi = [\alpha]$ that minimizes the objective function

$$\Phi(\phi) = \sum_{t_m} (T_{FEM,50}(\phi, t_m) - \bar{T}_{m,50}(\phi, t_m))^2, \quad (3)$$

where $T_{FEM,50}$ and $\bar{T}_{m,50}$ represent the simulated and the measured temperatures at $x = 50$ cm and a particular measurement time $t_m \in [t_1, t_2, \dots, t_N]$, where N is the total number of simulated measurements.

It's important to note that $\bar{T}_{m,50}$ is also dependent on the vector of the design variables ϕ . In our study, we observed that the agreement between the measurement and simulation

was rather poor when neglecting measurement inaccuracies. These inaccuracies arise from various sources, such as the accuracy of the thermal sensors, the accuracy of sensor placement, and the quality of the bond between the sensor and the rock material. Therefore, we extended the vector of design variables

$$\phi = [\alpha, \Delta T], \quad (4)$$

to include a measurement inaccuracy lump parameter ΔT , which compensates for the deviations in the measured temperatures at $x = 50$ cm

$$\bar{T}_{m,50}(\phi, t_m) = T_{m,50}(t_m) + \Delta T. \quad (5)$$

An example of the results is presented in Fig. 4, which shows the measured and the simulated temperatures versus time curves for the measurement site Planina. Note the two contrasting behaviors of the temperatures at a depth of 50 cm. When the thermal diffusivity is the only design parameter (blue line), the curve shows very little daily variation and only follows the general trend of temperature increase. If the design parameter ΔT is also taken into account in the optimization procedure (green line), a curve is obtained that is very similar to the black (measured) curve, except that it is shifted by an offset of $\Delta T \approx 0.5$ K above the black curve.

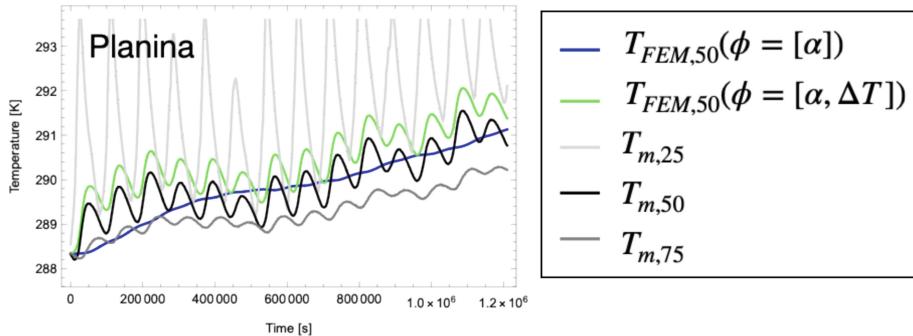


Fig. 4. The measured and the simulated temperatures versus time curves for the measurement site Planina.

4.3 Determination of the Surface Temperature

The objective of the second model is to determine the surface temperature. The model (see Fig. 3 middle) employed Dirichlet boundary conditions at depths of 0 cm and 50 cm. The applied values at the rock's surface $T_{BC,0}(t_m)$ are unknown and are determined with the optimization procedure while the applied values at $x = 50$ cm are the simulation values obtained with the optimal value of parameter α in Sect. 4.2

$$T_{BC,50}(t_m) = T_{FEM,50}(t_m)|_{\text{section 4.2}}. \quad (6)$$

Note that $T_{BC,50}(t_m)$ indirectly considers the deviations in measurements.

The optimization problem is in this context defined as: determine the value of the design variable vector

$$\phi = [T_{BC,0}(t_1), T_{BC,0}(t_2), \dots, T_{BC,0}(t_N)], \quad (7)$$

that minimizes the objective function

$$\Phi(\phi) = \sum_{t_m} (T_{FEM,25}(\phi, t_m) - T_{m,25}(t_m))^2. \quad (8)$$

It is noteworthy that the temperature measurements taken 5 cm above the rock were utilized to initialize the design variable vector $\phi^{(0)}$. This strategy provided a starting point for the optimization algorithm, ensuring that the initial values were not arbitrary.

4.4 Determination of the Temperature Profile

The objective of the third model is to determine the complete temperature profile. The model (see Fig. 3 below) utilized Dirichlet boundary conditions at depths of 0 cm and 75 cm. The applied values at the rock's surface were obtained from the optimization procedure described in Sect. 4.3

$$T_{BC,0}(t_m) = \phi|_{\text{Section 4.3}}. \quad (9)$$

Meanwhile, the values at 75 cm were taken directly from the temperature measurements

$$T_{BC,75}(t_m) = T_{m,75}(t_m). \quad (10)$$

Note that in this case, we are not dealing with an optimization problem but rather with a single finite element simulation.

5 Conclusions

This study presents a novel method for determining thermal diffusivity and creating a continuous temperature profile of a rock formation based on temperature measurements at multiple depths. The methodology uses a combination of finite element modeling and optimization techniques to accurately simulate the heat transfer processes in the rock. The optimization procedure is divided into three successive steps using three different finite element models. The first model uses optimization to calculate the thermal diffusivity by minimizing the difference between simulated and measured temperatures. The second model uses optimization to determine the unknown surface temperatures of the rock. The optimization procedures use gradient-based methods and consider sensitivity analysis to determine the derivatives of the objective functions. The third model uses finite element analysis to generate a continuous temperature profile over the entire rock depth. The optimized thermal diffusivity and surface temperatures from the two previous models are used. Further results and conclusions will be presented in an upcoming conference presentation.

Acknowledgments. The authors gratefully acknowledge the financial support provided by the Slovenian Research Agency through Research Project J1-3024 and the ARIS research and infrastructure program P2-0210.

References

1. Ibrahimbegovic, A.: Nonlinear Solid Mechanics: Theoretical Formulations and Finite Element Solution Method, vol. 160. Springer Science & Business Media (2009)
2. Stanić, A., Hudobivnik, B., Brank, B.: Economic-design optimization of cross laminated timber plates with ribs. *Compos. Struct.* **154**, 527–537 (2016)
3. Kegl, M., Brank, B.: Shape optimization of truss-stiffened shell structures with variable thickness. *Comput. Methods Appl. Mech. Eng.* **195**(19–22), 2611–2634 (2006)
4. Korelc, J.: *AceGen, AceFem*, <http://www.fgg.uni-lj.si/Symech>, last accessed 2023/11/15
5. Korelc, J.: Automation of primal and sensitivity analysis of transient coupled problems. *Comput. Mech.* **44**, 631–649 (2009)
6. Wolfram Research, Inc. Mathematica, Version 13.0.0, Champaign, IL (2021)

Virtual Reality and Augmented Reality (VR/AR)



Augmented Reality Technologies in Education and Training: A Pathway to Enhancing the Built Environment

Opeoluwa Akinradewo¹(✉), Clinton Aigbavboa², Douglas Aghimien³, Devon Rocha², Samuel Adekunle⁴, and John Aliu⁵

¹ Department of Quantity Surveying and Construction Management, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa

opeakinradewo@gmail.com

² Cidb Centre of Excellence, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

³ Department of Civil Engineering Technology, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

⁴ Department of Civil Engineering, College of Engineering and Technology, William V.S. Tubman University, East Harper, Maryland County, Republic of Liberia

⁵ Engineering Education Transformations Institute, College of Engineering, University of Georgia, Athens, GA, USA

Abstract. Integrating cutting-edge technologies has become pivotal in fostering innovation, efficiency, and safety in the ever-evolving landscape of the built environment. The adoption of augmented reality (AR) technologies has garnered considerable attention in recent years due to its potential to revolutionise education and training in the built environment. Hence, this study aims to examine the impact of augmented reality on the built environment, with a particular focus on its potential benefits in education and training. Data was collected from a sample of built environment professionals in South Africa using a structured survey instrument. The survey responses were analysed using descriptive and inferential statistics to identify trends and correlations among the variables. The findings of this study reveal that AR technology holds significant promise for the construction industry. Professionals rated its potential to assist in education before site experience and training before site work as the highest benefits. Additionally, AR was seen as a tool that can enhance task accuracy, improve collaborative teams, and contribute to reducing misinformation. The study concluded that augmented reality technologies offer numerous benefits for education and training in the South African built environment. Based on the study's findings, it was recommended that construction companies and educational institutions in South Africa consider integrating augmented reality technologies into their practices and curricula.

Keywords: Augmented Reality · Built Environment · Education · Interactive Learning · Training

1 Introduction

Augmented Reality (AR) has been identified as a potent tool in enhancing teaching effectiveness and assisting in practical training, especially in higher vocational education courses (Jiang 2018). It creates an interactive learning environment that goes beyond traditional methods, making education more engaging and effective. In the realm of IT and microelectronics, for instance, AR environments like the Proteus virtual environment have been shown to significantly improve the mastery of complex subjects and increase motivation for learning (Shamonia et al. 2019).

The application of AR technologies in education not only contributes to the development of the field but also significantly increases students' professional levels and motivation to learn (Iatsyshyn et al. 2020). By creating a virtual world for students to interact with, AR makes training modern, interactive, and practice-oriented, thereby improving student learning and engagement (Shopova and Dimitrov 2021). Furthermore, AR technology strengthens the effectiveness and attractiveness of teaching and learning, making the training process more active, effective, and meaningful (de Souza Cardoso et al. 2020). The integration of Virtual Reality (VR) and AR in teaching and learning enhances interactivity, increases effectiveness, efficiency, and learning retention, and enhances psychological and emotional training (Wang et al. 2020). AR elements in workplace training enable experiences that cannot be simulated with traditional media, benefiting motivation and learning performance (Sommerauer 2021). Moreover, AR-based software using visual cues can significantly enhance the learning process and outcome perception in complex mechanisms (Sidhu 2013).

In corporate environments, AR enhances training by providing visual, spatial, and sensorimotor feedback, thus improving the learning experience (Larson 2019). However, the successful implementation of AR in education requires continuous training and cooperation among teachers to enhance the learning process, motivation, and effectiveness (Tzima et al. 2019). In specialised fields like geoscience, AR sandboxes and constructs effectively enhance experiential learning by providing realistic simulations of earth processes (Kundu et al. 2017). In medical education, AR improves knowledge, understanding, practical skills, and social skills for students (Dhar et al. 2021), while in physical education, it improves motivation and acquisition of spatially oriented content (Moreno-Guerrero et al., 2020). AR also significantly improves immediate academic achievement and content retention in students with and without special educational needs (Badilla-Quintana et al. 2021). Gamified AR training for industrial tasks leads to better performance and greater user engagement compared to non-gamified designs (Nguyen and Meixner 2019).

Furthermore, AR increases student motivation, particularly in the Attention and Confidence dimensions of the ARCS model (Acosta et al. 2019). It also increases students' academic achievement, with perceived ease of use and benefit positively influencing attitude levels (Kaya and Bicen 2019). In museum settings, VR for kinaesthetic distance learning allows learners to perform tasks based on kinaesthetic learning characteristics, providing an engaging and mobile environment (Paulauskas et al. 2023). While AR has shown potential in enhancing learning across various domains, it's important to note that its full validation process in medical training is still lacking, indicating the need for further research and development (Barsom et al. 2016).

This article delves into the myriad ways AR technologies are enhancing education and training, particularly within the built environment, and how they pave the way for more effective, engaging, and immersive learning experiences.

2 AR Technologies for Education and Training

In higher education, AR has been recognised as an innovative technology that significantly enhances learning, particularly noted in Colombia, where it promotes interest and motivation in acquiring knowledge (Bayona Gómez 2022). This suggests that the interactive and immersive nature of AR can significantly increase student engagement and motivation across diverse educational settings. Similarly, in the training of future IT professionals, AR, exemplified by the Proteus Physics Laboratory, has been shown to enhance the visualisation of micro-level computing system functions, making complex and abstract concepts more tangible and understandable, thereby improving learning outcomes in technical fields (Shamonia et al. 2019).

In the realm of medical education and training, AR, mixed, and virtual reality-based head-mounted devices (HMDs) have shown promise, particularly for training in surgery and anatomy (Barteit et al. 2021). These technologies offer potential benefits for low-resource settings, suggesting that AR can democratise access to high-quality medical training. However, the potential for large-scale use of AR in education and training varies across different technologies and pedagogical approaches (Wang et al. 2018), highlighting the need for educators to carefully consider how AR can be integrated into their teaching strategies to maximise its benefits. In special education, AR has been found to improve motivation, learning performance, and promote self-directed and collaborative learning while being easy to use and efficient (Koppiat 2021). This underscores AR's potential to cater to diverse learning needs and styles. Similarly, in emergency medicine, AR has shown utility and feasibility in clinical care delivery, education, and training, with potential for telehealth applications (Munzer et al. 2019), indicating that AR can be a valuable tool in fast-paced and dynamic medical environments.

In engineering education, AR improves students' spatial ability, but there is a need for more learning content and personalisation for further advancements (Papakostas et al. 2021). This suggests that while AR is beneficial, its full potential can be unlocked with tailored content and approaches. In surgical education and training, AR shows validity and effectiveness, but more high-quality studies are needed before widespread implementation in surgical curricula (Kovoov et al. 2021), reflecting the current state of AR in surgical education as promising yet still emerging. In neurosurgery training, AR is a promising tool, improving technical skills and understanding of neuroanatomy (Olexa et al. 2022), demonstrating AR's potential to enhance highly specialised and complex training. Similarly, augmented and virtual reality training programs for health and care providers may improve quality of care and promote evidence-based practice but require contextual factors for successful implementation (Gasteiger et al. 2021), highlighting the importance of considering the specific context in which AR is used. However, according to Militello et al. (2019), AR adaptive training principles should focus on fidelity, realism, engagement, and scaffolding for effective recognition skills training in medical simulations, providing a framework for designing effective AR training experiences.

2.1 Benefits of Adopting AR Technologies for Education and Training

The adoption of Augmented Reality (AR) technologies in education and training, particularly within the built environment, is revolutionising traditional methods and outcomes with a multitude of transformative benefits. AR technologies create an interactive and practice-oriented educational environment, significantly improving student learning and engagement before they even set foot on the actual site. This pre-site experience, as noted by Shopova and Dimitrov (2021), is crucial in preparing students for the real-world challenges they will face. Similarly, AR in construction engineering education provides an engaging and immersive environment, improving training performance and ensuring that students are well-prepared before they begin actual site work (Wang et al. 2018).

Furthermore, AR improves learning motivation, attention, and establishes a better interactive learning environment, enhancing students' cognition and understanding of operating procedures, which in turn improves the accuracy of tasks performed in the built environment (Mao et al. 2017). It also enhances educational outcomes and promotes collaborative learning strategies, crucial for improving teamwork in the built environment where collaborative efforts are essential (Bistaman et al. 2018). The dynamic and interactive instructions for training students in various engineering sectors, including electrical, mechanical, and renewable energy engineering, are leading to its widespread implementation in the construction industry (Yousif 2022). The quality of learning is also significantly improved with AR-based learning, increasing motivation, engagement, and academic performance in the computer programming classroom, which can be extrapolated to other areas of the built environment (Larson and Chambers 2020). AR contributes to successful learning and increased motivation in classroom applications and teacher training, which is fundamental in fostering an environment of effective teamwork (Buchner and Zumbach 2020). Its ability to provide clarity and simulate processes effectively aids in a smoother handover by ensuring all parties have a clear understanding of the project components and status (Shamonia et al. 2019).

Moreover, AR enhances, motivates, and stimulates learners' understanding of events, allowing for quick learning in a quick mode, thereby reducing the chances of misinformation (Gutiérrez and Fernández 2014). It has a medium effect on learning effectiveness, with the most reported advantages being "learning gains" and "motivation," which is crucial in maintaining high levels of motivation among students and trainees (Garzón et al. 2019). By improving the understanding and accuracy of tasks, AR indirectly contributes to reducing site accidents, as better-prepared individuals make fewer errors (Paulauskas et al. 2023). It also improves learning and helps students understand their subjects better, combining the actual and virtual worlds, leading to fewer snags and smoother project execution (Ali et al. 2022). Lastly, AR training positively impacts students' spatial ability and learning for basic engineering graphics contents, which is crucial in reducing mistakes made by site personnel (Martín-Gutiérrez et al. 2010), marking a significant stride in the evolution of educational methodologies.

3 Research Methodology

The rationale of the current study is to contribute to the body of knowledge on the benefits of Augmented Reality (AR) technologies in education and training, particularly within the built environment. This study adopted a quantitative research approach to achieve the set objective. A quantitative research survey is a systematic method for collecting information from a sample of individuals by asking questions pertinent to the research. When conducting a study that makes use of quantitative research, it is crucial to numerically measure specific aspects of the phenomena with precision. The study collected data through a well-structured questionnaire distributed to respondents who are professionals in fields related to the built environment, such as architects, civil engineers, construction managers, and educational technologists, primarily focusing on those utilising or familiar with AR technologies in their practice. These respondents were selected from Gauteng province of South Africa. A 5-point Likert scale questionnaire was developed utilising knowledge obtained from literature to gather data relevant to the intent of the research. The questionnaire aimed to assess the perceived benefits, applications, and effectiveness of AR technologies in enhancing learning, improving task accuracy, and fostering collaboration among teams in the built environment.

The study distributed 225 questionnaires to professionals within the targeted area, and out of these, 187 questionnaires were returned and found suitable for analysis. The data obtained from the questionnaire were analysed using descriptive statistics. To determine the normality of the retrieved data, the Shapiro-Wilk test was engaged, while Cronbach's alpha was adopted to determine the reliability coefficient of the data collection instrument. The adopted cutoff alpha for this study was 0.70 and an alpha value of 0.925 was obtained, indicating that the data retrieved is reliable and valid for making inferences about the benefits and applications of AR in education and training within the built environment.

4 Findings

4.1 Demographic Information of Respondents

The analysis results show that the largest group of participants, accounting for 48%, are employed at consulting firms. Contractors represent 28% of the sample population. Additionally, both tertiary students and government employees constitute 12% each. Regarding educational qualifications, the most prevalent was a Bachelor of Science degree, held by 32% of respondents. This was closely followed by those with an Honour's Degree and Diploma, each at 28%. The Master's degree holders were next, making up 12%, while no respondents reported having a Doctorate. A significant 60% of respondents have 0–3 years of work experience, suggesting that the sample population is relatively inexperienced. This is followed by those with 4–8 years of experience, making up 28%, and a smaller 8% have 9–15 years of experience. Only 4% have worked for more than 15 years. These demographics suggest that, while the respondents have a moderate level of knowledge, they are relatively new to the field, which could influence their capacity to provide substantial insights into the research question.

4.2 Descriptive and Normality Analysis Result

Table 1 shows the descriptive and normality analysis result for the identified variables. The variable with the highest value to the respondent is ‘Assisting in education before site experience’ with a mean score of 3.14 and a standard deviation of 0.710. The variable with ranked second is ‘Assisting in training before site work’ with a mean score of 3.05 with a standard deviation of 0.722. The third highest ranked variable is ‘Improving the accuracy of tasks’ with a mean score of 2.95 with a standard deviation of 0.925. Ranked fourth is ‘Improving collaborative teams’ with a mean score of 2.93 and a standard deviation 0.799. Ranked fifth is the ‘Being widely implemented in the construction industry’ with a mean score of 2.88 and a standard deviation of 0.851. The table further presents the results of the Shapiro-Wilk test for normality, showing that the significance values for all 13 evaluated benefits are well below the 0.05 threshold required for normality. This suggests that the collected data is non-parametric. From the table, the significance value (Sig.) for all the identified variables was 0.000. This indicates that for all benefits, the p-value is less than 0.05, which is interpreted as a non-normal distribution. Thus, the null hypothesis (that the data is normally distributed) can be rejected for each benefit.

5 Discussion of Findings

The utilisation of Augmented Reality (AR) technologies in education and training is increasingly recognised for its transformative potential, particularly in preparing individuals for real-world scenarios before they encounter them. The highest-ranked benefit, “Assisting in education before site experience,” with a mean of 3.14, underscores the significant role of AR in pre-site preparation. This aligns with findings that AR can enhance heritage preservation and education by allowing learners to interact with 3D visuals and perform tasks in a virtual environment, thus providing a rich, contextual understanding of the subject matter before any physical interaction occurs (Paulauskas et al. 2023). This pre-emptive interaction is particularly beneficial in fields that require a high degree of spatial understanding and situational awareness.

In construction engineering, for example, AR’s capacity to provide immersive and interactive training experiences is crucial. Visualisation, health, safety training, and structural analysis are vital components of construction education, and AR’s ability to simulate these aspects in a controlled environment allows for a deeper understanding and better preparation. It enables students to visualise complex structures, understand the intricacies of architectural designs, and practice safety protocols in a risk-free setting, thereby enhancing their readiness for actual site work (Wang et al. 2018). Moreover, AR’s interactive and immersive nature not only enhances individual task performance but also fosters better teamwork and collaboration. By enabling a shared virtual space, team members can interact with each other and the project in a more cohesive manner, leading to improved communication and collaborative problem-solving (Lee 2012; Mao et al. 2017).

The benefit “Being widely implemented in the construction industry” reflects the growing recognition and adoption of AR in practical fields. It is enhancing teaching and learning by increasing interactivity, effectiveness, and efficiency. As AR technology continues to evolve, its applications in the construction industry are becoming more

Table 1. Benefits of AR technologies in education and training.

| Benefits | Mean | Std. Deviation | Rank | Shapiro-Wilk | |
|---|------|----------------|------|--------------|-------|
| | | | | Statistic | Sig |
| Assisting in education before site experience | 3.14 | 0.710 | 1 | 0.841 | 0.000 |
| Assisting in training before site work | 3.05 | 0.722 | 2 | 0.841 | 0.000 |
| Improving the accuracy of tasks | 2.95 | 0.925 | 3 | 0.798 | 0.000 |
| Improving collaborative teams | 2.93 | 0.799 | 4 | 0.813 | 0.000 |
| Being widely implemented in the construction industry | 2.88 | 0.851 | 5 | 0.844 | 0.000 |
| Improving quality | 2.86 | 0.804 | 6 | 0.776 | 0.000 |
| Improving overall teamwork | 2.86 | 0.833 | 6 | 0.831 | 0.000 |
| Assisting in a smoother handover | 2.79 | 0.888 | 8 | 0.789 | 0.000 |
| Reducing misinformation | 2.77 | 0.782 | 9 | 0.811 | 0.000 |
| Increasing motivation | 2.72 | 0.908 | 10 | 0.759 | 0.000 |
| Reducing site accidents | 2.70 | 0.773 | 11 | 0.808 | 0.000 |
| Assisting in reducing snags | 2.65 | 0.842 | 12 | 0.729 | 0.000 |
| Less mistakes made by site personnel | 2.53 | 0.909 | 13 | 0.777 | 0.000 |

sophisticated, ranging from virtual walkthroughs of architectural designs to detailed simulations of construction processes. This not only improves the learning experience but also enhances the quality of construction projects by allowing for more precise planning and execution (Wang et al. 2020).

Similarly, “Improving quality” and “Improving overall teamwork” are acknowledged as significant benefits. These benefits resonate with the understanding that AR can lead to better learning outcomes and foster a collaborative environment. By integrating AR into educational curricula, institutions can provide students with a more engaging and effective learning experience. This not only improves the quality of education but also prepares students to work effectively in team settings, an essential skill in many professional fields (Garzón et al. 2019; Shopova and Dimitrov 2021).

However, lower-ranked benefits such as “Reducing misinformation,” “Increasing motivation,” and “Reducing site accidents,” while still recognised as important, indicate areas where AR’s impact is acknowledged but perhaps not as strongly felt or as widely researched. This suggests a need for further exploration and improvement in these areas. Reducing misinformation through accurate and timely data presentation, increasing motivation by making learning more engaging and interactive, and reducing site accidents through better training and preparation are all areas where AR has shown potential but requires more focused research and development to fully leverage its capabilities (Ali et al. 2022; Dhar et al. 2021).

6 Conclusion and Recommendation

The exploration of augmented reality (AR) technologies in education and training, particularly within the built environment, has revealed a multifaceted landscape of benefits and applications. The study’s findings underscore the significant role of AR in enhancing pre-site educational experiences, improving the accuracy of tasks, fostering collaborative teamwork, and increasing overall motivation and quality in learning environments. The quantitative analysis further solidified the understanding that AR is not just a futuristic concept but a present-day tool that is being increasingly implemented across various sectors, including the construction industry. However, the study also highlights the relative infancy of widespread AR adoption, with a significant portion of the respondents being relatively new to the field. This indicates a burgeoning area ripe for exploration and expansion. The non-parametric nature of the data suggests a diversity in the adoption and perception of AR technologies, reflecting the varied stages of implementation and experience of respondents across different professions in the built environment.

Given the promising benefits of AR, it is recommended that educational institutions and industries related to the built environment invest in the development and integration of AR technologies. This includes providing training for educators and professionals, developing more intuitive and user-friendly AR applications, and conducting longitudinal studies to assess the long-term impact of AR on learning outcomes and industry practices. Furthermore, it is crucial to address the limitations identified in the study. The study’s focus on specific regions calls for more extensive research that includes a variety of geographical contexts to understand the global applicability and impact of AR technologies.

Areas for further research include exploring the specific types of AR applications that are most effective for different learning objectives and industry tasks, understanding the long-term retention of skills and knowledge acquired through AR, and investigating the cost-benefit analysis of implementing AR technologies in educational and industry settings. Moreover, research should also focus on overcoming the challenges associated with AR, such as cognitive overload, technical issues, and ensuring accessibility and inclusivity in AR-based learning environments.

References

1. Acosta, J.L.B., Navarro, S.M.B., Gesa, R.F., Kinshuk, K.: Framework for designing motivational augmented reality applications in vocational education and training. *Australas. J. Educ. Technol.* **35**(3) (2019).
2. Ali, N.A., Sadiq, M.H., Albabawat, A.A., Salah, R.M.: Methods and applications of augmented reality in education: a review. In: 2022 International Conference on Computer Science and Software Engineering (CSASE) (2022) <https://doi.org/10.1109/CSASE51777.2022.9759807>
3. Badilla-Quintana, M.G., Sandoval-Henríquez, F.J.: Students' immersive experience in initial teacher training in a virtual world to promote sustainable education: interactivity, presence, and flow. *Sustain.* **13**(22), 12780. (2021)
4. Barsom, E.Z., Graafland, M., Schijven, M.P.: Systematic review on the effectiveness of augmented reality applications in medical training. *Surg. Endos.* **30**, 4174–4183. (2016)
5. Bartelt, S., Lanfermann, L., Bärnighausen, T., Neuhann, F., Beiersmann, C.: Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. *JMIR Serious Games* (2021). <https://doi.org/10.2196/29080>
6. Bayona Gómez, S.: Augmented reality as a teaching and learning strategy in higher education in Colombia. *Int. J. Res. Publication Rev.* (2022). <https://doi.org/10.55248/gengpi.2022.3.9.42>
7. Bistaman, I.N.M., Idrus, S., Rashid, S.A.: The use of augmented reality technology for primary school education in Perlis, Malaysia. *J. Phys: Conf. Ser.* (2018). <https://doi.org/10.1088/1742-6596/1019/1/012064>
8. Buchner, J., Zumbach, J.: Augmented reality in teacher education. A framework to support teachers' technological pedagogical content knowledge. *J. Educ. Technol.* (2020). <https://doi.org/10.17471/2499-4324/1151>
9. de Souza Cardoso, L.F., Mariano, F.C.M.Q., Zorزال, E.R.: A survey of industrial augmented reality. *Comput. Ind. Eng.* **139**, 106159 (2020)
10. Dhar, P., Rocks, T., Samarasinghe, R.M., Stephenson, G., Smith, C.M.: Augmented reality in medical education: students' experiences and learning outcomes. *Med. Educ. Online* (2021). <https://doi.org/10.1080/10872981.2021.1953953>
11. Garzón, J., Pavón, J., Baldiris, S.: Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality* (2019). <https://doi.org/10.1007/s10055-019-00379-9>
12. Gasteiger, N., van der Veer, S.N., Wilson, P., Dowding, D.: Upskilling health and care workers with augmented and virtual reality: protocol for a realist review to develop an evidence-informed programme theory. *BMJ Open* (2021). <https://doi.org/10.1136/bmjopen-2021-050033>
13. Gutiérrez, J.M., Fernández, M.D.M.: Augmented reality environments in learning, communicational and professional contexts in higher education. *Digital Educ. Rev.* (2014) <https://doi.org/10.1344/DER.2014.0.61-73>
14. Iatsyshyn, A.V., Kovach, V.O., Lyubchak, V.O., Zuban, Y.O., Piven, A.G., Sokolyuk, O.M., ... Shyshkina, M.P.: Application of augmented reality technologies for education projects preparation. In: CTE Workshop Proceedings Vol. 7, pp. 134–160. (2020)
15. Jiang, W., Ma, L., Zhang, B., Fan, Y., Qu, X., Zhang, X., Liao, H.: Evaluation of the 3D augmented reality-guided intraoperative positioning of dental implants in edentulous mandibular models. *Int. J. Oral Maxillofac Implant* **33**(6), (2018)
16. Kaya, O.S., Bicen, H.: Study of augmented reality applications use in education and its effect on the academic performance. *Int. J. Distance Educ. Technol. (IJDET)*. **17**(3), 25–36. (2019)
17. Κορφιάτη, Η.: Συγκριτική μελέτη και ανασκόπηση της βιβλιογραφίας για την επανζημένη πραγματικότητα στην ειδική αγωγή και εκπαίδευση (2021). <https://doi.org/10.26265/POLYNOE-775>

18. Kovoov, J.G., Gupta, A.K., Gladman, M.A.: Validity and effectiveness of augmented reality in surgical education: a systematic review. *Surg.* **170**(1), 88–98. (2021)
19. Kundu, A.S., Mazumder, O., Dhar, A., Lenka, P.K., Bhaumik, S.: Scanning camera and augmented reality based localization of omnidirectional robot for indoor application. *Procedia Comput. Sci.* **105**, 27–33. (2017)
20. Larson, K.: Using augmented reality in the corporate environment: A proposed study protocol. In: 2019 Eighth International Conference on Educational Innovation through Technology (EITT) pp. 245–249. IEEE. (2019)
21. Larson, K.M., Chambers, R.: AR in the computer programming classroom: a review of the literature. In: 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE) (2020). <https://doi.org/10.1109/TALE48869.2020.9368329>
22. Lee, K.: Augmented reality in education and training. *TechTrends* (2012). <https://doi.org/10.1007/S11528-012-0559-3>
23. Mao, C.-C., Chen, C.-H., Sun, C.: Impact of an augmented reality system on learning for army Military Decision-Making Process (MDMP) course (2017). https://doi.org/10.1007/978-3-319-41685-4_58
24. Martín-Gutiérrez, J., Saorín, J.L., Contero, M., Raya, M.A.: AR_Dehaes: an educational toolkit based on augmented reality technology for learning engineering graphics. In: 2010 10th IEEE International Conference on Advanced Learning Technologies (2010). <https://doi.org/10.1109/ICALT.2010.45>
25. Militello, L., Sushereba, C.E., Hernandez, O.K., Patterson, E.: Augmented reality adaptive training principles. In: Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare (2019). <https://doi.org/10.1177/2327857919081016>
26. Moreno-Guerrero, A.J., Alonso Garcia, S., Ramos Navas-Parejo, M., Campos-Soto, M.N., Gomez Garcia, G.: Augmented reality as a resource for improving learning in the physical education classroom. *Int. J. Environ. Res. Pub Health*, **17**(10), 3637. (2020)
27. Munzer, B., Khan, M.M., Shipman, B., Mahajan, P.: Augmented reality in emergency medicine: a scoping review. *J. Med. Internet Res.* (2019). <https://doi.org/10.2196/12368>
28. Nguyen, D., Meixner, G.: Gamified augmented reality training for an assembly task: a study about user engagement. In: 2019 Federated Conference on Computer Science and Information Systems (FedCSIS) pp. 901–904. IEEE. (2019, September)
29. Olexa, J., Cohen, J., Alexander, T., Brown, C., Schwartzbauer, G., Woodworth, G.: Expanding educational frontiers in neurosurgery: current and future uses of augmented reality. *Neurosurgery* (2022). <https://doi.org/10.1227/neu.0000000000002199>
30. Papakostas, C., Troussas, C., Krouskas, A., Sgouropoulou, C.: Exploration of augmented reality in spatial abilities training: a systematic literature review for the last decade. *Informatics Educ* (2021). <https://doi.org/10.15388/INFEDU.2021.06>
31. Paulauskas, L., Paulauskas, A., Blažauskas, T., Damaševičius, R., Maskeliūnas, R.: Reconstruction of industrial and historical heritage for cultural enrichment using virtual and augmented reality. *Technologies* (2023). <https://doi.org/10.3390/technologies11020036>
32. Shamonina, V., Semenikhina, O., Proshkin, V., Lebid, O., Kharchenko, S., Lytvyn, O.: Using the Proteus virtual environment to train future IT professionals (2019). <https://doi.org/10.31812/EDUCDIM.V53I1.3842>
33. Shopova, V., Dimitrov, I.: The place of augmented reality in the learning process. *Educ. Technol. J.* (2021). <https://doi.org/10.26883/2010.211.3261>
34. Sidhu, M.S.: Enhancing a multi-body mechanism with learning-aided cues in an augmented reality environment. In IOP Conference Series: materials Science and Engineering Vol. 46, No. 1, p. 012032. IOP Publishing. (2013)
35. Sommerauer, P.: Augmented reality in VET: benefits of a qualitative and quantitative study for workplace training. In: 54th Hawaii International Conference on System Sciences (HICSS), Grand Wailea, HI, USA, 5–8, pp. 1623–1632. University of Hawai'i at Manoa. (2021, January)

36. Tzima, S., Styliaras, G., Bassounas, A.: Augmented reality applications in education: teachers point of view. *Educ. Sci.* **9**(2), 99. (2019)
37. Wang, M., Callaghan, V., Bernhardt, J., White, K., Peña-Ríos, A.: Augmented reality in education and training: pedagogical approaches and illustrative case studies. *J. Ambient Intell. Humanised Comput.* (2018). <https://doi.org/10.1007/s12652-017-0547-8>
38. Wang, N., Magagna, W., Peck, K.L., Pfeifer, C., Wang, C.: Virtual reality and augmented reality: perceptions of early adopters. *Int. J. Smart Technol. Learn.* (2020). <https://doi.org/10.1504/ijsmartl.2020.10034287>
39. Yousif, M.: VR/AR environment for training students on engineering applications and concepts. *Artif. Intell. Robot. Dev. J.* (2022). <https://doi.org/10.52098/airdj.202254>



Architectural Design Revolution: Enhancing Student Learning Through VR

Jean-Pierre Basson^(✉) and Ashvin Manga

Nelson Mandela University, Port Elizabeth, South Africa
jp.basson@mandela.ac.za

Abstract. Architectural education is on the brink of a transformative leap, and Virtual Reality (VR) technologies have emerged as a catalyst for innovation in design. The study investigates the user experience, satisfaction and implementation of Bachelor of Architecture (Honours) students. The paper's driving question is how integrating VR tools in design studios shapes students' spatial learning experiences and acceptance levels compared to the traditional design process. Through a mixed-method and action research approach, we explore how VR enhances design fluency, spatial development, and creative innovation. Primary data is collected using a survey questionnaire and observational study. The juxtaposition of these two design approaches sets the stage for an exciting exploration of pedagogical innovation. VR tools are scrutinized on how they impact students' design fluency, spatial development, and creative innovation. VR technologies can transform the conventional design critique into an interactive, three-dimensional experience, enabling the learning process and pushing the boundaries of architectural education, especially in spatial understanding. The survey captured students' firsthand experience and provided testimonials about their experience creating and exploring their designs using VR technology. The observational study uncovers moments of epiphany, where traditional barriers dissolve and architectural concepts come alive. Students' satisfaction is also examined using VR tools, identifying elements that ignite their passion for design. Iteration and refinement capabilities are refined using VR. The study underscores the importance of addressing challenges in effectively implementing VR, emphasizing the necessity for a paradigm shift in institutional mindset, infrastructure enhancements, and targeted training initiatives.

Keywords: Architectural Education · Design Process · Spatial Understanding · Studio · Virtual Reality

1 Introduction

Green and Bonollo [1] consider the design studio to be the centre of architectural design education, and it distinguishes architectural education from other academic fields. According to Green and Bonollo [1], students learn to visualize and graphically portray architectural challenges inside this instructional area. The design studio is both a

learning centre and a complex social organization, emphasizing its importance in curricular alignment. Therefore, fostering a collaborative, multi-sensory, experimental, and problem-based environment conducive to the free interchange of ideas for the design studio to thrive, is essential.

Demirbas and Demirkhan [2] emphasize the studio as a critical component for learning new skills, mastering design language, and developing architectural thinking. Understanding the dynamics of a design studio during a project is crucial in understanding its influence. A design studio's physical features include an open and flexible area with movable tables, chairs, and computer workstations, allowing for various instructional methods such as presentations, discussions, and workshops [1].

The design process in architectural education, as defined by Kvan [3] and the BIM Handbook [4], is a dynamic activity that defines critical project information. The educational design project concerns lecturer-defined design issues that students must answer within a deadline. The typical architectural design studio project follows the architectural work stages appraisal and definition of the project (stage 1), design concept (stage 2), design development (stage 3) and technical documentation (stage 4). Therefore, the studio design project is conceptual in nature. The design process is grouped into four stages: the definition of the project (information development stage), conceptual design stage, design development stage and finally, the documentation and presentation stage. Students must produce documentation demonstrating the project's function, appearance, sustainability and design intent. The function includes a site plan, floor plans, sections, and outline specification (also referred to as the accommodation schedule). Appearance includes elevations, 3D physical models, and illustrative material (computer simulations). Sustainability aspects include energy calculations and simulations whereas the design intention talks to diagrams that discuss layout relationship and function, massing, construction types and building placement and orientation.

According to Green and Bonollo [1] and Kvan [3], the activity of the design studio is divided into three separate periods. During the first part, the lecturer explains the project by defining the design challenge, objectives, site specifics, and technical submission requirements. These issues are typically addressed in a hypothetical environment, allowing students to delve into specific technical, functional, or aesthetic concerns within a particular political, socio-cultural, and economic framework. In the second phase, the studio develops into formal and informal settings, encouraging interaction. Students participate in conversations centred on the questions provided by the brief, presenting their proposed design solutions through drawings, sketches, or models. Informal settings include small group or individual critique sessions with the speaker, and formal settings include posting up work for panel or group debates. This stage is distinguished by conceptual and rough work, representing the evolving process. The final phase concludes with students presenting their design suggestions within a time limit. The iterative design studio process is completed when the studio leader reviews the presentations, provides written or vocal feedback, and issues a grade.

The study aims to explore the transformative effects of VR technologies on students' learning experiences, design fluency, spatial development, and creative innovation. By employing a comprehensive survey method, the research investigates the perceptions and insights of students who have experienced both the conventional design process and the

immersive world of VR in their undergraduate and postgraduate studies. The paper seeks to understand how VR tools enhance the traditional design critique, transforming it into an interactive, three-dimensional experience that pushes the boundaries of architectural education. Ultimately, this study contributes to the ongoing discourse on the role of VR in shaping the future of architectural education and design pedagogy.

2 Literature Review

2.1 Industry 4.0

In the historical context, industrial-age education was primarily concerned with the recall of information and facts, which corresponded to the lowest level of Bloom's taxonomy. Towards the end of the industrial period, an emphasis on higher-order thinking evolved, forcing educators to include critical thinking skills in teaching, such as analyzing evidence and synthesizing knowledge [5]. The fourth industrial revolution (4IR) differs from its predecessors in that it is characterized by the overall development and use of cyber-physical systems based on intelligence, networking, and virtual-physical environments [6, 7]. Digital technologies, at the forefront of the 4IR, connect billions of people daily, increasing worldwide competitiveness for jobs [8]. Constant access to information, quicker computers, automation replacing mundane employment, and networked communication characterize the 4IR [9]. During the 4IR, however, issues in social factors, technological misuse, erroneous information, politics, and insufficient human resources in education were found [9].

The current generation, nicknamed the “copy and paste generation” by Richards [10], has easy access to knowledge but struggles with critical thinking. Whilst technology can help people express themselves, it is unclear if it fosters critical thinking or only expedites the completion of work in an acceptable style [11]. Educational institutions must build appropriate learning environments to educate students for the dynamic working world moulded by Industry 4.0. There is a shift in study focus away from the use of technology and towards surroundings that encourage students' critical thinking tendencies [12]. They suggest that technology benefits students by allowing them to generate knowledge, employ sophisticated information-gathering tools, participate in collaborative procedures, and apply higher-order thinking abilities to complicated evaluations.

To meet the expectations of Industry 4.0, defined by rapid technological breakthroughs, educational institutions must prioritize critical thinking as a fundamental soft skill. Critical thinking allows for systematic problem-solving by analyzing and reasoning from the quantity of information made available by technological processes. This obligation extends beyond general education to specialized sectors, such as architecture, where revolutionary developments occur. The architecture profession is actively pursuing new technologies such as Building Information Modelling (BIM) and virtual reality, which radically transform the traditional design process and the character of design studios. These advancements not only improve the efficiency and precision of architectural workflows, but they also need a deep integration of critical thinking abilities to traverse and exploit the potential of these cutting-edge tools. As the architectural landscape evolves with technological breakthroughs, cultivating a critical thinking culture becomes vital

to equip architects to creatively connect with, adapt to, and drive forward the industry's dynamic changes.

2.2 Virtual Reality

Virtual reality (VR) technology consists primarily of simulated environments, perception, natural skills, and sensing equipment [13]. The simulation environment is a computer-generated three-dimensional realistic image that is real-time and dynamic. Aside from virtual perception provided by computer graphics technology, there is also the perception of hearing, touch, force, motion, smell, and taste, known as multi-perception. VR allows users to have immersive experiences while moving freely. VR is ultra-high-definition and changes in real-time with the user's head and eye movements [14]. VR in education has the potential to create fresh and innovative learning experiences. While VR has historically been used in technical higher education, it has grown into various sectors, including health-related fields, general education, engineering, and sciences. This broader integration reflects a growing trend of employing VR technology to improve learning across multiple educational fields [15]. VR systems can be divided into three categories [16]:

- Imagery is displayed in 2D on the monitor via sophisticated computer-stored data, providing practically authentic graphics, sound, and responsiveness;
- Immersive systems, and
- Telepresence.

2.3 VR, the Architecture Design Studio and Learning and Teaching

The 21st-century learning and teaching landscape saw the Internet of Things (IoT) increase, such as e-learning, MOOCs, VR and Augmented Reality [17]. The application of VR in architecture stretches over various applications, such as the planning and ideation phase, concept development, detail design, and visualization [18, 19]. While the effectiveness of VR technology has been well validated, there is still significantly low adoption and usage, indicating a lack of understanding about people's attitudes towards embracing VR technologies [20]. The advantages of VR technology are clear when visualizing complicated construction environments and implementing a gamification strategy to improve learner participation. Despite these advantages, virtual reality technology has not yet become a standard tool in architecture education and the use of VR in a dynamic learning environment has not been extensively researched [21].

Numerous factors influence the architectural design process, as designers frequently use their spatial experiences to solve design problems. The information gained from previous spatial experiences is the foundation for developing a conceptual framework for solving design difficulties. Furthermore, an individual's cognitive style, impacted by factors such as education and environment, can majorly impact design problem-solving. Designers must be spatially proficient, manipulating spatial information in two-dimensional (2D) and three-dimensional (3D) environments. Spatial aptitude includes visualizing, mentally rotating, and changing spatial information [22]. Current educational tools and courses are inadequately geared to assist students with developing their spatial visualization skills, highlighting the need for a innovative approach. Online and

virtual reality (VR) applications show potential as teaching aids for improving the spatial visualization skills of computer-aided design (CAD) students. These technologies can potentially improve students' understanding of three-dimensional things and enhance their spatial visualization skills [23].

3 Methodology

This research employed a mixed-method approach to depict, interpret, and scrutinize a phenomenon within its natural context. Using action research (AR) methodology facilitated the generation of primary data. As per Costello [24], action research is characterized by a practical, problem-solving orientation undertaken by individuals, professionals, and educators to enhance educational practices. Action research comprises a set of features, including a flexible design and active participant engagement, rather than being a singular approach. Action research allows for incorporating diverse data collection methods, such as observational studies, interviews, and document analysis.

The research employed a questionnaire survey to acquire primary quantitative data and an observational strategy to gather qualitative data concerning students' spatial cognitive understanding by applying VR as a design tool in the Bachelor of Architectural Studies (Honours) program. The investigative approach embraced a descriptive analysis with coding methodologies. The survey questionnaire systematically probed into the students' antecedent encounters with VR, their training in VR utilization, the application of VR in design scenarios, their expectations, and contributions to the overall design experience.

Between May 18 and June 8, 2023, the class was partitioned into groups of eight students, and a series of weekly VR training sessions were administered in the Architectural Computer Usage module. The training regimen comprised three integral components. Firstly, a presentation elucidated the conceptualization of VR and its relevance to both architecture and architectural education. The discourse delved into the application of VR for comprehending three-dimensional spatial configurations in contrast to its commercial utility as a visualization tool. Subsequently, the students underwent training on the operational aspects of the Oculus Quest 2, encompassing an understanding of the interface and proficiency in controller usage. Lastly, the students were acquainted with SimLAB, receiving guidance on the transference of their Autodesk Revit modules to SimLab Studio. They underwent the necessary modifications to facilitate uploading onto the SimLAB viewer platform for immersive engagement in VR. The practical application of VR headsets was then integrated into the training program. The students had access to the Oculus Quest 2 headsets during the second semester.

The survey employed Likert-type questions to determine the frequency of students' engagement in specific activities. In cases where students were uncertain about their response, an 'unsure' option was provided to accommodate such instances. Consequently, mean scores (MSs) were calculated to rank fixed-response items, considering the central tendency of responses (Table 1):

$$MS = \frac{1n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5}{(n_0 + n_1 + n_2 + n_3 + n_4 + n_5) - n_0}$$

Table 1. Definition of Likert scale points and related variables

| Likert scale point | | Variable |
|--------------------|-------------------|----------|
| Unsure | Unsure | n_0 |
| Never | Strongly disagree | n_1 |
| Rarely | Disagree | n_2 |
| Sometimes | Neutral | n_3 |
| Often | Agree | n_4 |
| Always | Strongly agree | n_5 |

The average class size for the Bachelor of Architecture Honours program at Nelson Mandela University is typically 25 students. However, for this study, 16 students participated, constituting 62% of the total cohort.

4 Results and Findings

Table 2 indicates the extent to which students agree or disagree with statements about students' prior engagement with VR technologies in terms of MS, ranging between 1.00 to 5.00 based upon percentage responses to a scale of strongly disagree to strongly agree. Notably, all the mean MSs are below the midpoint of 3.00, which indicates that the Bachelor of Architectural Studies (Honours) students disagree, as opposed to agreeing with the statements about exposure to VR technologies. The MS of 2.06 indicates that students strongly disagree to disagree / disagree that they had previous exposure to VR technologies. The findings align with the current curriculum where VR technologies is not introduced to the undergraduate students, nor are the students exposed to VR. It is only at the postgraduate level that architectural students are exposed to VR technologies.

Table 3 indicates the extent to which students agree or disagree with statements about the student's expectations in VR training in terms of MS, ranging between 1.00 to 5.00 based on percentages responses to a scale of strongly disagree to strongly agree. The MS is above the midpoint 3 three, which generally indicates the students agree, as opposed to disagreeing with the statement about meeting the expectations of VR training. The MS of 4.38 indicates the students agree to strongly agree / strongly agree to the statement that the training was clear and effective. The findings indicate that the current implementation framework of VR training, which is in the form of a two-week winter workshop, meets the student's expectations and the timeframe is adequate for training students using VR technologies.

Table 4 indicates the extent to which students agree or disagree with statements about the application of VR technologies in the design studio and studio design project in terms of MS ranging between 1.00 and 5.00 based upon percentages responses to a scale of strongly disagree to strongly agree. It is notable that all the mean MSs are above the midpoint scale of 3.00, which indicates that, in general, the Bachelor of Architectural Studies (Honours) students agree, as opposed to disagreeing with the statements about VR applications in studio design projects. However, it is notable that 3 / 5 (60%) of the

Table 2. Prior VR exposure

| Statement | Response (%) | | | | | MS | Rank |
|--|-------------------|----------|---------|-------|----------------|------|------|
| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree | | |
| My familiarity with VR technologies at the beginning of my postgraduate studies was adequate | 6.2 | 18.8 | 56.2 | 12.6 | 6.2 | 2.94 | 1 |
| Before my postgraduate studies, I had exposure to Virtual Reality (VR) technologies | 43.8 | 25.0 | 18.8 | 6.2 | 6.2 | 2.06 | 2 |

Table 3. VR Training

| VR Training | Response (%) | | | | | MS | Rank |
|---|-------------------|----------|---------|-------|----------------|------|------|
| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree | | |
| The introduction of VR tools in my postgraduate studies was clear and effective | 0.0 | 6.2 | 6.2 | 31.3 | 56.3 | 4.38 | 1 |

mean MSs are $>3.40 \leq 4.20$, and thus the agreement can be deemed to be experienced between neutral to agree / agree. ‘VR technology greatly enhanced my ability to visualize and communicate dressing ideas’ predominates with a MS of 4.19, followed by ‘VR Tools enhanced my learning experience compared to the traditional design process’ and ‘VR tools had a notable impact on my understanding of spatial relationships’. The findings align with Fussell’s that state VR holds various advantages to architectural education, especially visualization. However, the use of VR to investigate spatial relationships and improve design processes is ranked second to last and last. This is an indicator that VR as a design tool is not yet fully adopted, which could be a result of students not trusting the technologies as stated by Zhang [20].

Table 5 indicates the extent to which students agree or disagree with statements about the impact of VR on architectural education in terms of MS ranging between 1.00 and 5.00 based upon percentages responses to a scale of strongly disagree to strongly

Table 4. VR Applications

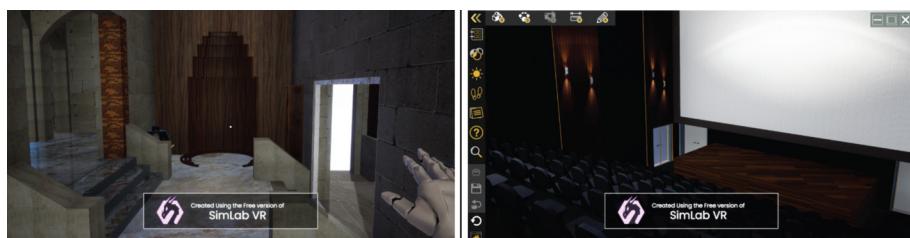
| VR Applications | Response (%) | | | | | MS | Rank |
|---|-------------------|----------|---------|-------|----------------|------|------|
| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree | | |
| VR technology greatly enhanced my ability to visualize and communicate dressing ideas | 0.0 | 0.0 | 18.8 | 43.7 | 37.5 | 4.19 | 1 |
| VR Tools enhanced my learning experience compared to the traditional design process | 0.0 | 6.2 | 25.0 | 50.0 | 18.8 | 3.81 | 2 |
| VR tools had a notable impact on my understanding of spatial relationships | 0.0 | 12.6 | 31.2 | 31.2 | 25.0 | 3.69 | 3 |
| VR Tools significantly improved my design fluency | 0.0 | 25.0 | 50.0 | 18 | 6.2 | 3.06 | 4 |
| VR facilitated my creative exploration during the design task | 6.2 | 6.2 | 62.6 | 6.2 | 18.8 | 3.25 | 5 |

agree. It is notable that all the mean MSs are above the midpoint scale of 3.00, which indicates that, in general, the Bachelor of Architectural Studies (Honours) students agree, as opposed to disagreeing with the statements about the impact of VR on architectural education. However, it is notable that 2 / 2 (100%) of the mean MSs are $>4.20 \leq 5.00$, and thus the agreement can be deemed to be experienced between agree to strongly agree / strongly agree. ‘VR technologies have the potential to reshape architectural education’ predominates with a MS of 4.50, followed by ‘VR technologies have the potential to reshape architectural education’. The findings indicate that the students are well aware of the potential benefits of VR technologies within architectural education, which speaks about the students’ openness and acceptance of the new Technologies.

Table 5. VR and the impact on architectural education

| Impact | Response (%) | | | | | MS | Rank |
|---|-------------------|----------|---------|-------|----------------|------|------|
| | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree | | |
| VR technologies have the potential to reshape architectural education | 6.2 | 0.0 | 6.2 | 12.6 | 75.0 | 4.50 | 1 |
| VR technologies have the potential to reshape architectural education | 0.0 | 6.2 | 12.5 | 18.8 | 62.5 | 4.38 | 2 |

Three key themes are prominent from the observational study, namely: Aesthetic appeal and immersive experience, spatial understanding and effectiveness of design concept. In the first theme, aesthetic appeal and immersive experience, students fail to express the appeal of the virtual spaces in terms of the use of colour, textures, lighting and overall visual composition, as seen in Fig. 1. The viewer is not engaged and immersed in the virtual environment due to a lack of elements that stimulate and enhance the emotional and sensory experience of the space. The ambience and atmosphere created in the virtual environment were not thought through, which hindered the total immersive experience.

**Fig. 1.** Students 1 and 4 demonstrate the lack of aesthetic appeal and immersive experience

In the second theme, spatial understanding, students lack clarity of spatial arrangements within the virtual environment. Spaces feel random, with no consideration of any architectural design principles. Consideration as to the function of the space is not clearly demonstrated within the virtual environment. The Autodesk Revit model is often

brought into the SimLab platform, material or colour is assigned, and some random furniture is placed from the SimLab library. The spaces were an afterthought and not part of a rigorous design process that explores 3D space making, as seen in Fig. 2. There is a misalignment between the intended functionality and user experience.

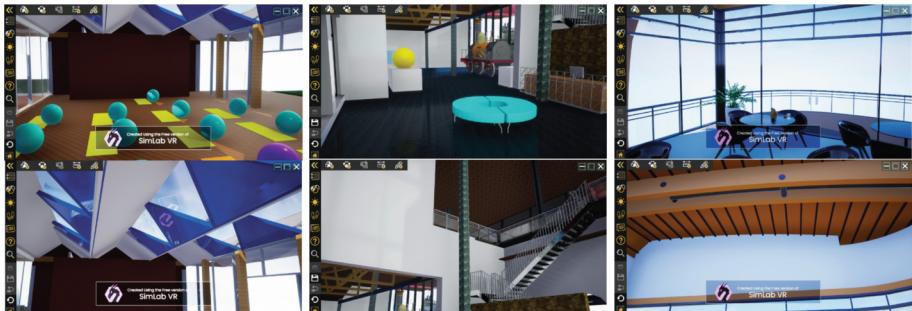


Fig. 2. Students 2, 3 and 5 demonstrate the lack of spatial understanding in the virtual environment.

The last theme, the effectiveness of the design concept, refers to the clarity of the virtual environment to demonstrate the design's intent towards the design strategy, material strategy, and design concept. The projects lack clarity into what is the driving concept for the creation of the space. No principles such as space-defining elements and space-enclosing elements are used, which is concerned with the design strategy. Furthermore, no strategic evidence of material usage indicates the use of material to inform space-making and the quality of space, as seen in Fig. 3. It is evident that the virtual environments created by the students do not reflect a clear understanding of architectural principles and goals.

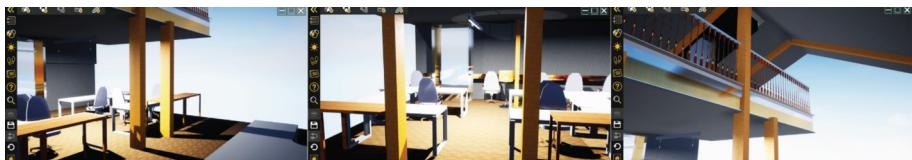


Fig. 3. Student 6 indicates a lack of indicating the effectiveness of the design concept.

Two key themes are derived from the data: VR as a presentation tool versus a design tool and design studios' readiness for VR technologies.

4.1 VR as a Presentation Tool Versus a Design Tool

Although the students had access to eight Oculus Quest 2 headsets, none of the students used the headsets during the semester. The headsets were only asked for a week before the projects were evaluated. This indicates that the students did not use VR during the design project but reverted to the traditional design process of printed 2D and 3D

materials to explore their designs. Therefore, VR was seen as an afterthought, which resulted in VR being used as a marketing tool to sell their design (presentation tool) rather than a design tool to understand their spatial design better. This aligns with the survey where students ranked VR's ability to assist with visualization first and that VR and spatial exploration and design fluency are ranked fourth and fifth. The findings align with Zhang [20], who states that the effectiveness of VR technology is well validated, but due to the attitudes toward embracing VR technology, there is a low adoption and usage. Therefore, it is evident that the students acknowledge and understand the benefits of using VR technologies. However, when implementing new technologies, the students revert to the traditional design processes they are more familiar with and trust.

4.2 Design Studios' Readiness for VR Technologies

The current traditional design project that is taking place in the design studio is not yet ready for adopting VR technologies as a process. If VR is not promoted within the design studio as part of the design process, VR will always be seen as the presentation tool; in some cases, people also refer to it as a 'gimmick'. Staff and students must be trained in the VR process and enhance the collaborative opportunity in a virtual studio for improve architectural design communication. The survey indicates that the students know VR's potential to reshape architectural education. However, as stated by Brazley [23], our current design studios are not geared towards transforming virtual architectural education. One can predict that there will be many challenges, of which resistance to change is one. The entrenched resistance from institutional staff over the years may be rooted in their established methods, and adapting to new technologies could bring difficulty in embracing digital and virtual design processes and technologies without traditional studio settings. Along with resistance comes barriers such as physical infrastructure allowing for virtual collaboration, hardware and software challenges and inadequate technical support.

5 Conclusion and Further Research

In conclusion, this study emphasizes the design studio's critical role in architectural education, emphasizing its importance as a collaborative and experimental learning platform. While students recognized VR's revolutionary potential, they had limited exposure and had difficulty to integrate VR into their design processes effortlessly.

The thematic analysis of the observational study brought to light significant gaps in aesthetic appeal, spatial understanding, and the effectiveness of design concepts within virtual environments. Notably, students tended to perceive VR more as a presentation tool than an intrinsic component of the design process, indicating the need for institutional readiness and comprehensive training.

The study underscores the importance of addressing challenges in effectively implementing VR, emphasizing the necessity for a paradigm shift in institutional mindset, infrastructure enhancements, and targeted training initiatives. Future research could delve into pedagogical strategies that facilitate the seamless integration of VR into traditional architectural design studios, aiming to enhance student engagement and improve the overall efficacy of VR in architectural education.

Furthermore, potential avenues for further research may include exploring strategies to overcome institutional resistance to VR adoption, analyzing the effectiveness of enhanced VR training modules, and conducting comparative studies across diverse architectural education programs. Long-term studies tracking students' attitudes and performance as VR becomes a consistent element in architectural design projects could provide valuable insights into the evolving landscape of architectural education.

References

1. Green, L.N., Bonollo, E.: Studio-based teaching: history and advantages in the teaching of design. *World Trans. Eng. Technol. Educ.* **2**(2), 269–272 (2003)
2. Demirbas, O.O., Demirkhan, H.: Focus on architectural design process through learning styles. *Des. Stud.* **24**(5), 437–456 (2003)
3. Kvan, T.: The pedagogy of virtual design studios. *Autom. Constr.* **10**, 345–353 (2001)
4. Eastman, C., Sacks, R., Lee, G., Teicholz, G.: BIM Handbook a Guide to Building Information Modelling for Owners, Designers, Engineers, Contractors and Faculty Managers, 3rd edn. Wiley, New Jersey (2018)
5. Hay, L.: Thinking skills for the information age. In: Costa, A.L. (ed.) *Developing Minds A Resource Book for Teaching Thinking*, 3rd edn. Hawker Brownlow Education, California, pp. 7–10 (2001)
6. Guoping, L., Yun, H., Aizhi, W.: Fourth industrial revolution: Technological drivers, impacts and coping methods. *Chin. Geogra. Sci.* **27**(4), 626–637 (2017)
7. Bloem, J., Van Doorn, M., Duivestein, S., Excoffier, D., Maas, R., Van Ommeren, E.: The fourth industrial revolution. *Things Tighten* **8**(1), 11–15 (2014)
8. Peters, M.A.: Technological unemployment: educating for the Fourth Industrial Revolution. *J. Self-Governance Manag. Econ.* **5**(1), 25–33 (2017)
9. Rafzan, R., Budimansyah, D., Fitriasari, S., Rahmat, R.: Development of critical thinking skills through the citizenship education course in the era of industrial revolution 4.0. *Adv. Soc. Sci. Educ. Humanities Res.* **418**, 256–418 (2019)
10. Richards, K.: *Qualitative Inquiry in TESOL*. Palgrave Macmillan, New York (2001)
11. Dixon, F., Cassady, J., Cross, T., Wiliams, D.: Effects of technology on critical thinking and essay writing among gifted adolescents. *J. Second. Gift. Educ.* **16**(4), 180–189 (2005)
12. Hopson, M.H., Simms, R.L., Knezek, G.A.: Using a technology-enriched environment to improve higher-order thinking skills. *J. Res. Technol. Educ.* **34**(2), 109–119 (2001)
13. Ding, Y., Li, Y., Cheng, L.: Application of internet of things and virtual reality technology in college physical education. *IEEEAccess* **8**, 96065–96074 (2020)
14. Hu, M., Luo, X., Chen, J., Lee, Y.C., Zhou, Y., Wu, D.: Virtual reality: a survey of enabling technologies and its applications in IoT. *J. Netw. Comput. Appl.* **178**, 102970 (2021)
15. Asfarian, A., Nurhadryani, Y., Ardiansyah, F., Hermadi, I., Ramadhan, D.A.: From immersive to metaverse: the gap of learning and technology in agriculture education application. *Jurnal Ilmu Komputer dan Agri-Informatika* **9**(2), 127–136 (2022)
16. Latif Rauf, H., Shareef, S.S., Najim Othman, N.: Innovation in architecture education: collaborative learning method through virtual reality. *J. High. Educ. Theory Pract.* **21**(16), 33–40 (2021)
17. Abd Majid, F., Mohd Shamsudin, N.: Identifying factors affecting acceptance of virtual reality in classrooms based on technology acceptance model (TAM). *Asian J. Univ. Educ.* **15**(2), 1–10 (2019)
18. Fakahani, L., Aljehani, S., Baghdadi, R., El-Shorbagy, A.M.: The use and challenges of virtual reality in architecture. *Civil Eng. Architect.* **10**(6), 2754–2763 (2022)

19. Vlah, D., Čok, V., Urbas, U.: VR as a 3D modelling tool in engineering design applications. *Appl. Sci.* **11**(16), 7570 (2021)
20. Zhang, M., Shu, L., Luo, X., Yuan, M., Zheng, X.: Virtual reality technology in construction safety training: extended technology acceptance model. *Autom. Constr.* **135**, 1–13 (2022)
21. Fussell, S.G., Truong, D.: Using virtual reality for dynamic learning: an extended technology acceptance model. *Virtual Reality* **26**(1), 249–267 (2022)
22. Cho, J.Y.: An investigation of design studio performance in relation to creativity, spatial ability, and visual cognitive style. *Think. Skills Creat.* **23**, 67–78 (2017)
23. Costello, P.J.M.: Effective Action Research: Developing Reflective Thinking and Practice. Bloomsbury Publishing, London (2011)
24. Welman, C., Kruger, F., Mitchell, B.: Research Methodology, 3rd edn. Oxford University Press, Cape Town, South Africa (2005)



Ink or Pixels: Exploring the Impact of Hand-Drawing Skills on Learning Abilities in a Tech-Savvy Generation Z

Jean-Pierre Basson^(✉) and Riëtte Kotzé

Nelson Mandela University, Port Elizabeth, South Africa

jpbasson@mandela.ac.za

Abstract. Architectural education is transforming significantly in response to the profession's demands and evolving societal needs. Studio-based learning remains central, as a hub for design exploration and skill development. However, the landscape of architectural education is shifting with the emergence of Generation Z, a cohort deeply immersed in digital technology. This study investigates the impact of reintroducing hand drawing into architectural education on Generation Z students' learning capabilities and design proficiency. Unlike previous generations, Generation Z's exposure to pervasive technology has led to declining traditional drawing skills, raising concerns about their ability to engage critically in the design process. Through a mixed-method research approach, this study compares the learning outcomes of first-year students engaging in hand drawing with third-year students utilizing computer-aided design (CAAD). Findings reveal that while both approaches meet learning attributes, Generation Z demonstrates higher levels of critical thinking when utilizing digital design tools. The study recommends the integration of workshops combining digital and manual drawing approaches to enhance students' cognitive processes and bridge the gap between traditional and digital methodologies. Further research is warranted to explore students' perspectives and experiences and to understand subsequent cohorts' attitudes and aptitudes towards drawing in architectural education. In conclusion, architectural education must adapt to the demands of the Fourth Industrial Revolution by embracing technology while preserving the essence of critical thinking and creativity fostered through traditional hand-drawing skills.

Keywords: Architectural Design · Architectural Education · BIM · Hand-drawing · Computer-Aided Architectural Design

1 Introduction

Architectural education is constantly changing to meet the needs of the profession and society. It is based on the belief that design should be considered critically on both an emotional and intuitive level. Studio-based learning is a common teaching method in design fields, and it is especially important in architectural education. The studio is a place where students can work on design projects, learn, and share their skills.

Communication is essential for architectural students to succeed in the design studio, and the tools they use to communicate are critical to their ability to learn and collaborate [1]. Communication, problem-solving, critical thinking, creativity, and collaboration are also seen as the key 21st century learning skills required by students studying to enable them to become global citizens and succeed in this more competitive world [2].

Contemporary architectural education still follows the basic principles of the Beaux-Arts [3, 4] (the studio) and Bauhaus [5, 6] (curriculum), with a focus on design education in the design studio and supporting modules separate from the design studio [3]. However, the contemporary university student, often categorized as Generation Z, is a tech-immersed individual born after 1995 [4–8]. Students enter the classroom with computers, tablets, and cellphones that connect them to a world of knowledge that they can access quickly and easily [8]. This ease of access to information allows for procrastination, leaving little opportunity to critically scrutinize the content [9]. They feel they must be connected 24/7 and want everyone else to be online and available quickly. In the classroom, Generation Z relies heavily on technology to learn. They favor PC-recorded lectures over taking notes, regard lecturers as ‘come and entertain me’, ask questions using an online medium such as a blog or email, and do not want to wait for a lecturer’s response [7]. This prompts an inquiry into the relevance of current traditional architectural education models in the context of Industry 4.0 and a generation that is deeply connected and fluent in technology such as virtual reality, augmented reality, Building Information Modelling, and Artificial Intelligence (AI). Previous research on the debate of hand-drawing versus digital design studies conducted by authors such as Hannibal, Brown and Light [9] investigated the importance of sketching as an integral part of the design process to promote creativity and effective communication but noted the challenge to seamlessly integrate sketching with digital design. Lastly, Brown’s [10] research focused on the importance of visualization for effective communication and to close the gap between artistic expression and scientific rigor. Therefore, the importance of traditional hand-drawing skills for architectural education and the need to integrate digital technologies is well documented in the literature.

This study delves into the question: How does the reintroduction of hand-drawing into architecture education influence the learning capabilities and design proficiency of Generation Z students? Unlike previous generations, where proficient drawing skills were ubiquitous due to the absence of pervasive technology, Generation Z has grown up in a digital era where manual drawing is less emphasized. Drawing, once a daily activity, has now been largely supplanted by digital mediums. This shift raises concerns about the erosion of traditional drawing skills among Generation Z students, whose familiarity with technology surpasses their proficiency in hand-drawing. The aim of this study is to evaluate the role of hand-drawing in Industry 4.0 architectural education for Generation Z and assess its impact on their learning abilities to promote a context of inclusivity in skillsets and aligned to student success based on the 21st learning skills, especially communication. Therefore, the study wishes to identify any challenges or limitations associated with the integration of hand-drawing in architecture education for Generation Z students and propose strategies for addressing them. This study was undertaken within the Department of Architecture at Nelson Mandela University, located

in Gqeberha, South Africa. It was conducted as part of the Bachelor of Architectural Studies program.

2 Literature Review

Whilst considerable literature debates the merits of hand-drawing versus computer-aided drawing in architecture, there remains a gap in understanding the transition between these two approaches, both involving the left-brain's neurological processes. Exploring this transition is essential to uncover how both methods and tools complement each other and contribute to the overall design process.

2.1 Architectural Hand-Drawing

Many written works emphasize the importance of hand-drawing in the field of architectural design. According to Pallasmaa [11], the hand could grasp the palpable essence of an idea and translate it into a concrete image. Akin [12] underscores the significance of drawing in the design process, characterizing drawing methods as avenues to communicate the architect's design intentions. Olsberg [13] tells the story of "*Carlo Scarpa opening all his courses in design at the University of Venice by demonstrating the art of sharpening a pencil. That was the precise point, he claimed, from which all architecture proceeds*".

In the realm of creative fields, there is a growing focus on investigating the complex relationship between manual dexterity and cognitive processes, emphasizing the dynamic interplay between physical actions and mental functions. This concept extends to diverse creative domains where the act of shaping and producing takes a pivotal role in the creative process. Pallasmaa [11] describes the pencil in hand as a link between the imagination and the paper, suggesting that during the creative act, the metaphorical 'bridge' may be overlooked, and the resulting image appears almost as an automatic projection of the imaginative mind.

Design drawing is an iterative and collaborative process that encompasses capturing ideas, discerning functions and meanings within the drawings, and discovering novel forms while integrating them into the design [2]. Edwards [14] emphasizes that the process of drawing holds significance not only as a means of communicating with others but also as a tool that aids designers in visually perceiving and comprehending the forms they are working with. The intriguing dynamic between the brain's uncertainty and the hand's attempt to interpret has been a subject of research. Despite efforts to replicate this phenomenon using digital tools, success has been limited [2].

Gallov [15] rightfully states that self-discovery occurs through the act of drawing by hand. He elaborates on the significance of discerning when and how to leverage software for visual communication, underscoring that maintaining a distinctive voice is paramount for designers.

2.2 The Impact of CAAD on Architectural Education

In contemporary architectural practice, CAAD has become increasingly indispensable. The benefits of CAAD have been widely researched and experienced, and the need

for architecture graduates to enter the profession with the proficiencies to use CAAD software has become increasingly urgent in recent years [16]. Consequently, the onus rests on the Architectural Learning Sites (ALSs) to equip architecture graduates with the necessary CAAD skills to become valuable employees.

It was not until the 1980s that CAAD began to be widely adopted in architectural schools [17]. One of the main drivers for the transition to CAAD was the increasing complexity of architectural design. As buildings became more complex, it became more difficult to produce accurate and detailed drawings by hand. CAAD software offered several advantages over hand-drawing, including increased accuracy, speed, and flexibility [18].

The integration of CAAD in architectural education emerged gradually. Initially, hand-drawn sketches and drawings remained primary, serving as the foundation for design exploration [19]. However, the advent of computer software for drafting and design, such as AutoCAD, 3D modeling programs and finally BIM, gradually found their way into curriculums [17, 20].

The introduction of CAAD in the university design studio remains a much-deliberated issue. Whilst some academics believe that students should not be exposed to this digital tool in the undergraduate program, others are of the opinion that it should be introduced as soon as possible for students to become proficient and use the software as a design aid [20].

Computer-generated visualizations are more informative and easier to understand than traditional design methods [3] but designers need CAAD skills to use them effectively [21]. Information visualization can be used to transfer knowledge and provide feedback to designers and non-designers [21, 22]. However, the process is difficult to manage and requires consideration of the information's depth, time constraints, cognitive differences, and relevance. Despite these challenges, information visualization is an essential part of the communication process in design. Therefore, the pencil remains a popular conceptual design tool due to its low cognitive demand [21, 22].

2.3 Critical Thinking and Industry 4.0 (4IR)

Critical thinking represents the pinnacle of human cognitive abilities, enabling the formulation of judgments—a distinct trait setting humans apart from other species. Communication involves the effective articulation of thoughts and ideas, encompassing both verbal and nonverbal forms. Collaboration, closely linked to communication, necessitates the adept functioning within diverse teams, involving information sharing, flexibility and a collective pursuit of common objectives. Creative thinking, identified by Bloom, Krathwohl, and Masia [23] as the highest-ranking thinking skill, is a rational process crucial for adapting to challenges by generating innovative and effective responses. Creative thinking essentially builds upon critical thinking skills [23–28].

In the realm of 21st-century learning, skills are categorized into hard skills such as digital literacy, humanistic skills encompassing career and life skills, and a set of soft skills (collaboration, communication, and creativity) along with cognitive processing skills like problem-solving and critical thinking. These cognitive abilities allow engagement with the higher echelons of educational taxonomies, notably analyzing, evaluating, and creating within Bloom's framework [23, 29], thus recognized as higher-order

thinking skills (HOTS). Consequently, this study directs its focus towards the cognitive processing skills, particularly critical thinking, within the spectrum of 21st-century learning skills.

The shift from the industrial age to the information age has transformed learning methods. Initially, education focused on memorization and low-level thinking skills, suited for assembly-line work [30]. The advent of the information age and the subsequent Fourth Industrial Revolution (4IR) altered this landscape, demanding higher order thinking due to technological advancements [30–32]. The 4IR emphasizes digital, physical, and biological technologies, reshaping job requirements and enhancing global competition [33]. This revolution introduces constant information access, faster computing, automation replacing routine jobs, and enhanced interconnectedness [31]. As the 4IR unfolds, challenges emerge. The focus must shift to incorporating technological systems effectively, empowering citizens, and addressing values and ethics in technology usage. The infusion of technology in problem-solving remains an underexplored area in education, raising questions about the significance of memorization and the development of information literacy [30]. Education faces challenges in adapting to the 4IR, including social changes, misinformation, political influences, and limited resources [31]. The shift to a ‘copy and paste’ generation has raised concerns about the depth of understanding and critical thinking skills [34].

3 Methodology

This study used a mixed-method research approach to portray, decode, translate, and focus on a phenomenon in its natural setting by applying an action research (AR) methodology to generate primary data. According to Costello [35], action research has a practical and problem-solving emphasis and is carried out by individuals, professionals, and educators with the goal of improving education practice. Action research is not a single approach but rather a collection of features that include adaptable design and participant involvement. Welman et al. [36] emphasize the necessity of involving participants in the study in action research. Various data collection strategies, such as observational studies, interviews, and documents, can be used in action research.

This study uses a comparison strategy to collect primary data, comparing first-year architectural students completing their first architectural presentation by hand, and third-year architectural students completing their first computer-aided design presentation in a longitudinal observational study. This permitted the distinction between the traditional approach of learning how to produce hand presentations and the CAAD method of making presentations. The cohort comparison was not intended to assess design quality but rather the ability to transfer knowledge and track the learning curve in each mode of presentation. Before the first-year students present their first design project, they are taught the basic drawing conventions in architectural visualization, including how to technically draw plans, sections, and elevations, followed by workshops in pencil, ink and color rendering. After mastering the foundational skills, students embark on their first design project. Similarly, the third-year students engage in a one-week workshop unpacking Autodesk Revit as a modelling and presentation tool by working with an existing building before they model and present their design project. In this study, the

same project, the design of a personal living unit, was used for both the hand-drawn presentation and computer-aided presentation.

The study used a descriptive analysis approach with coding [36, 37]. The South African Council of the Architectural Profession's (SACAP) guidelines outlines the different work stages were the minimum criteria for the student's projects: Workstage three, which speaks to the design development relating to the form of the building, structure resolution, finishes, and the performance of the building, was used to develop the three deliverables, namely function, aesthetics, and decoration. A matrix was delivered from the information to ensure that the extent of the work stage was analyzed universally to remove bias in the research. The study employed Wei and Yuen's [38] framework for informative and summative assessment to delineate the learning trajectory experienced by students within each of the two methodologies. Table 1 provides a synopsis of the attributes.

Table 1. Informative and summative assessment framework adopted from Wei and Yuen (2016).

| | |
|----------------------------|--|
| Professional documentation | Documentation is the level of preparation and effort placed into effectively presenting project materials |
| Content Completion | The extent to which the project comprises the specified deliverables, ranging from less than 50% to 90–100% coverage, is referred to as completion |
| Content accuracy | Accuracy relates to the precision and correctness of the information delivered in project deliverables, with coverage levels ranging from bad to outstanding |
| Formatting | Formatting evaluates conformance to formatting criteria, with scores ranging from minimal effort to almost flawless uniformity |

The Nelson Mandela University's first-year architecture average class size is fifty students, while the third-year average class size is thirty-five students. Fourteen students participated in the study, which stretched over their first year (2021) and third year (2023), which, on average, represents a third to half of the class.

4 Results and Findings

The student work (hand drawings, physical and digital models and digital presentations) was graded using a rubric that emphasized the qualitative aspects of visualization (professional document and formatting) and design function (content accuracy and content completion).

Tables 2 and 3 indicate the extent to which the students meet the learning attributes. These tables offer insights into the learning curve performance, on a scale ranging from poor to good, and quantified through a MS ranging between 1.00 to 5.00. Notably, all the MSs are higher than the midpoint score of 3.00, indicating that both methods, on

average, meet the learning attributes frequently as opposed to infrequently. However, it is prominent that 4 / 4 (100%) of the MSs for the use of CAAD are $>4.20 \leq 5.00$, and thus can be deemed to be meet expectations to be between above average to good / good compared to the hand-drawing method 2 / 4 (50%). Notable 2 / 4 (50%) MSs for the hand-drawing method are $>3.40 \leq 4.20$ and thus be deemed to meet expectation to be between average to above average / above average. In comparing the two cohorts' attributes that is deemed to be above average to good / good, content completion and professional document MSs differ by 0.15 to 0.21 indicating that the students are on par. The difference lies within the formatting and content accuracy MSs that differs by 1.07 demonstrating that the CAAD method allows for a more critical engagement with the design content compared to the hand-drawing method where the individual's background and skills influence the formatting (visualization) that influence the content accuracy.

Table 2. 2021 Cohort using hand-drawing method to meet expectation levels.

| Learning attribute | Response (%) | | | | | MS | Rank |
|-----------------------|--------------|---------------|---------|---------------|------|------|------|
| | Poor | Below Average | Average | Above Average | Good | | |
| Content completion | 0.0 | 7.1 | 7.1 | 0.0 | 85.8 | 4.64 | 1 |
| Professional document | 0.0 | 14.3 | 0.0 | 21.4 | 64.3 | 4.36 | 2 |
| Formatting | 7.1 | 7.1 | 14.3 | 28.6 | 42.9 | 3.93 | 3 |
| Content accuracy | 7.1 | 7.1 | 43 | 21.4 | 21.4 | 3.43 | 4 |

Table 3. 2023 Cohort using CAAD for meeting expectations levels.

| Learning attribute | Response (%) | | | | | MS | Rank |
|-----------------------|--------------|---------------|---------|---------------|-------|------|------|
| | Poor | Below Average | Average | Above Average | Good | | |
| Formatting | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 | 5.00 | 1 |
| Content completion | 0.0 | 0.0 | 0.0 | 21.4 | 78.6 | 4.79 | 2 |
| Professional document | 0.0 | 0.0 | 0.0 | 42.9 | 57.1 | 4.57 | 3 |
| Content accuracy | 0.0 | 0.0 | 7.1 | 35.7 | 57.2 | 4.50 | 4 |

The learning attribute **professional documentation** relates to accessibility information quality present in work produced by the students. Accessible information is

information that is useable, reliable, and readily available to people who require it [39]. Although both methods achieved a MS of meeting the expectation of above average to good, the hand-drawing method's ranking is higher than the CAAD method, but MS is lower. The nature of the drawings presented by the hand-drawing cohort does lack information for the reader to fully understand the extend of the design. The design sections speak to 3D space making, while 2D space making is lacking in plans. The hand-drawing submissions also lack materiality and technological information compared to the CAAD method that expresses architectural space making in both 2D and 3D, indicating materiality and technological information as indicated in Fig. 1.

The learning attribute **content completion** relates to contextual information quality present in the work produced by the students. Contextual information is the background knowledge offered that serves as a context for understanding and interpreting a specific piece of information [40]. Similar to professional documentation, the hand-drawing method ranked higher than the CAAD method, but the CAAD method obtains a higher MS of 4.79 compared to the 4.64. Drawings produced by the hand-drawn cohort lack clarity which provide context for each of the individual drawings. Plans do not contain any expression or understanding of the material. The students employed the CAAD method to present information that is contextualized where one drawing provides background to another, for example the plan drawing correlate to the section and elevation, as indicated in Fig. 2.

The learning attribute **content accuracy** relates to intrinsic information quality present in the work produced by the students. Intrinsic information is defined as the meaning of a given piece of information regardless of context or relevance to the situation, but the features of the information determine it [41]. It is notable that both cohorts scores reflect that accuracy is ranked fourth. However, The CAAD method obtained a MS of good, compared to the above average of the hand-drawing method. This indicates that both methods act as a tool to complete the task, and that the tool is only as good as the user. The hand drawn method cohort's information is more intrinsic, with each individual drawing (plan, section, and elevation) considered as a separate drawing with a set of information specific to that drawing on it, rather than as a holistic building compared to the CAAD method which allowed the 2023 cohort to develop the project holistically, combining intrinsic, contextual, representational, and accessible information quality.

The last learning attribute, **formatting**, relates to representational information quality in the work produced by the students. The ability of information to represent a specific notion or idea, as well as how well the information represents the 'object' it is supposed to depict, is referred to as representational information [42]. The data indicates that there is a direct link between the designer and the expression of ideas on the paper using the hand-drawing method. The CAAD method acts as a mediator of skill that can assist students to clearly communicate their ideas and intentions, as indicated in Fig. 3.

From this research, the following theme is prominent: digital environment and creativity erosion. The literature indicates a concern of the impact of information technology on problem-solving which is a major concern for educators teaching in Industry 4.0 [30, 43]. Professionals in the built environment are also afraid that technology may reduce productivity, disrupt established workflows, and fix fundamental design problems automatically. The drawings in both cohorts are comparable in nature, ranging from 2D

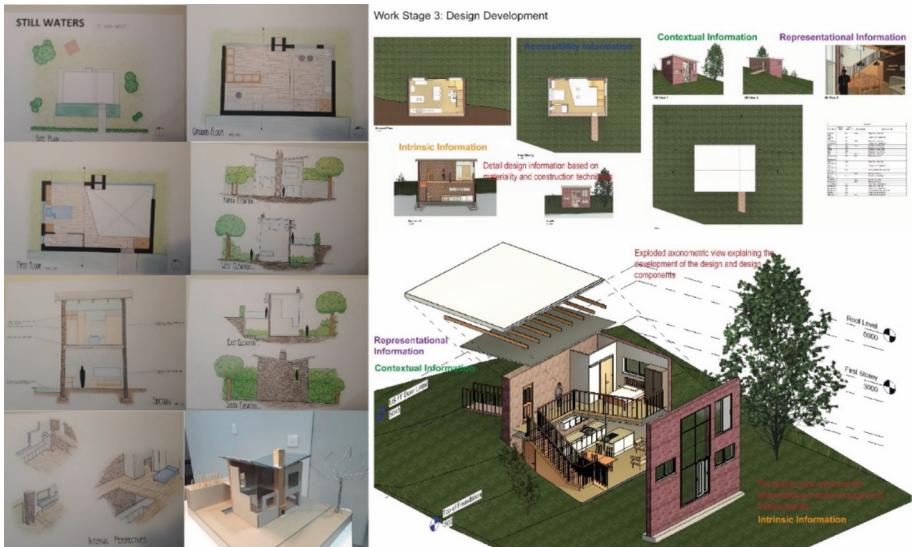


Fig. 1. Professional documentation, through accessible information from the hand-drawing (left) and CAAD (right).



Fig. 2. Project completion as seen by the hand(left) and CAAD (right) method.

plans, sections, and elevations to 3D axonometric and isometric representations. However, when the drawing is examined based on qualitative learning attributes by looking at the quality of the information, which the literature defines as intrinsic, contextual,



Fig. 3. Formatting based on the hand (left) and CAAD (right) method.

representational, and accessible information quality, similarities and differences occur. Both cohorts share above-average to good professional documents and content completion, therefore using the medium (hand-drawing or CAAD) as a communication tool as indicated in the work of Brown [10]. The literature indicated that hand drawings plays an important role in design development and must be seamlessly incorporated in digital design training [9]. However, it is evident that Generation Z is more digitally inclined compared to the Generations prior, and lacking basic drawing skills. The learning attributes formatting, and content accuracy are lower in the hand-drawing cohort than the CAAD drawing indicating that the act of drawing is not a natural talent and trait of Generation Z. Due to the personal limitation, it is limiting students to engage critically in the design process.

These findings offer substantial implications for architectural education, emphasizing a generation where the focus has shifted from drawing with a pencil on a page to a digital ‘pull and push’ of objects. Carnevale [44] reinforces the difference between the terms *designing* and *drawing*. She goes on to say that many tutors in design courses agree that students should be motivated to engage in manual drawing to acquire design skills rather than relying on computers. To craft a masterpiece in design, it necessitates possessing a creative mindset, originality, and a clear concept, typically neurologically dependent on the right lobe of the brain. To translate it into a visual representation requires a specific skill set, typically dependent on the left lobe of the brain [44] which Generation Z is lacking.

5 Conclusion, Recommendation and Further Research

Architectural education, deeply rooted in historical models such as the Beaux-Arts and Bauhaus, has faced transformative challenges in the era of Industry 4.0. This study delves into the reintroduction of hand-drawing into architectural education influence the learning capabilities and design proficiency of Generation Z, who is currently occupying our classrooms and studios.

The results illustrate a nuanced and complex picture. The comparison of the 2021 hand-drawing cohort and the 2023 CAAD cohort reveals that while both approaches meet the learning attributes, Generation Z shown higher levels of critical thinking when using digital design tools as design method than using the hand-drawing method compared to the pre-digital or pre-technology generations. The study concluded that the introduction and promotion of digital design is well aligned to Industry 4.0 and the current student studying architecture, but where the focus were in 2005 on the integration of digital tools into the curriculum, the focus for Generation Z will be to upskill in hand-drawing, since many ALSs require first-year architecture students complete design projects using hand-drawing techniques, and only during their senior years are introduced to digital design process. This may hinder students from succeeding in their studies, and cause learning frustrations that can lead to the discontinuation of studies or loss of interest in architecture.

Based on the findings, the study recommends that workshops combining digital and hand-drawing techniques in order to effectively communicate design intent, may be provided. These courses might show students how to combine digital tools like tablets and styluses with analogue sketching. By integrating the advantages of both methodologies, students can improve their digital talents while sharpening their manual drawing abilities promoting cognitive processes.

While this study provides valuable insights into the learning curve Generation Z experience using both a hand-drawing and CAAD methodology, there is a need to conduct further research into the student's voice by unpacking the students' perspective, experience, and challenges. Furthermore, given the rapid evolution of technology and the emergence of new generations, future research could delve deeper into how subsequent cohorts attitude and aptitude to drawing entering architectural education. In conclusion, architectural education must adapt to the demands of the 4IR by embracing technological tools, especially CAAD, while preserving the essence of critical thinking and creativity using traditional hand-drawing skills.

References

1. Basson, J.-P.: The use of Building Information Modelling as a tool to improve informed design communication between student and lecturer during critique sessions. In: Construction Management. Nelson Mandela University, Port Elizabeth, p. 363 (2018)
2. Do, E.Y.-L.: Design sketches and sketch design tools. *Knowl. Based Syst.* **18**(8), 383–405 (2005)
3. Salama, A.: Pedagogical approaches in architectural education. *J. Arch. Educ.* **58**(3), 162–173 (2005)

4. Cilliers, E.: The challenge of teaching Generation Z. *PEOPLE: Int. J. Soc. Sci.* **3**, 188–198 (2017)
5. Dangmei, J., Singh, A.: Understanding the Generation Z: the future workforce **3**, 1–5 (2016)
6. Seemiller, C., Grace, M.: *Generation Z Goes to College*. Wiley, Hoboken, New Jersey (2015)
7. Turner, A.: Generation Z: technology and social interest. *J. Individ. Psychol.* **71**(2), 103–113 (2015)
8. Igel, C., Urquhart, V.: Generation Z, meet cooperative learning: properly implemented cooperative learning strategies can increase student engagement and achievement. *Middle Sch. J.* **43**(4), 16–21 (2012)
9. Hannibal, C., Brown, A., Knight, M.: An assessment of the effectiveness of sketch representations in early stage digital design. *Int. J. Arch. Comput.* **3**(1), 107–125 (2005)
10. Brown, A.G.: Visualization as a common design language: connecting art and science. *Autom. Constr.* **12**(6), 703–713 (2003)
11. Pallasmaa, J.: Embodied and existential wisdom in architecture: the thinking hand. *Body Soc.* **23**(1), 96–111 (2017)
12. Akin, Ö.: *Psychology of Architectural Design*. Pion, London (1986)
13. Olsberg, N.: The evolving role of the drawing. *Arch. Rev.* **233**(1395) (2013)
14. Edwards, B.: *The New Drawing on the Right Side of the Brain*. HarperCollins, New York (2001)
15. Gallov, G.: Observations on drawing: the art of architecture. *Arch. Des.* **89**(6), 52–57 (2019)
16. Chohan, A.H., Awad, J.: Shaping the architects of tomorrow, interplay of teaching philosophies and practice requirements: an empirical taxonomy of professional architectural practice in the UAE. *Buildings* **13**(5), 1231 (2023)
17. Woodbury, R.: *Elements of Parametric Design*. Routledge, London (2010)
18. Kensek, K., Noble, D.: *Building Information Modeling: BIM in current and future practice* (2014)
19. Pallasmaa, J.: *The Thinking Hand: Existential and Embodied Wisdom in Architecture*, 1st ed. AD primers. Wiley, Chichester (2009)
20. Salman, H., Laing, R., Conniff, A.: The changing role of CAAD in the architectural design studio. *Built Hum. Environ. Rev.* **1**, 25 (2008)
21. Croser, J.: Is the solution also the problem? *Arch. J.* **7**(12), 51 (2001)
22. Burkhard, R.A. Learning from architects: the difference between knowledge visualization and information visualization. In: *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004* (2004)
23. Bloom, B.S., Krathwohl, D.R., Masia, B.B.: *Taxonomy of Educational Objectives*. Pearson Education, London (1984)
24. Saputra, M.D., et al.: Developing critical-thinking skills through the collaboration of jigsaw model with problem-based learning model. *Int. J. Instr.* **12**(1), 1077–1094 (2019)
25. Bekteshi, E.: The ‘Four Cs’—collaboration, communication, critical thinking and creativity at the Faculty of Arts (FLUP), University of Porto, Porto, Portugal. *J. Int. Soc. Res.* **10**(50) (2017)
26. Guo, Z.: The cultivation of 4C’s in China: critical thinking and communication. *Manag. Appl. Soc. Sci.* 11–16 (2016)
27. Khoshneshin, Z.: Collaborative critical thinking in online environment. *Procedia Soc. Behav. Sci.* **30**, 1881–1887 (2011)
28. Wang, S.-Y., et al.: Socrates, problem-based learning and critical thinking—a philosophic point of view. *Kaohsiung J. Med. Sci.* **24**(3), S6–S13 (2008)
29. Sivaraman, S.I., Krishna, D.: Blooms taxonomy—application in exam papers assessment. *Chem. Eng. (VITU)* **12**(12), 5–9 (2015)
30. Hay, L.: Thinking skills for the information age. In: *Developing Minds: A Resource Book for Teaching Thinking*, pp. 7–10 (2001)

31. Budimansyah, D., Fitriasari, S.: Development of critical thinking skills through the citizenship education course in the era of industrial Revolution 4.0. In: 2nd Annual Civic Education Conference (ACEC 2019). Atlantis Press (2020)
32. Swart, R.: Critical thinking instruction and technology enhanced learning from the student perspective: a mixed methods research study. *Nurse Educ. Pract.* **23**, 30–39 (2017)
33. Peters, M.A., Jandrić, P.: Education and technological unemployment in the Fourth Industrial Revolution. *Educ. Philos. Theory* **49**, 1–6 (2016)
34. Richards, K.: Qualitative Inquiry in TESOL. Springer, New York (2003)
35. Costello, P.J.: Effective Action Research: Developing Reflective Thinking and Practice. Bloomsbury Publishing, London (2011)
36. Welman, C., Kruger, F., Mitchell, B.: Research Methodology. OUP Southern Africa, Cape Town (2006)
37. Creswell, J.W.: Qualitative Inquiry and Research Design: Choosing Among Five Approaches. SAGE Publications, Thousand Oaks (2013)
38. Wu, W., Luo, Y.V.: Pedagogy and assessment of student learning in BIM and sustainable design and construction. *J. Inf. Technol. Constr. (ITcon)* **21**(15), 218–232 (2016)
39. Hersh, M.A., Johnson, M.A.: Assistive Technology for Visually Impaired and Blind People, vol. 1. Springer (2008)
40. Holland, J.: “Looking behind the veil” Invisible corporate intangibles, stories, structure and the contextual information content of disclosure. *Qual. Res. Financ. Mark.* **1**(3), 152–187 (2009)
41. Collier, J.: Intrinsic information. In: Information, Language and Cognition: Vancouver Studies in Cognitive Science, vol. 1, pp. 390–409 (1990)
42. Shea, N.: Naturalising representational content. *Philos Compass* **8**(5), 496–509 (2013)
43. Hopson, M.H., Simms, R.L., Knezek, G.A.: Using a technology-enriched environment to improve higher-order thinking skills. *J. Res. Technol. Educ.* **34**(2), 109–119 (2001)
44. Carnevale, V.: To design or to draw. Two different verbs, two different abilities, one result. *Int. J. Des. Objects* (2015)



VR-Based Computer Gaming Application in Digital Rehearsing of Mobile Crane Construction Operations

Tan Qu^(✉) and Marwan Shaban

School of Construction, Design, Engineering and IT, Seminole State College of Florida,
Sanford, FL 32771, USA
qut@seminolestate.edu

Abstract. With the advancement of virtual design and construction (VDC) technologies and the increasing adoption of such technologies by construction companies, animation, visualization and simulation of the construction process continues to be a domain requiring more research efforts. Users of n-dimensional (3+ dimensions) models of construction projects constantly identifies potential uses of model data in planning and management of construction processes and drive the innovations in existing workflow or integrated solutions to sustain technology adoptions that will result in long-term positive returns on the investment in such VDC technologies. This paper first examines the approaches explored in recent years to animate, visualize, or simulate various construction operations and discusses the differences and level of complexities/limitations of these various approaches. The paper then proceeds to introduce an approach by the authors of using Unity (a computer gaming application technology) and SteamVR (virtual reality system) to simulating various mobile crane operations in concrete tilt wall construction process. Operations simulated in the VR environment include various modes of crane maneuvers and lifting/setting operations of concrete tilt wall panels to identify potential constructability issues that may present in real construction operations. Lessons learned and industry user feedback will be discussed along with the directions identified for future research efforts.

Keywords: VDC · VR · Unity · Crane · Constructability · Gaming

1 Introduction

1.1 Construction Industry Trend and Existing Research in Crane Lift Planning and Optimization (LPO)

Market studies and forecasts by leading construction market analytic and research firms such as Allied Market Research [1], Dodge Construction Network [4], FMI [5], and McKinsey and Company [13] indicate modular, prefabrication, precast and other highly productive construction methods will continue to be a growing industry trend for the foreseeable future. The advantages of these construction methods in enhancing construction efficiency and quality while shortening the onsite construction time are well

recognized [16]. Due to the substantial dimensions and weights of the prefabricated building components or modules, cranes have become an indispensable equipment in these construction methods. The constructability of these projects therefore to a great extent depends on the operational feasibility of the selected crane in executing tasks such as offloading offsite-produced components or modules from delivery vehicles, deploying the components to the planned lifting locations, and installing these components in the final positions, all in the ever-changing and dynamic construction site environment that is unique to each project. Because of the high hourly equipment rates, analyzing and optimizing path planning and lift operations are essential in the overall success of the project cost and schedule performance.

Importance of crane selection, lift planning and optimization (LPO) is well recognized in the construction research community. In a study of cataloging the existing research in crane lift planning and optimization between 1990 and 2019, Zhang and Pan [22] identified 108 journal publications in this research subject area. Researched topics ranged from crane selection, path planning, collision analysis, location optimization, cycle time estimation, and lifting schedule optimization. By grouping existing research work into three categories: entity-related (crane selection and location optimization), process-related (cycle time estimation, path planning, collision analysis, and lifting schedule optimization), and integrated issues, Zhang and Pan found the top three researched issues for the last decade have been location optimization, path planning, and collision analysis. This study also found the majority (more than 2/3) of the existing research effort focused on mathematically based approach (artificial intelligence-based, analytic models, and computer simulations), while virtual model integrated computer animation approach accounted for the remaining 30% research work. Although Zhang and Pan study did not break down existing research work into categories of tower/fixed crane versus mobile crane, their study identified a critical need for further work in the area of path planning integrated with collision analysis/optimization for congested and dynamic construction sites that are common on most construction projects.

1.2 Building Information Modelling (BIM) Integration in Mobile Crane Simulation Research

Technically speaking, telescopic mobile cranes (wheeled) and lattice crawler cranes (tracked) are two different types of cranes, but for the purpose of this paper they are collectively considered as mobile cranes to make a distinction from fixed tower cranes that are also widely utilized in the construction industry. Widespread adoption of BIM and virtual design and construction (VDC) technologies in the construction industry has given researchers and practitioners new tools in simulations of various aspects of the mobile crane construction process. More and more engineering, design and construction firms are incorporating VDC technologies in their business operations and workflow. Design modeling and constructability analysis of the project incorporating essential construction means and methods have become even more so relevant in the concept of Digital Twin (DT) model validation and development process. Kalantari et al. developed and tested a DT model that allows adjustments in a physical floor plan model to be mirrored and analyzed in a digital platform [9]. Liu et al. presented a safety risk control framework for prefabricated building lift operations based on a DT model that incorporated BIM model

and field sensors [12]. Tu et al. presented a proof-of-concept version of a mixed reality application and controlling and monitoring a DT based overhead crane [20]. Lin et al. developed a communication and simulation system integrated with BIM, game engine and VR technologies for healthcare design in a semi-immersed VR environment [11]. Tak et al. explored the use of Autodesk Navisworks program for approximate mobile crane deployment interference checking in a simulated construction site environment that incorporated site layout and boundaries, temporary facilities, existing objects, and obstructions [19]. Han et al. introduced a methodology of applying a dynamic graphical description of 3D visualization to simulate various scenarios with different crane models to select the most effective and efficient crane operation [8]. Boton presented a framework using 4D BIM model and VR technologies in Autodesk Navisworks software in conducting constructability analyses in an immersive virtual environment [2].

In a study aimed to analyze the architecture/engineering/construction (AEC) industry paradigm shift, Brozovsky et al. defined Construction 4.0 as a “transformative framework” consisting of three main transformative trends on direct digital construction: industrial production (prefabrication, modularization, robotics, etc.), cyber-physical systems (sensors, robots, drones, etc.), and digital technologies (BIM, AI, VR, simulation, etc.) [3]. To achieve the objectives of Construction 4.0, one can recognize importance of research advancements in BIM, VDC, VR, and simulation in prefabrication and modular construction methods.

2 Comparisons of Existing 3D-Based Crane Selection and Lift Plan Approaches

2.1 Existing Research in 3D-Based Crane Selection and Lift Plan Optimization

With prevalent 3D-based modelling and simulation software and technologies, pure mathematical or algorithm-based approaches in crane selection and lift plan research have seen decreased uses. Recent research has largely transitioned to 3D-based crane selection and lift planning processes. For example, Tantisevi and Akinci presented an approach that determines possible locations of mobile cranes based on discrete-event simulation of crane operations incorporating dynamic behaviors of cranes in a 3D visualization environment [17, 18]. Pan et al. also proposed an approach for automated method for optimizing feasible locations of mobile cranes based on 3D visualization [14]. Zi et al. experimented with a parallel robot method to determine the localization of multiple cranes based on the multi-point localization method in a 3D grid environment [23]. Pooladvand et al. presents a crane simulator system developed in the virtual reality (VR) environment integrated with a database of comprehensive lift studies and a detailed crane path planning system [15]. Kayhani et al. presented a VR environment based on Unity game engine in which the user can experience lifting process in an immerse virtual environment [10]. Guo et al. proposed an approach for the simulation of lifting operation at the candidate points, feasible location points and crane types through three constraint checks (environment constraint, operation constraint, and safety constraint) and tested the approach in Unity 3D gaming system that included input module, selection module, and optimization module [7].

2.2 Crane Selection and Lift Planning Practices in the Industry

Historically, crane selection practices in the industry had been a chart-based manual approach. Construction managers/planners of a project would first consider the maximum weight and dimensions (length, width, and depth) of the load object to be lifted, required maximum operating radius, and hook height required to clear identified obstructions. This information is then compared to the operating data charts of various crane models. A particular crane model with a boom/jib attachment combination meeting the operating parameters is then selected. Nowadays, with the advancement of computer technologies, crane selection has become an algorithm-based “human-in-loop” interactive visualization/simulation process. Leading crane manufacturers and lift planning software developers have developed sophisticated programs for crane selection and lift simulation in 3D environment. Figure 1 shows the crane planning process in Liebherr Crane Planner 2.0 program. A user can define the workspace, select specific crane model with variations (boom, jib, etc.), configure rigging, define load objects, select and rig the load object, and define obstruction objects. The user can then interactively simulate the lifting process by moving the crane within the 3D workspace, rotating upper carriage for the desired orientation, adjusting boom angle and jib angle, adjust hook angle, and adjusting hook height to determine if the desired lift operation can be successfully completed, all at the same time as crane real-time operating/feedback data (such as total load, radius, maximum ground force/pressure, etc.) being dynamically checked/displayed. 3D BIM objects from a project design model can be imported as either load objects or obstructions to further enhance the realism of the analysis. 3D Lift Plan is a lift planning software widely used in the concrete tilt wall industry. A user can go through the load wizard to define the load object, rigging wizard to define parameters such as sling leg length, leg height, and sling angle, jobsite wizard to define crane location and the obstructions. Based on the input the program can search from the crane database for the crane model and configuration meeting the operating requirements. Figure 2 shows a screenshot from the 3D Lift Plan program which is widely used by crane and industrial rigging companies. At the present time, the problem for single location-based crane selection and lift planning has largely been solved by commercially available 3D based programs from the crane manufacturers and lift plan software developer companies.

2.3 Abstract Versus Realistic Crane Models

As load capacity and working radius of a specific crane is determined by the specific combination of boom and jib options, boom/jib operating angles, and upper carriage rotation/orientation angle, use of abstract crane models will have limited values in the real-world project lift planning processes, especially when highly detailed digital models of the real cranes are available from the crane manufacturers. Crane manufacturers have made significant progresses in developing detailed 3D crane models and attachments in real dimensions using advanced 3D manufacturing modeling software. However, it should be noted that real crane models with intricate geometrical details will increase the computing load in the simulation process due to graphic processing and rendering

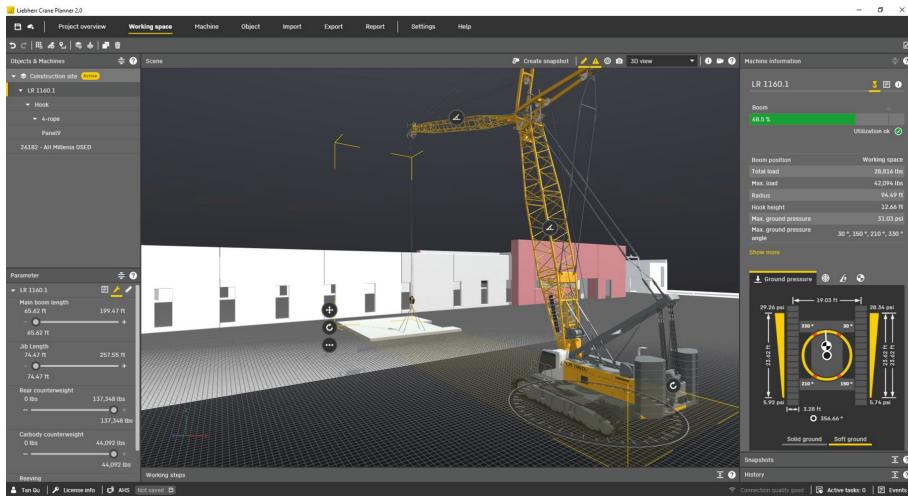


Fig. 1. Crane selection and lift simulation in Liebherr Crane Planner 2.0

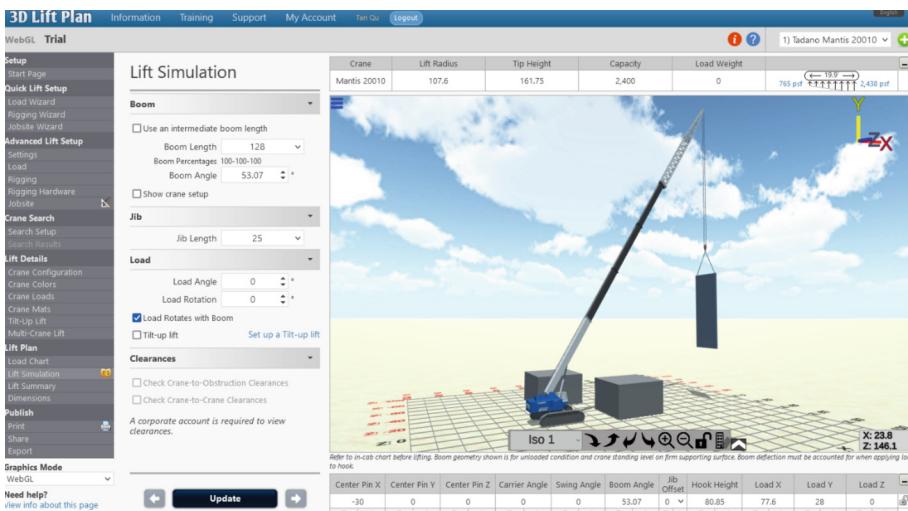


Fig. 2. Screenshot of lift planning simulation in 3D Lift Plan Trial Software

of such details. As such, a balanced approach should be adopted to condition/simplify the detailed crane model to a level that is sufficient for the simulation purpose at the same time omitting unnecessary fine details and small parts that do not contribute to the validity of the simulations.

2.4 Rigged Versus Static Crane Models

Mobile crane lift operations are a series of continuous steps of moving crane to the planned staging location, adjusting the boom angle/the jib angle/ hook to connect to the rigging system installed on the load object, turning upper carriage to move the load object to intended location, all in the while of making continuous adjustments of boom/jib/hook to clear obstructions. These complex steps often have many possible variations/sequences to achieve the operation. This requires the crane models used for the simulations to be rigged, i.e., the movable/rotating parts of the crane such as upper carriage, boom, jib, and hook be enabled to accomplish such simulation. While most computer game engines such as Unity support user real-time control of rigged equipment models, 4D BIM simulation software such as Autodesk Navisworks, Bentley Synchro Pro, and Kallocotech Fuzor do not readily support such user intervention during a simulation session.

2.5 Crane Lift Planning Simulation Environment—4D BIM Simulation Software Versus Computer Gaming Engine

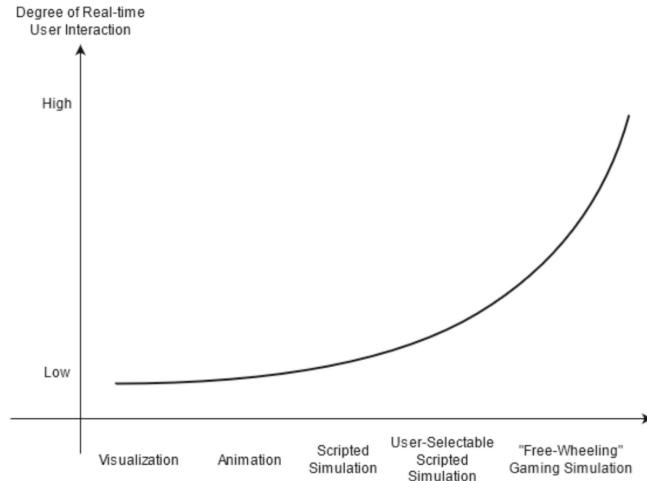
Existing research in 3D-based crane simulation in construction environment to a large extent relied on using 4D BIM simulation software such as Autodesk Navisworks and Bentley Synchro Pro [18, 19]. This approach generally consists of programming and running scripts to simulate transition of the key scene frames of the visualization process. Real-time user interaction/intervention during the simulation session is not supported. Should a new lift path need to be evaluated, the script will need to be modified and the simulation session rerun. Additionally, the complex movements/translation/rotation of the various crane components (upper carriage, boom, jib, hook, etc.) are not supported in the software environment. The limitation of this type of 3D visualization approach in crane lift planning has been identified [8]. In contrast, game engines such as Unity and Unreal Engine are oriented to provide the real-time user interaction and therefore can provide more realistic simulation of crane lift operations [7, 10, 15]. Computer game engine-based simulation system can provide the user with a “freewheeling” experience that resembles the real-world crane operations where the crane operators can execute unscripted/unlimited combinations of crane maneuvers based on the operators’ best judgement, in comparison to the limited scripted scenarios that 4D BIM simulation software can offer.

2.6 Summary of 3D Model-Based Crane Selection and Lift Plan Simulation Approaches

Using the degree of user interaction, simulation fidelity, and computer programming requirements, 3D model based crane selection and lift plan simulation approaches can be summarized in Table 1. Figure 3 shows the degree of real-time user interaction in ascending order, consistent with the research progresses and trends during the last 30 years.

Table 1. Summary of 3D model-based crane selection and lift plan simulation approaches to date

| | Degree of user interaction | Programming requirement |
|------------------------------------|--|-------------------------|
| Visualization | None | None |
| Animation | None, supported by developed application | None |
| Scripted simulation | Low/limited, non-real time, follows pre-set sequence | Basic |
| Pre-set user selectable simulation | Some, real-time but limited to preset | Moderate |
| “Freewheeling” gaming simulation | High, real-time, user driven | Intensive |

**Fig. 3.** Degree of real-time user interaction in simulation process in various approaches

2.7 Single Location Based Versus Continuous Lift Path Computer Gaming Simulations

Research work to date has made significant contribution to the body of knowledge in this area, but it has not addressed crane relocation/repositioning as commonly taking place in typical project lift operations. The available commercial crane selection and lift plan programs from the crane manufacturers and software developers generally focus on one lift operation in the lift plan analysis but not the entire lift process on a project. With the AEC industry embracing DT applications and the benefits of digitally rehearsing construction operations, it will demand more realistic lift operation simulations. As shown in Fig. 4, a typical construction project would have tens if not hundreds of individual building components/load objects that need to be installed. As the lift process progresses, the installed building components and associated temporary shoring and support systems

can then become potential obstructions to the subsequent crane maneuvers and lifting operations. Because there are often many different lift sequences and combinations, creating individual scenario-based lift plan simulations to validate the project construction methods could quickly become overwhelming. Therefore, continuous lift path gaming simulation system is essential.



Fig. 4. Crane lift operations on a multi-story concrete tilt wall project (*photo credits: Robins & Morton*)

2.8 Other Obstructions and Site-Specific Environmental Hazards

Many spatial conflicts and environmental hazards exist on construction sites that can significantly impact mobile crane operations, either in physical or virtual forms. Above-ground obstructions and hazards typically include existing buildings, structures, partially installed building components, trees, overhead power lines, walls, etc. Often not considered in crane lift planning but very important are underground features such as buried structures or utilities that are susceptible to damage exerted by cranes, especially in loaded condition. Other physical hazards can include ground drop-off conditions such as the proximity to stormwater pond or retaining wall structures where the soil pressures exerted by the crane can cause ground stability issues and in turn lead to loss of crane stability. The physical obstructions and hazards can typically be identified on the project survey or through the site visit. Virtual obstructions or constraints affecting mobile crane operations include the project or site property boundaries that dictate the limits/envelope of crane operating space. If the required crane operations cannot be contained on the project site, legal permission/instrument such as temporary construction easement will need to be obtained well in advance to allow the crane to operate from adjacent properties. Constructability issues from the failure to identify the required temporary construction easement beforehand can significantly impact the project schedule and costs.

3 VR-Based Continuous User-Driven Computer Gaming Application in Mobile Crane Simulations

3.1 Reloadable Project-Based VR Computer Gaming System

Construction projects and site conditions are often unique problems. As such, the gaming environment for mobile crane simulations on a given project site will also need to be specifically configured. In this paper, the authors present a VR-based gaming prototype system that supports reloadable projects at the user level without additional programming needs.

3.2 System Architecture

Software/Application Layer

The gaming system was developed in Unity environment and incorporates SteamVR virtual reality module for VR functions. Figure 5 shows the application architecture in Unity that depicts the asset groups/definitions, key modules, and game interaction mechanisms. Figure 5 shows a typical programming environment in Unity.

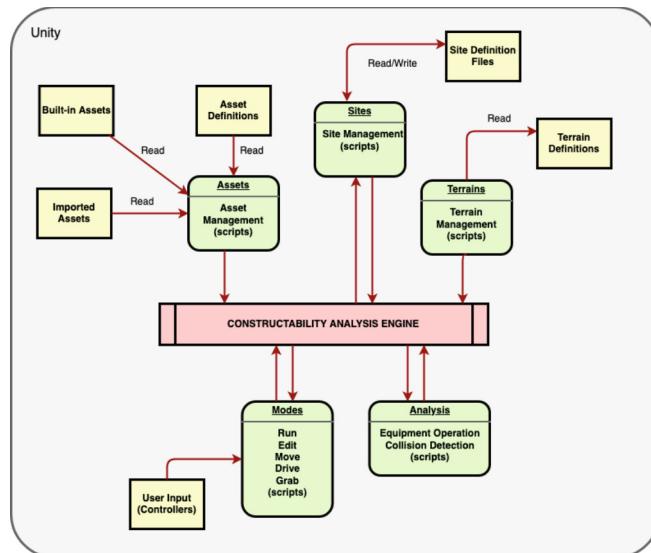


Fig. 5. Degree of real-time user interaction in simulation process in various approaches

Unity Asset Group Definitions

Four main groups of assets were used in the game application. Load object group includes the assets to be used for load objects and are FBX file format objects exported from a 3D building model in Autodesk Revit program. Final Reference Building group is the wireframe style 3D model of the final building used during the game session to

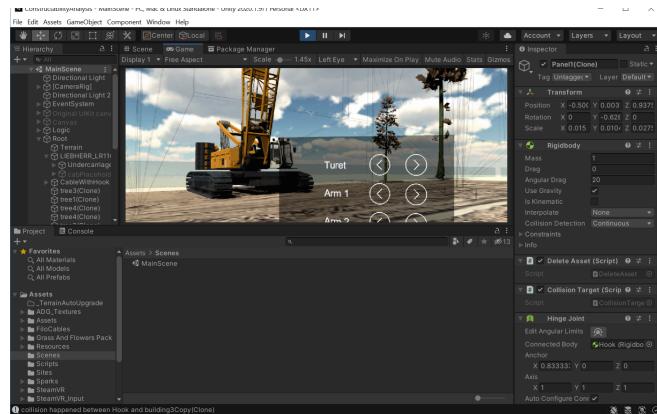


Fig. 6. Unity Game Programming Environment

provide users with visual cues where the load object needs to be installed. In real world crane lift operations, the load objects are often guided by installation crews. Equipment asset group is for cranes and other construction equipment to be used during the game. Equipment is generally in high-resolution obj files that have been rigged for the moving parts/pieces to be user controllable during the game sessions. Lastly, obstruction/hazard asset group include common obstructions/hazard such as power lines, trees, walls, and virtual barrier such as property lines.

Hardware

As Unity game development is specific to the VR head mounted display (HMD), HTC Vive Pro (Version 1) and Pro 2.0 HMD's were adopted for the test system. The primary reason for adopting HTC Vive Pro HMD's is the minimum display refresh rate of 90 Hz (11.1 ms or less) that authors set to ensure a smooth gaming experience for the users. The computer used for testing was an Alienware m17 R5 laptop that was configured with AMD Ryzen 9 6900HX CPU, Nvidia GeForce RTX 3080 Ti GPU with 16 GB GDDR6 VRAM, 64 GB DDR5 4800 MHz RAM, and 1 TB PCIe SSD.

Game Interaction

HTC Vive Pro game controllers (left and right) have been programmed to perform specific functions/tasks such as activate/deactivate menu's and sub-menu's, teleport movement, individual controls for crane driving (forward, reverse, left turn, right turn) and lift commands (boom angle adjustment, jib angle adjustment, raising/lowering hook cable, upper carriage turns). Figure 6 shows button/function assignments (Fig. 7).

First Person View (FPV) and Third Person View (TPV) Modes

The gaming system supports both TPV mode and FPV mode. The view angle is controlled and determined by the game teleport button/function. When a user teleports to certain distance to the driver's cab, the game mode will switch to FPV mode and the HMD displays the view from the driver's seat inside the cab. Figure 8 shows the typical FPV and TPV in game mode.

Game Menu Functions and Process Flow

When a game session is launched, there are two modes available. Under the “Edit” mode, the user can define/create a game simulation project by importing/selecting

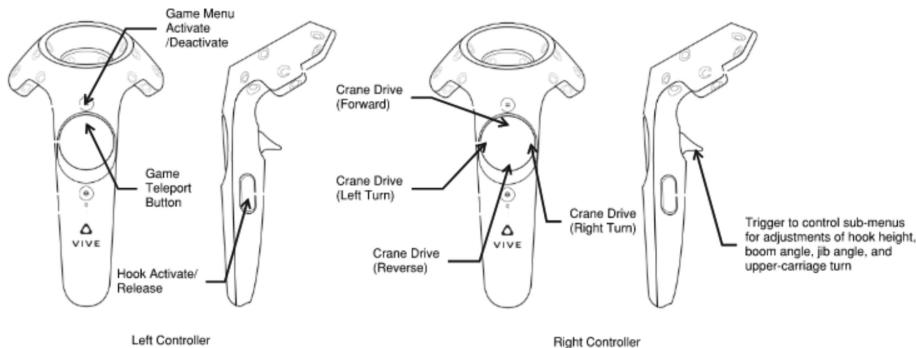


Fig. 7. HTC Vive Pro Game Controller Button Function Assignments



Fig. 8. Game FPV (left) and TPV (right) Modes

game terrain, final reference building, load objects, equipment, and define site obstructions/constraints. User can then switch to the “Run” mode to play the game. Figure 9 shows some of the game menu interfaces. Under terrain definition interface, image files derived from project site survey or aerial image can be used to simulate a realistic ground condition and define the obstructions and hazards that are particular to the site.

4 Case Study

4.1 Test Project

The Revit model from a one-story healthcare facility concrete tilt wall project was obtained from the industry and used to test the system concept. The 3D construction model in Autodesk Revit program was exported to a fbx file for the building reference frame for the game. The building model was rendered in wire frame mode in the game to give the users the visual cues of the target positions of the lift operations. Individual panels (load objects) were generated from the panel casting layout model/plan that was created in Autodesk Dynamo program. Figure 10 shows the panel casting layout plan generated in Autodesk Dynamo Studio.

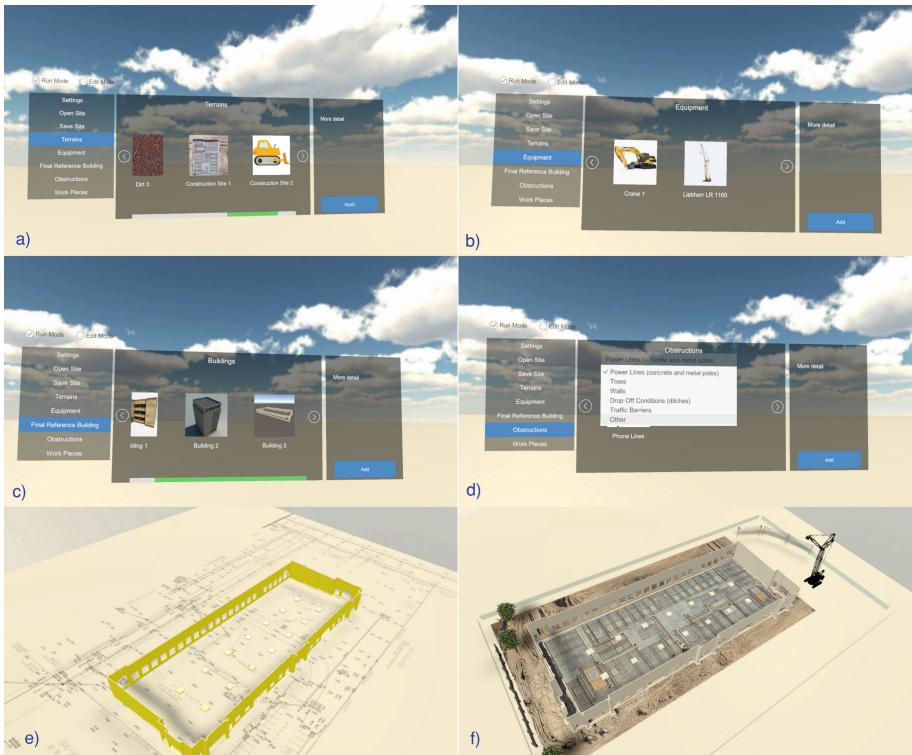


Fig. 9. Game menu interface: a) terrain selection. b) equipment selection. c) final reference building selection. d) define obstructions. e) project survey as terrain background. f) aerial image as background

4.2 Test Crane

For the test game, the same Liebherr 1160 crane as shown in Fig. 2 (Liebherr Crane Planner 2.0) was used. This crane model was rigged in Unity to provide full drive functions (forward, backward, left turn, and right turn) and user controls for upper carriage turning, boom angle, jib angle and hook height.

4.3 Results

The test game was created and provided to the construction company for testing. In the test game, site obstructions such as trees, power lines, and virtual project boundary were added. Test users were first provided with an introduction of the game user interface and a demonstration of using game controllers to control crane movements and panel lift process. Users were able to drive the crane around the site and simulate panel lifting/tilting operations with some practices. VR system performance was able to achieve greater than 90 Hz refresh rates with no reported motion sickness or discomfort by the test users. Figure 11 shows some example scenes of the test game sessions. Users were able to simulate various constructability issues such as load objects colliding with the

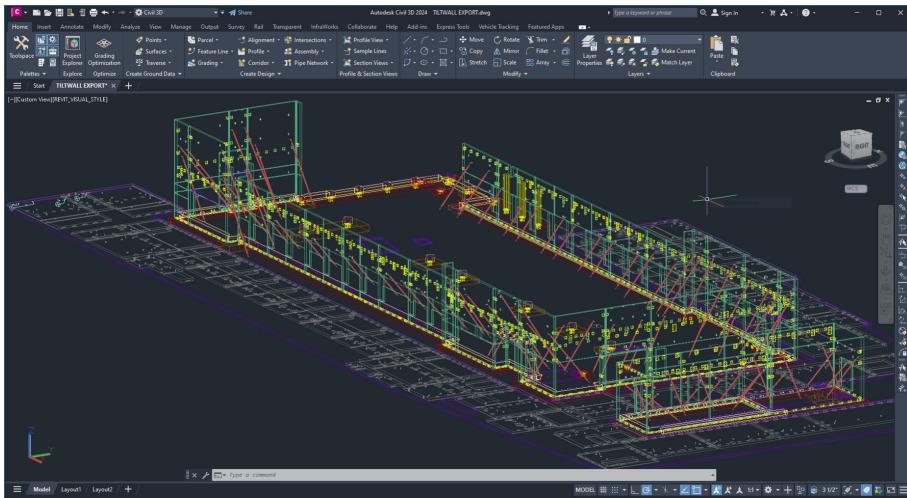


Fig. 10. Test concrete tilt wall building model and layout plan generated by Dynamo as shown in Revit program

wire building frame (installed building objects) as illustrated in subfigures 11.f and 11.g and exceeding site operating boundary as illustrated in subfigure 11.h, which typically indicates the need to obtain temporary construction easement/use permission from the neighboring site/property well in advance of the planned crane lift operations.

4.4 Test User Feedback and Lessons Learned

Crane Operator Controls and Usability Enhancement Suggestions

During the testing by construction industry users, various user feedback and usability enhancement suggestions were received. Submenus for controlling the boom, jib, and hook posed challenges to use by some test users. As the gaming application system uses the VR system game controllers instead of the joysticks and lever controls often found in real crane cabs, user experiences would be somewhat different from real operating conditions. However, it should be noted that user control mechanisms in real world crane equipment are also going through continuous innovation and transformation, e.g., old fashioned multi-lever controls are phased out in favor of multi-function joysticks and programmable user interface. Crane operator controls also often vary from manufacturer to manufacturer and vary from model to model even within the same manufacturer's model line-up. Therefore, it is not feasible to create gaming simulation user controls that will match specific crane make/model nor it is relevant to do so in a game simulation environment. Nonetheless, the approach/concept of using gaming simulation to test constructability issues was accepted as a valid approach by most test users.

Programming complexities

Computer gaming application development requires extensive and advanced programming knowledge and skills. The VR game simulation system presented in this paper was a result of several years' ongoing work by the authors. Even with the substantial efforts invested in the project, certain modeling details/nuances were omitted in the

current version of the application. For example, the load spreader bar rigging, realistic load object position spatial transformation, and dynamic response from slewing and luffing during the lift operations were far too complex to model and these implementations could be separate research projects by themselves.

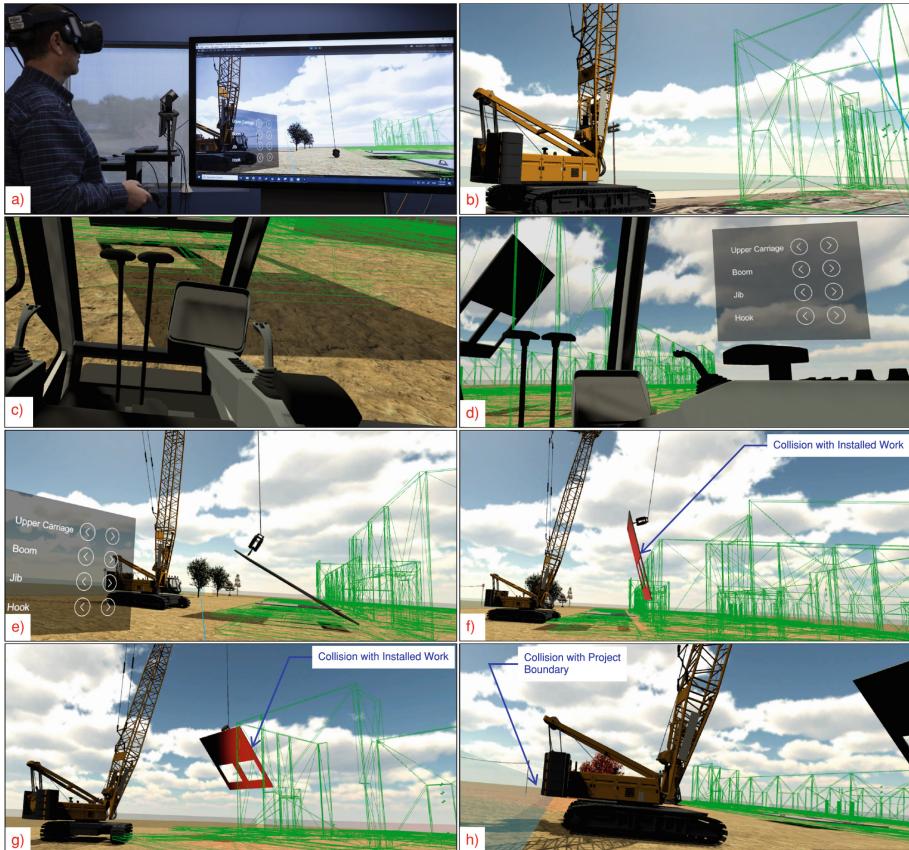


Fig. 11. Computer graphics from test gaming simulation

Model Complexities

Real world construction project BIM models especially at LOD 350+ levels contain many fine details/parts such as embeds and small attachments which significantly increase the file size and were found to create issues in Unity simulation environment. As a result, the project model used for testing had to be simplified to omit certain unnecessary fine details/features in order to keep the number of surfaces in the 3D model to below certain threshold. Conditioning and preparation of real-world project models for gaming simulation environments are an important topic that deserves further research attention and exploration.

5 Conclusions and Future Works

Crane selection and lift plan simulations have been a domain that have received ongoing and extensive research efforts because of the importance of crane operations in construction methods for prefabricated/modular buildings and structures. This paper provided a review of existing research efforts on 3D model-based crane selection and lift plan simulation approaches. Currently, crane manufacturers and lift plan software developers have developed mature technologies in crane selection and lift simulations for individual case-based lift operations in highly capable 3D environment. This paper introduced a conceptual framework for a reloadable computer gaming simulation application that can accommodate different projects and continuous crane lift operation simulations in highly realistic VR gaming environment using real-world crane models and construction project Revit models. Initial industry user testing indicated a general enthusiasm and acceptance of the concept of using gaming simulation to identify crane lift operation constructability problems. Future work planned by the authors includes implementing untethered HMD implementation, evaluating efficient porting and conditioning building BIM model elements to create final reference building and load objects, and programming/incorporation of realistic kinematic crane behaviors.

References

1. Allied Market Research Precast Concrete Market Forecast 2021–2030 Report, <https://www.alliedmarketresearch.com/precast-construction-market>, last accessed 12 Jan. 2024
2. Boton, C.: Supporting constructability analysis meetings with Immersive Virtual Reality-based collaborative BIM 4D simulation. *Autom. Constr.* **91**, 1–25 (2018). <https://doi.org/10.1016/j.autcon.2018.08.020>
3. Brozovsky, J., Labonne, N., Vigren, O.: Digital technologies in architecture, engineering, and construction. *Autom. Constr.* **158**(105212) (2024). <https://doi.org/10.1016/j.autcon.2023.105212>.
4. Dodge Construction Network/Data and Analytics Prefabrication and Modular Construction Smart Market Report 2020, <https://www.construction.com/reports/major-growth-expected-prefabrication-permanent-modular-construction-new-study-finds/>, last accessed 12 Jan. 2024
5. FMI 2023 North American Engineering and Construction Industry Overview, <https://fmi-corp.com/insights/construction-outlook/2023-north-american-engineering-and-construction-overview-first-quarter>, last accessed 12 Jan. 2024
6. Guo, H., Zhou, Y., Pan, Z., Lin, X.: Automated lift planning methods for mobile cranes. *Autom. Constr.* **132**(103982) (2021)
7. Guo, H., Zhou, Y., Pan, Z., Zhang, Z., Yu, Y., Li, Y.: Automated selection and localization of mobile cranes in construction planning. *Buildings MDPI* **12**(580). <https://doi.org/10.3390/buildings12050580> (2022)
8. Han, S. H., Hasan, S., Boufenguene, A., Al-Hussein, M., Kosa, J.: Utilization of 3D visualization of mobile crane operations for modular construction on-site assembly. *J. Manag. Eng.* **31**(45) (2015)
9. Kalantari, S., Pourjabar, S., Xu, T., Kan, J.: Developing and user-testing a “Digital Twins” prototyping tool for architectural design. *Autom. Constr.* **135**(104410) (2022)
10. Kayhani, N., Taghaddos, M., Mojtaba, N., Hermann, U.: Utilization of virtual reality visualizations on heavy mobile crane planning for modular construction. In: 35th International Symposium on Automation and Robotics in Construction. Berlin, Germany (2018)

11. Lin, Y., Chen, Y., Yien, H., Yien, H., Huang, C., Su, Y.: Integrated BIM, game engine and VR technologies for healthcare design: a case study in cancer hospital. *Adv. Eng. Inf.* **36**(130–145) (2018)
12. Liu, Z., Li, A., Sun, Z., Shi, G., Meng, X.: Digital twin-based risk control during prefabricated building hoisting operations. *Sensors* **22**, 2522 (2022)
13. McKinsey & Company Modular Construction 2019 Report, <https://www.mckinsey.com/capabilities/operations/our-insights/modular-construction-from-projects-to-products>, last accessed 12 Jan. 2024
14. Pan, Z., Guo, H., Li, Y.: Automated method for optimizing feasible locations of mobile cranes based on 3D visualization. *Procedia Eng.* **196**, 36–44 (2017)
15. Pooladvand, S., Taghaddos, H., Eslami, A., Tak, A., Hermann, U.: Evaluating mobile crane lift operations using an interactive virtual reality system. *J. Constr. Eng. Manag.* **147**(11) (2021)
16. Rocha, P.F., Ferreira, N.O., Pimenta, F., Pereira, N.B.: Impacts of prefabrication in the building construction industry. *MDPI Encyclopedia* **3**, 28–45 (2023). <https://doi.org/10.3390/encyclopedia3010003>
17. Tantisevi, K., Akinci, B.: Automated generation of workspace requirements of mobile crane operations to support conflict detection. *Autom. Constr.* **16**, 262–276 (2007)
18. Tantisevi, K., Akinci, B.: Simulation-based Identification of possible locations for mobile cranes on construction sites. *J. Comput. Civil Eng.* **22**(1), 21–30 (2008)
19. Tak, A., Taghaddos, H., Mousaei, A., Balourani, A., Hermann, U.: BIM-based 4D mobile crane simulation and onsite operation management. *Autom. Constr.* **128** (2021). <https://doi.org/10.1016/j.autcon.2021.103766>
20. Tu, X., Autiosalo, J., Jadid, A., Tammi, K., Klinker, G.: A mixed reality interface for a digital twin based crane. *Appl. Sci.* **11**(20) (2021). <https://doi.org/10.3390/app11209480>
21. Wang, X., Liu, J., Liu, F., Gao, S.: Collision-free locating of mobile cranes in 3D lifting system. In: Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 479–486, Montreal, QC, Canada (2010)
22. Zhang, Z., Pan, W.: Lift planning and optimization in construction: a thirty-year review. *Autom. Constr.* **118**, 103271 (2020). <https://doi.org/10.1016/j.autcon.2020.103271>
23. Zi, B., Lin, J., Qian, S.: Localization, obstacle avoidance planning and control of a cooperative cable parallel robot for multiple mobile cranes. *Robot. Comput. Integrat. Manufact.* **34**, 105–123 (2015)



Pre-design Stage: Computerizing Architect-Client Communication Using Virtual Reality and Artificial Intelligence

Omar Bagasi¹(✉), Nawari Nawari¹, and Fazil Najafi²

¹ School of Architecture, University of Florida, Gainesville, FL 32611-5703, USA
obagasi@ufl.edu

² Engineering School of Sustainable Infrastructure & Environment, University of Florida,
Gainesville, FL 32611-5703, USA

Abstract. Virtual reality (VR) is an emerging technology used in the design process. It has been utilized as a showcase tool that enhances communication between a designer and a client. It visualizes ideas and models more tangibly from a first-person point of view. In many cases, conventional variable and textual methods are not efficient enough. Mistakes often happen due to miscommunication, which leads to a result that may be good but not what the client wants. The study's aim is to use VR in the pre-design stage, which allows a designer to recruit this technology to improve his work. The study conducts an experiment that uses VR with eye-tracking and artificial intelligence AI during the pre-design stage. The extracted data determines the client's preferences. VR technology during the pre-design stage affords a better understanding of communicating a concept or idea. Eye-tracking senses what grabs the client's attention, which can confirm what the client said. Overall, VR and eye-tracking technologies enhance communication during architectural programming. In addition, it helps partly automate the pre-design stage, reducing the time and cost of the services, which benefits the designer and the client.

Keywords: Architecture Programming · Clients' Needs · Improve Design · Designer · Artificial Intelligence · Residential Building

1 Introduction

In the last two decades, several advanced technologies have been developed for communication between designers and clients during the design process. During the pre-design stage, client and architect communication is frequently verbal and/or textual, with little use of technology. Most cutting-edge technologies were utilized during the design process; however, it is not applied enough in the pre-design stage. The research focuses on advanced technologies used in architectural programming, such as artificial intelligence (AI) and virtual reality (VR). Virtual reality (VR) is a technology that enhances the design process. It provides a good visualization of the project to stakeholders. Prospective households can take virtual tours of their future projects to give them a realistic representation of the space and features.

Furthermore, virtual reality stimulates better collaboration between the various stakeholders involved in a building (Davila Delgado et al. 2020). Not only VR but also AI has played a significant role in the design process. Some AI models can manage tasks, such as producing floor plans and designing furniture layouts for living rooms or bedrooms (Peters 2022).

Another advanced technology that helps in the design process is eye-tracking. It has been used in visual preference experiments to track people's preferred choices (Batool et al. 2021). It is critical to understand the experience aspect of providing goods or services: how people are satisfied with the service. Typical methods include questioning individuals about their emotions; however, this strategy has limitations since emotions are not always the same as remembered. Descriptions of experiences may unconsciously differ from actual events. These technologies can be utilized in the design process.

The primary role of architects is to develop designs for projects. Design can be defined as a problem-solving conscious effort at imposing meaningful order. Architectural programming, defined by the Whole Building Design Guide (WBDG), established by the National Institute of Building Sciences, as the research and decision-making process specifying the work scope to be designed, helps architects better understand design problems (Pena and Parshall 2012). Architectural programming is However, solving the problem without a detailed understanding can lead to suboptimal results.

This research investigates the pre-design phase of architectural projects, emphasizing the communication between architects and clients in the context of residential design and a focus on the exterior of the building and its facade. The study explores the application of technology, particularly (VR), to facilitate this process in contemporary settings. The primary objective is to analyze clients' role as the building's main users while incorporating AI and computational methods to enhance the architectural design process.

Many architects face the problem of changing the plans and ideas of the client in almost every project. This study will reduce these changes, enhance design quality, and enable a better understanding of the client's needs. Miscommunication between the architect and client leads to a loss of information, resulting in a lack of understanding of client need in the design. Due to that problem, here comes the importance of architecture programming which provides most of the data in the design process. Being clear at the beginning saves time and money. Usually, some clients' requirements are vague and subjective, and the architect's role is to discuss them with the client to get a clear idea. In the end, the client's requirements and needs should be objective and clear.

Another problem in architecture programming: The software needs to be better. General-purpose programs are used to do the work. Architectural programming should have digital tools. We have many software and digital tools in other design phases, like in design and construction documents. However, the software used in architecture programming is generic software, like Microsoft Office Excel. Also, communication between the architect and client is typically verbal, so implementing advanced technology enhances workflow and facilitates tasks.

In the process of designing, getting information from clients is essential. Some clients are fully aware of their preferences; however, some may remain doubtful or need help to express what they want. Furthermore, there may be situations where clients pretend to understand, and others find verbal communication difficult. On the other hand,

communication problems could come from the architect's end as well, perhaps as a result of a lack of experience or poor soft skills, but this is outside of the scope of this study.

The aim of the study is to use advanced technologies, VR, AI, and Eye-tracking to improve communication between architects and clients in pre-design stage, objectively determine the clients' design requirements, and partly automate the architectural programming process during the pre-design phase. In addition, virtual reality is a fertile area for research in architectural pre-design field. It is a great tool to be part of the solution in the design process.

2 Literature Review

VR has been widely used in the architecture, construction, and engineering (ACE) industry, such as in urban design and architectural design. VR can be used to project proposed 3D model designs, which can then be shared with stakeholders for feedback (Heydarian et al. 2015). This allows stakeholders to experience the proposed design from a first-person point of view like a person inside the building. This allows architects and clients to see different design ideas and to get feedback from each other before the building is constructed (Chen et al. 2019).

2.1 VR in Architectural Design Process

VR is an interactive technology that takes place within a simulated environment. It can be used to generate immersive experiences similar to physical reality. VR has been used extensively in the design sector, such as urban planning and architectural design. VR can improve design quality by allowing users to identify design issues within the immersive virtual environment (IVE) that two-dimensional drawings or building information models cannot (Osorio Carrasco and Chen 2021). In scientific study, VR can be used to examine environmental elements such as material, color, and lighting environments (Maffei et al. 2015).

Research indicates that VR technologies demonstrate effectiveness in various aspects of construction, including safety training (Li et al. 2018), control of project schedules (Fu et al. 2018), and optimization of construction project site layouts (Muhammad et al. 2020). Additionally, VR technologies can facilitate improved collaboration among stakeholders (Alizadehsalehi et al. 2019). Enhance comprehension of intricate designs (Sutcliffe et al. 2019), pinpoint design issues (Romano et al. 2019), represent building geometry to assist users in project understanding and decision-making (Bille et al. 2014), and support collaborative decision-making processes (Du et al. 2018).

This paper observes how VR and interactive technology are used to explore scale, lighting experience, and material iterations (Gegana et al. 2019). Another study showed that the participants rated the perceived quality of the lighting in the VR display to be similar to the perceived quality of the lighting in the real world. This suggests that VR can be used as a tool for evaluating the perceived quality of lighting systems (Rockcastle et al. 2021).

2.2 AI in Generative Design

In recent years, AI has been progressively playing a big role in the field of architecture, especially in generating architectural designs. AI integration in architecture opens up new possibilities for creativity, efficiency, and optimization throughout the design process. AI can be used to automate many of the tasks related to architectural design, such as generating 3D models, creating renderings, and analyzing spatial data (As 2018). In addition, it creates new design tools and workflows that are not possible with traditional methods (As 2018).

Generative adversarial network (GAN) is a generative AI machine learning framework. It is also used in architectural design. For example, in a study, Nate (2020) developed a collaborative platform using neural networks and a prototype interface. He describes neural networks for design as an AI pattern-recognition method. The study employs the Pix2pix approach to translate images in order to understand them. The main idea is to use an AI model to create a floor plan for a house. Based on a data set of houses designed by professional architects, this AI model is built on training a model on how to design a house (Peters 2022). In parallel, scholars have explored AI models that facilitate the translation of textual input into images, exhibiting significant progress in the area of picture creation (Ramesh et al. 2022). Midjourney and DALL.E models are powerful tools in this area.

Utilizing generative systems presents an opportunity to craft designs that are both complex and adaptive, better aligning with the demands of the 21st century. The generative system played a pivotal role in shaping the form, structure, and material composition of a design (Henriques et al. 2019). This method introduces a novel and auspicious avenue, asserting our conviction that generative systems can be harnessed to produce designs that are more complex, responsive, and intricate.

AI is an effective tool that can be used to automate a variety of architectural design-related tasks. However, AI is a developing technology that must be carefully planned for and used. More research is needed on the use of AI for architectural design.

2.3 Data Capture: Eye-Tracking

The eye-tracking method, tracking gaze and eye movements, proves instrumental in identifying points of attention and pinpointing sources of information, contributing not only to consumer studies but also to design analysis in environmental design studies (Gero and Milovanovic 2020). Moreover, client and neuromarketing academics have embraced VR-integrated with eye-tracking systems since 2017, marking a significant development in virtual reality's potential within marketing to deliver consumer experiences akin to those in physical stores (Clay et al. 2019). In the relevant literature, the eye-tracking method also has the ability to provide gaze specifics in the client interest (Santos et al. 2015).

This literature review has shown that the application of AI in architectural programming and design offers a variety of possible benefits, including increased efficiency and accuracy. However, there is still much to learn about the usefulness of AI-based design tools in many contexts. Also, further research is needed to fill the gaps highlighted in this review. The proposed research framework, which is based on VR experiments with

clients, could be a valuable new strategy for improving architectural pre-design stage communication. From the literature review, there is not satisfying finding that talk about using AI and VR during pre-design stage. Many studies exist about AI and VR during design stage however, there is a lack of studies in pre-design stage. This is the gap that this study will be investigating.

3 Methodology

The methodology employed in this research is fundamental to understanding the role of VR in the pre-design stage. It will employ two methods, both quantitative. It integrates both survey and experiment as data-collection techniques to provide a holistic understanding of the topic.

The primary objective of this methodology is to investigate the current utilization of VR in the pre-design stage to improve client-architect communication, assess its impact on design outcomes and user experiences, and identify the challenges and opportunities associated with the integration of VR into the pre-design process.

This research will utilize these two methods. The first one, a survey, will provide data from the participants as a reference, i.e., ground truth, for the sake of this methodology. The second method, a VR experiment, will provide the primary data that will be investigated. The study seeks to have almost the same data from both methods; however, the purpose is to compare the methods and validate the data of the VR experiment. The rationale behind this choice is to gain a comprehensive understanding of the multifaceted aspects of virtual reality in pre-design stage communication.

Several research design alternatives were considered before settling on these two methods. One alternative was an experimental approach, focusing on one aspect of the data and statistical analysis. While this approach could provide two layers of validation in addition to valuable insights into VR in pre-design communication, instead of just having the lack of depth and context that one-method data can offer.

Ultimately, the two-method approach was chosen because it strikes both the regular method of communication, oral and textual, and a new technological method, VR technology, aligning with the research objectives to investigate VR in pre-design stage communication comprehensively.

The number of participants will be 25 people. The study targets people with potential to design a house and build it in the state of Florida. The age range is between 25–60 years and the gender is both men and women. The researcher will recruit research subjects who want to have a house in the future, not necessarily in the near future. The survey will take 12–15 min, and the experiment lasts 30–40 min. A consent form will be provided.

First, participants will start with an electronic version survey distributed to a diverse sample of potential clients. The survey instrument will be designed to capture essential information regarding the client requirement about the aesthetic of the building exterior. The survey will be processed via online survey platforms, allowing for efficient data collection and analysis.

The survey contains two parts. The first part has demographic questions about age, gender, marital status, city, income, and level of education. The second part of the survey is about building aesthetic aspects from the building exterior. The questions are

about the client's preference of architectural styles, exterior materials, and colors with multiple-choice questions.

The answers are going to be used to create prompts to be used in generative AI tools to generate designs or ideas that represent client requirements or desired ideas. In this study, midjourney.com is utilized to execute these prompts.

The participants who completed the survey are qualified for the second part. Before starting the VR experiment, participants will be prepared for the experiment. The participants will wear VR headsets. The experiments will show buildings that were generated based on the data that the participants provide from the survey previously. Again, the data will be collected in this time during this experience. The data will be the clients' preferences of the architectural styles, exterior materials, and colors; however, this method will get the data in an indirect way. It will collect data based on staring time on buildings which means that the more staring time the more the client likes the design that represents these criteria, architectural styles, exterior materials, and color (Fig. 1).

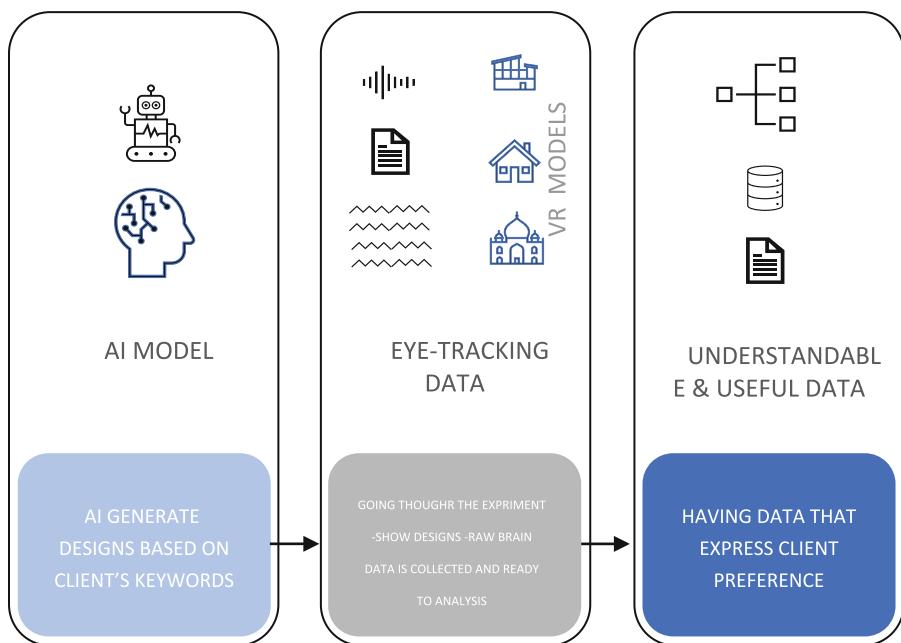


Fig. 1. A proposed computable framework for the pre-design stage.

Data will be collected through the built-in eye-tracking device in the VR headset and the participant selection through the VR set controller sticks. These experiments will be designed to explore the details of the topics covered in the surveys. The experiment will be screen-recorded and analyzed, then linked with eye-tracking data.

Survey data will be analyzed using descriptive statistics to summarize and present the survey responses. The experiment data will be analyzed using descriptive statistics to summarize and present the survey responses. Both data sets will be compared. Then,

for all data, the study will use inferential statistical analysis to draw the relation between the study variables.

A pilot study will be conducted for this research. The survey will be pre-tested and refined to ensure clarity and relevance. Moreover, the survey questions are derived from well-known manuals utilized in related studies. The experiment will be pre-tested and refined to ensure precision and relation.

In conclusion, the chosen methodology for this research employs a survey and experiment, combining both in sequence to have a comprehensive investigation of the role of VR in communication during the pre-design stage. This methodology aligns with the research objectives, which aim to explore communication using VR in the pre-design stage and assess its effectiveness, impact, and challenges (Fig. 2).

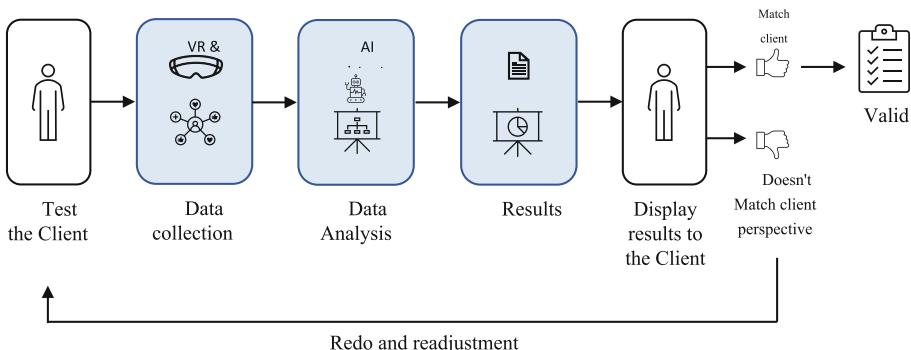


Fig. 2. Data Validation

4 Anticipated Findings

The primary anticipated result is that the tool will enhance pre-design architect-client communication. The study will conduct a method that contains a survey and an experiment. The expected results from the survey represent verbal communication, which will be confirmed later by using new technology methods that occur in the experiment, representing the validation data, which is eye-tracking data. VR is a great tool to communicate in a medium that allows for easy data transactions. The study will show to what extent this method can help in the pre-design stage.

In our VR experiment, participants engaged in a simulated environment designed to assess facade cognition. Initial observations indicate a noteworthy variance in participants' reaction times when navigating virtual 3D models with different architectural styles. Surprisingly, participants consistently demonstrated heightened focus and task performance in scenarios requiring intricate visual reasoning. Additionally, subjective feedback suggests a positive correlation between perceived architectural style and overall satisfaction with the VR experience. These preliminary findings suggest a potential link between architectural style, cognitive abilities, and user engagement in VR environments, also highlighting opportunities for further investigation and refinement of VR experiments.

By providing insights into the application and usefulness of immersive technology, eye-tracking, and AI modules in addressing the communication challenges between architects and clients during the pre-design stage, the research findings are relevant to each research question. The findings offer useful data that may be used to guide implementation and decision-making in the field of architecture.

5 Conclusion

This study creates a method to enhance communication between architect and client during the pre-design stage using VR. The study anticipates that utilizing VR during meetings with the client can enhance the quality of the data and reduce loss of information that usually happens in communication throughout architecture programming. The method that has been used documents and validates the data transaction between shareholders. As a result, the client gets high-quality design that is a desired product, and the architect works efficiently. This study can be implicated not just the pre-design stage but also the whole design process and even more, like other building types.

References

1. Heydarian, A., Carneiro, J.P., Gerber, D., Beccerik-Gerber, B., Hayes, T., Wood, W.: Immersive virtual environments versus physical built environments: a benchmarking study for building design and user-built environment explorations. *Autom. Constr.* **54**, 116–126 (2015). <https://doi.org/10.1016/j.autcon.2015.03.020>
2. Chen, Y., Cui, Z., Hao, L.: Virtual reality in lighting research: comparing physical and virtual lighting environments. *Light. Res. Technol.* **51**(6), 820–837 (2019). <https://doi.org/10.1177/1477153518825387>
3. Osorio Carrasco, M.D., Chen, P.H.: Application of mixed reality for improving architectural design comprehension effectiveness. *Autom. Constr.* **126**, 103677 (2021). <https://doi.org/10.1016/j.autcon.2021.103677>
4. Maffei, L., Masullo, M., Pascale, A., Ruggiero, G., Puyana Romero, V.: On the Validity of Immersive Virtual Reality as Tool for Multisensory Evaluation of Urban Spaces. *Energy Procedia*, Elsevier Ltd, pp. 471–476 (2015). <https://doi.org/10.1016/j.egypro.2015.11.703>
5. Li, X., Yi, W., Chi, H.L., Wang, X., Chan, A.P.C.: A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Autom. Constr.* **86**, 150–162 (2018). <https://doi.org/10.1016/j.autcon.2017.11.003>
6. Fu, M., Liu, R., Rinker, M., Professor, A.: The application of virtual reality and augmented reality in dealing with project schedule risks (2018)
7. Muhammad, A.A., Yitmen, İ., Alizadehsalehi, S., Celik, T.: Adoption of Virtual Reality (VR) for site layout optimization of construction projects. *Teknik Dergi* **31**(2), 9833–9850 (2020). <https://doi.org/10.18400/tekderg.423448>
8. Alizadehsalehi, S., Asce, S.M., Hadavi, A., Asce, M., Huang, J.C.: Virtual reality for design and construction education environment (2019)
9. Sutcliffe, A.G., Poullis, C., Gregoriades, A., Katsouri, I., Tzanavari, A., Herakleous, K.: Reflecting on the design process for virtual reality applications. *Int. J. Hum. Comput. Interact.* **35**(2), 168–179 (2019). <https://doi.org/10.1080/10447318.2018.1443898>
10. Romano, S., Capece, N., Erra, U., Scanniello, G., Lanza, M.: On the use of virtual reality in software visualization: the case of the city metaphor. *Inf. Softw. Technol.* **114**, 92–106 (2019). <https://doi.org/10.1016/j.infsof.2019.06.007>

11. Bille, R., Smith, S.P., Maund, K., Brewer, G.: Extending building information models into game engines. In: Proceedings of the 2014 Conference on Interactive Entertainment, ACM, New York, NY, USA, December 2014, pp. 1–8. <https://doi.org/10.1145/2677758.2677764>
12. Du, J., Zou, Z., Shi, Y., Zhao, D.: Zero latency: real-time synchronization of BIM data in virtual reality for collaborative decision-making. *Autom. Constr.* **85**, 51–64 (2018). <https://doi.org/10.1016/j.autcon.2017.10.009>
13. Gegana, G.A., Theodore, J., Gunawan, F.: Study of lighting and material iterations in full scale model using virtual reality and interactive architectural representation. In: IOP Conference Series: Earth and Environmental Science, vol. 238, p. 012025 (2019). <https://doi.org/10.1088/1755-1315/238/1/012025>
14. Rockcastle, S., et al.: Comparing perceptions of a dimmable LED lighting system between a real space and a virtual reality display. *Light. Res. Technol.* **53**(8), 701–725 (2021). <https://doi.org/10.1177/1477153521990039>
15. As, I., Pal, S., Basu, P.: Artificial intelligence in architecture—generating conceptual design via deep learning—2018. **16**(4), 306–327 (2018)
16. Peters, N.: Enabling alternative architectures. In: Machine Learning and the City: Applications in Architecture and Urban Design, pp. 193–200 (2022). <https://doi.org/10.1002/9781119815075.CH16>
17. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with CLIP Latents (2022)
18. Henriques, G.C., Bueno, E., Lenz, D., Sardenberg, V.: Generative systems: intertwining physical, digital and biological processes, a case study. In: Proceedings of the International Conference on Education and Research in Computer Aided Architectural Design in Europe, vol. 1, pp. 25–34 (2019). https://doi.org/10.5151/PROCEEDINGS-ECAADESIGRADI2019_100
19. Gero, J.S., Milovanovic, J.: A framework for studying design thinking through measuring designers' minds, bodies and brains. *Design Science* **6**, e19 (2020). <https://doi.org/10.1017/dsj.2020.15>
20. Clay, V., König, P., König, S.U.: Eye tracking in virtual reality. *J. Eye Mov. Res.* **12**(1) (2019). <https://doi.org/10.16910/jemr.12.1.3>
21. dos Santos, R.D.O.J., de Oliveira, J.H.C., Rocha, J.B., Giraldi, J.D.M.E.: Eye tracking in neuromarketing: a research agenda for marketing studies. *Int. J. Psychol. Stud.* **7**(1) (2015). <https://doi.org/10.5539/ijps.v7n1p32>



Implementation of VDC and Virtual Reality to Meet the Deadline and Budget of Work. Case Study: UTEC CAMPUS BARRANCO—FASE 02

Luis L. Espinoza Cautí^(✉) and Enrique A. Juarez Aquino

Universidad de Ingeniería y Tecnología—UTEC, Lima, Peru
lespinozac@utec.edu.pe

Abstract. For decades, the construction sector has suffered from a very low productivity compared to other sectors such as manufacturing, which has the highest growth. It is for this reason that the sector is currently seeking to increase productivity based on the implementation of new methodologies and technologies. One of the current projects aligned to this goal is UTEC CAMPUS BARRANCO—FASE 02, which seeks to improve productivity by avoiding rework during the design phase to meet the deadline and budget of work with the use of Virtual Design and Construction (VDC) with Virtual Reality (VR).

First, the VDC framework was designed to meet the deadline and budget of work with the use of VR as main tool. Finally, a survey of the use of VR in the project was made according to evaluate the impact of its use compared with other current tools such as 2D plans and 3D models.

As a result, the redesign of the new laboratories was completed satisfactorily, avoiding unnecessary rework in the design phase and it was demonstrated that the use of VR was extremely useful in the design phase of the project compared to other current tools such as 2D plans and 3D models.

Keywords: VDC · ICE sessions · BIM · Virtual Reality

1 Introduction

The construction sector is considered one of the major supports of the global economy, which is the reason why around \$10 trillion is invested annually in infrastructure development [1]. However, this contribution to the economy could be much greater if it were not for poor productivity in the sector, which has experienced a slow growth of 1% per year compared to the higher growth registered in the manufacturing sector of 3.6% [1].

One of the reasons for low productivity in the construction sector is rework, which is responsible of interruptions on construction projects [2, 3]. The great negative impact that rework has on low productivity is due to the strong relationship between them [4]. It is for this reason that, based on the findings of different investigations, it is proposed that rework should be considered in the construction-schedule analysis since it has a direct

relationship with the failure to meet construction deadlines [5–7]. It should be noted that this is one of the most important problems in Peru, where the number of suspended works was 2,066 in September 2023 [8].

In addition, direct rework costs are a major contributor to cost growth and schedule overruns in building construction projects [9], which average were 20.7% [2]. This recurring problem occurs at both the design and construction phase. Rework in both stages is often related in complex projects since undiscovered rework in the design phase can produce a significant amount of rework in the construction phase [10]. Actually, in the construction industry, rework in the construction phase could increase costs by 4% to 12% of the construction contract amount [11–13].

According to reduce rework in the design phase, new technologies such as Virtual Reality (VR) are being used in recent years due to the benefits offered by its application, such as the detection of design flaws before construction begins through immersive visualization in the design review [14, 15]. In addition, the Virtual Design and Construction (VDC) methodology is also used for this purpose by improving management as possible so future conflicts, changes, rework and constructability issues can be minimized [16, 17].

Therefore, this study focuses on implementing VDC with the use of VR to meet the deadline and budget of work of “UTEC CAMPUS BARRANCO—FASE 02” project by reducing rework in the redesign of new university laboratories before the construction of these begins.

2 Literature Review

2.1 VDC in Construction Industry

In the last decade, in order to improve the performance in the construction industry, VDC emerged as a new Informational Technological (IT) approach [17]. VDC is described as the use of integrated multi-disciplinary performance models of design-construction projects to support explicit and public business objectives [18–21]. VDC consists of five main components: Building Information Modelling (BIM), Project Production Management (PPM), Integrated Concurrent Engineering (ICE), Client goals and Project objectives and Metrics [22, 23]. Table 1 describes the five components of VDC.

Those organizations that have used VDC in their projects agree that it allows the improvement of profitability, efficiency of work processes, as well as the reduction of construction time, and based on these, the quality of the project and the final results are improved [24]. In addition, it should be noted that the general perception of the organizations is that VDC has a positive impact on information management, project communication and number of problems to be solved on site [24].

As an example of the implementation of VDC in construction projects, there is the construction of the Ovalo Monitor Bridge where the VDC methodology was implemented with Prefabricated Reinforcement Cages (PRC) elements [25]. The reason for the implementation is related to the fact that reinforcing steel together with pouring concrete represents 50% of the total construction cost and with PRC elements, costs and time can be reduced, while improving fabrication quality and safety [26, 27]. As a result

Table 1. VDC components.

| VDC component | Explanation |
|-------------------------------------|--|
| BIM | The creation and use of 3D models for communication and informed decision making [23] |
| PPM | Organization and control of the physical work activities in the project by viewing the project as a production system [22, 23] |
| ICE | Information is shared, actions are coordinated, problems are solved, and decisions are made in ICE sessions [22] |
| Client goals and Project objectives | The project team defines project objectives considering the total cost and building performance according to support the client's goals [22, 23] |
| Metrics | The client's goals are measured by metrics considering operations, use, sustainability performance, safety, schedule, and costs [22] |

of the implementation, the assembly times of the structural item were reduced by 31% and 100% of buildable design of the PRC elements was achieved [25].

2.2 VR in Construction Industry

VR is an Information Technology (IT) which simulates a virtual environment that immerses users to the extent that they have the feeling of “being there” by the use of a head-mounted display for a single individual or room-sized systems that enable VR experiences for many individuals [28–30]. The quality of any VR system depends on the main features of VR: Immersion, Interaction and Imagination [31]. Table 2 describes the three main features of VR.

Table 2. Main features of VR.

| VR feature | Explanation |
|-------------|---|
| Immersion | Feeling of being present of the computer-generated world by the stimulation of human senses [32] |
| Interaction | Communication with the system based on real time interaction and human participation [32] |
| Imagination | Capability of the system designer to execute a particular goal for complex problem solving in diverse fields [32] |

VR has multiple benefits, like in the design phase where it facilitates decision-making, coordination and the communication of client's requirements in the design review process by clarifying many aspects of the design that is difficult from traditional

design documents, such as clashes and lack of space [33, 34]. Also, among other current uses of VR are construction safety planning and training [35, 36], as well as production planning [37].

The MK3, a new pelletizing plant located in the north of Sweden, is one of the projects where VR walkthrough environments have been implemented for monitoring the design, planning and construction process of the plant [33]. The implementation in this project shows that the use of VR has increased the value for the client and the reliability in the design process, and minimized the waste in the production phase by eliminating collisions between the different designs [33].

3 Methodology

3.1 VDC Framework

In the beginning, the client's objective (CO) is defined based on the client's needs and expectations with the project. Then, based on the CO, the project objective (PO) is established, which serve as goals or objectives to be followed during the project.

Once the CO and PO have been defined, the scope of the other VDC components (ICE, BIM, PPM) are defined, which must be oriented to the fulfillment of the CO and PO. Table 3 describes the scope of each VDC component of the project.

Table 3. Scope of each VDC component of the project.

| VDC component | Scope |
|---------------|--|
| CO | Redesign the new laboratories with the guarantee of meeting the deadline and budget of work |
| PO | Use VR in the redesign of the new laboratories to avoid rework at the design and construction phase |
| ICE | Realize weekly ICE sessions with key stakeholders (laboratory directors, architect, structural engineer, supervisor, contractor) to resolve incompatibilities detected in the current design |
| BIM | Realize a 3D model with LOD 350 of the new laboratories to detect interferences, show them in VR to key stakeholders and update them weekly with the solutions proposed in the ICE sessions |
| PPM | Realize a weekly report of proposed changes to the current design using VR with the director of each new laboratory and his work team |

3.1.1 ICE Metrics

In the ICE component, 2 production metrics and 2 controllable factor metrics have been defined, these metrics are used to achieve the PO and CO. Table 4 describes each ICE metric.

Table 4. ICE Metrics.

| ICE | Objective | Metric | Goal | ID | Reference Source |
|----------------------|--|---|--------------|-------|------------------|
| Production Metrics | Identification and resolution of interferences in the development of the project | # of interferences resolved / # of interferences found | > 90% | ICE-1 | [38] |
| | Approve proposals to advance the project collaboratively in ICE session | # of proposals accepted / # of proposals presented in session | > 80% | ICE-2 | [39] |
| Controllable Factors | Preparation of the topics to be addressed at the ICE session | # of days before the ICE session on which the agenda is available | 1 day before | ICE-3 | [40] |
| | Project participants attend and participate in the ICE session | % Guest attendance | 100% | ICE-4 | [41] |

The selection of these ICE metrics was based on previous works which implemented VDC in order to achieve the same satisfactory results. Those metrics related to the use of VR do not have a reference source as there are not many works linking VDC with VR.

3.1.2 BIM Metrics

In the BIM component, 2 production metrics and 2 controllable factor metrics have been defined, these metrics are used to achieve the PO and CO. Table 5 describes each BIM metric.

The selection of these BIM metrics was based on previous works which implemented VDC in order to achieve the same satisfactory results. Those metrics related to the use of VR do not have a reference source as there are not many works linking VDC with VR.

3.1.3 PPM Metrics

In the PPM component, 1 production metric and 1 controllable factor metric have been defined, these metrics are used to achieve the PO and CO. Table 6 describes each PPM metric.

Table 5. BIM Metrics.

| BIM | Objective | Metric | Goal | ID | Reference Source |
|----------------------|---|--|----------------|-------|------------------|
| Production Metrics | Update the BIM model with the accepted project proposals | LOD = 350 | 350 | BIM-1 | [41] |
| | Show the updated VR of the new laboratories to the key stakeholders before the ICE session | % of key stakeholders who used the VR | 100% | BIM-2 | Own elaboration |
| Controllable Factors | Update the model of the new laboratories based on the proposals accepted at the ICE session | # of days of updated model after the ICE session | < 4 days later | BIM-3 | Own elaboration |
| | Convert the updated models into VR | # of days of updated VR after the ICE session | < 4 days later | BIM-4 | Own elaboration |

3.2 Use of VR

As the implementation is focused on the use of VR as the main technology, after the VDC implementation, a survey was made to the directors of the new laboratories about the use of VR in the redesign of these. A Likert scale (scale from 1-Not at all important to 5-Extremely important) was used to compare the use of VR in comparison to other current tools such as 2D plans and 3D model in the design. The criteria were chosen based on previous work on comparing the use of VR with other tools and the project objective of designing the new laboratories. Table 7 gives a summary of the criteria used in the survey.

4 VDC Implementation

4.1 ICE Metrics and Controllable Factors

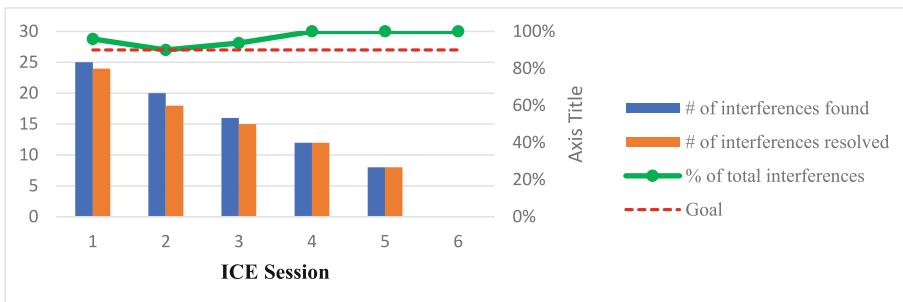
At the end of the design of the new laboratories, the following results were obtained for the evolution of the ICE metrics and controllable factors (Figs. 1, 2, 3 and 4).

Table 6. PPM Metrics.

| PPM | Objective | Metric | Goal | ID | Reference Source |
|----------------------|---|---|----------------|-------|------------------|
| Production Metrics | Realize a report with proposed changes to the redesign to be discussed at the ICE session | # of days that the report is ready before the ICE session | > 1 day before | PPM-1 | Own elaboration |
| Controllable Factors | Conduct team meetings under the leadership of the director of each new laboratory | Frequency = weekly | Once a week | PPM-2 | Own elaboration |

Table 7. Criteria used in the survey.

| Criteria | Explanation | Reference source |
|----------------------------|--|------------------|
| Easy to use | Simplicity of use | [42] |
| Space distribution | Helps to understand the distribution | [43, 44] |
| Visualize real environment | Helps to visualize the internal and external environment objects as they will be constructed | [43, 44] |

**Fig. 1.** The # of interferences resolved at the ICE session.

4.2 BIM Metrics and Controllable Factors

At the end of the design of the new laboratories, the following results were obtained for the evolution of the BIM metrics and controllable factors (Figs. 5, 6, 7 and 8).

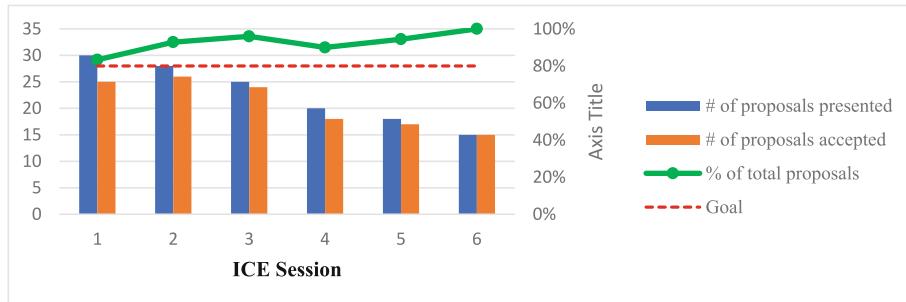


Fig. 2. The # of days before the ICE session on which the agenda is available.

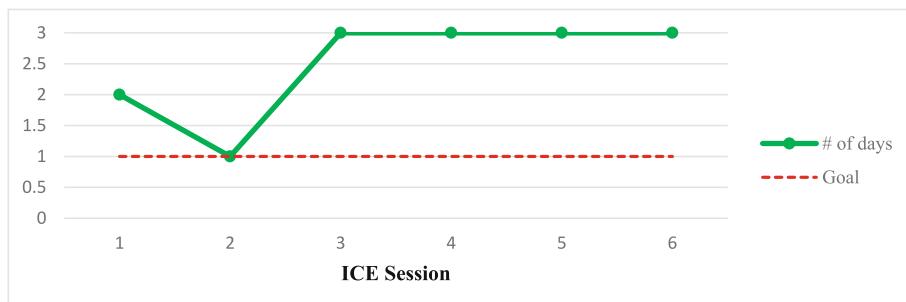


Fig. 3. The # of days before the ICE session on which the agenda is available.

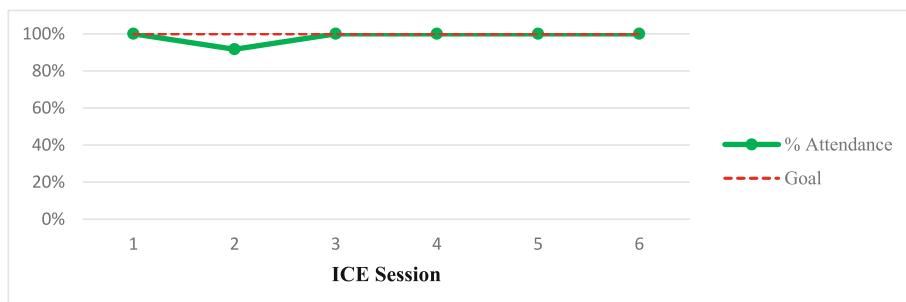
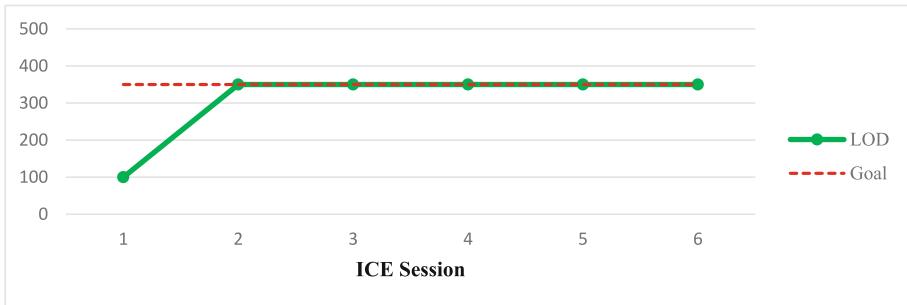
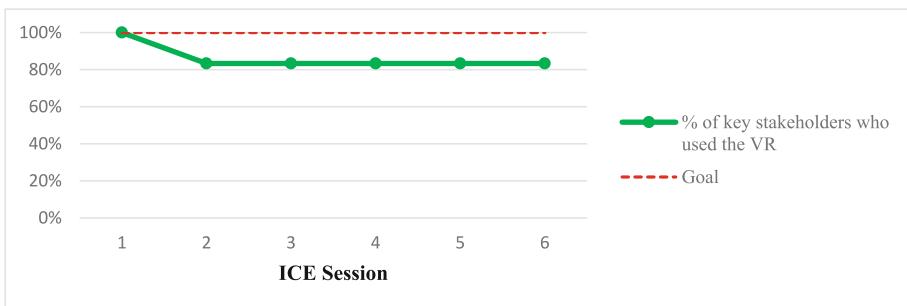
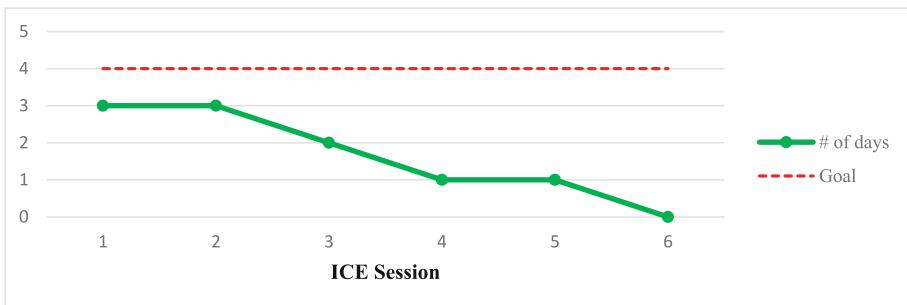


Fig. 4. The % Guest attendance.

4.3 PPM Metrics and Controllable Factors

At the end of the design of the new laboratories, the following results were obtained for the evolution of the PPM metrics and controllable factors (Fig. 9).

**Fig. 5.** LOD in each ICE session.**Fig. 6.** The % of key stakeholders who used the VR.**Fig. 7.** The # of days of updated model after the ICE session.

5 Use of VR

After the survey on the use of VR was completed by the director of each laboratory and his work team, who were the ones who used VR the most during the project, the results are shown in Table 8 and Fig. 10.

In order to perform a more in-depth analysis of these results, the statistical parameters (median, mean) were obtained. Table 9 gives a summary of these statistical parameters.

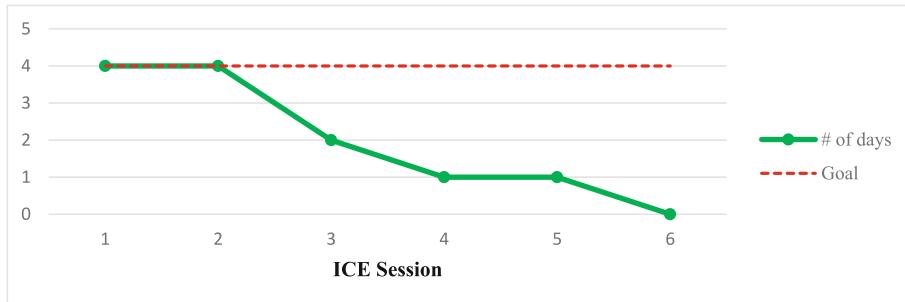


Fig. 8. The # of days of updated VR after the ICE session.

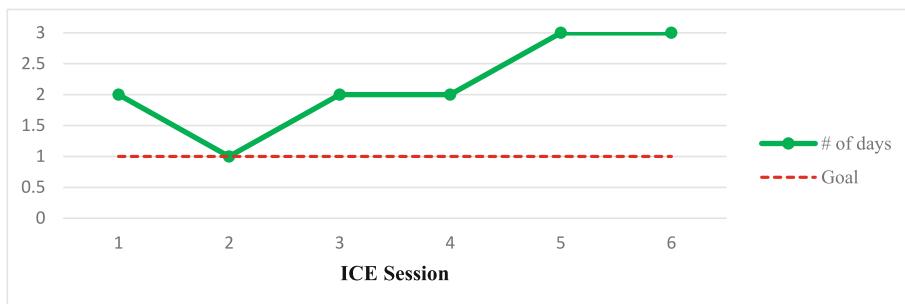


Fig. 9. The # of days that the report is ready before the ICE session.

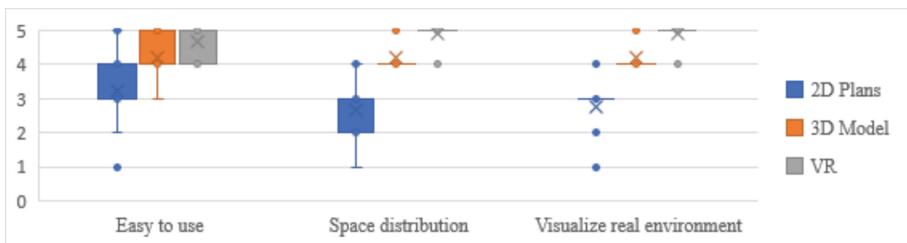
6 Discussion

Regarding the ICE-1 and ICE-2 metrics, these had an increasing evolution over time, and meet the established goal, reaching 100% in the last ICE session. Additionally, the total number of interferences and proposals decreased because as the redesign of the new laboratories progressed, the final changes that these would have, were gradually implemented. In the case of the controllable factors ICE-3 and ICE-4, in both cases there was a drop in the second ICE session and this was due to the fact that at that time the laboratory directors were busy with academic activities (final evaluations), which led to this drop and even caused the attendance percentage goal not to be met in the case of ICE-4.

Regarding the BIM metrics and controllable factors, the BIM-1 met the goal of the LOD, going from the level at which the project was received of 100 to the desired LOD 350. On the other hand, in BIM-2 the goal of 100% was not achieved and this was due to the fact that after the first ICE session, two of the directors refused to continue using VR in the following sessions because it was uncomfortable and even made him dizzy. In the case of BIM-3 and BIM-4, the goal was achieved in both cases, but it should be noted that although the VR was updated the same day the model was updated, this was not the case in the first weeks because the team in charge did not fully master the creation of the VR based on a 3D model. Regarding the PPM metrics and controllable factors,

Table 8. Results of the survey of use of VR.

| Laboratory | Easy to use | | | Space distribution | | | Visualize real environment | | |
|---------------------------|-------------|----------|----|--------------------|----------|----|----------------------------|----------|----|
| | 2D Plans | 3D Model | VR | 2D Plans | 3D Model | VR | 2D Plans | 3D Model | VR |
| Civil Engineering | 4 | 4 | 5 | 4 | 4 | 5 | 3 | 4 | 5 |
| | 3 | 4 | 5 | 2 | 4 | 5 | 3 | 4 | 5 |
| Computer Science | 1 | 4 | 5 | 1 | 4 | 5 | 1 | 4 | 5 |
| Mechanical Engineering | 2 | 5 | 4 | 2 | 5 | 4 | 2 | 5 | 4 |
| Environmental Engineering | 3 | 5 | 4 | 3 | 4 | 5 | 3 | 4 | 5 |
| Bioengineering | 5 | 4 | 5 | 2 | 4 | 5 | 3 | 4 | 5 |
| Energy Engineering | 5 | 4 | 5 | 4 | 5 | 5 | 4 | 5 | 5 |
| Science Area | 3 | 5 | 4 | 3 | 4 | 5 | 3 | 4 | 5 |
| | 3 | 3 | 5 | 3 | 4 | 5 | 3 | 4 | 5 |

**Fig. 10.** Survey results by criteria.**Table 9.** Statistical parameters of the survey.

| Laboratory | Easy to use | | | Space distribution | | | Visualize real environment | | |
|------------|-------------|----------|------|--------------------|----------|------|----------------------------|----------|------|
| | 2D Plans | 3D Model | VR | 2D Plans | 3D Model | VR | 2D Plans | 3D Model | VR |
| Median | 3 | 4 | 5 | 3 | 4 | 5 | 3 | 4 | 5 |
| Mean | 3.22 | 4.22 | 4.67 | 2.67 | 4.22 | 4.89 | 2.78 | 4.22 | 4.89 |

PPM-1 and PPM-2 met the established goals. In the case of PPM-1, as in other metrics, it had an increasing evolution but in the second session it had a drop due to academic activities.

When analyzing the results of the survey on the use of VR, it can be determined that VR was very useful in the present project based on the three criteria used according to the objective of redesigning the new laboratories since the mean and the median for VR is higher than those of the other tools, which shows that most of the responses to the use of VR tend to be higher than those of the other tools. However, in the case of the last two criteria, atypical data are observed for the 3D model and VR, which is due to the fact that two directors did not feel comfortable using VR and preferred the use of 3D model because this does not make them dizzy.

7 Conclusion

The research shows that the use of the VDC methodology in conjunction with VR allows the project objectives to be achieved and the client's needs to be met, such as compliance with the deadline and budget of this project by avoiding rework in the design phase and its impact on future rework during construction, as well as an improvement in the design phase with the use of VR compared to other current tools such as 2D plans or 3D models due to the improved understanding of the final result of the project.

Being the benefits of VDC implementation with VR stated, there are obstacles that must be considered in future projects, such as the fact that not all those involved in the project will be willing to use VR since it is a new technology in the construction industry for which a way must be found to be comfortable to use for everyone or a method in which the use of this is complemented with other tools such as 3D models in this project where 3D models were used as a common language for all and VR for specific objectives during ICE sessions.

References

1. Barbosa, F.: Reinventing construction: a route to higher productivity. *McKinsey Global Institute* (2017)
2. Love, P.E.D.: Influence of project type and procurement method on rework costs in building construction projects. *J. Constr. Eng. Manag.* **128**(1), 18–29 (2002)
3. Arashpour, M., Wakefield, R., Blismas, N., Lee, E.W.M.: A new approach for modelling variability in residential construction projects. *Aust. J. Constr. Econ. Build.* **13**(2), 83–92 (2013)
4. Mahamid, I.: Study of relationship between rework and labor productivity in Building Construction Projects. *Rev. Construcción* **19**(1), 30–40 (2020)
5. Arashpour, M., Wakefield, R., Blismas, N., Lee, E.W.M.: Analysis of disruptions caused by construction field rework on productivity in residential projects. *J. Constr. Eng. Manag.* **140**(2), (2014)
6. Dalton, T., Wakefield, R., Horne, R.: Australian suburban house building: industry organization, practices and constraints. *Australian Housing and Urban Research Institute (AHURI)*, Melbourne, VIC (2011)

7. Hegazy, T., Said, M., Kassab, M.: Incorporating rework into construction schedule analysis. *Autom. Constr.* **20**(8), 1051–1059 (2011)
8. Subgerencia de Seguimiento y Evaluación del Sistema Nacional de Control (SESNC): Reporte de Obras Paralizadas en el Territorio Nacional a setiembre 2023. Contraloría General de la República (2023)
9. Love, P.E.D., Edwards, D.J., Watson, H., Davis, P.: Rework in civil infrastructure projects: determination of cost predictors. *J. Constr. Eng. Manag.* **136**(3), 275–282 (2010)
10. Li, Y., Taylor, T.R.B.: The impact of design rework on construction project performance. In: Annual Conference of the Canadian Society for Civil Engineering, pp. 2659–2668. Ottawa, Canada (2011)
11. Burati, J.J.: Causes of quality deviations in design and construction. *J. Constr. Eng. Manag.* **118**(1), 34–49 (1992)
12. Josephson, P., Hammarlund, Y.: Causes and costs of defects in construction: a study of seven building projects. *Autom. Constr.* **8**(6), 681–687 (1999)
13. Love, P.E.D., Li, H.: Quantifying the causes and costs of rework in construction. *Constr. Manag. Econ.* **18**(4), 479–490 (2000)
14. Alizadehsalehi, S., Hadavi, A. and Huang, J.: Virtual reality for design and construction education environment. In: Architectural Engineering Conference, pp. 193–203. Virginia, United States (2019)
15. Haggard, K.: Case study on virtual reality in construction (2017)
16. Belsvik, M.R., Lædre, O., Hjelseth, E.: Metrics in VDC projects. In: 27th Annual Conference of the International Group for Lean Construction, pp. 1129–1140. Dublin, Ireland (2019)
17. Aslam, M., Gao, Z., Smith, G.: Integrated implementation of Virtual Design and Construction (VDC) and lean project delivery system (LPDS). *J. Build. Eng.* **39**, (2008)
18. Kunz, J., Fischer, M.: Virtual design and construction: themes. Center for Integrated Facility Engineering (CIFE), Stanford University, Case Studies and Implementation Suggestions (2012)
19. Alarcon, L.F., Mandujano, M.G., Mourgues, C.: Analysis of the implementation of VDC from a lean perspective: literature review. In: 21th Annual Conference of the International Group for Lean Construction, pp. 781–790. Fortaleza, Brazil (2013)
20. Khanzode, A., Fischer, M., Reed, D., Ballard, G.: A guide to applying the principles of Virtual Design and Construction (VDC) to the lean project delivery process. Stanford University, Center for Integrated Facility Engineering (CIFE) (2006)
21. Gilligan, B., Kunz, J.: VDC use in 2007: significant value, dramatic growth, and apparent business opportunity. Stanford University, Center for Integrated Facility Engineering (CIFE) (2007)
22. Rischmoller, L., Reed, D., Khanzode, A., Fischer, M.: Integration enabled by virtual design & construction as a lean implementation strategy. In: 26th Annual Conference of the International Group for Lean Construction, pp. 240–249. Chennai, India (2018)
23. Breistein, H., Lædre, O.: Implementing VDC. In: 30th Annual Conference of the International Group for Lean Construction, pp. 937–948. Edmonton, Canada (2022)
24. Gustafsson, M., Gluch, P., Gunnemark, S., Heinke, K., Engström, D.: The role of VDC professionals in the construction industry. *Procedia Econ. Financ.* **21**, 478–485 (2015)
25. Tuesta, R., Vicuña, M., Savio, A.A.D., Palpan, A., Valle, E., Quiroz, F.: Prefabricated reinforcement in construction using VDC: case study Ovalo Monitor Bridge. In: 30th Annual Conference of the International Group for Lean Construction, pp. 1008–1019. Edmonton, Canada (2022)
26. Simonsson, P., Emborg, M.: Industrialization in Swedish bridge engineering: a case study of lean construction. In: 15th Annual Conference of the International Group for Lean Construction, pp. 244–253. Michigan, USA (2007)

27. Devine, R.D., Barbachyn, S.M., Thrall, A.P., Kurama, Y.C.: Effect of tripping prefabricated rebar assemblies on bar spacing. *J. Constr. Eng. Manag.* **144**(11) (2018)
28. Bowman, D.A., McMahan, R.P.: Virtual reality: how much immersion is enough? *Computer* **40**(7), 36–43 (2007)
29. DeFanti, T.A., et al.: The StarCAVE, a third-generation CAVE and virtual reality OptIPortal. *Futur. Gener. Comput. Syst.* **25**(2), 169–178 (2009)
30. Wohlgenannt, I., Simons, A., Stieglitz, S.: Virtual reality. *Bus. Inf. Syst. Eng.* **62**(5), 455–461 (2020)
31. Burdea, G.C., Philippe, C.: *Virtual Reality Technology*, 2nd edn. Wiley-IEEE Press, New York (2003)
32. Bamodu, O., Ye, X.: Virtual reality and virtual reality system components. *Adv. Mater. Res.* **765–767**, 1169–1172 (2013)
33. Woksepp, S., Olofsson, T., Jongeling, R.: Design reviews and decision-making using collaborative virtual reality prototypes: a case study of the large-scale MK3 project. In: 13th Annual Conference of the International Group for Lean Construction, pp. 145–152. Sydney, Australia (2005)
34. Zaker, R., Coloma, E.: Virtual reality-integrated workflow in BIM-enabled projects collaboration and design review: a case study. *Vis. Eng.* **6**(4), (2018)
35. Azhar, S.: Role of visualization technologies in safety planning and management at construction jobsites. *Procedia Eng.* **171**, 215–226 (2017)
36. Hafnia, M., Monacelli, E., Martin, H.: Virtual reality simulator for construction workers. In: Proceedings of the Virtual Reality International Conference (VRIC 18), pp. 1–7. Laval, France (2018)
37. Ahmad, A., Yitmen, I., Alizadehsalehi, S., Celik, T.: Adoption of Virtual Reality (VR) for site layout optimization of construction projects. *Tek. Dergi* **31**(2), (2019)
38. Alvarez, J., Bárcena, V., Chunga, E., Jaliri, J.: Implementación de la metodología VDC en la etapa de planeamiento. Caso de estudio: mejoramiento del servicio institucional de la sede central del Gobierno Regional de Tacna. Universidad Peruana de Ciencias Aplicadas (UPC), Perú (2021)
39. Vergara, M.: VDC en la implementación de un modelo de inteligencia artificial para la supervisión remota de obra CEBUL al 65% de desarrollo. Universidad de Lima, Center for Integrated Facility Engineering (Standford University), Standford Center for Professional Development (2020)
40. Vidal, J.F.: VDC framework applied to the ‘Centro de Bienestar Universitario’ project. Universidad de Lima, Center for Integrated Facility Engineering (Standford University), Standford Center for Professional Development (2020)
41. Del Savio, A.A.: VDC Methodology Applied to the Conceptual Design of Civil Engineering Laboratories. Universidad de Lima, Standford Center for Professional Development (2020)
42. Soliman-Junior, J., Awwal, S., Tzortzopoulos, P., Ayo-Adejuyigbe, M., Kagioglou, M.: Eliciting requirements in social housing retrofit projects: tools and processes within a living lab setting. In: 30th Annual Conference of the International Group for Lean Construction, pp. 468–479. Edmonton, Canada (2022)
43. Liu, C., González, V.A., Liu, J., Rybkowski, Z., Schöttle, A., Mourgues, C., Pavez, I.: Accelerating the Last Planner System® (LPS) uptake using virtual reality and serious games: a sociotechnical conceptual framework. In: 28th Annual Conference of the International Group for Lean Construction, pp. 481–492. Berkeley, California, USA (2020)
44. Orihueta, P., Noel, M., Pacheco, S., Orihueta, J., Yaya, C., Aguilar, R.: Application of virtual and augmented reality techniques during design and construction process of building projects. In: 27th Annual Conference of the International Group for Lean Construction, pp. 1105–1116. Dublin, Ireland (2019)



Evaluating Impact of Smart Technologies Within Construction Site and Office Operations Among South African Contractors

Johannes S. Mchunu¹, Iruka C. Anugwo¹(✉), and Jachike C. Anugwo²

¹ Department of Construction Management & Quantity Surveying, Durban University of Technology (DUT), Durban, South Africa
IrukaA@dut.ac.za

² Department of Infrastructure Planning and Maintenance, Olam OK Foods, Lagos, Nigeria

Abstract. The purpose of this research study is to assess the impact of the implementation of emerging ICT technologies within the construction sites and offices among Durban South African contractors. The study explores various emerging Information and Communication Technology (ICT) technologies and how they can be used to improve construction project delivery in South African construction industry. The questionnaires were formulated to evaluate the level of implementation among construction contractors towards advancing their business operations with the emerging technologies presently linked to the Fourth Industrial Revolution (4IR). This study adopted a quantitative research, whilst data collection was done through a questionnaire survey among 145 contractors (105 medium contractors and 40 large contractors) in Durban, South Africa. The data was analysed using descriptive statistics, frequencies and percentages. The data were analysed by employing percentages, mean score, standard deviation, and ranking each question using the SPSS Statistical Package. Findings using Mann-Whitney test, in which the study revealed that the adequate implementation of smart construction technologies embedded in the elements of fourth industrial revolution applications would be significantly beneficial for their overall organisations' productivity in terms of project design, procurement and construction delivery. The study findings reveal that "electronic procurement (e-procurement), Building Information Modelling (BIM), robotic technology, and "Three-dimensional (3D) technology", usage of the Internet of Things (IoT), and the utilisation of Cyber Physical System (CPS), and usage of Radio Frequency Identification (RFID) technology were ranked high as most significantly elements of 4IR that would address the potential and actual challenges associated construction, and advance construction project delivery. This study concluded that the implementation of smart construction technologies by the South African construction companies will significantly improve their business performance, construction site operation, office productivity and overall competitiveness of the industry.

Keywords: Construction industry · Contractors · Smart technologies · South Africa

1 Introduction

Globally, there is an evolving transition of industries activities from the Third to the Fourth Industrial Revolution (4IR), referred to as Industry 4.0. However, it seems that the construction industry activities is still based on traditional concepts and techniques which are time-consuming, expensive, and potential for error during project delivery (Pereira da Silva et al. 2022). It is evident that many countries and economic sectors around the world are not fully prepared for the 4IR and South Africa (SA) and its construction sector are no exception. The elements of 4RI has the potential to technological transform construction processes and delivery, thereby introducing more efficient and sustainable methods empowered by emerging technologies (Pereira da Silva et al. 2022). According to the Minister of International Relations and Cooperation (Sisulu, 2019), South African industries are far behind in terms of competitiveness due to the advent of the Fourth Industrial Revolution. South Africa's Minister of Communication, Ndabeni-Abrahams (2019) also acknowledged that Artificial Intelligence (AI) technologies are developing at such as a fast pace that the country would not be able to cope. As such there is need to create awareness and foster readiness towards harnessing the benefits of the Fourth Industrial Revolution. Ndabeni-Abrahams (2019) further asserted that the government was striving earnestly to create an enabling Information, Communication and Technology (ICT) infrastructure to enhance awareness, capacity and eagerness among South Africans to deal with the emerging technologies. The fourth industrial revolution is a fusion of all current and emerging technologies (Chung and Kim 2016). The fourth industrial revolution confluence of technologies ranging from a variety of digital technologies e.g. 3D printing, Internet of Things, advanced robotics, artificial intelligence etc. These technologies have the potential to revolutionise operations and supply chain management within construction industry (Dallasega et al. 2018; Koh et al. 2019). Industry 4.0 is an emerging concept deriving from technological advancement and disruptive developments in the industrial sector worldwide in the past few years (Dallasega et al. 2018; Koh et al. 2019). The 21st century economic environment has ushered both opportunities and challenges, particularly if one considers the disruptive potential of the 4IR. In construction industry, the labour costs are high and manual operations are prone to errors, health and safety (H&S) and inaccuracies. The application of automated solutions that based on robots and robotic arms would reduce the labour crisis and challenges on construction sites, which essentially means that safety on sites can be improved (Li 2017). The emerging technologies in embed in the 4RI offers practical alternatives to conventional construction project process and delivery that are currently used in the Architecture, Engineering, and Construction (AEC) industry (Cabibihan et al. 2023).

According to Lacovidou, Purnell and Lim (2018), the last decades, a number of advanced “smart” technologies have emerged including radio frequency identification (RFID) tags, optical character recognition, 3D scanning laser, building information modelling (BIM), etc., and these technological elements has becoming important tools in the construction sector. Thus, the smart technologies leverage the industry 4.0 dimensions and foster the interconnectivity, which provide intelligence to the new construction system, and enable construction site participants to track and monitor the site activities (Frank et al. 2019). These technologies according to Osunsanmi et al. (2018), are the

roadmap for the implementation of construction industry 4.0 concepts in the South African construction industry. Nevertheless, the South African construction industry is notoriously resistant to change due to its unwavering commitment to protect jobs and unpreparedness to abandon familiar operations (Pillay et al. 2017). The resistance to change in the construction industry has contributed to the slow paced adoption and implementation of emerging innovative technologies (Pillay et al. 2017). Notwithstanding the general negative attitude towards new technologies within the industry, the Fourth Industrial Revolution presents immense opportunities and potential disruptions to the construction business, especially in developing countries (You and Feng 2020; Aghimien et al. 2021).

1.1 Research Hypothesize

H1a: The implementation of Smart construction ICT technologies in project design, procurement and delivery of construction projects in Durban is at an average level.

H1o: The implementation of Smart construction ICT technologies in project design, procurement and construction delivery in Durban is at an advanced level.

2 Literature Review and Theoretical Background

2.1 An Overview of the Smart Construction Technologies

In recent years, as a result of the 4th industrial revolution and rapid development and maturity of emerging information technology, an agile construction system has become increasingly important for built environment project process, services and product delivery and management systems (Yang et al. 2018). The elements of Smart Construction Systems (SCS) such as Virtual Reality (VR), Argument Reality (AR), sensor technology, robots, Building Information Modelling (BIM), air drones, and Internet of Things (IoT), are undoubtedly making construction a must-have among hi-tech industries (Liu et al. 2018). Smart Construction Objects (SCOs) are construction resources (e.g. machinery, equipment and building materials) that are made “agile” by developing themselves with technologies that provide independence, awareness, and interpersonal skills. This “intelligence” can make it easier to make informed decisions, and understanding SCOs (Niu et al., 2016). In addition, Li (2017) states that the elements of Smart Work Environments (SWE) would enable the construction industry to effectively monitor its operations, personnel, tools, and equipment in the workplace, aimed at improving environmental safety management and sustainability practice. Smart technology such as robotic tool can significantly reduce challenges of physical demands, health and safety risks on construction workers thus improving both the productivity and safety of construction projects (Wang et al. 2023).

Li (2017) stated that the provision of technological infrastructure promotes virtual and physical communication based on ubiquitous technology, as well as IoT. Therefore, this approach offers a promising opportunity within SWE and physical cyber systems. Besides, in the construction industry, labour costs are high and handicrafts tend

to be flawed and inaccuracies. The use of automated solutions based on robots and robotic arms would reduce personnel problems and challenges in construction sites, which means that site safety on site can be improved. The utilization of computerized solutions that are based on automation and robotic arms would lessen the labour crisis and challenges on construction sites, substantially meaning that safety on sites could be improved (Li 2017: 302). According to Iacovidou, Purnell and Lim (2018), in recent decades, advanced “smart” technologies have emerged, including radio frequency identifications tags (RFID), visual character recognition, 3D scanning lasers, building information model (BIM), etc., and these technological elements have become important tools in the construction industry. In order for construction processes to qualify as smart, technology components such as elements of fourth industrial revolution are crucial. According to Wang et al. (2016), the basic technologies are composed of a so-called Cyber Physical System (CPS), and Computer-Aided Technology (CAT), such as Radio frequency identification (RFID) and Internet of Things (IoT). The smart technologies expand the 4IR magnitudes and promote linkage, which provides intelligence to the new construction system and enables construction site participants to track and monitor the site activities (Frank et al. 2019). These technologies, according to the studies (Osunsanmi et al. 2018) provide a guide to the introduction of construction industry 4.0 in the South African construction industry. The name Construction 4.0 is coined from the concept of industry 4.0, which is linked to the 4IR.

2.2 Understanding the Concept of Smart Supply Chain in the Construction Industry

Benton and McHenry (2010) define Supply Chain Management (SCM) as a system for continuously integrating a variety of business functions and processes across a range of value-added customer needs, with the aim of improving performance. According to Benton and McHenry (2010), the Construction Supply Chain Management (CSCM) is a key management tool to manage the flow of business information, tasks, activities, and processes involving various networks of private organizations and communications. It produces an improved value that is delivered to customers in the form of a completed project (Benton and McHenry 2010). The construction sector has a flexible and uncontrollable environment, where tracking and accessing information related to components/materials, plant and equipment can be a major challenge (Ergen and Akinci 2007). Thus, traditional methods of managing materials and equipment require more workers in terms of data collection methods for construction processes and their components (Ergen et al. 2007). Many of the current field processes on construction sites still rely on manual processes for asset tracking and information management (Jang and Skibniewski 2009). Thus, construction data collected using manual methods often has errors and is unreliable or has incomplete results due to its dependence on motivation and human skills (workers in general) (Ergen et al. 2007).

According to Ergen, Akinci and Sacks (2007), data collected by traditional methods is often transmitted and stored in a paper-based format, difficult to locate and access at a glance, and which makes the processing of useful information expensive and unreliable. Therefore, part of the information ends up being not readily available to the people involved, who need to access the information in time to make effective decisions. Thus,

failure to successfully track building materials and information related to machinery and equipment during planning and construction work may result in project delays and increased labour costs (Ergen and Akinci 2007; Sardroud 2012). Therefore, successful supply chain management is essential for effective competition in today's global markets (Papadopoulos et al. 2016). In addition, the advent of advanced information technology and sensor solutions has created an understanding of some shortcomings of the manual process (Grau et al. 2009). Papadopoulos et al. (2016), state that one of the best ways to improve CSCM is to adopt and use information technology (IT). This approach will improve the supply chain operation, assist the supply chain performance, and enable supply chain integration. In a combined construction supply chain, information is shared and made available to members and thus improving the visibility of the supply chain, which avoids information delays and distortions (Liu and Chua 2016). The flow of information in current building settings may not work properly because most of the information goes through an inefficient and unorganised system (Liu and Chua 2016). According to Araszkiewicz and Tryfon-Bojarska (2017), the use of digital tools could greatly enhance the preparation and execution of building projects. Therefore, digitalisation is much needed to improve the flow of information, regarded as the lifeblood of the construction supply chain (Liu and Chua 2016). The construction sites are very vigorous as building materials, which could arrive daily, are staged in different locations and their locations might change several times before being placed in their final locations within the facility (Ergen and Akinci 2007). Unlike factory industries, construction sites have unregulated areas, where pieces of equipment do not follow prearranged routes when moving or do not operate in a predetermined manner (Ergen and Akinci 2007).

3 Research Methodology

3.1 Research Approach Adopted

This research paper adopted quantitative research method “utilising questionnaire survey (online and hand delivery)”. Primary data was sourced through research questionnaires. The reason for using the quantitative method is that the techniques of the quantitative approach helped to maximise the knowledge yielded of the research endeavour (Creswell 2003). Additionally, quantitative techniques permit the gathering of data that is rich in detail as required by the current study (Creswell 2003). Thus, the collected data in this study was analysed using percentages, mean score, standard deviation, and ranking each question with the support of using the SPSS statistical package. The target population for this research study was medium and large construction contractors, within civil engineering and general building. The calculation of sample was based on CIDB grading of contractors between grades 4 to 9 on the Construction Industry Development Board (CIDB) Register of Contractors (RoC).

3.2 Data Collection Techniques Adopted

According to Chaleunvong (2009), data gathering techniques allow the researcher to systematically gather knowledge about his or her study material (people, objects, events),

and the settings in which it occurs. In data collection, the researchers should be organised. If data is collected randomly, it will be difficult to answer the research questions in a comprehensive way (Chaleunvong 2009). The researcher administered research questionnaires through an online survey using the Google form-online based approach as well as normal paper-based questionnaires through mail and hand delivery. The questionnaire was distributed among Durban construction contractors in grades 4 to 9 (as respondents) who are duly registered with the Construction Industry Development Board (CIDB) Register of Contractors (RoC), within civil and general building in the eThekwini region (See Table 1). The responses were recorded as yes or no, as well as on a Likert-type 5-point scale, in which 1 represents strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 total agreement, and reliability and validity issues were also assessed.

$$\text{MIS} = \frac{1n1 + 2n2 + 3n3 + 4n4 + 5n5}{\Sigma N}$$

where;

n1 = number of respondents for strongly disagree

n2 = number of respondents for disagree

n3 = number of respondents for neutral

n4 = number of respondents for agree

n5 = number of respondents for strongly agree

N = Total number of respondents

Source CIDB (2020)

Grade 1–3 = small contractors; Grade 4–6 = medium contractors & 7–9 = Large contractors.

3.3 Sampling Size

Based on the estimated population provided by the Construction Industry Development Board (CIDB), which is 194 (medium contractors) and 56 (large contractors), the researcher used Slovin's formula with a margin of error of $\pm 5\%$, and the confidence level of 95%. The calculated sample size was as follows:

$$n = \frac{N}{1 + N(e^2)}$$

where n = sample size, N = number of target population, e^2 = error of margin.

According to the CIDB register of contractors' site, there are 194 contractors listed under CIDB level 4 to 6 for GB and CE classes.

Based on the above formula, therefore, the population is:

$$n = \frac{194}{1 + 194(0.05^2)} \\ n = 130.6$$

Say = 131 contractors (Sample size of medium contractors)

Table 1. The Durban CIDB Register of Contractors—Research Population

| Contractors Grading Designation | Maximum Tender Value (R) | Total Contractor (Active—Currently in Durban) | Population Proportion (%) | Size of contractor |
|------------------------------------|-----------------------------|---|---------------------------|--------------------|
| 1 | R200 000.00 | 7141 | 89.34 | Small |
| 2 | R500 000 to R1000 000 | 404 | 5.05 | Small |
| 3 | R1000 000 to R3000 000 | 198 | 2.48 | Small |
| 4 | R3000 000 to R6000 000 | 102 | 1.28 | Medium |
| 5 | R6000 000 to R10 000 000 | 53 | 0.66 | Medium |
| 6 | R10 000 000 to R20 000 000 | 39 | 0.49 | Medium |
| 7 | R20 000 000 to R60 000 000 | 33 | 0.41 | Large |
| 8 | R60 000 000 to R200 000 000 | 20 | 0.25 | Large |
| 9 | R200 000 000 to No limit | 3 | 0.04 | Large |
| Grand Total | | 8118 | 100.00 | |
| Overall Sample size for this study | | | | |
| 1 | Medium | 131 | 73 | Medium |
| 2 | Large | 49 | 27 | Large |
| Grand Total | | 180 | 100 | |

Note Grade 1–3 = small contractors; Grade 4–6 = medium contractors & 7–9 = Large contractors.
Source CIDB (2023)

According to the CIDB register of contractors' site, there are 56 contractors listed under CIDB level 7 to 9 for GB and CE classes

$$n = \frac{56}{1 + 56(0.05^2)}$$

$$n = 49.1$$

Say = 49 contractors (Sample size of large contractors)

The research questionnaire was distributed randomly among selected research respondents (grade 4–9 contractors), based on the sample size through two main strategies, namely: online (survey monkey), emailing and when necessary, direct drop-offs.

4 Research Result and Discussion

4.1 Respondents' Demographics Profile

The research result shows that males (76%) are the majority compared to females (24%) respondents. It was found that from the medium contractors 78.6% were males compared to large contractors who had 72.5% males. Females from the medium contractors were 21.4% compared to large contractors who had 27.5%. It is evident from this study that the construction industry remains male dominated. Findings indicate that respondents between 18 to 30 years of age from the medium contractors were 5.8% and from the large contractors were 20.0%. Respondents between 31 to 45 years of age were 51.5% from medium contractors and 57.5% from large contractors. Respondents aged between 46 to 55 years were 34.0% from medium contractors and 20.0% from large contractors. Respondents aged 56 and above were 8.7% from medium contractors and 2.5% from large contractors.

Thus, it is evident that the respondents within ages of 31 to 45 years were in strategic positions as dominating figures in the industry, as the sample indicated 51.5% were from medium contractors and 57.5% from large contractors, followed by those aged between 46 to 55 years at 34.0% from medium contractors, and 20.0% from large contractors. It can therefore be concluded as that the respondents that make up the survey sample are mature and have a high possibility of having a vast wealth of experience in the construction industry. The designation of the sampled respondents, shows that 40.8% from medium contractors and 15.0% from large contractors were business owners/directors, whilst 11.7% respondents from medium contractors and 2.5% from large contractors were in strategic job positions as managers. In addition, 32.0% respondents from medium contractors and 17.5% from large contractors were in an executive firm's representative's position, whilst 15.5% from medium contractors and 65% from large contractors were noted as other designations.

The results show that medium contractors in their everyday work are managed by owners/directors, whereas the opposite applies to the large contractors where the majority of the respondents were the employees. This make sense because large contractors are more capable of handling large-scale projects as per their grade than the medium-size contractors. Finding shows that 58.3% from medium contractors and 72.5% from large contractors had between 5 to 15 years of experience, 32.0% from medium contractors and 15.0% from large contractors had between 16 to 25 years of experience, 6.8% from medium contractors and 7.5% from large contractors had 26 to 35 years of experience, while 2.9% from medium contractors and 5.0% from large contractors had 36 to 45 years of experience. 0% experience for both medium and large contractors between 46 to 55 and 56 above. The results show that the respondents from large contractors were slightly more experienced when compared to medium contractors' respondents. Looking at 26 to 35 and 36 to 45 years of experience on the above graph, it is evident that respondents from large contractors were marginally more experienced compared to respondents from medium contractors.

The Potential and Actual Benefits Implementation of Smart Technology Within Construction Site and Office with Elements of Fourth Industrial Revolution (4IR).

Tables 2 and 3 reveal the respondents' ranking on the perceived and actual benefits on implementation of smart technology within construction site and office as emanating from elements of Fourth Industrial Revolution (4IR). The implementation of ICT technologies in project design, procurement and construction delivery were ranked as follows (Comparison between medium and large contractors):

Medium Contractors' response on implementation of Smart Technologies.

The "electronic procurement (e-procurement)", which means it would be most beneficial to utilise the electronic communication technologies (ICT) in conducting business operations, was ranked first with a mean score of 4.41 and standard deviation (SD) = 0.910. Agreement among medium contractors that utilising Building Information Modelling (BIM) in planning and procuring a construction project would be most beneficial in all phases of the facility's life cycle ranked second with a mean score of 3.98 and standard deviation (SD) = 1.036.

The hypotheses that implementation of "robotic technology", and "Three-dimensional (3D) technology" would advance construction delivery and operations were ranked third with a mean of 3.90 and standard deviation (SD) = 0.890; and fourth with a mean score of 3.86 and standard deviation (SD) = 1.051; respectively. Thus, the "application of Virtual Reality (VR) was ranked fifth with a mean score of 3.76 and standard deviation (SD) = 1.089; usage of the Internet of Things (IoT), and the utilisation of Cyber Physical System (CPS) system operations to monitor (plant and equipment) were ranked sixth, with mean score of 3.63 and standard deviation of 1.018 and seventh, with mean score of 3.49 and standard deviation of 1.040 respectively. However, the usage of Radio Frequency Identification (RFID) technology to address the potential and actual challenges associated with tracking resources on construction sites was ranked eighth with mean score of 3.46 and standard deviation of 1.174.

Large Contractors' response on implementation of Smart Technologies.

Table 3, the result reveals that utilisation of "electronic procurement (e-procurement)" i.e. the electronic communication technologies (ICT) in conducting construction business operations is most beneficial and it was ranked first with a mean score of 4.51 and standard deviation (SD) = 0.644. However, the result shows that "implementing of robotic technology", and utilisation of Building Information Modelling (BIM) in planning and procuring a construction project would be most beneficial in the current era of industry revolution as they were ranked respectively second with a mean of 4.13 and standard deviation (SD) = 0.741; and third with mean score of 4.09 and standard deviation (SD) = 0.919.

Thus, it was perceived that implementation of Three-Dimensional (3D) technology, and application of Virtual Reality (VR) would significantly enhance construction delivery as they were ranked fourth with mean score of 3.97 and standard deviation (SD) = 0.890; and fifth with mean score of 3.93 and standard deviation (SD) = 0.842 respectively.

The usage of Radio Frequency Identification (RFID) technology to address the challenges associated with tracking resources on construction sites in the large contractors were ranked sixth, with mean score of 3.90 and standard deviation of 0.860. In addition, the result emanating from the study reveals that usage of the Internet of Things (IoT), and utilisation of Cyber Physical System (CPS) system operations to monitor (plant

Table 2. Response Analysis on Benefit of Implementing of Smart Construction ICT Technologies within Construction Site and Office: Medium Contractors' Perspectives

| Questions | Mean Score | Std. Dev | Asymp. Sig (2-sided) | Rank |
|---|------------|----------|----------------------|------|
| Electronic procurement (e-procurement) utilising the electronic communication technologies (ICT), is used in transaction processes to buy services, goods, and works to improve delivery, reduce paperwork and lower administrative costs | 4.41 | 0.91 | 0.000 | 1 |
| The use of Building Information Modelling (BIM) in planning and procuring construction project is benefiting all phases of the facility's life cycle | 3.98 | 1.04 | 0.000 | 2 |
| Implementing of Robotic technology has the capability to generate higher output at a lower unit cost, with better quality product | 3.9 | 0.89 | 0.000 | 3 |
| Implementing Three-Dimensional (3D) technology brings significant benefits to the organisation, in terms of increased customisation, reduced construction time, reduced manpower and construction cost | 3.86 | 1.05 | 0.000 | 4 |
| Application of Virtual Reality (VR) allows costly mistakes to be identified and rectified before they occur and easy communication among site staff and office staff | 3.76 | 1.09 | 0.000 | 5 |
| Safety systems in construction sites have integrated the usage of the Internet of Things (IoT) | 3.63 | 1.02 | 0.000 | 6 |

(continued)

Table 2. (*continued*)

| Questions | Mean Score | Std. Dev | Asymp. Sig (2-sided) | Rank |
|--|------------|----------|----------------------|------|
| Cyber Physical System (CPS) system operations are used for monitoring (plant and equipment), coordinating, progress tracking, construction process control, and as-built documentation | 3.49 | 1.04 | 0.002 | 7 |
| Radio Frequency Identification (RFID) technology is used to address the challenges associated with tracking resources on construction sites | 3.46 | 1.17 | 0.003 | 8 |

Source Researchers' construct (2023)

and equipment) has significant potential to advance construction delivery as they were ranked respectively sixth with a mean score of 3.90 and standard deviation of 0.772; and seventh with a mean score of 3.87 and standard deviation of 0.548.

Table 3. Response Analysis on Benefit of Implementing of Smart Construction ICT Technologies within Construction Site and Office: Large Contractors' Perspectives

| Questions | Mean Score | Std. Dev | Asymp. Sig (2-sided) | Rank |
|---|------------|----------|----------------------|------|
| Electronic procurement (e-procurement) utilising the electronic communication technologies (ICT), is used in transaction processes to buy services, goods, and works so as to improve delivery, reduce paperwork and lower administrative costs | 4.41 | 0.91 | 0.000 | 1 |
| Implementing of robotic technology has the capability to generate higher output at a lower unit cost, with better quality product | 4.13 | 0.74 | 0.000 | 2 |

(*continued*)

Table 3. (*continued*)

| Questions | Mean Score | Std. Dev | Asymp. Sig (2-sided) | Rank |
|--|------------|----------|----------------------|------|
| The use of Building Information Modelling (BIM) in planning and procuring a construction project is benefiting all phases of the facility's life cycle | 4.09 | 0.92 | 0.000 | 3 |
| Implementing Three-Dimensional (3D) technology brings significant benefits to the organisation, in terms of increased customisation, reduced construction time, reduced manpower and construction cost | 3.97 | 0.89 | 0.000 | 4 |
| Application of Virtual Reality (VR) allows costly mistakes to be identified and rectified before they occur and fosters easy communication among site staff and office staff | 3.93 | 0.84 | 0.000 | 5 |
| Safety systems in construction sites have integrated the usage of the Internet of Things (IoT) | 3.9 | 0.77 | 0.000 | 6 |
| Cyber Physical System (CPS) system operations are used for monitoring (plant and equipment), coordinating, progress tracking, construction process control, and as-built documentation | 3.87 | 0.55 | 0.000 | 7 |
| Radio Frequency Identification (RFID) technology is used to address the challenges associated with tracking resources on construction sites | 3.9 | 0.86 | 0.000 | 8 |

Source Researchers' construct (2023)

5 Conclusion and Recommendations

The study revealed that a significant number of construction contractors that affirmed and opined that the adequate implementation of smart construction technologies embedded in the elements of fourth industrial revolution applications would be significantly beneficial for their overall organisations' productivity in terms of project design, procurement and construction delivery. Thus, from analysis and emanated findings it is notable that the

level of utilisation of smart technology and elements of fourth industrial revolution is high within the office operations whilst it is significantly low in construction site operations. Thus, this paper recommends that the adoption of smart technologies by construction industry organisations should be expedited in order to improve the overall performance project delivery within South Africa. Finally, construction sector professionals should adopt a proactive thinking approach to their daily construction activities.

References

- Aghimien, D., Aigbavboa, C., Matabane, K.: Impediments of the Fourth Industrial Revolution in the South African Construction Industry. In: *Collaboration and Integration in Construction, Engineering, Management and Technology*, pp. 223–227. Springer (2021)
- Araszkiewicz, K., Tryfon-Bojarska, A.: Modern information management throughout a construction project life cycle—selected issues concerning digitization in construction and a case study. *Czas. Tech.* **8**, 36–51 (2017)
- Benton, W., McHenry, L.F.: *Construction purchasing & supply chain management*. McGraw-Hill New York (2010)
- Cabibihan, J., et al.: A guided approach for utilizing concrete robotic 3D printing for the architecture, engineering, and construction industry. *Constr. Robot.* (2023). <https://doi.org/10.1007/s41693-023-00103-9>
- Chaleunvong, K.: Data collection techniques. Training Course in Reproductive Health Research Vientine (2009)
- Chung, M., Kim, J.: The internet information and technology research directions based on the Fourth Industrial Revolution. *KSII Trans. Internet Inf. Syst.* **10**(3), 2 (2016)
- Creswell, J.: *Research design: Qualitative, quantitative and mixed methods approach* (2nd ed.) (2003)
- Dallasega, P., Rauch, E., Linder, C.: Industry 4.0 as an enabler of proximity for construction supply chains: a systematic literature review. *Comput. Ind.* **99**, 205–225 (2018)
- Ergen, E., Akinci, B.: An overview of approaches for utilizing RFID in construction industry. In: *Proceedings of 2007 1st Annual RFID Eurasia*, pp. 1–5. IEEE (2007)
- Ergen, E., Akinci, B., Sacks, R.: Life-cycle data management of engineered-to-order components using radio frequency identification. *Adv. Eng. Inform.* **21**(4), 356–366 (2007)
- Frank, A.G., Dalenogare, L.S., Ayala, N.F.: Industry 4.0 technologies: Implementation patterns in manufacturing companies. *Int. J. Prod. Econ.* **210**, 15–26 (2019)
- Grau, D., Caldas, C.H., Haas, C.T., Goodrum, P.M., Gong, J.: Assessing the impact of materials tracking technologies on construction craft productivity. *Autom. Constr.* **18**(7), 903–911 (2009)
- Iacovidou, E., Purnell, P., Lim, M.K.: The use of smart technologies in enabling construction components reuse: a viable method or a problem creating solution? *J. Environ. Manage.* **216**, 214–223 (2018)
- Jang, W.-S., Skibniewski, M.J.: Embedded system for construction asset tracking combining radio and ultrasound signals. *J. Comput. Civ. Eng.* **23**(4), 221–229 (2009)
- Koh, L., Orzes, G., Jia, F.J.: The fourth industrial revolution (Industry 4.0): technologies disruption on operations and supply chain management. *Int. J. Oper. & Prod. Manag.* (2019)
- Li, R.Y.M.: Smart construction safety in road repairing works. *Procedia computer science*, pp. 301–307 (2017)
- Liu, D., Lu, W., Niu, Y.: Extended technology-acceptance model to make smart construction systems successful. *J. Constr. Eng. Manag.* **144**(6), 04018035 (2018)
- Liu, R., Chua, V.C.: Theoretical digitalization of information flow in the construction supply chain. *Int. J. Manag. Res. Bus. Strat.* **5**(1), 10–27 (2016)

- Ndabeni-Abraham, S.: Impact of Fourth Industrial Revolution (2019). https://www.google.com/search?rlz=1C1CHFX_enZA783ZA783&ei=hQl2XvfyIY6flwTx35iIDw&q=Nda
beni-Abrahams+on+4ir&oq=Ndabeni-Abrahams+on+4ir&gs_l=psy-ab.3..33i160.16660.
22191..22526...0.0..403.2733.3-7j1.....0....1..gws-wiz.....0i7i30j0i67j0j0i30j0i5i30j0i22
i30.Y0Gfw6RZ6lg&ved=0ahUKEwi3yajWyavoAhWOz4UKHfEvBvEQ4dUDCAs&uact=5
(Accessed 15 April 2020)
- Niu, Y., Lu, W., Chen, K., Huang, G.G., Anumba, C.: Smart construction objects. *J. Comput. Civ. Eng.* **30**(4), 04015070 (2016)
- Osunsanmi, T.O., Aigbavboa, C., Oke, A.: Construction 4.0: the future of the construction industry in South Africa (2018)
- Papadopoulos, G.A., Zamer, N., Gayialis, S.P., Tatsiopoulos, I.P.: Supply chain improvement in construction industry. *Univers. J. Manag.* **4**(10), 528–534 (2016)
- Pereira da Silva, N., Eloy, S., Resende, R.: Robotic Construction analysis: simulation with virtual reality. *HELIYON* (2022). <https://doi.org/10.1016/j.heliyon.2022.e11039>
- Pillay, K., Ori, A., Merkofer, P.: Industry 4.0: is Africa ready for digital transformation (2017)
- Sardroud, J.M.: Influence of RFID technology on automated management of construction materials and components. *Sci. Iran.* **19**(3), 381–392 (2012)
- Sisulu, L.: South Africa not ready for fourth industrial revolution. SABC Digital News, South Africa (22 Jan. 2019)
- Wang, S., Wan, J., Li, D., Zhang, C.: Implementing smart factory of industry 4.0: an outlook. *Int. J. Distrib. Sens. Netw.* **12**(1), 3159805 (2016)
- Wang, X., Yu, H., Wes McGee, W., Menassa, C.C., Kamat, V.R.: Enabling BIM-driven robotic construction workflows with closed-loop digital twins. eprint arXiv:2306 (2023)
- Yang, Z., Wang, Y., Sun, C.: Emerging information technology acceptance model for the development of smart construction system. *J. Civ. Eng. Manag.* **24**(6), 457–468 (2018)
- You, Z., Feng, L.: Integration of industry 4.0 related technologies in construction industry: a framework of cyber-physical system. *IEEE Access* **8**, 122908–122922 (2020)



Virtual Reality Experience and Motion Sickness in Construction Human-Robot Collaboration Learning

Adetayo Onososen^(✉), Innocent Musonda, and Thembani Moyo

Centre for Applied Research and Innovation in the Built Environment (CARINBE), Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

221046022@student.uj.ac.za

Abstract. VR offers an immersive and interactive environment that enhances learning experiences, particularly in complex domains like HRC. By simulating realistic scenarios, individuals can gain practical skills without real-world risks. Customizable training scenarios cater to learners' specific needs, and remote collaboration facilitates participation regardless of physical location, thereby enhancing learning outcomes, mitigating risks, and improving cost-effectiveness. However, the adoption of VR in sectors like architecture, engineering, and construction remains limited, partly due to challenges such as motion sickness. Symptoms like nausea, dizziness, headache, and eyestrain can arise from the sensory conflicts inherent in VR experiences. While technological advancements have mitigated some side effects, understanding how users with varying levels of VR experience respond differently to these technologies remains crucial. This study investigates motion sickness issues in HRC learning to enhance user experiences. Through a mixed-methods approach, combining qualitative and quantitative analyses, the research identifies and prioritizes symptoms experienced during VR interactions. Results reveal a predominance of male skilled operators with limited prior VR experience. General discomfort, dizziness, headache, difficulty focusing, and nausea emerge as top symptoms, highlighting challenges in VR usage for HRC learning. Efforts to design VR devices and experiences that minimize discomfort are essential to improve user acceptance and engagement. Strategies such as incorporating breaks and limiting exposure duration can alleviate symptoms and enhance overall user experience. By addressing motion sickness and associated discomforts, VR technology can realize its full potential in enhancing human-robot collaboration and learning in the construction industry and beyond.

Keywords: Virtual Reality · Training · Learning · Robotics · Human-Robot · Motion Sickness · Collaboration

1 Background

Virtual reality (VR) is a visualization technology with the potential to alter how humans engage with visual data profoundly [1, 2]. Up to this point, primary utilization has been within the gaming and entertainment industries, yet sectors such as tourism, marketing,

sports, education, and training have also seen significant expansion [3]. Also emerging and being driven is the application of VR for construction industry use cases, including Design Visualization, Training and Simulation, virtual site tours and learning, etc. [4, 5]. Numerous companies across diverse industries, such as aerospace, logistics, and retail, employ VR technologies to enhance education, training, and productivity. Similarly, a use case with an emerging interest in the construction industry uses virtual reality to undertake human-robot collaboration learning. Virtual reality (VR) offers a highly immersive and interactive environment that can significantly enhance the learning experience, particularly in complex domains such as human-robot collaboration. By simulating realistic scenarios within a controlled virtual environment, individuals can gain practical experience and skills in human-robot collaboration without the risks associated with real-world experimentation.

Moreover, VR allows for the creation of dynamic and customizable training scenarios tailored to learners' specific needs and skill levels. This flexibility enables learners to practice various tasks and interactions with robots, from basic cooperation to more advanced collaborative behaviours, all while receiving real-time feedback and guidance. Furthermore, VR facilitates remote collaboration and training, allowing individuals to participate in collaborative learning experiences regardless of physical location. This is especially valuable in scenarios where access to physical robots or specialized training facilities may be limited. Hence, VR is valuable for HRC Enhanced Learning Outcomes, Risk Mitigation, and Cost-Effectiveness in human-robot training and design [6, 7]. Technological advancements in robotics have enabled humans to collaborate with robots.

Nonetheless, individuals frequently experience a sense of insecurity when working alongside robots [8]. This collaboration allows for humans to delegate repetitive and mundane tasks to robots, which is especially advantageous as such tasks frequently lead to physical injuries known as repetitive motion injuries [9]. By utilizing robots, humans can allocate their attention to tasks less suitable for automation [10]. However, humans are reluctant to engage with or work alongside robots if they perceive it as unsafe, irrespective of the actual level of safety. This has further driven the need to adopt VR to enhance human-robot learning.

Although AR and VR technologies possess immense potential, similar to other digital technologies, their adoption within the architecture, engineering, and construction (AEC) sectors remains notably limited [1]. Studies such as Kashmiri et al. [11] have attributed these technologies' minimal uptake to many intricate and interconnected factors, such as motion sickness. "Simulator sickness" refers to the symptoms experienced by individuals when using virtual reality (VR) or simulator systems. These symptoms can include Nausea, Dizziness, Headache, Eyestrain, Sweating, Fatigue, Disorientation, Vertigo, Increased salivation, Loss of balance and general discomfort [12].

Motion sickness frequently arises in the VR context when there is a disparity between the anticipated and actual motion. On the other hand, VR sickness occurs when exposure to a virtual environment induces symptoms akin to those of motion sickness [13]. VR sickness differs from motion sickness in that it can be triggered by the visually-induced sensation of movement, even in the absence of actual physical motion. Sensory conflict theory posits that sickness arises when a user's perception of movement conflicts with

their prior expectations, stemming from incongruent sensory inputs from the visual system, vestibular system, and non-vestibular proprioceptors [14]. VR locomotion methods can be categorized based on interaction, movement, spatial aspects, and the devices used. This VR locomotion technique involves a synthetic interaction approach where users manipulate controllers to navigate within the VR environment. The interaction space in VR is unrestricted, allowing for continuous movement. This method is user-friendly and can be implemented through controllers equipped with a joystick, thumb stick, or trackpad. This approach serves as a solid foundation for locomotion testing since the controllers directly govern movement.

Nonetheless, this locomotion method is susceptible to causing motion sickness [13]. VR locomotion is categorized into motion-based, controller-based, and teleportation-based techniques according to HMD devices with inside-out tracking. Using VR applications can potentially cause discomfort for users, leading to side effects that may include symptoms not typically associated with sickness, such as fatigue, sweating, and difficulty focusing. However, advancements in VR hardware, including display, tracking, and processing technologies, have led to fewer side effects for users.

While studies investigating Virtual Reality (VR) have revealed numerous technical benefits of VR compared to traditional media, there remains a gap in understanding how individuals with diverse levels of prior experience in VR respond differently to the technology. Also, as stated by Chatta et al. [15], according to a US National Library of Medicine report, approximately one in three individuals has a predisposition to motion sickness, and the exact cause remains unidentified. Despite the substantial advantages and reduced costs offered by VR technology, many users cannot fully benefit from it due to the side effects of motion sickness. This study critically examines motion sickness issues in HRC learning to enhance the user experience. This study's objectives were to first pinpoint and prioritize the most significant symptoms experienced during VR learning and, secondly, to offer an understandable and concise elucidation of these symptoms, serving as a foundation for devising measures to alleviate them. A mixed-method approach was utilized to attain the objectives, incorporating qualitative and quantitative data collection and analysis techniques. Statistical methods were applied to rank the factors accordingly.

2 Research Method

The broader investigation utilizes a mixed research approach, integrating both qualitative data collection and analysis along with quantitative data collection and analysis. However, this paper focuses solely on the initial phase of quantitative analysis. This particular mixed research methodology has demonstrated its effectiveness as a robust tool for exploring intricate processes and systems in various domains, including engineering, mining, healthcare etc. [12]. Therefore, it was chosen for implementation in this research. Such mixed methods prove highly valuable in uncovering fundamental factors within intricate systems by complementing and directing quantitative data collection and analysis through qualitative research endeavors. The integration of both qualitative and quantitative analyses aids in the elucidation, classification, and generalization of findings.

The initial stage, employing a quantitative methodology, involved conducting an experiment where participants interacted with robots within a virtual environment. Prior to and following the experiment, participants answered quantitative questions. Log files from the virtual environment were also recorded to assess participants' performance. The robots designed for interaction with skilled operators included Robot Dumpers, Robot Forklifts, Robot Cranes, Robot Dump Trucks, and drones, showcasing various use cases within the construction sector. A qualitative approach involving focus group discussions with participants was also conducted. The study developed a framework for enhancing human-robot collaboration. However, this paper examines the motion sickness of the participants. This is presented and discussed in relation to their prior VR experience. The factors were identified and ranked. A reliability analysis was carried out on the questionnaire results to validate the internal consistency of the results. Using the results of the study, relevant insights into how motion sickness affects human-robot collaboration learning were drawn out. Forty-three skilled operators were involved in the study with the intention to assemble varied expertise in construction equipments/machinery.

3 Results & Discussion

3.1 Demographic Profile

By employing descriptive statistical analysis to examine the demographic data presented, the results based on convenience sampling revealed that all participants filled out the questionnaires after the experiment, signifying a 100% response rate. The distribution of the respondents revealed that 72% were male while 28% were female skilled operators. Derived from the participants' responses, a factor assessing expertise level was formulated to offer insight into the pertinent experience of the respondents. The study revealed that 53.5% had less than five years experience. 32.6% of the participants had 5–10 years, while 9.3% had 11–15 years experience. 4.7% of participants had more than 20 years of experience and were the most experienced in the group (Figs. 1 and 2).

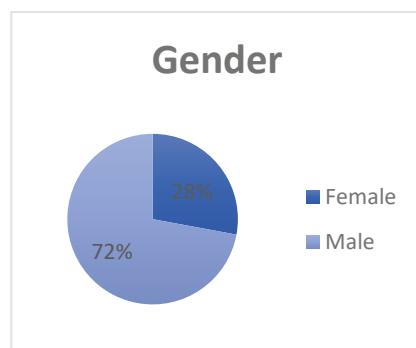


Fig. 1. Gender of Participants

As seen in Fig. 3, the trade categories of the operators show there were more Crane operators in the experiment, followed by dumper operators, drone operators, forklift

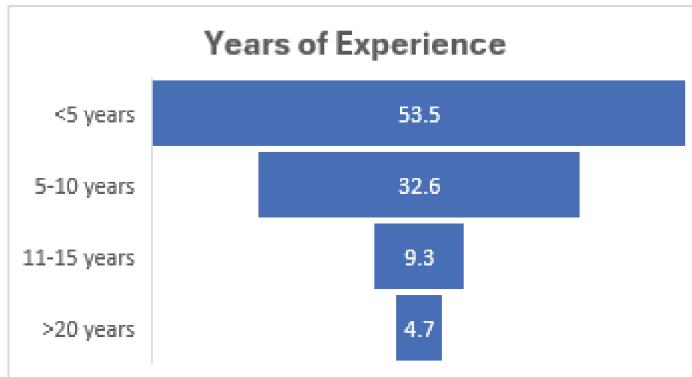


Fig. 2. Years of Experience

operators and dump truck operators. The age analysis reveals that most respondents were between 18 and 29, making up 65% of the population. Followed are those aged between 30 and 39 years, making up 30% of the population, while 5% were aged 50 years and above. This is presented in Fig. 4.

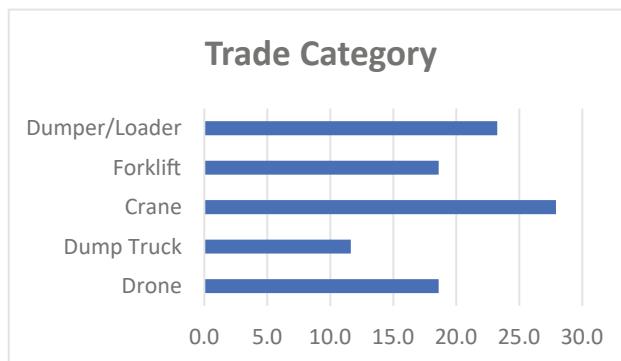
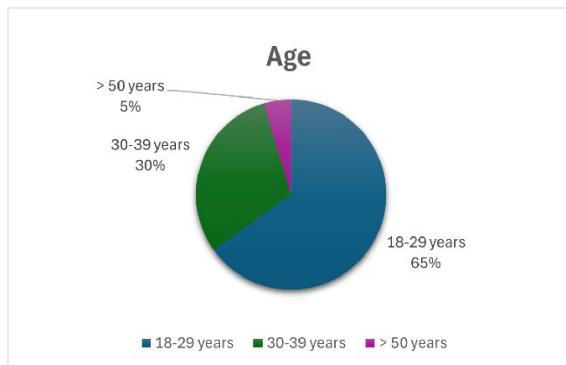
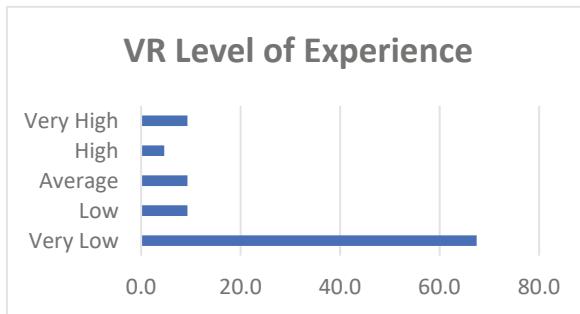


Fig. 3. Trade Category

3.2 Previous & Level of Experience with VR, Video Games

In understanding the previous experience the participants have had with virtual reality and how it affects their performance in the experiment, the study revealed, as shown in Fig. 5, that 66% of the respondents have had no VR experience. In comparison, 34% have had a VR experience. On their experience level, 67.4% had very low experience, 9.3% had low experience, 9.3% had average experience, 4.7% had high experience, and 9.3% had very high experience, as shown in Fig. 6.

In ascertaining the familiarity of the participants with video games and their level of experience, it was discovered, as shown in Fig. 7, that 83.7% of the respondents spent

**Fig. 4.** Age**Fig. 5.** Virtual Reality Experience**Fig. 6.** VR level of experience

0–9 h on video games, 4.7% spent 10–19 h, while 2.3% spent 20–29 and 30–39 h. Those who played more than 40 h were 7% of the population. The experience level with video games revealed that most participants had very little experience (44.2%), 20.9% had an average level of experience, 16.3% had very high experience, and 9.3% had high and low levels of experience. This is presented in Figs. 7 and 8.

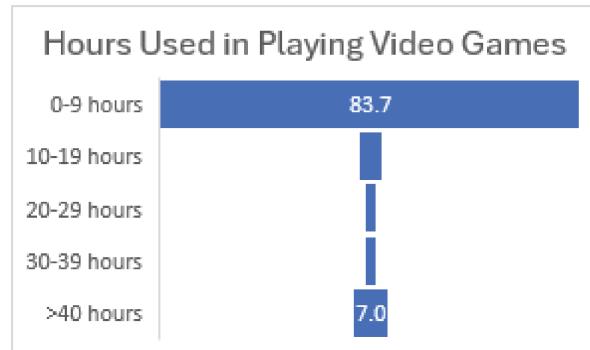


Fig. 7. Hours used in playing video games

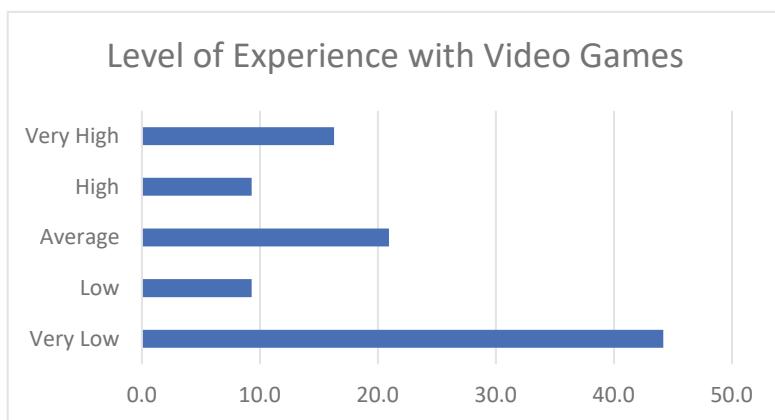


Fig. 8. Level of experience with Video Games

3.3 Identified Motion Sickness

Table 1 presents the identified motion sickness according to the mean value of the responses from the participants. The standard deviation (SD) is presented as well. The SD gives insight into the dissimilarity between the factors. In this instance, it denotes the extent of disparity in responses among the participants. A small standard deviation suggests a consensus among respondents regarding the significance of the factors, while a large standard deviation indicates disagreement. The mean reveals the balance of the respondents based on the 5-scale. Most of the means tend towards 1 and 2, indicating that most respondents agree they were only mildly affected by motion sickness.

3.4 Relative Importance Index

The Relative Importance Index (RII) is calculated for each item to pinpoint the most crucial variables for achieving desired results. Items are grouped based on their RII values, enabling the identification of the most significant symptoms through assigned

Table 1. Identified Motion sickness

| Symptoms | Mean | Std. Deviation |
|--------------------------|------|----------------|
| Sweating | 1.12 | 0.391 |
| Increased salivation | 1.19 | 0.546 |
| Burping | 1.19 | 0.546 |
| Stomach awareness | 1.21 | 0.600 |
| Vertigo | 1.35 | 0.752 |
| Nausea | 1.37 | 1.001 |
| Fatigue | 1.40 | 0.760 |
| Difficulty concentrating | 1.42 | 0.794 |
| Headache | 1.47 | 0.960 |
| Dizziness (eyes closed) | 1.47 | 0.667 |
| Fullness of head | 1.49 | 0.883 |
| Blurred vision | 1.51 | 0.798 |
| Dizziness (eyes open) | 1.74 | 1.071 |
| Difficulty focusing | 1.74 | 1.026 |
| General discomfort | 2.07 | 1.078 |
| Eyestrain | 2.12 | 0.905 |

ratings for each aspect. The RII formula is employed to ascertain each item's relative index, aiding in determining their relative importance. The RII method is employed for data analysis and ranking computation following the receipt of responses. Researchers utilize the RII technique to assign weights to each item in the questionnaire for the purpose of obtaining rankings. According to Table 2, general discomfort ranked 1 with the highest RII of 0.9778. This is followed by Dizziness (eyes open) (0.9711), headache (0.95988), difficulty focusing (0.92569), Nausea (0.90055), Fullness of head (0.88296), Eyestrain (0.80526), difficulty concentrating (0.79380), Fatigue (0.76031), Vertigo* (0.75226), Blurred vision (0.72798), dizziness (eyes closed) (0.72722), Stomach awareness (0.71993), Burping (0.71580), Increased salivation (0.70580), and sweating (0.70009).

3.5 Top 5 Symptoms of Motion Sickness

The top five symptoms as shown in Table 3 are General discomfort with an RII of 0.97781 and ranked no 1. 39.5% of the respondents on being asked if they experienced this symptom indicated not at all. 25.6% indicated mild and moderate when asked if they experienced it. Ranked 2nd is the feeling of dizziness amongst the participants during the experiment. 58.1% of the respondents indicated they were not experiencing this but 13% felt it was severe while only 1% considered it very severe. The symptom of headache is ranked 3rd with an RII of 0.95988. Only 4.7% of the respondents experience this while

Table 2. Relative Importance Index (RII)

| Motion sickness | RII | Rank |
|--------------------------|---------|------|
| General discomfort | 0.97781 | 1 |
| Dizziness (eyes open) | 0.97111 | 2 |
| Headache | 0.95988 | 3 |
| Difficulty focusing | 0.92569 | 4 |
| Nausea | 0.90055 | 5 |
| Fullness of head | 0.88296 | 6 |
| Eyestrain | 0.80526 | 7 |
| Difficulty concentrating | 0.79380 | 8 |
| Fatigue | 0.76031 | 9 |
| Vertigo* | 0.75226 | 10 |
| Blurred vision | 0.72798 | 11 |
| Dizziness (eyes closed) | 0.72722 | 12 |
| Stomach awareness** | 0.71993 | 13 |
| Burping | 0.71580 | 14 |
| Increased salivation | 0.70580 | 15 |
| Sweating | 0.70009 | 16 |

2.3% experienced difficulty focusing (0.92569, ranked 4th) and 7% had severe nausea (0.90055, ranked 5th).

Table 3. Top 5 Symptoms

| Motion sickness | Not at all | Mild | Moderate | Severe | Very Severe | RII | Rank |
|-----------------------|------------|------|----------|--------|-------------|---------|------|
| General discomfort | 39.5 | 25.6 | 25.6 | 7 | 2.3 | 0.97781 | 1 |
| Dizziness (eyes open) | 58.1 | 23.3 | 4.7 | 13 | 1 | 0.97111 | 2 |
| Headache | 74.4 | 14 | 4.7 | 4.7 | 2.3 | 0.95988 | 3 |
| Difficulty focusing | 58.1 | 16.3 | 20.9 | 2.3 | 2.3 | 0.92569 | 4 |
| Nausea | 86 | 2.3 | 2.3 | 7.0 | 2.3 | 0.90055 | 5 |

4 Discussion

The study utilized descriptive statistical analysis to examine demographic data gathered through convenience sampling. Results indicated a 100% response rate, with 72% male and 28% female skilled operators. The majority (53.5%) had less than five years of experience, while 4.7% had over 20 years. Crane operators were most common, followed by dumper, drone, forklift, and dump truck operators. In terms of age, 65% of respondents were 18–29, 30% were 30–39, and 5% were 50 and above. Regarding virtual reality (VR) experience, 66% had none, while 34% had some. Most had very little experience with VR (67.4%). In terms of video game familiarity, 83.7% spent 0–9 h playing, with varying levels of experience. In the initial stages of designing HMD devices, only a VR headset was available, requiring direct interaction with the virtual environment through gaze. Additional peripherals, like joysticks or Leap Motion for hand gestures, were necessary for controlling movements in VR. Consequently, various locomotion techniques were developed, each affecting users differently [13]. While these innovations have been important in improving the experience in VR towards learning, symptoms such as motion sickness still persists and other motion sickness heavily influenced by VR locomotion. This phenomenon occurs when motion images do not align with actual movements, leading to the need for a locomotion technique that minimizes side effects and health issues for users. Considering the substantial investments necessary for the implementation of human-robot collaboration learning through virtual reality, in terms of equipment, space, time, and skill enhancement, these investments can only be deemed reasonable when users can use these learning approaches without symptoms [17]. This is further necessitated by the limited numbers of people currently interested in using VR for learning and lack of implementation standards [12, 18]. Furthermore, symptoms such as General discomfort, Dizziness (eyes open), Headache, Difficulty focusing, Nausea, Fullness of the head, Eyestrain, Difficulty concentrating, Fatigue, Vertigo, Blurred vision, Dizziness (eyes closed), Stomach awareness, Burping, Increased salivation and Sweating are a major limitation to the adoption of VR. Because of the intricacy and duration of construction projects, extended use of these devices can cause discomfort for the majority of users. To mitigate side effects stemming from hardware, the focus lies on designing locomotion systems. Furthermore, the classification of VR sickness is complex due to various factors such as challenges in symptom measurement, rapid technological evolution, and sensitivity influenced by content.

Also, we should strive to enhance the design and advancement of VR devices to lessen discomfort and mitigate their adverse effects on users. Moreover, offering regular breaks during extended usage and restricting the duration of exposure to the virtual environment can aid in decreasing symptoms linked to motion sickness and other discomforts.

Hence, VR developers must comprehend these side effects to minimize or eradicate them for users. The theories regarding VR sickness are linked to motion sickness and simulator sickness.

5 Conclusion

The study delved into the crucial issue of motion sickness within the context of human-robot collaboration (HRC) learning using virtual reality (VR) technology. By employing a mixed-methods approach, combining quantitative analysis with qualitative insights, the research aimed to identify and prioritize symptoms experienced by participants during VR interactions. Through descriptive statistical analysis, the demographic profile of the participants was examined, revealing insights into their experience levels, age distribution, and prior exposure to VR and video games. Results indicated a predominance of male skilled operators, with varying levels of experience in the construction industry. Most participants had limited prior experience with VR, highlighting the need for tailored approaches to address potential discomfort and adverse effects associated with VR usage. The Relative Importance Index (RII) methodology was employed to rank symptoms of motion sickness experienced by participants during VR interactions. General discomfort, dizziness, headache, difficulty focusing, and nausea emerged as the top symptoms, shedding light on users' challenges when engaging with VR technology for HRC learning. The findings underscore the importance of understanding and mitigating motion sickness and associated discomforts to enhance the usability and effectiveness of VR technology in HRC learning. Despite advancements in VR hardware and locomotion techniques, symptoms such as motion sickness persist, posing barriers to widespread adoption and utilization. Efforts to design VR devices and experiences that minimize discomfort and adverse effects are essential to improve user acceptance and engagement. Strategies such as incorporating regular breaks during VR sessions and limiting exposure duration can help alleviate symptoms and enhance overall user experience.

References

1. Davila Delgado, J.M., Oyedele, L., Beach, T., Demian, P.: Augmented and virtual reality in construction: drivers and limitations for industry adoption. *J. Constr. Eng. Manag.* **146**(7), 04020079 (2020). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001844](https://doi.org/10.1061/(asce)co.1943-7862.0001844)
2. Sabat, M.: Cognitive load , fatigue and aversive simulator symptoms but not manipulated Zeitgebers affect duration perception in virtual reality. *Sci. Rep.* 1–8 (2022). <https://doi.org/10.1038/s41598-022-18520-1>
3. Froehlich, M.A., Azhar, S.: Investigating virtual reality headset applications in construction. In: 52nd ASC Annual International Conference Proceedings, pp. 1–8 (2016)
4. Wen, J., Gheisari, M.: Using virtual reality to facilitate communication in the AEC domain: a systematic review. *Construction Innovation*. Emerald Group Publishing Ltd., pp. 509–542 (2020). <https://doi.org/10.1108/CI-11-2019-0122>
5. Adami, P., Rodrigues, P.B., Woods, P.J., Becerik-Gerber, B., Soibelman, L., Copur-Gencturk, Y., Lucas, G.: Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation. *Adv. Eng. Informatics* **50**(August), (2021). <https://doi.org/10.1016/j.aei.2021.101431>
6. Yun, H., Jun, M.B.G.: Immersive and Interactive Cyber-Physical System (I2CPS) and virtual reality interface for human involved robotic manufacturing. *J. Manuf. Syst.* **2022**(62), 234–248 (Nov. 2021). <https://doi.org/10.1016/j.jmsy.2021.11.018>
7. Dallel, M., Havard, V., Dupuis, Y., Baudry, D.: Digital twin of an industrial workstation: a novel method of an auto-labeled data generator using virtual reality for human action

- recognition in the context of human-robot collaboration. *Eng. Appl. Artif. Intell.* **2023**(118), 105655 (Dec. 2022). <https://doi.org/10.1016/j.engappai.2022.105655>
- 8. Green, S.A., Billinghamurst, M., Chen, X., Chase, J.G.: Human-robot collaboration: a literature review and augmented reality approach in design. *Int. J. Adv. Robot. Syst.* **x**, 18 (2002)
 - 9. Sebo, S.S.: Developing robot teammates that enhance social dynamics and performance in human-robot teams. Yale University (2020)
 - 10. You, S., Kim, J.H., Lee, S.H., Kamat, V., Robert, L.P.: Enhancing perceived safety in human-robot collaborative construction using immersive virtual environments. *Autom. Constr.* **96**(September), 161–170 (2018). <https://doi.org/10.1016/j.autcon.2018.09.008>
 - 11. Kashmiri, D., Taherpour, F., Namian, M., Ghiasvand, E.: Construction Research Congress 2020 809. *Constr. Res. Congr.* **1994**, 809–818 (2020)
 - 12. Hafnia, M., Monacelli, E., Martin, H.: Virtual reality simulator for construction workers. *Assoc. Comput. Mach. No. VRIC '18*, April 4–6, 2018, LAVAL, France issue (2018)
 - 13. Khundam, C.: A study on usability and motion sickness of locomotion techniques for virtual reality, pp. 347–361. <https://doi.org/10.37936/ecti-cit.2021153.240834>
 - 14. Chattha, U.A., Shah, M.A.: Survey on causes of motion sickness in virtual reality. 2018 24th Int. Conf. Autom. Comput. (September), 1–5 (2018). <https://doi.org/10.23919/IConAC.2018.8749071>
 - 15. Chattha, U.A., Janjua, U.I., Anwar, F.: Motion sickness in virtual reality: an empirical evaluation 8 (2022). <https://doi.org/10.1109/ACCESS.2020.3007076>
 - 16. Gregory, K.: Intermediate IBM SPSS basic statistical techniques, pp. 1–14 (2013)
 - 17. Maqsoom, A., Zulqarnain, M., Irfan, M., Ullah, F., Alqahtani, F.K., Khan, K.I.A.: Drivers of, and barriers to, the adoption of mixed reality in the construction industry of developing countries. *Buildings* **13**(4), 1–17 (2023). <https://doi.org/10.3390/buildings13040872>
 - 18. Potseluyko, L., Pour, F., Dawood, N., Elghaish, F.: Game-like interactive environment using BIM-based virtual reality for the timber frame self-build housing sector. *Autom. Constr.* **142**(July), 104496 (2022). <https://doi.org/10.1016/j.autcon.2022.104496>

Technologies and Smart Data Management



Integrating Privacy-Preserving Occupancy Estimation Using Thermal Camera with a Digital Twin Platform

Benjamin Rafatian^(✉), Pezhman Sharafdin, and Ali Motamed

Quebec University, Montreal, QC, Canada

benjamin.rafatian.2@ens.etsmtl.ca

Abstract. Optimizing building energy consumption based on occupancy information is a crucial approach for improving energy management in buildings. Various studies have proposed different solutions for estimating occupant count. However, high costs and privacy concerns are still significant barriers that limit the practice of occupant count estimation. To address these issues, this paper proposes an integrated solution that combines privacy-preserving occupancy estimation using a low-cost and low-resolution thermal camera within a digital twin platform. The system captures images of occupants using a thermal camera installed above an entrance and uses image processing techniques to estimate the number of occupants passing the doorway. The occupancy data is then integrated into a digital twin platform to provide real-time occupancy information. The paper presents an overview of various methods of occupancy estimation and discusses their benefits and challenges. Additionally, it outlines the proposed system's architecture and presents a use case to demonstrate the capability of the solution. The proposed solution has numerous applications, including building automation, security, and energy management. The findings of this study can assist building owners and operators in optimizing their buildings energy consumption while preserving occupants' privacy.

Keywords: Occupancy Estimation · Digital Twin · Thermal Camera

1 Introduction

Buildings consume a significant portion of the world's energy, and their energy usage is a major contributor to greenhouse gas emissions [1]. Therefore, reducing energy consumption in buildings is crucial to move towards global energy sustainability goals. Additionally, it has gained increasing attention due to its environmental and economic benefits. An efficient strategy for reducing energy usage in buildings involves estimating occupancy levels in each space and accordingly adjusting heating, ventilation, and air-conditioning (HVAC) systems [2].

Occupancy estimation is the process of determining the number of people present in a building or a room. There are several methods for occupancy estimation, which use various types of sensors and cameras. Occupancy sensors can detect human presence by

measuring physical parameters, such as movement, temperature, and sound. Cameras can also be used to track people's movements and estimate occupancy levels. In addition to the use of sensors, a study by Chidurala and Li [3], shows that, machine learning algorithms can be trained to estimate occupancy based on historical data on the number of occupants. These algorithms can consider various factors, such as the time of day, day of the week, and season, to estimate occupancy levels [4].

Furthermore, the concept of the Digital Twin (DT) has gained significant attention in recent years as a powerful tool for optimizing the design, operation, and maintenance of complex systems, including buildings. A DT is a virtual replica of a physical asset, process, or system that is connected to its real-world counterpart through sensors and data streams [5]. The DT allows for real-time monitoring, simulation, and optimization of the physical asset or system, enabling better decision-making and improved performance [6]. Consequently, a DT can be effectively utilized for energy optimization, occupant monitoring, amongst other purposes. Therefore, this paper presents a solution that integrates the data of a developed occupancy estimation system using low-cost thermal camera within a DT platform. The camera detects human presence and movement using low-resolution thermal images; hence, the developed solution ensures that the privacy of the building's occupants is protected by not capturing any information that can be used to individually identify the occupants. The proposed solution aims to provide real-time occupancy data to the DT platform, which can then be used to optimize the energy consumption and occupants' comfort in a building. Additionally, it can provide insights on occupancy patterns for managing shared resources and addressing emergency scenarios based on real-time occupancy data.

2 Background Work

In this research, a literature review in the field of occupancy detection using thermal cameras was conducted across various scientific databases (e.g., IEEE Xplore, Scopus, Engineering Village, and Google Scholar). Total of 368 papers were found in IEEE Xplore database using a query that combined keywords related to thermal cameras and occupancy detection (i.e., (“Thermal Camera” OR “Thermal Image*”) AND (“Occupancy detect*” OR “Occupant monitoring” OR “human detect*” OR “human monitoring”)). Similarly, 110 papers were found on Scopus, while 130 papers were found in Engineering Village. Another 46 papers were added from Google Scholar by including the keyword “MLX90640” (type of the IR array sensor that is being used).

The research papers found from above mentioned databases were filtered by date and those published after 2008 were selected for further analysis. These papers were downloaded into the Zotero app and the duplicates were removed. In the first step, irrelevant papers were removed based on their titles, that narrowed down the total to 118. Finally, the abstracts of the remaining 118 papers were read, and the most relevant papers were categorized into three main categories: Review papers (6), line-based counting methods (8), and scene-based counting methods (18) (explained in Sects. 2.3, and 2.4).

2.1 Applications of Occupancy Estimation

Understanding the number of occupants in a building is important for a wide range of scenarios, including optimizing energy usage, improving building security, emergency response, and managing shared resources.

For energy management, knowing the number of occupants in each section of a building can be used to optimize the HVAC systems to reduce energy consumption. HVAC systems can account for up to 50% of a building's energy usage [1], hence optimizing them based on occupancy levels can lead to significant energy savings [7]. For example, a study by Nikdel et al. [8], shows a significant reduction of 22–50% and 47–87% in electricity and natural gas usage, respectively, by applying occupancy-based strategies to control HVAC systems. Another study done by de Bakker et al. [9], shows that there could be up to 60% savings in electrical energy consumed in office buildings for lighting through an occupant presence lighting control system.

In commercial buildings and offices, understanding the number of occupants in each section of the building is critical for emergency situations and security purposes. In the event of an emergency, having an accurate count of occupants in each area of the building is crucial.

In offices and shared spaces, the number of occupants can be used for managing resource allocation. For example, occupancy sensors can be used to optimize cleaning schedules based on actual usage of a space, rather than cleaning on a fixed schedule. This can reduce cleaning costs and increase efficiency. In a study done by Doherty et al. [10], toilet usage is monitored by a privacy preserved system to optimize cleaning schedule. In addition, occupancy data can be used to optimize the use of other shared resources, such as conference rooms and kitchen facilities.

2.2 Methods of Occupancy Estimation

There are various methods for occupancy estimation in buildings, and each has its own benefits and drawbacks. The common sensors used for occupancy estimation are Passive Infrared (PIR) sensors, Wi-Fi indoor positioning, RFID readers, environmental sensors such as CO₂, humidity, temperature, and air pressure, and cameras. However, the privacy of individuals may be compromised when using some of these sensors.

PIR sensors can detect the presence or absence of humans in a room or office through sensing the changes in infrared radiation caused by the motion of humans [11]. It also can be used as a means for counting the number of present occupants when used in pairs [12]. However, as these sensors are only able to detect a heat source and no other characteristics of a human, they cannot distinguish between people and pets or other moving sources with a heat signature.

Wi-Fi indoor positioning relies on wireless access points to determine the location of individuals by analyzing the signal strength and signal propagation delay [13]. Wi-Fi positioning systems can estimate the location and the number of occupants in a defined zone. One of the major drawbacks of using Wi-Fi indoor positioning is its dependency on the availability and the setup of wireless access points and receivers. Most of the methods require a sufficient number of access points to accurately estimate the location of occupants, which can be costly and time-consuming to install and calibrate. Additionally,

the accuracy of Wi-Fi positioning systems is affected by numerous factors such as interference from other wireless devices, signal blockages from walls and obstacles, and changes in the environment that can affect signal strength and propagation delay [14]. As a result, the accuracy of Wi-Fi indoor positioning for occupant counting may not be consistent, and it may not be suitable for applications that require high precision, such as security or emergency response systems. Finally, this method may raise privacy concerns as it relies on monitoring the signals emitted from individuals' devices.

RFID readers can also identify the location of individuals using radio frequency identification tags. The readers are short range, and the users should carry their tag at the entrance/exit point. Although it is a relatively accurate method, it fails to count visitors who are not carrying an RFID tag. There are a number of research projects that focused on only using RFID readers for occupancy estimation. Wang et al. [15] proposed a solution for occupancy monitoring which combines four complimentary data sources, i.e., relative humidity sensor, passive infrared sensor, camera, and carbon dioxide concentration sensor together with the RFID reader.

CO_2 sensors can estimate the number of occupants in a room by measuring the concentration of carbon dioxide. Primarily, they can detect the presence of humans if the amount of CO_2 reaches a certain threshold [16]. Additionally, by applying neural networks and using Bayesian methods on the derived data from the sensors, these methods can provide an estimate of the number of occupants in a room [17]. In addition to CO_2 , other environmental variables such as room temperature, and humidity are valuable sources of data that can be used for estimating the number of people occupying a certain area [2]. Although these environmental sensors are easy to install and cost-effective, they may exhibit lower accuracy compared to alternative methods. Notably, they face challenges in distinguishing between human and non-human sources influencing room temperature, humidity, and CO_2 concentrations. Moreover, these sensors have limitations in identifying occupants in scenarios where HVAC systems are in operation, or in rooms with open windows.

Another group of methods for occupancy estimation is vision-based approaches. In general, these methods use a form of image data, e.g., RGB or thermal image, and a series of image processing and Computer Vision (CV) algorithms to extract meaningful data for estimating occupancy in an area. RGB cameras provide high quality images, that when combined with a robust CV method can provide accurate results in counting people. However, as the faces of individuals may be identifiable, these methods may raise privacy concerns. On the other hand, thermal cameras use a similar process of data capturing and analysis. However, as they only capture the heat signature of the subjects, they can be considered as privacy preserving approaches.

The main goal of this research is to develop a privacy-preserved solution for occupancy estimation. While all the above methods can provide occupancy estimation, thermal camera-based methods are most suited, as they are relatively accurate and privacy preserving.

2.3 Vision-Based Occupancy Estimation

As illustrated in Fig. 1, estimating occupancy using cameras involves four essential steps before counting. The initial step is data acquisition, where various camera types, such as

RGB, thermal, depth sensors, or Lidar, are employed based on the specific use case. In the subsequent stage, raw input images from cameras undergo preprocessing to prepare them for computer vision algorithms. The preprocessing algorithms may vary depending on the input data type and specific methods used in the following steps.

In the third step, occupants within the scene are detected, and finally, in the last step, the identified human subjects are tracked. Tracking these individuals enables the calculation of the total number of occupants in real-time.



Fig. 1. Main steps of vision-based occupancy estimation systems.

Every camera has a limited Field of View (FoV), a crucial consideration when counting the number of people in each space. This factor is pivotal in choosing between Scene-based Counting Methods (SCM) and Line-based Counting Methods (LCM) [18, 19], two subcategories of vision-based methods for occupancy counting.

When the camera's FoV covers the entire area where occupancy estimation is required, scene-based methodologies are applied. In this approach, the camera captures the entire room, and individuals are initially detected by CV algorithms and then tracked to determine the overall number of occupants. Alternatively, when a single camera's FoV cannot cover the entire area, two solutions may be implemented. Multiple cameras with SCM can cover the entire scene, or a single camera with LCM can be used. For LCMs, a camera is positioned at a doorway or a similar passage area through which all occupants must pass. By capturing and tracking individuals as they pass, the trajectory of movement can be detected, allowing for the estimation of the number of occupants.

As indicated in Table 1, Maaspuro [20] implemented an LCM utilizing a Panasonic GridEye thermal camera for data collection. The GridEye sensor's native 8×8 pixel resolution is enhanced to 71×71 pixels through interpolation to facilitate subsequent processing stages. Background subtraction, a technique shared by the studies of Cokbas et al. [21], Garaza et al. [23], and Sun et al. [18], is employed in the pre-processing phase to differentiate dynamic elements (foreground) from static ones (background) by comparing consecutive frames. This method is particularly effective when the camera is stationary. Perra et al. [24] also adopted the Panasonic GridEye sensor in their LCM system, confirming its recurrent application in such studies. Both Cokbas et al. [21], and Zhu [19] utilized the Melexis MLX90640 IR array sensor, which is used in our own case study for the data acquisition layer.

During the occupant detection phase, Blob Detection is the predominant technique employed by most studies. This process involves isolating moving objects within the foreground pixels and encapsulating them within defined boundaries. The tracking phase, which presents significant challenges, has seen various approaches. While some studies have applied image processing techniques to track the centroid of the detected person-shaped blob, others such as Sun et al. [18] have harnessed Deep Learning (DL) algorithms, specifically DeepSORT, for real-time object tracking [25].

For the tracking of human movement, Maaspuro [20] utilized Kalman filtering, whereas Cokbas et al. [21] leveraged Markov Random Field (MRF) methods to enhance

Table 1. Vision-based occupancy estimation papers.

| Data acquisition | Sensor name | Pre-processing | Occupant detection | Occupant tracking | References |
|-----------------------|-----------------------------|--------------------------------|------------------------------|------------------------------|------------|
| Thermal Camera | GridEye 8 × 8 | Interpolation + BG subtraction | Blob Detection | Kalman Filtering | [20] |
| Thermal Camera | MLX90640 | RGA BG Subtraction | Blob Detection | Euclidean Distance + MRF | [21] |
| Fusion RGB Camera | N/A | BG Subtraction | YOLOX | DeepSORT | [18] |
| Fusion Thermal Camera | Lepton + GridEye + MLX90621 | Adaptive Thresholding | Blob Detection | Image Processing | [22] |
| Thermal Camera | D6T 44L | BG Subtraction | Blob Detection | Image Processing | [23] |
| Thermal Camera | GridEye 8 × 8 | N/A | Movement Pattern Recognition | Movement Pattern Recognition | [24] |
| Thermal Camera | MLX90640 | Data Augmentation and Cleaning | SORT | SORT | [19] |

tracking accuracy. Moreover, both Perra et al. [24], and Zhu [19] innovatively combined the two stages of occupant detection and tracking. Perra et al. [24] achieved this by integrating movement pattern recognition, while Zhu adopted the Simple Online Real-time Tracking (SORT) algorithm for both detection and tracking.

3 Proposed Method

3.1 System Architecture

In our literature review, we identified diverse approaches and methods employed by researchers for each of the four key steps of occupant counting systems: data acquisition, pre-processing, occupant detection, and occupant tracking. Figure 2 encompasses the identified methods in the literature. Subsequently, for each step in our proposed solution, we selected and integrated the methods highlighted in green.

The proposed LCM, in the broader context, builds upon the method introduced by Cokbas et al. [21] with modifications made to certain components. Our approach leverages thermal cameras in the data acquisition layer, facilitating the capture of thermal images for subsequent analysis.

For the preprocessing step, binary thresholding is chosen, owing to its compatibility with thermal images [26]. Since thermal images are single-channel representations where pixel values correspond to temperature, the normalization of values followed by the

application of binary thresholding yields an image mask. The resulting mask highlights pixels with higher temperatures, which correspond to moving humans that we aim to detect.

For occupant detection, the technique of blob detection is adopted. After applying binary thresholding on the raw image, clear white blobs become discernible in the image mask. This method effectively identifies potential human occupants within the scene. It is noteworthy that alternative approaches, such as employing deep learning object detection models such as the YOLO family [27] and ResNet [28], are viable alternatives for this stage, providing further avenues for exploration.

In the tracking stage, image processing techniques are used to compute the center point of the detected objects and track their movements as they pass through the doorway line. In addition to this method, calculation of Euclidean distance between objects, using Kalman filtering, or the deployment of deep learning tracking solutions such as Deep SORT [25] have been considered by others for tracking the occupant. Finally, if the tracked person passes the doorway line, the system will update the number of occupants for the corresponding area.

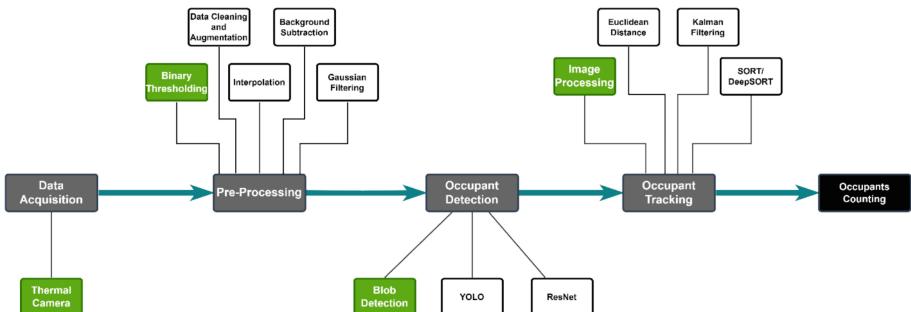


Fig. 2. Occupancy counting systems steps and methods.

4 Implementation

The developed solution is installed in GRIDD lab at École de technologie supérieure (ÉTS), which is a space with an approximate area of 151 m². The 3D view of the BIM model is shown in Fig. 3. This space has a single entrance/exit from the hallway, therefore the calculation of the total number of occupants inside the area is possible via LCM. As shown in Fig. 4, the camera is mounted to the wall on top of the doorway at the height of 230 cm facing downwards.

4.1 Hardware Setup

The hardware setup for the occupant estimation system has two main components: a Raspberry Pi 4b with 8GB RAM serving as the central hub and a Melexis MLX90640 IR array sensor. The selected sensor has a wide FoV of 110° × 75° and captures a

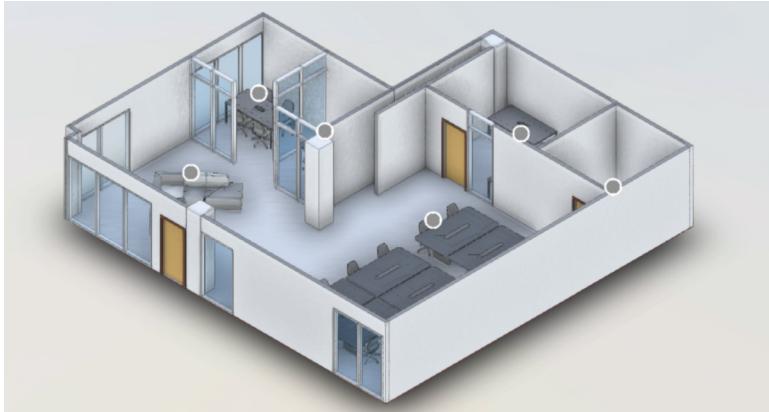


Fig. 3. BIM model of GRIDD Lab.

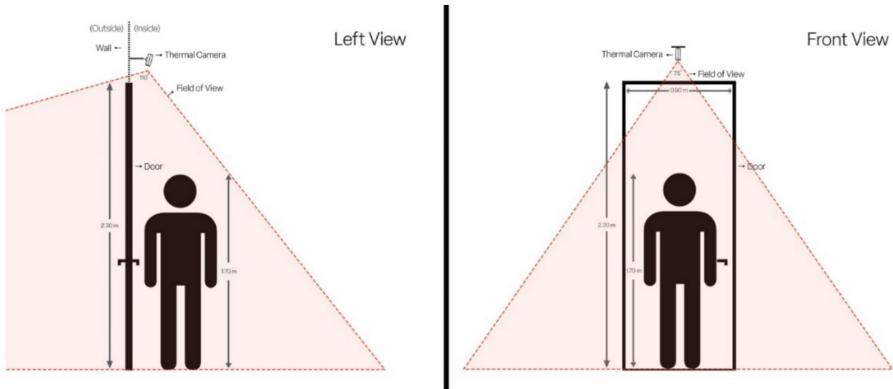


Fig. 4. Schematic of the developed hardware setup at GRIDD Lab.

thermal image of 32×24 pixels. It has a temperature range from -40°C to 300°C and can operate in a temperature range of -40°C to 85°C . The sensor has a programmable refresh rate that can range from 0.5Hz to 64Hz. Previous studies (e.g., [20, 21]) shown that the average speed of human movement is between 1.3 m/s to 3 m/s under normal conditions, which suggests that a refresh rate of around 10 FPS is adequate. Therefore, to further cover the possible exceptions, the refresh rate is set to 16 FPS.

4.2 Software Setup

The DT platform used in this study was developed using a web application that utilized a backend created with Node.js and working with Autodesk Forge APIs, along with a frontend that was developed with the React framework. Subsequently, the web application was deployed on Azure cloud services, making it available to the users via the internet. To ensure that the BIM model of the building remains up to date, the web

application is integrated into BIM 360 using Autodesk BIM 360 API. This allows all stakeholders to work on the same model in a collaborative environment while ensuring that all changes are applied to the central model. To visualize the BIM model of the GRIDD lab uploaded into BIM360, the Autodesk Forge Viewer was used.

For the occupant estimation system, Python 3 programming language with several key libraries is used for providing various functions. OpenCV is used to manage and process the input video stream from the IR array sensor, while NumPy is used for matrix functions. The Raspberry Pi and IR array sensor are connected using busio and adafruit_mlx90640 libraries, respectively. The calculated results are then sent to Azure IoT Hub. The stream of data is subsequently routed from IoT Hub to the Event Hub endpoint and passed to a Power BI database in real-time via a Stream Analytics module. Finally, a real-time dashboard is created using Power BI and added to the DT platform. Figure 5 illustrates the cloud architecture for integrating the occupant estimation system with the DT platform.

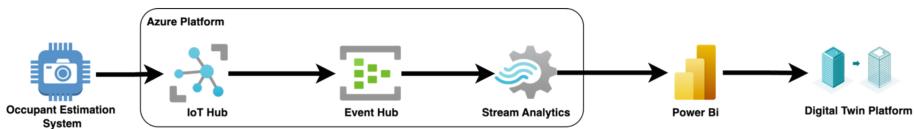


Fig. 5. Cloud architecture.

5 Results

Figure 6 illustrates both raw and processed images captured by the thermal camera during an exit event. The person enters the frame from left side, which is inside the lab area, opens the door and then exits from the right side of the frame. A binary threshold is applied to the raw image, and the occupant is detected by a contour, then it is enclosed within a bounding box. The center point of the contour is marked with a point. An entry event is recognized and counted when the center point of the detected occupant crosses the central line.

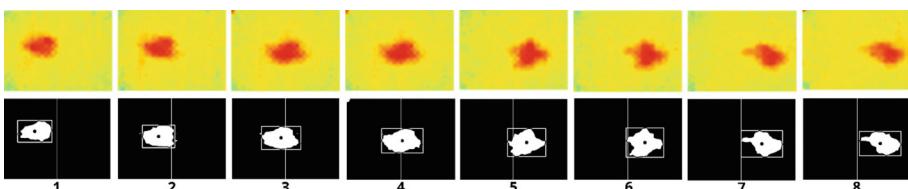


Fig. 6. Example of an exit event.

As discussed in 4.2, the output of the occupant's estimation system, in the form of a data stream, is integrated into the Power BI platform for visualization. Figure 7 shows

the dashboard of the implemented system, which is incorporated into the DT platform. The dashboard employs a bar chart to display the historical data of entry and exit events. The top left corner displays the timestamp of the last entry/exit event, along with the current count of occupants inside the room.

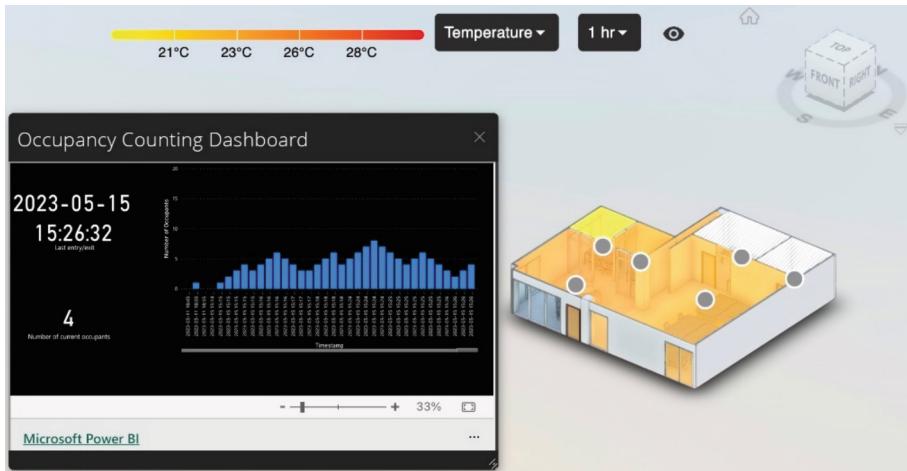


Fig. 7. Result: The occupancy counting dashboard integrated into the DT.

6 Discussion

This study investigated an approach for occupancy monitoring by integrating the occupancy data calculated using a low-resolution thermal camera within a DT platform. The case study underwent evaluation through continuous monitoring of the data flow originating from the developed occupant counting system to the DT platform, with real-time visualization on the Power BI dashboard. The focus of this study was to propose a proof of concept; however, the accuracy of the counting system needs to be evaluated as future work. Although the developed system succeeded in proving the possibility of integration of an occupancy estimation system within a DT platform while maintaining individuals' privacy, it highlights certain limitations. For example, the thermal camera may mistakenly count non-human heat sources, such as pets, leading to inaccurate occupancy estimates. Thus, further research is required to differentiate between heat sources more accurately.

This study integrated occupancy estimation data into DT technology implementation, demonstrating the potential for smart building management, particularly concerning energy efficiency and occupant comfort. It serves as a step towards incorporating various other metrics within the DT framework for enhanced building management.

For future work, the study suggests incorporating advanced detection algorithms, such as deep learning object detection models (e.g., the YOLO family and ResNet [28]). These models have the potential to improve the accuracy of distinguishing various heat

sources, addressing the limitations of the current solution and paving the way for more sophisticated applications in smart building systems.

7 Conclusions

The study provided an overview of diverse methods and recent advancements in the field of occupancy estimation, with a particular emphasis on vision-based approaches. It illustrated the applications of occupancy estimation and explored the merits and challenges associated with each. Furthermore, it delved into the architecture and implementation details of the proposed system.

In response to the identified challenges and with an eye on optimizing energy consumption in buildings, this paper introduced a solution that integrates a privacy-preserving occupancy estimation system within a DT platform, using a low-cost thermal camera. By using images captured by the thermal camera installed above a door, the system can estimate the number of occupants passing through the doorway. Notably, the proposed system not only offers a more cost-effective alternative but also addresses privacy concerns inherent in traditional occupancy estimation methods. This research primarily focused on presenting a proof of concept for the proposed system, paving the way for future developments in the realm of efficient and privacy-conscious occupancy estimation systems.

References

1. Pérez-Lombard, L., Ortiz, J., Pout, C.: A review on buildings energy consumption information. *Energy Build.* **40**(3), 394–398 (Jan. 2008). <https://doi.org/10.1016/j.enbuild.2007.03.007>
2. Chen, Z., Zhu, Q., Masood, M.K., Soh, Y.C.: Environmental sensors-based occupancy estimation in buildings via IHMM-MLR. *IEEE Trans. Industr. Inf.* **13**(5), 2184–2193 (Oct. 2017). <https://doi.org/10.1109/TII.2017.2668444>
3. Chidurala, V., Li, X.: Occupancy estimation using thermal imaging sensors and machine learning algorithms. *IEEE Sens. J.* **21**(6), 8627–8638 (Mar. 2021). <https://doi.org/10.1109/JSEN.2021.3049311>
4. Leech, C., Raykov, Y.P., Ozer, E., Merrett, G.V.: Real-Time Room Occupancy Estimation with Bayesian Machine Learning Using a Single PIR Sensor And Microcontroller. In 2017 IEEE Sensors Applications Symposium (SAS), pp. 1–6 (Mar. 2017). <https://doi.org/10.1109/SAS.2017.7894091>
5. Davari, S., Shahinmoghadam, M., Motamed, A., Poirier, E.: Demystifying the Definition of Digital Twin for Built Environment (2022)
6. Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B.: Characterising the digital twin: a systematic literature review. *CIRP J. Manuf. Sci. Technol.* **29**, 36–52 (May 2020). <https://doi.org/10.1016/j.cirpj.2020.02.002>
7. Erickson, V.L., et al.: Energy efficient building environment control strategies using real-time occupancy measurements. In: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, in BuildSys '09. New York, NY, USA: Association for Computing Machinery, pp. 19–24 (Nov. 2009). <https://doi.org/10.1145/1810279.1810284>

8. Nikdel, L., Janoyan, K., Bird, S.D., Powers, S.E.: Multiple perspectives of the value of occupancy-based HVAC control systems. *Build. Environ.* **129**, 15–25 (Feb. 2018). <https://doi.org/10.1016/j.buildenv.2017.11.039>
9. de Bakker, C., Aries, M., Kort, H., Rosemann, A.: Occupancy-based lighting control in open-plan office spaces: a state-of-the-art review. *Build. Environ.* **112**, 308–321 (Feb. 2017). <https://doi.org/10.1016/j.buildenv.2016.11.042>
10. Doherty, B., Gardner, N., Ray, A., Higgs, B., Varshney, I.: The shakedown: developing an indoor-localization system for quantifying toilet usage in offices. *Archit. Sci. Rev.* **63**(3–4), 325–338 (Jul. 2020). <https://doi.org/10.1080/00038628.2020.1748869>
11. Andrews, J., Kowsika, M., Vakil, A., Li, J.: A Motion Induced Passive Infrared (PIR) Sensor for Stationary Human Occupancy Detection. In: 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS), pp. 1295–1304 (Apr. 2020). <https://doi.org/10.1109/PLANS46316.2020.9109909>
12. Wahl, F., Milenkovic, M., Amft, O.: A Distributed PIR-based Approach for Estimating People Count in Office Environments. In: 2012 IEEE 15th International Conference on Computational Science and Engineering, pp. 640–647 (Dec. 2012). <https://doi.org/10.1109/ICCSE.2012.92>
13. Wang, Y., Shao, L.: Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning. *Build. Environ.* **114**, 106–117 (Mar. 2017). <https://doi.org/10.1016/j.buildenv.2016.12.015>
14. Xie, T., Jiang, H., Zhao, X., Zhang, C.: A Wi-Fi-based wireless indoor position sensing system with multipath interference mitigation. *Sensors* **19**(18), Art. no. 18 (Jan. 2019). <https://doi.org/10.3390/s19183983>
15. Wang, M., Wang, X., Zhang, G., Li, C.: Occupancy detection based on spiking neural networks for green building automation systems. In: Proceeding of the 11th World Congress on Intelligent Control and Automation, pp. 2681–2686 (Jun. 2014). <https://doi.org/10.1109/WCICA.2014.7053149>
16. Ansanay-Alex, G.: Estimating Occupancy Using Indoor Carbon Dioxide Concentrations Only in an Office Building: A Method and Qualitative Assessment (2013)
17. Rahman, H., Han, H.: Occupancy estimation based on indoor CO₂ concentration: comparison of neural network and Bayesian methods. *Int. J. Air-Cond. Ref.* **25**(03), 1750021 (Sep. 2017). <https://doi.org/10.1142/S2010132517500213>
18. Sun, K., Liu, P., Xing, T., Zhao, Q., Wang, X.: A fusion framework for vision-based indoor occupancy estimation. *Build. Environ.* **225**, 109631 (Nov. 2022). <https://doi.org/10.1016/j.buildenv.2022.109631>
19. Zhu, S.: Privacy-preserving Building Occupancy Estimation via Low-Resolution Infrared Thermal Cameras (2021)
20. Maaspuro, M.: A low-resolution IR-array as a doorway occupancy counter in a smart building. *Int. J. Onl. Eng.* **16**(06), Art. no. 06 (May 2020). <https://doi.org/10.3991/ijoe.v16i06.13915>
21. Cokbas, M., Ishwar, P., Konrad, J.: Low-Resolution Overhead Thermal Tripwire for Occupancy Estimation. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 398–406. IEEE, Seattle, WA, USA (Jun. 2020). <https://doi.org/10.1109/CVPRW50498.2020.00052>
22. Mikkilineni, A.K., Dong, J., Kuruganti, T., Fugate, D.: A novel occupancy detection solution using low-power IR-FPA based wireless occupancy sensor. *Energy Build.* **192**, 63–74 (Jun. 2019). <https://doi.org/10.1016/j.enbuild.2019.03.022>
23. Garaza, C.R.A., Pedrasa, J.R.: Development of a doorway occupancy counter based on thermal array sensing technology. In: 2016 IEEE Region 10 Conference (TENCON), pp. 3506–3510 (Nov. 2016). <https://doi.org/10.1109/TENCON.2016.7848708>

24. Perra, C., Kumar, A., Losito, M., Pirino, P., Moradpour, M., Gatto, G.: Monitoring indoor people presence in buildings using low-cost infrared sensor array in doorways. *Sensors* **21**, 4062 (Jun. 2021). <https://doi.org/10.3390/s21124062>
25. Wojke, N., Bewley, A., Paulus, D.: Simple Online and Realtime Tracking with a Deep Association Metric. *arXiv*, Mar. 21, 2017. <http://arxiv.org/abs/1703.07402> (Accessed: Mar. 15, 2023)
26. Roy, P., Dutta, S., Dey, N., Dey, G., Chakraborty, S., Ray, R.: Adaptive thresholding: a comparative study. In: 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 1182–1186 (Jul. 2014). <https://doi.org/10.1109/ICCICCT.2014.6993140>
27. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. *arXiv*, May 09, 2016. <https://doi.org/10.48550/arXiv.1506.02640>
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. *arXiv*, Dec. 10, 2015. <http://arxiv.org/abs/1512.03385> (Accessed: Dec. 04, 2023)



Automated Registration of Ground 3D Point Cloud Data for Individual Buildings

Huaiyuan Weng¹(✉), Chul Min Yeum¹, Derek T. Robinson², and Bruce Macvicar¹

¹ Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada

h22weng@uwaterloo.ca

² Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada

Abstract. As urban development recently, the need for detailed 3D city models grows exponentially. Central to these models are the 3D depictions of buildings, indispensable for hazard assessments such as flood and earthquake risks, and post-disaster evaluations. Central to these models are the 3D depictions of buildings, indispensable for hazard assessments such as flood and earthquake risks, and post-disaster evaluations. These models are primarily derived from 3D point cloud data, sourced from LiDAR scans or photogrammetry. While aerial mapping is a prevalent method to collect this point cloud data, the task of generating thorough point cloud maps for buildings becomes challenging. These involve maintaining a balance between coverage and resolution, as well as occlusions like trees or neighboring buildings. This study presents a sophisticated method that enhances building point cloud maps by seamlessly merging aerial data with 3D ground point cloud data. By harnessing building footprint data, point clouds from both aerial and ground sources are extracted, focusing on shared static points. Using these points, a transformation is computed to precisely align the ground map to its aerial counterpart, enriching the overall building point cloud data. To validate our approach, we conducted an experiment in a residential region to achieve a dense and accurate point cloud of individual residential buildings. We also introduced a novel bike cargo scanner, designed for rapid, close-range ground data collection. Our results conclusively demonstrated the successful integration of rich and accurate building point cloud data into aerial data.

Keywords: Reality capture · 3D point cloud · 3D reconstruction · Ground mapping · Simultaneous localization and mapping

1 Introduction

Starting from the late 1990s, automatic reconstruction of urban 3D models has become an important part of photogrammetric research [1]. Available 3D building mapping methods usually record 3D mapping information at the required detail and accuracy and then use this information to generate a geometric representation of the building in a subsequent step [2]. Different from traditional photogrammetry methods, 3D building

mapping generates a point cloud map without the need for known control points or manual matching. However, as pointed out by [3], digital building model generation of complex structures remains to be a challenging issue. Semi-automatic components are often necessary to support the recognition of complex buildings by a human operator. Despite the challenges, 3D building mapping still demonstrates excellence in various aspects, including 3D building realistic renderings and infrastructure for visualization and inspection. For example, with the combination of 2D flood simulations, 3D visual techniques and 3D point cloud from laser scanning, 3D representation of flood inundation is used to enhance 3d realistic risk perception and communication [4].

Aerial mapping, a methodology reliant on drones equipped with air-borne LIDAR and cameras for data collection and 3D mapping, stands as the most widely adopted approach for 3D building mapping. The advantages of aerial mapping offer a comprehensive view and access to areas inaccessible from ground-level perspectives. However, despite these merits, aerial mapping has limitations [5]. Despite the completeness of the whole map, map quality on buildings degrades heavily. In detail, the geometric representation of building facades, particularly for vertical structures, encounters serious deficiencies. In contrast, ground mapping provides an accurate representation of building facade geometry at close range and has flexible scanning times to compensate for the shortcomings of aerial mapping.

Therefore, to achieve a comprehensive 3D mapping of buildings, it is essential to integrate both ground mapping and aerial mapping techniques. Ground-to-aerial integration allows us to preserve the strengths inherent in aerial mapping, including the complete coverage and accurate terrain height of 3D aerial maps. Furthermore, this integrated approach would address missing information and low map quality of vertical building facades in aerial mapping maps.

However, the direct registration of full ground 3D maps to their aerial counterparts is a complex task, presenting significant challenges. One key issue lies in the global positioning errors. In the context of 3D building mapping, GPS positioning of aerial maps lacks sufficient accuracy, while ground maps exhibit more pronounced errors. Another challenge arises from the inherent differences between ground and aerial mapping—specifically, the less overlap between the two maps. Ground mapping excels in capturing detailed information about building facades and front yards, whereas aerial mapping focuses more on building roofs and the surrounding environment. Additionally, potential time gaps between ground and aerial mapping sessions add complexity, with the asynchrony possibly resulting in variations in environmental conditions. Addressing all these challenges is crucial for achieving a seamless and accurate integration of ground and aerial 3D mapping.

Therefore, this article aims to address challenges within the realm of 3D building mapping by providing a technique that integrates ground mapping with aerial mapping. In the context of aerial 3D point cloud maps, generated using a commercial aerial mapping system, the article emphasizes that the focus is not on aerial mapping. Instead, ground 3D models are created through a custom mobile data collection platform. To facilitate clarity, in this study, the building point cloud map, a point cloud representation of buildings and their surroundings, refers to the ground and the aerial 3D map. The building point cloud

model, in this study, means individual building point clouds with rich geometric and spatial information after ground-to-aerial integration.

The objective of this study is to enhance the completeness and quality of 3D building mapping by augmenting the aerial 3D map with ground-level building information. This augmentation is anticipated to yield a higher point density for building information and a broader scanning coverage compared to the conventional ground or aerial mapping approach. Furthermore, the technique is intricately designed to target specific locations of interest, with a primary focus on building information. This involves leveraging building footprint data to extract relevant information from both ground 3D maps and aerial 3D maps. The aerial 3D map assumes the role of a reference map, providing context, while the ground 3D map serves as the primary source for ground-view building information.

2 Related Works

2.1 Ground 3D Mapping

Ground 3D mapping methods broadly fall into two categories: vision-based and lidar-based methods. The vision-based method uses visual features or landmarks from consecutive images to estimate the camera's motion and generate maps. Structure from motion (SfM) is known as the most popular technique for the process of generating point clouds from monocular imagery, which has been applied to civil infrastructure modelling and inspection [6]. 3D SfM models are prone to omit crucial information when images lack texture. Using the SfM algorithm to generate point clouds presents challenges, especially when dealing with poor or variable lighting conditions and a lack of visual features for image tracking, which are common in outdoor environments. Previous studies, exemplified by [6], have recognized the quality issues associated with SfM-generated point clouds and recommended exploring alternative sensor types. Consequently, we refrain from using this vision-based mapping approach in this article due to the acknowledged limitations.

On the other hand, lidar-based mapping makes use of lidar as well as the embedded internal measurement unit (IMU) to create 3D point cloud mapping based on lidar odometry which is typically performed by finding the relative transformation between two consecutive frames. Lidar (Light Detection and Ranging), insensitive to illumination changes, facilitates detailed environmental mapping, particularly in complex urban areas, ensuring accurate localization and producing high-resolution 3D building maps. As the paramount lidar-based mapping methodology, lidar-based SLAM technologies have evolved significantly since the introduction of LOAM [7] and continue to undergo enhancements. FASTLIO2 [8], a fast and robust lidar-inertial odometry package, can obtain a precise point cloud of the structure with tightly coupled iterated Kalman filter on lidar and IMU measurements. Utilizing Fast-LIO2, R3live [9] achieves precise odometry, resulting in a more accurate point cloud map through the integration of lidar, inertial, and visual constraints. This lidar-based mapping technique extends its application to practical civil infrastructure inspection, as demonstrated by HG-SLAM, which employs a tilted 3D lidar and a monocular camera to locate the UAV and map the target bridge, specifically for bridge inspection purposes [10].

In this article, based on the aim of robust map building, rapid scanning from mobile platforms is more reliable than vision-based systems while surveying structural accuracy because depth estimation is directly measured but may not be accurately estimated by vision systems. On the other hand, the enhanced precision of range sensing requires the system to have a larger computation capacity. Therefore, ground mapping methods based on lidars are more feasible for applications requiring precise and reliable 3D mapping in the study.

2.2 Ground-To-Aerial Point Cloud Integration.

The integration between ground mapping and aerial mapping encompasses two main methods: registration-based and fusion-based approaches. In the registration-based method, the objective is to achieve registration by calculating correspondences, typically utilizing building features like outlines, to align ground and aerial LIDAR data. For instance, Yang et al. [11] introduced a robust registration method for TLS and ALS data. Their approach involves determining potential matching pairs of outlines based on geometric constraints between building outlines. However, their experiments rely on terrestrial laser scanning and aerial laser scanning data achieved without occlusion and in fully overlapped scenarios, which is not supported in neighborhoods characterized by high housing density and occlusion.

On the other hand, the fusion-based method involves blending ground and aerial models to generate a self-consistent model. For instance, Gao et al. proposed a method for the 3D digitalization of ancient Chinese architecture, where two ground and aerial mappings were merged via bundle adjustment [12]. Another noteworthy example is the collaborative mapping system [13], which integrates the ground-aerial point cloud registration algorithm with a pose graph-based collaborative SLAM pipeline. However, in fusion-based methods including two previous examples, it is important to acknowledge a limitation: they require meta-data from both ground and aerial mapping for post-processing tasks, either odometry fusion or mapping optimization [14].

In this paper, several known constraints are considered in terms of ground-to-aerial building point cloud integration methodology. Firstly, there is limited access to meta-data of aerial mapping, where only aerial 3D maps are available from the aerial mapping software, which makes fusion-based methods inapplicable. Another constraint involves the inaccuracy and inconsistency of GPS measurements, leading to poor coarse alignment. Moreover, the incompleteness of ground 3D building maps, caused by occlusion challenges, introduces difficulties when seeking correspondence through feature matching, as extracting outlines may not be universally applicable to all buildings. Therefore, to overcome these constraints, we employ a point cloud registration-based ground-to-aerial integration method in a coarse-to-fine manner. The ground and aerial maps are geo-referenced by GPS measurements for coarse alignment, after which the coarsely aligned models undergo refinement based on 3D point correspondences.

3 Overview

The proposed approach aims to achieve a 3D point cloud map with building detail information enhanced, as illustrated in previous sections. In the whole technique, the input involves an aerial point cloud map and collected ground 3D mapping data at the same site, including LIDAR, IMU, RGB images, GPS measurements, and building footprint information. The implementation of the technique is divided into three key modules: ground mapping module, Individual point cloud extraction module and ground-to-aerial integration module (see Fig. 1).

We now define the notations and coordinates that are used in this article. We consider $(\cdot)^w$ as world coordinate, which we define to be the global GPS coordinate for 3D global point cloud maps or global sensor measurements. $(\cdot)^l$ is local coordinate, the coordinate for local sensor's measurements. Transformations between different coordinates are represented by T , comprising the rotation matrix and the translation vector. We understand $(\cdot)_l^w$ as a transformation from local coordinate to world coordinate. To distinguish ground and aerial data, we consider $(\cdot)_g$ as ground information and $(\cdot)_a$ as aerial information. A point cloud is presented by a capital letter, for example, $X = [x_0, x_1, \dots, x_n]$. As the point cloud could be understood as a vector, each lowercase letter with subscripts in brackets represents the individual points in the point cloud vector. Similarly, the point vectors are also represented in the same way.

In the first step, ground 3D mapping data, including 3D LIDAR scans, IMU measurements, images, and GPS measurements, is collected by our custom mobile data collection platform. Then, a ground 3D point cloud map M_g^l in local coordinate will be generated by a lidar-based SLAM algorithm. Under a global geo-reference transformation, the ground 3D point cloud map will be transformed into M_g^w in global coordinate. Details are illustrated in the “Ground 3D mapping using a custom mobile mapping platform” section. In the second module, point clouds, including ground 3D point cloud map M_g^w and aerial 3D point cloud map M_a^w , are extracted to get individual building point clouds. In the third module, ground and aerial individual building point clouds are processed and aligned through a registration procedure. The alignment result is then used to merge ground and aerial point cloud maps. Details are provided in the following sections.

4 Ground 3D Mapping Using a Custom Mobile Mapping Platform

In this work, a custom mobile mapping platform is motivated to be developed due to the limitations of existing open-source datasets, like Google Street View, which provide sparse images and lack raw data availability. Another significant factor is the drawbacks of traditional vehicle data collection platforms, including occlusion issues with passing cars, pedestrians, and trees. Also, the high cost associated with commercial data collection platforms reinforces the need for our data collection solution. Hardware and software about our custom mobile mapping platform are introduced, and data is used for 3D ground mapping and coordinate geo-reference.

4.1 Hardware Development

A custom mobile mapping platform has been developed, utilizing a cargo foundation for enhanced functionality. This platform is specifically designed to collect data along

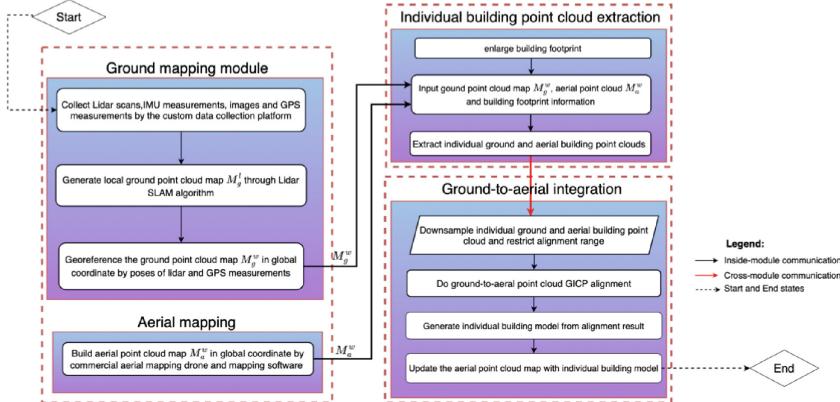


Fig. 1. Overall flow chart of the proposed methodology in this article.

sidewalks, which focuses on building facades in close range, and minimizes the impact of passing cars and interference from pedestrian trees. Figure 2(a) shows the onsite data collection platform and details of the sensors we use:

- **Lidar:** The Avia lidar, known for its multi-line and high-speed scanning capabilities, facilitates non-repetitive scanning with an $70.4^\circ \times 77.2^\circ$ FOV field of view. Enhancing coverage and scanning efficiency, three Livox Avia lidars are mounted on the platform.
- **Camera:** The platform incorporates the Teledyne Flir BFS-PGE-16S2C-CS camera. This GigE camera boasts a resolution of 1440×1080 and a lightweight design of 36 g.
- **GPS:** The GPS features an RTK GPS board with an antenna, developed by ArduSimple. This GPS module establishes RTK connection by connecting to the nearest public NTRIP base station.
- **Onboard computer:** The platform is equipped with an NUC 11, with an Intel i7-1165G7 CPU and 8 GB RAM.

As shown in Fig. 2(b), all the sensors are centrally controlled and managed by the onboard computer in the Robot Operating System (ROS), which allows sensor control and data collection to be performed easily. During data collection, the rosbag data package, including LIDAR, IMU, RGB images, and GPS measurements, is recorded with timestamps associated with each frame from different sensor topics.

The calibrations involve three main parts: (1) camera calibration; (2) lidar-to-lidar calibration; (3) camera-to-lidar calibration. Firstly, for the camera, we compute its intrinsic parameters using the Matlab Camera Calibration Toolbox [15]. This process, specific to a single camera, utilizes a checkerboard target. Secondly, considering the three lidars on the data collection platform, we determine the extrinsic parameters between lidars to integrate them into a unified lidar topic for mapping. Lidar-to-lidar calibration begins by measuring a coarse transformation matrix from the CAD model. This coarse extrinsic result is then refined using the MLCC package [16], which optimizes extrinsic results by aligning point cloud maps from each lidar. Lastly, for lidar-to-camera calibration, a similar procedure is followed. An initial coarse extrinsic transformation is obtained from

the CAD model, and then the livox-camera-calib package [17], a targetless camera lidar calibration tool, is used to align the transformation matrix between the camera image and lidar scan precisely.



Fig. 2. Mobile data collection system: (a) Data collection system on a cargo bike and (b) Hardware connection diagram of the platform

4.2 Ground 3D Mapping

The ground 3D mapping algorithm necessitates specific requirements for its successful implementation. Firstly, as discussed in the Sect. 2 a lidar-based SLAM system is essential. Secondly, RGB-colored information is a requisite for the point cloud map, ensuring that the ground map is consistent with aerial maps in terms of colour style. Thirdly, the algorithm mandates a pose trajectory with low error derived from the SLAM process for both mapping and geo-transforming purposes. It is noteworthy to mention that the focus is not on developing a new algorithm but rather on selecting an algorithm that meets these requirements.

Under such circumstances, R3live proves to be a fitting SLAM solution for these requirements. Functioning as an RGB-colored Lidar-internal-visual SLAM algorithm, it outperforms other alternative lidar-based SLAM algorithms by exhibiting a lower relative pose error, as indicated in the R3live paper. In the deployment, R3live takes as input Lidar scans, IMU measurements, and images; in output, it delivers an RGB-colored point cloud map and 6-Dof pose trajectory.

4.3 Map Georeferencing

Coordinate transformation plays a crucial role in the processing of the ground point cloud map. In this paper, the geo-referred coordinate, the internal coordinate system of a digital map, is specifically identified as the global GPS coordinate. In contrast, the ground point cloud map M_g^l generated from the SLAM algorithm, is based on the sensor's local coordinates. To facilitate further stages in the analysis, such as individual building point cloud extraction and aerial-to-ground integration, it becomes imperative that the ground point cloud maps transform into the global GPS coordinate.

To address the alignment challenge, timestamps are first matched into pairs between the SLAM poses in local coordinate and GPS measurements in global coordinate by the ROS TimeSynchronizer. But only GPS measurements with high accuracy, calculated by RTK correction based on NTRIP, are kept. Next, pairs of SLAM poses and GPS measurements are selected near the targeted individual buildings. Then, the SVD (Singular Value Decomposition) [18] is applied to SLAM odometry and GPS data for transformation calculation. Finally, the local ground point cloud map M_g^l is transformed into the ground point cloud map M_g^w in global coordinates using the obtained transformation T_l^w .

5 Extraction of Building Point Clouds

To get complete and comprehensive building models and enhance the map with the region of interest, extraction of the building point clouds is deployed utilizing the building footprint. Besides building footprint, inputs to this module include both ground 3D map M_g^w from the previous section and the aerial 3D maps M_a^w in global coordinate, then the module outputs individual building point clouds. The point cloud extraction methodology is achieved basically by the geometric relationship between a point and a polygon building footprint. For each point inside a point cloud, iteratively judge whether the point lies within the building footprint polygon. Whether the point is inside the polygon is determined by the count of intersections with the polygon's outlines. The point sends a ray to the right and then counts the number of intersections with the polygon. If the count is odd, the point is inside the polygon. Conversely, it lies outside the polygon.

However, problems with inaccurate and incomplete point clouds of individual buildings occur. The position of ground 3D maps is not precise enough due to global georeferencing transformation, while aerial 3D maps are also subject to GPS position drift, despite being optimized using markers. To address these problems, the solution lies in extending the building footprint. The extended building footprint enhances the spatial tolerance of the point cloud map coordinates, ultimately ensuring the quality and integrity of the extracted individual house point clouds.

The method for extending the building footprint involves treating the extension of an individual building's footprint as polygon offsetting. This process, proposed by Chen et al. [19], creates parallel curves offset by a specified distance from their primary curves. In this article, the building footprint, represented by corners of the polygon, is a polygon with all corners at right angles. As shown in Fig. 3, exemplified by a corner of a footprint polygon, point e is calculated to be the new corner of building footprint polygon. Implementation of the polygon extension is assisted with Clipper2.

6 Ground-To-Aerial Integration

In this section, ground-to-aerial alignment is achieved for each building by GICP (Generalized Iterative Closest Point algorithm) [20]. In detail, individual ground and aerial building point clouds are firstly downsampled to reduce computational costs and accelerate matching speed. Then, a transformation is obtained through the ground-to-aerial

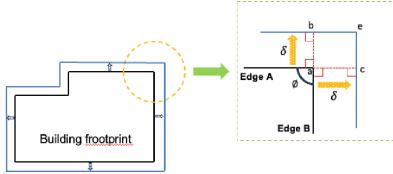


Fig. 3. Geometric relationship for building footprint extension. Original building footprints are depicted by black lines, while extended footprints are represented by blue lines. δ is the offset distance, θ is the angle between two edges, *edgeA* and *edgeB*. Point a and point e are the pointed before and after extension.

alignment. With the alignment result, the building point cloud model is generated by merging the ground and aerial building point clouds.

For ground-to-aerial alignment, in standard ICP [21], was designed to precisely align point clouds by iteratively computing the rigid transformation between the source and target. This process involves computing correspondences between the two scans, followed by the computation of a transformation that minimizes the distance between the target and the source point cloud, as outlined in Algorithm 1. Despite its general applicability, standard ICP faces challenges in our case. During the experiment, alignment results from standard ICP prove to be neither robust nor accurate, which is raised by two reasons: firstly, the distribution of points in the ground and aerial cloud maps displays different characteristics with limited overlap, leading to inaccurate correspondence computation; secondly, the minimum distance transformation and point-to-point correspondence in standard ICP fails to achieve seamless alignment due to the presence of noise in both ground and aerial point clouds, especially for vertical geometry structures.

Algorithm 1. Standard ICP algorithm

Input: Two point clouds: $A = \{a_1, \dots, a_m\}$ $B = \{b_1, \dots, b_m\}$; An initial transformation: T_0
 Output: The correct transformation, T, which aligns A and B

| | | | |
|----|------------------------|--|---|
| 1 | $T \leftarrow T_0$ | | |
| 2 | while not converged do | | |
| 3 | | for $i \leftarrow 1$ to N do | |
| 4 | | | $m_i \leftarrow \text{FindClosestPointInA}(Tb_i)$ |
| 5 | | | if $\ m_i - Tb_i\ _2 \leq d_{max}$ then |
| 6 | | | |
| 7 | | | else |
| 8 | | | |
| 9 | | | $w_i \leftarrow 1$ |
| 10 | | | $w_i \leftarrow 0$ |
| 11 | | | end if |
| 12 | end for | | |
| | | $T \leftarrow \underset{T}{\operatorname{argmin}} \sum_{i=0}^n w_i \ Tb_i - m_i\ _2$ | |
| | end | | |

Therefore, GICP, adding a probabilistic model in the distance minimization step, is used in this article to solve the problems discussed. In the inputs A and B, a covariance

matrix is used for a point with the surface normal, assuming inputs A and B are Gaussian distributions. Then, inputs A and B can be presented by Eq. (1). Line 11 in Algorithm 1, the distance minimization formula is formulated as Eq. (2). ϵ represents the Gaussian covariance, C_i^B and C_i^A then denote the input point clouds with gaussian distribution.

$$C_i^B = R_{\mu_i} \bullet \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \bullet R_{\mu_i}^T; C_i^A = R_{v_i} \bullet \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \bullet R_{v_i}^T \quad (1)$$

$$T = \underset{T}{\operatorname{argmin}} \sum_{i=0}^n d_i (C_i^B + T C_i^A T^T - m_i)^{-1} d_i^{(T)} \quad (2)$$

7 Experiment and Results

7.1 Ground Data Collection

In the experiment, our data collection focused on a neighboring area containing 80 houses in 4 blocks, all falling under the category of individual residential houses. The choice of experiment site was driven by the relevance of the area to our research projects, providing data collection authentication for research purposes. With the developed ground 3D data collection platform, data is collected in 24 min of scanning into a data package of size 52GB, with respective frequencies of 30 Hz for images, 10Hz for LIDAR scans, and 1Hz for GPS measurements.

7.2 3D Mapping & Building Point Cloud Extraction

First, R3live, the RGB-colored Lidar-internal-visual SLAM algorithm, is used for 3D mapping with an output of an RGB-colored point cloud map. Then, based on the ground 3D point cloud map and the given aerial point cloud map from a commercial aerial mapping service, the proposed building point cloud extraction is performed by leveraging the extended building footprints extended by 0.5 m. As we can see from Fig. 4, individual building point clouds from both ground and aerial 3D point cloud maps are built and extracted ideally.

7.3 Ground-To-Aerial Integration

In order to systematically evaluate the proposed ground-to-aerial integration method, both qualitative and quantitative evaluations are done on the experiment results. The qualitative evaluation results reveal crucial insights, providing a more intuitive way to validate the effectiveness of the proposed technique. The quantitative evaluation is performed by distance measurements on the individual building ground-to-aerial integration results. It is pertinent to note that our technique primarily focuses on ground and aerial map alignment; thus, point cloud map alignment is evaluated, excluding assessment of the quality of the ground or aerial point cloud map itself. Also, because of limited access to the raw data of the aerial point cloud input, we are unable to make comparisons with other ground and aerial point cloud merging approaches.



Fig. 4. Individual building point clouds extracted from ground and aerial point cloud maps. Pairs (a) to (c) exhibit four pairs of randomly selected building point clouds. Each pair consists of a ground building point cloud (left) and an aerial point cloud (right).



Fig. 5. The qualitative ground-to-aerial integration results, focusing on the same four pairs of building point clouds as depicted in Fig. 4. For each pair, the figure displays the ground and aerial building point clouds before integration (left), and the integrated individual building model (right).

Qualitative Integration Results. First, ground-to-aerial integration is performed on each building and results are shown in Fig. 5. Under the condition that there is a certain overlapping area, the ground building point cloud aligns with the aerial point cloud well. Figure 5 (a)–(c) are successful integration examples, Fig. 5(d) is failed integration examples, caused by a poor overlap prior between ground and aerial. In Fig. 6, we can see from the point cloud map before and after ground-to-aerial integration that the performance of the proposed technique is evident on the global map. Individual building information is enhanced from the ground 3D map into the aerial map. Details of building facades are clear, highlighting an enhancement in the completeness of the original aerial map.



Fig. 6. Point cloud map before and after ground-to-aerial integration: (a) original aerial point cloud map and (b) full map with ground-to-aerial integration.

Quantitative Evaluation. To quantitatively evaluate our ground-to-aerial alignment results of the proposed method, we use the nearest point distance method on the overlap area of each two point clouds. First of all, for ground and aerial point cloud maps, we find their overlap area by the distance between points. And then, for each point in the ground point cloud, its nearest aerial point is found by Euclidean distance. Finally, a cumulative error distribution is calculated with a distance less than nu , where u is a minimum judgement distance, n would iteratively grow from $n = 1$ to $n = 9$. We treat the distances, which are greater than $9u$ as the same and cumulate them at $10u$. Although this quantitative evaluation method is not super accurate, it can provide an evaluation insight into our alignment method. The proposed quantitative method is evaluated on five randomly picked ground-to-aerial integration results. As shown in Fig. 7, for different buildings, the percentage of cumulative nearest point rates reaches above 80% on the overlap area, revealing the effectiveness of the proposed ground-to-aerial integration method.

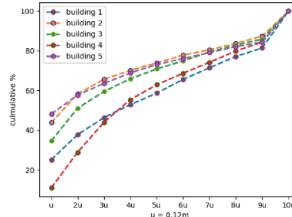


Fig. 7. Nearest point distance quantitative evaluation on five randomly selected buildings: The curves are the nearest point distance distribution with the y-axis being the percentage of the ground points with errors $< nu$ for different buildings.

8 Conclusion

In conclusion, we have presented a novel approach to tackle the challenge of missing geometric information and low-resolution representations of building facades. Our technique integrates ground mapping with aerial mapping, facilitated by a custom mobile data collection platform designed for ground data collection on sidewalks and 3D mapping.

Leveraging building footprint information, we successfully extract individual building point clouds from both ground and aerial maps. Through ground-to-aerial integration, we obtain individual building models as well as a more comprehensive point cloud map. Our experimental results validate the effectiveness of our approach, offering promising prospects for enhancing the accuracy and detail of building facade representations.

References

1. Gruen, A., Baltsavias, E., Henricsson, O.: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Springer Science & Business Media (1997)
2. Haala, N., Kada, M.: An update on automatic 3D building reconstruction. *ISPRS J. Photogramm. Remote Sens.* **65**(6), 570–580 (Nov. 2010). <https://doi.org/10.1016/j.isprsjprs.2010.09.006>
3. Habib, A.F., Zhai, R., Kim, C.: Generation of complex polyhedral building models by integrating stereo-aerial imagery and lidar data. *Photogramm. Eng. Remote Sens.* **76**(5), 609–623 (May 2010). <https://doi.org/10.14358/PERS.76.5.609>
4. Costabile, P., Costanzo, C., De Lorenzo, G., De Santis, R., Penna, N., Macchione, F.: Terrestrial and airborne laser scanning and 2-D modelling for 3-D flood hazard maps in urban areas: new opportunities and perspectives. *Environ Model Softw.* **135**, 104889 (Jan. 2021). <https://doi.org/10.1016/j.envsoft.2020.104889>
5. Luo, L., et al.: Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: a review of the century (1907–2017). *Remote Sens. Environ.* **232**, 111280 (Oct. 2019). <https://doi.org/10.1016/j.rse.2019.111280>
6. Khaloo, A., Lattanzi, D., Cunningham, K., Dell'Andrea, R., Riley, M.: Unmanned aerial vehicle inspection of the Placer River Trail Bridge through image-based 3D modelling. *Struct. Infrastruct. Eng.* **14**(1), 124–136 (Jan. 2018). <https://doi.org/10.1080/15732479.2017.1330891>
7. Zhang, J., Singh, S.: LOAM: Lidar Odometry and Mapping in Real-time. In: *Robotics: Science and Systems X, Robotics: Science and Systems Foundation* (Jul. 2014). <https://doi.org/10.15607/RSS.2014.X.007>
8. Xu, W., Zhang, F.: FAST-LIO: a fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter. *IEEE Robot. Autom. Lett.* **6**(2), 3317–3324 (Apr. 2021). <https://doi.org/10.1109/LRA.2021.3064227>
9. Lin, J., Zhang, F.: R3LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package. In: *2022 International Conference on Robotics and Automation (ICRA)*, pp. 10672–10678 (May 2022). <https://doi.org/10.1109/ICRA46639.2022.9811935>
10. Jung, S., Choi, D., Song, S., Myung, H.: Bridge inspection using unmanned aerial vehicle based on HG-SLAM: hierarchical graph-based SLAM. *Remote. Sens.* **12**(18), Art. no. 18 (Jan. 2020). <https://doi.org/10.3390/rs12183022>
11. Yang, B., Zang, Y., Dong, Z., Huang, R.: An automated method to register airborne and terrestrial laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* **109**, 62–76 (Nov. 2015). <https://doi.org/10.1016/j.isprsjprs.2015.08.006>
12. Gao, X., Shen, S., Zhou, Y., Cui, H., Zhu, L., Hu, Z.: Ancient Chinese architecture 3D preservation by merging ground and aerial point clouds. *ISPRS J. Photogramm. Remote Sens.* **143**, 72–84 (Sep. 2018). <https://doi.org/10.1016/j.isprsjprs.2018.04.023>
13. He, J., Zhou, Y., Huang, L., Kong, Y., Cheng, H.: Ground and aerial collaborative mapping in urban environments. *IEEE Robot. Autom. Lett.* **6**(1), 95–102 (Jan. 2021). <https://doi.org/10.1109/LRA.2020.3032054>

14. Gao, X., Shen, S., Hu, Z., Wang, Z.: Ground and aerial meta-data integration for localization and reconstruction: a review. *Pattern Recogn. Lett.* **127**, 202–214 (Nov. 2019). <https://doi.org/10.1016/j.patrec.2018.07.036>
15. Bouguet, J.-Y.: Camera calibration toolbox for MATLAB (2015). http://www.vision.caltech.edu/bouguetj/calib_doc/ (Accessed Jan 20, 2024)
16. Liu, X., Yuan, C., Zhang, F.: Targetless extrinsic calibration of multiple small FoV LiDARs and cameras using adaptive voxelization. *IEEE Trans. Instrum. Meas.* **71**, 1–12 (2022). <https://doi.org/10.1109/TIM.2022.3176889>
17. Yuan, C., Liu, X., Hong, X., Zhang, F.: Pixel-level Extrinsic Self Calibration of High Resolution LiDAR and Camera in Targetless Environments. *arXiv*, Jun. 25, 2021. <https://doi.org/10.48550/arXiv.2103.01627>
18. Klema, V., Laub, A.: The singular value decomposition: Its computation and some applications. *IEEE Trans. Autom. Control* **25**(2), 164–176 (Apr. 1980). <https://doi.org/10.1109/TAC.1980.1102314>
19. Chen, X., McMains, S.: Polygon Offsetting by Computing Winding Numbers, presented at the ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, pp. 565–575 (Jun. 2008). <https://doi.org/10.1115/DETC2005-85513>
20. Segal, A., Haehnel, D., Thrun, S.: Generalized-icp. *Robot.: Sci. Syst.* **2**(4), (2009)
21. Besl, P.J., McKay, N.D.: Method for registration of 3-D shapes. In: *Sensor Fusion IV: Control Paradigms and Data Structures*, SPIE, pp. 586–606 (Apr. 1992). <https://doi.org/10.1117/12.57955>



Automated Stakeout Survey for Efficient Earthwork Construction Project Control

Sida Wang, Qingying Wu, Ming Lu^(✉), and Chen Wang^(✉)

Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB,
Canada

{sida1, cwang20, mlu6, qingyin1}@ualberta.ca

Abstract. Stake-out survey checks the actual elevations of control points in a construction field against known coordinates as per the engineering design. Yet, stake-out survey relies highly on manual work in the construction field and thereby results in productivity losses, quality defects and safety hazards. This research presents a novel solution for conducting a stakeout survey in construction engineering. By developing and field-testing an automated stakeout survey system, comprising a survey robot (off-the-shelf Leica robotic total station) and a cost-effective custom-built rover, it substantially lowers the barrier to real-time data acquisition in the dynamic environment of a construction site. This system not only enhances efficiency and safety but also serves as a practical, low-cost bridge between the physical and virtual worlds, facilitating the implementation of Digital Twins for earthwork construction. The innovation of this work lies in its potential to overcome one of the primary obstacles to digital transformation in the construction industry—high data collection costs. By demonstrating a cost-effective method to accurately track changes in the construction field, this research lays the groundwork for the future development of Digital Twins for earthwork construction.

Keywords: Robotic total station · rover robot · Civil 3D

1 Introduction

As the construction industry evolves, the integration of high-tech innovations—ranging from Building Information Modeling (BIM) and the Internet of Things (IoT) to Digital Twins (DT)—has become increasingly crucial. Despite these advancements, the industry faces a persistent challenge: the high cost and complexity of data collection in the ever-changing environment of construction sites [1]. This research delves into a pioneering solution to this challenge, presenting an integration of robotic surveying techniques and digital terrain modeling, specifically tailored to mitigate the financial and logistical barriers in earthwork construction. Utilizing the established reliability of the Robotic Total Station, our approach introduces an innovative combination of precision surveying with a cost-effective, remote-controlled rover. This strategy significantly reduces the overhead associated with traditional survey methods while ensuring work progress measurement accuracy and quality assurance in rough grading.

Additionally, we address the labor-intensive and safety-concerned nature of traditional stake-out surveys, underscored by findings from the Center for Construction Research and Training [2]. In response to the pressing need for field crews to accommodate time constraints and safety hazards, our research proposes a method that not only minimizes safety risks but also streamlines the data collection process, making it both safer and more efficient. Fundamentally, our automated surveying system seeks to alleviate safety concerns and address the critical bottleneck of high data collection costs, paving the way for the adoption of Digital Twins technology in construction projects.

Traditional survey methods, while reliable, often aims at achieving higher efficiency and accuracy in catering to the complex demands of modern construction projects. Yet, the necessity for continuous trial and error to enhance precision inherently leads to inefficiencies in the “stake-out” field survey and disruptions in the operations of heavy equipment fleets in the field, thereby leading to potential productivity losses in construction activities [3]. Recent strides in Digital Twins technology provide real-time monitoring and control of the construction process, emphasizing progress and work quality [4]. The importance of field surveys becomes evident in the implementation of construction-oriented “Digital Twins,” involving the creation of a digital model mirroring physical reality and conducting simulations to guide immediate decisions in the real world. Our proposed automated field monitoring and work measurement system strikes a good balance between (1) delivering sufficient accuracy in terrain mapping and control point establishment and (2) lowering the technology application cost in the construction field. This approach not only promises to reduce the data collection costs drastically but also enhances operational efficiency, safety, and accuracy in progress measurement and quality monitoring. The new system’s core is a survey robot -a Leica Viva TS 15 robotic total station [5] equipped with automated functions such as power search, auto-targeting, and target lock. Complementing the survey robot is a rover that can be controlled remotely and is equipped with a prism that replaces a poleman. This combination allows for automated positioning and precise fixation of target points.

In addition, the system is seamlessly integrated with modern construction management software such as AutoCAD Civil 3D to enable data transfer and analysis for real-time decision-making on construction sites. This method contrasts the time-consuming, error-prone manual methods, significantly advancing survey accuracy and project management efficiency. The automated stake-out survey solution provides the cost-effective technology linking the virtual and real worlds to realize a Digital Twins solution for earthwork construction. It is worth mentioning that Autodesk Civil 3D emerges as a specialized tool that leverages the capabilities of BIM with Digital Elevation Model (DEM) technology, offering a comprehensive solution for civil engineering design and documentation. Its strength lies in its ability to handle large datasets and complex designs, making it an indispensable tool for modern construction projects. Civil 3D’s dynamic design functionality allows for real-time updates and design change synchronization, which is crucial for maintaining the integrity and accuracy of the project design [6].

2 Literature Review

While the total station has been instrumental in advancing construction survey accuracy and cost-effectiveness over the past three decades, the industry now faces the immediate challenge of integrating these technologies with the dynamic requirements of Digital Twins and real-time data acquisition. The total station falls short in addressing the high costs and logistical challenges associated with continuous data collection in fluctuating construction environments. In the realm of construction project management, the integration of robots introduces faster and more efficient alternatives for project control than conventional field operations. However, the integration often stumbles upon the high operational costs and the complexity of adapting these technologies to ever-changing construction sites, underscoring a gap our research aims to bridge. Experiments focusing on dynamic target tracking and structural deformation monitoring have demonstrated the total station's efficacy in achieving efficiency and precision. A higher sample rate consistently yielded better results and captured more data for analysis, while the influence of the target distance on survey outcomes was also observed [7]. Research showcased the applicability of robotic total stations in earthwork volume calculation; in comparison with 3D laser scanning and photogrammetry, the geodetic method utilizing a robotic total station achieved an accuracy of nearly 90%, while confining the operation process to acceptable time frames [8].

While the cooperation between robotic total stations and mobile robots, facilitated by Robotic Process Automation (RPA), marks a leap towards autonomous project planning and control, the innovation's full potential is hindered by the economic and practical barriers of implementing such technologies in the field [9]. This research focuses on leveraging mobile robots in execution of construction survey tasks; a mobile robot navigates on-site and collaborates seamlessly with digital survey equipment through wireless communication technology.

Although Building Information Modeling (BIM) technology represents a significant advancement in earthwork operations, its integration into daily construction activities remains limited by the prohibitive costs of continuous and comprehensive data collection efforts. To better understand the application of 3D model technology in earthwork calculations, one approach involved calculating earthwork volume using the cross-section method based on the Digital Elevation Model (DEM) data generated from drone surveys in road engineering [10]. Another approach compared the method accuracy to Global Navigation Satellite System (GNSS) surveying, showing comparable results from Unmanned Aerial Vehicle (UAV)-based and GNSS-based volumes [11]. Further, PhotoScan and Pix4D were utilized to calculate a volume based on a DEM created by a drone survey [12]. AutoCAD Civil3D provided the functionality for volume calculation based on point clouds and TIN (Triangulated Irregular Network) data created by the GNSS survey [13].

3 Analysis of Automated and Manual Survey Methods in Earthwork Survey

Traditional land survey is performed by trained professionals: a total station operator and a target finder. The former operates the total station, while the latter moves a telescopic pole in the field with a prism attached to it, aiming to locate the position of a target point in the field and hold up the pole for the total station operator to fix the coordinates. Where there is no landmark reference to the real position of a control point in the field, the worker needs to tediously move around the pole and repeat the survey cycle to locate the target point by trial and error. Although this practice has stood the test of time and provides adequate precision, to a certain extent, it presents hurdles to efficiency and productivity in the construction field while falling short to meet the dynamic project management needs in terms of providing quick updates.

The advent of Real Time Kinematics (RTK) GPS [14] lends a groundbreaking technology to field survey. RTK GPS provides centimeter-level accuracy in positioning contingent on the reception of signals from multiple satellites (within a 2.9 cm error margin), which is adequate for a wide range of applications, from construction management to automotive test preparations [15]. Nonetheless, RTK GPS's performance is heavily contingent upon environmental factors that facilitate or hinder satellite signal receptions; for instance, signal latency, obstructions to line-of-sight with satellites. Furthermore, RTK GPS typically needs professional equipment and licensed operators, which is expensive to implement in the field and often requires a temporary shutdown of construction operations.



Fig. 1. Survey robot (Leica Viva TS15) on survey tripod set up in the test field & The stakeout rover operated by a remote controller on 2.4 GH radio frequency

Recognizing the existing methods' limitations in cost and adaptability to changing construction environments, our research leverages the Leica Viva TS 15 [5] robotic total station, enhancing it with a cost-effective, dynamic approach to data collection. This robotic total station is equipped with Automatic Target Aiming (ATR) and automatic prism tracking capabilities, featuring a maximum tracking speed of 4 m per second at

a radius distance of 800 m. Typically, within its operational range, the total station can capture and store one point in its internal storage every 5–10 s when tracking a moving target. In ATR mode, the robotic total station achieves a maximum range of 1000 m, while in the regular round-style survey prism locking mode, it extends up to 800 m. Figure 1 displays an image of a robotic total station positioned on a survey tripod in the field. Note the line of sight between a target object and the RTS unit is required for tracking and survey operations. The application of total stations has primarily been in static survey scenarios in the construction industry. On the other hand, integrating moving target tracking and point survey automation features in RTS lends itself well to more dynamic applications.

The fundamental principle behind the operation of a total station involves utilizing laser to measure both the angle and distance from the (robotic) total station to a target object. Land surveying serves two primary purposes: first, it is conducted before construction commences to establish the site layout; second, after construction is completed, surveying is performed to verify the quality and quantity of the finished products [16]. Central to our methodology is a custom-designed wheeled stake-out robot (the rover), equipped with an Arduino minicomputer and various sensors. This design prioritizes not only functionality but also cost-effectiveness and ease of replication in real-world construction settings [17]. This stake-out robot collaborates seamlessly with a survey robot, namely the robotic total station, enabling it to reach any target point within a range of 1000 m. This capability is particularly beneficial in a typical rough grading field without conspicuous obstacles [18].

The foundation of the stake-out rover is the “Axial Wraith 4 × 4 electric remote-control vehicle (rock crawler)”, subject to custom enhancements [19]. Powering this 4 × 4 electric remote-control vehicle is a specialized 2300kV motor and a Spektrum radio control system [20]. The rover is operated remotely using the anti-interference 2.4 GHz radio frequency, boasting a communication range of 500 m under ideal open land conditions. To withstand the rough field conditions, the chassis is reinforced with durable metal components, and the tires are upgraded for optimal performance in mud and soil [21].

A standard round-style survey prism is mounted on the rear of the remote-controlled vehicle, positioned 26.6 cm above the ground for fixation of a target point in the field, as depicted in Fig. 1. Additionally, a universal 360-degree survey prism is shown in the top right corner of Fig. 1. This special prism substitutes for the regular round-style survey prism when the survey robot needs to track the rover across multiple target points over a vast field area.

With the prism affixed to its rear end, the stake-out rover is capable of real-time tracking of x, y, and z values of the prism by the survey robot. On each target point in the field, the survey robot conducts at least five measurements.

Comparing with the stake-out survey practice by RTK GPS [14] in both “spot checking” and “area mapping” in construction (referred to as “stake-out” and “topological survey” in survey engineering), features and advantages of the automated stake-out solution are illustrated in Fig. 2.

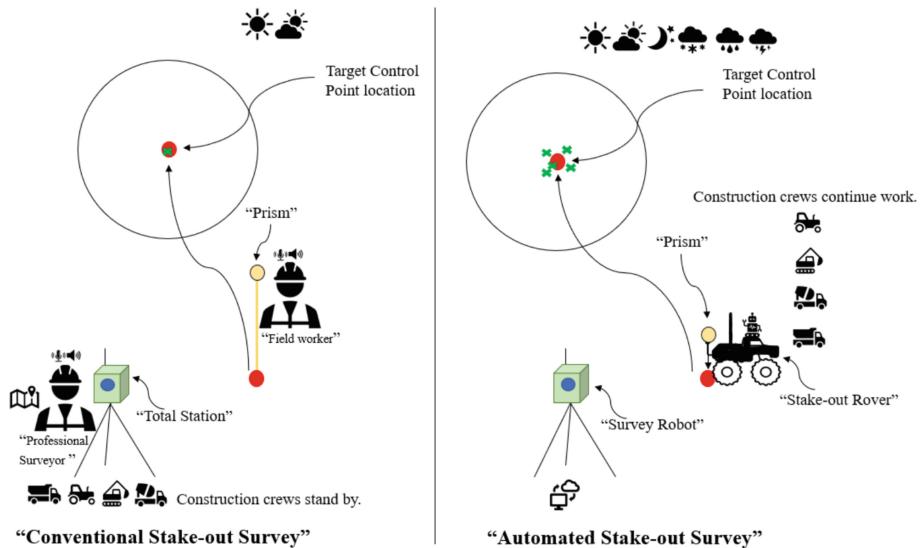


Fig. 2. Contrasting Methods: Conventional Stake-out Survey vs Automated Stake-out Survey

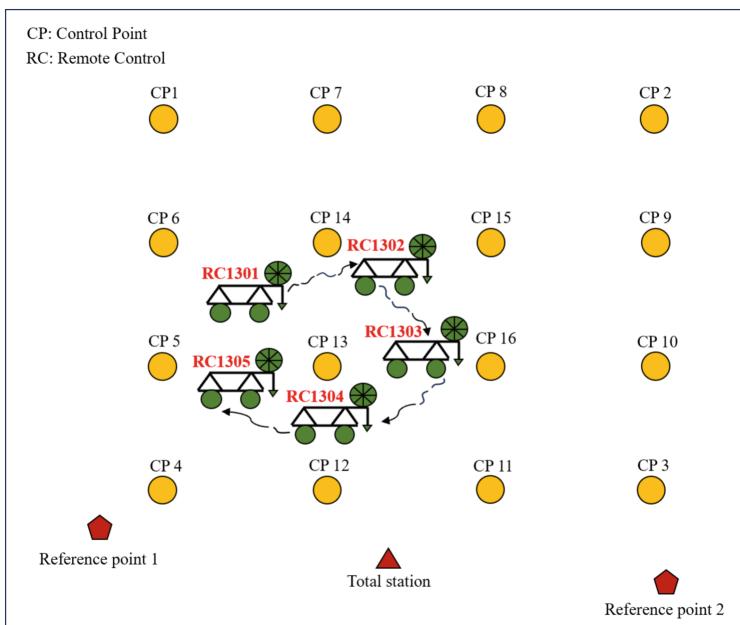


Fig. 3. Layout configuration in the field

In our innovative field experiment, we strategically marked sixteen control points in an earthwork field, setting the stage to test our automated survey system against traditional methods. The x and y coordinates were obtainable manually from the spikes on

the ground, and all the control points were about 3 m away from each other (Fig. 3). During this field test. The automated survey system not only demonstrated its efficiency by completing the elevation capture of sixteen control points in just eighteen minutes—a stark contrast to the forty-two minutes required by traditional stake-out surveys—but also highlighted the cost-effectiveness and practicality of our approach in real-world construction environments. Then a weighted average algorithm was applied to derive elevations based on the data from automated survey and a linear regression analysis performed to compare elevation data from the two methods, resulting in a general agreement in the two sets of results (the maximum deviation under 35 mm.) Hence, the automated stakeout survey solution is verified to be sufficiently accurate for work progress checks and quality assurance in rough grading scenarios.

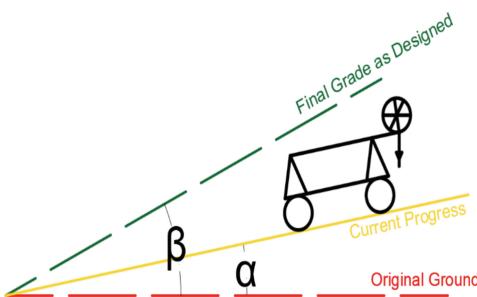


Fig. 4. Rover on tilted terrain

Notably, a tilt angle between the vertical prism alignment and the plumb line is conceivable, as illustrated in Fig. 4. As the tilt angle proved insignificant during field testing (remaining under 10 degrees), its impact on the vertical distance was ignored in the analysis. Future research may explore automatic accommodation for the tilt angle's effect upon the survey results for enhanced accuracy.

4 Advancing Earthwork Surveying with Digital Twin Integration

4.1 Enter Rover Collected Data Points into Civil 3D for Project Control

The survey robot continuously tracked the remote rover and captured the data for 16 control points in the CSV (comma-separated values) format, which is well-regarded for its simplicity and compatibility with various computing systems [22]. Once the survey data is converted into CSV format, it can be imported into AutoCAD Civil 3D -a BIM software program tailored for geodetic and civil engineering projects. It is worth mentioning that the importation of data into AutoCAD Civil 3D represents a critical step in the evolution of automated earthwork surveying methods. AutoCAD Civil 3D utilizes CSV files for data post-processing, facilitating efficient visualization and in-depth data analysis [23], which includes creating Digital Elevation Models (DEM), volume calculations, and alignment designs. Additionally, various specialized AutoCAD Civil 3D add-ons are available to enhance geospatial analysis.

After importation, the data undergoes cleaning and organization, which involves removing any outliers and ensuring data consistency. Then as seen in Fig. 5, a Digital Elevation Model (DEM) is generated in Civil 3D, allowing for a visual and quantitative analysis of the site. This analysis includes slope calculations, contour generation, and other terrain assessments.



Fig. 5. Create Digital Elevation Model (DEM) in Civil3D

4.2 Organization Difference

Aligning the DEM with the design surfaces is the next step to check the current ground against the intended design. By directly observing the overlay, any visible misalignments or inconsistencies are identified. This visual inspection method uses qualitative evaluation, which is different from quantitative analysis mentioned previously such as weighted average algorithm and linear regression. These combined analytical techniques offer a thorough insight into the accuracy of the automated survey system and ensure that the survey data is a true reflection of the data as designed.

4.3 To Facilitate the Planning of “Work to Do” for Immediate Future

AutoCAD Civil 3D enables calculating volumes and creating profile views in just a few seconds, as shown in Fig. 6 and Fig. 7. This ensures that takeoff on the quantity of work to do is precise for a particular day, instead of estimating from historical data [23]. By regularly comparing the actual earthmoving progress with the projected progress, project managers can track the current progress while planning for the future work. This real-time monitoring is crucial for project control in terms of timely interventions, adjustments in strategy, or resource reallocation to keep the project on track. These

technologies and tools shift from reactive to proactive project management in earthwork construction, allowing for formulation of a more dynamic and responsive construction management strategy and short-term work plans.

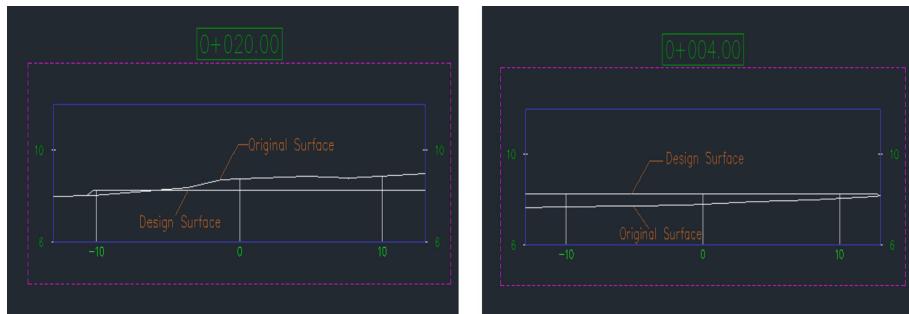


Fig. 6. Create Section View of Cut and Fill

| Volume Report | | | | | | | | | |
|--|-------------------------|---------------------------|--------------------------------|--------------------------|----------------------------|------------------------------|-----------------------------------|-------------------------------|------------------------------|
| Project: D:\OneDrive\Drawing4.dwg | | | | | | | | | |
| Alignment: Alignment - (4)
Sample Line Group: SL Collection - 6
Start Sta: 0+002.000
End Sta: 0+024.000 | | | | | | | | | |
| Station | Cut Area (Sq.m.) | Cut Volume (Cu.m.) | Reusable Volume (Cu.m.) | Fill Area (Sq.m.) | Fill Volume (Cu.m.) | Cum. Cut Vol. (Cu.m.) | Cum. Reusable Vol. (Cu.m.) | Cum. Fill Vol. (Cu.m.) | Cum. Net Vol. (Cu.m.) |
| 0+002.000 | 0.00 | 0.00 | 0.00 | 11.71 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0+004.000 | 0.00 | 0.00 | 0.00 | 11.71 | 4.15 | 0.00 | 0.00 | 4.15 | -4.15 |
| 0+006.000 | 2.03 | 11.59 | 11.59 | 8.36 | -3.75 | 11.59 | 11.59 | 0.40 | 11.19 |
| 0+008.000 | 4.83 | 10.83 | 10.83 | 8.74 | 14.36 | 22.42 | 22.42 | 14.76 | 7.65 |
| 0+010.000 | 0.46 | -19.53 | -19.53 | 8.45 | 40.13 | 2.89 | 2.89 | 54.90 | -52.01 |
| 0+012.000 | 0.04 | -1.81 | -1.81 | 5.95 | 10.99 | 1.09 | 1.09 | 65.89 | -64.81 |
| 0+014.000 | 0.77 | 3.11 | 3.11 | 3.07 | 13.96 | 4.19 | 4.19 | 79.85 | -75.66 |
| 0+016.000 | 10.30 | 52.52 | 52.52 | 5.78 | -19.14 | 56.72 | 56.72 | 60.71 | -4.00 |
| 0+018.000 | 10.90 | 21.19 | 21.19 | 4.08 | 9.86 | 77.91 | 77.91 | 70.57 | 7.34 |
| 0+020.000 | 0.00 | -11.23 | -11.23 | 0.00 | 13.77 | 66.68 | 66.68 | 84.34 | -17.66 |
| 0+022.000 | 11.00 | 11.07 | 11.07 | 0.00 | 0.00 | 77.75 | 77.75 | 84.34 | -6.59 |
| 0+024.000 | 12.42 | 23.42 | 23.42 | 0.00 | 0.00 | 101.17 | 101.17 | 84.34 | 16.83 |

Fig. 7. Create Volume Report in Civil3D

4.4 Short Turnaround Cycle for Updating

The utility of Civil 3D visualization is extended beyond one-time analysis. As the project unfolds, subsequent surveys can be rapidly integrated into the existing Civil 3D model.

This integration would allow for continuous monitoring and updating of the earthwork calculations. Repeated cycles of surveying and analysis ensure that the project stays aligned with its as-designed specifications, while allowing for dynamic adjustments and decision-making in project control. This iterative process, empowered by Civil 3D's built-in data handling and analytical capabilities, potentially delivers precision and efficiency in managing large-scale earthwork projects.

4.5 Advancements in Detecting Future Earthwork Scenarios

This research presents a novel approach to enhancing earthwork construction efficiency through detection and simulation of near future operational scenarios. This approach involves the creation of virtual environments that precisely model the operational dynamics of construction equipment, such as trucks and excavators, under a Digital Twins application framework. By utilizing rapid updates of rover data, project managers can predict and adjust the operational states of construction machinery on a near real-time basis, ensuring optimal performance and resource allocation. Key to this innovation will be the ability to explore the “rolling resistance” of temporary haul road with varying terrain types, which directly influences the optimal paths for heavy equipment transit and movement. This not only maximizes fuel efficiency but also minimizes the time spent in material transportation. Furthermore, through the application of precise algorithms, it's possible to determine the optimal load capacity for each vehicle, ensuring that each trip is as efficient as possible without compromising safety and capacity of the machinery. Another significant advancement is in the optimal supply quantity of equipment required on-site. By simulating different project scenarios and equipment configurations, construction planners can accurately forecast the quantity of equipment in a fleet needed to execute the earthwork and meet project deadlines. This not only helps in reducing the direct and indirect costs associated with over or under-supply of equipment but also ensures that projects can adapt to unforeseen circumstances with greater agility.

5 Conclusion

Our innovative rover solution not only extends the capabilities of traditional surveying instruments in the dynamic environment of construction fields but also marks a significant leap in reducing the application cost barriers in connection with precise data acquisition. This advancement directly tackles the critical challenge of high data collection costs, paving the way for widespread adoption of Digital Twins technology in earthwork construction. In comparison with UAV photogrammetry and GPS technologies, our system offers a pragmatic and cost-effective alternative. It empowers on-site personnel with the ability to conduct repeated, accurate surveys without requiring specialized training and expertise, thereby substantially lowering operational costs and democratizing access to advanced surveying technology.

Building on the success and insights gained from our current research, future work will delve into the integration of AI and machine learning algorithms to further enhance the accuracy and efficiency of our field surveying system. We aim to explore autonomous

navigation capabilities for the rover, reducing human intervention and accelerating the data collection process. Furthermore, the system's integration with AutoCAD Civil 3D leads to a potential creation of "Digital Twins"-like models for earthwork projects, featuring near real-time project monitoring and enhanced decision-making capabilities in project planning and control. Moreover, the project has set a foundational base for future technological enhancements in the construction engineering and management sphere. The envisioned expansions include the development of autonomous navigation and self-driving capabilities for the rover, leveraging more sophisticated sensor technology, applying simulation modeling for optimal fleet configuration and operation, and performing condition evaluation of temporary haul roads in earthmoving.

Acknowledgements. This work was supported by National Science and Engineering Council of Canada Alliance Grant Program (Grant number NSERC ALLRP 578484–22). The authors also thank the following engineering students at University of Alberta, Serhii Naumets for facilitating this research, Qiutong Wang and Chenhao Ma for assisting in stakeout rover prototyping design and field testing.

References

1. Cho, S.I., Lim, J.H., Lim, S.B., Yun, H.C.: A study on dem-based automatic calculation of earthwork volume for BIM application. *한국측량학회지* **38**(2), 131–140 (2020)
2. Kazan, E., Usmen, M.A.: Worker safety and injury severity analysis of earthmoving equipment accidents. *J. Safety Res.* **65**, 73–81 (2018)
3. Opoku, D.G.J., Perera, S., Osei-Kyei, R., Rashidi, M.: Digital twin application in the construction industry: a literature review. *J. Build. Eng.* **40**, 102726 (2021)
4. Feng, Y., Wang, J.: GPS RTK performance characteristics and analysis. *Positioning* **1**(13), (2008)
5. Leica Geosystems AG, Leica Viva TS15 Datasheet. https://www.gefos-leica.cz/ftp/Monitoring/TS15_DAT_en.pdf (accessed 28 August 2023)
6. Whitmeyer, S., Dordevic, M.: A new tool for producing 3D orientation symbology for Google Earth. *GSA Today* **33**(3–4), 28–29 (2023)
7. Yakar, M., Yilmaz, H., Mutluoglu, Ö.: Close range photogrammetry and robotic total station in volume calculation. *Int. J. Phys. Sci.* **5**, (2010)
8. Nyquist, H.: Certain topics in telegraph transmission theory. *Trans. Am. Inst. Electr. Eng.* **47**(2), 617–644 (1928)
9. Suarez, J.D., Correa, A.C.: Design of a holonomic robotic system for interior reconstruction using SLAM. In: 2020 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONITI) (pp. 1–5). IEEE (Sep. 2020)
10. Lee, H.K., Jung, K.Y.: Applicability evaluation of earth volume calculation using unmanned aerial images. *Asia Pac. J Multimed. Serv. Converg. Art Humanit. Sociol.* **5**(5), 497–504 (2015)
11. Han, S.M., Park, T.S.: A comparative analysis of the earth-work volume calculation by UAV (mapping-drone) survey. *Korean Soc. Geospatial Inf. Sci.* 115–116 (2018)
12. Hugenholtz, C.H., Walker, J., Brown, O., Myshak, S.: Earthwork volumetrics with an unmanned aerial vehicle and softcopy photogrammetry. *J. Surv. Eng.* **141**(1), 06014003 (2015)
13. Raeva, P.L., Filipova, S.L., Filipov, D.G.: Volume computation of a stockpile—a study case comparing gps and uav measurements in an open pit quarry. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **41**, 999–1004 (2016)

14. Pirti, A.: Performance analysis of the real time kinematic GPS (RTK GPS) technique in a highway project (stake-out). *Surv. Rev.* **39**(303), 43–53 (2007)
15. Visconti, P., Iaia, F., De Fazio, R., Giannoccaro, N.I.: A stake-out prototype system based on GNSS-RTK technology for implementing accurate vehicle reliability and performance tests. *Energies* **14**(16), 4885 (2021)
16. Mao, S., Lu, M., Shen, X., Hermann, U.: Multi-point Concurrent Tracking and surveying in construction field. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 30, p. 1. IAARC Publications (2013)
17. Banzi, M., Shiloh, M.: Getting started with Arduino. Maker Media, Inc. (2022)
18. Wehner, A.: Entwicklung und Erprobung von Systemen zur automatisierten präzisen Punktabsteckung (Doctoral dissertation, Universitätsbibliothek der Universität der Bundeswehr München) (2022)
19. Spektrum: DX5 Pro 5-Channel DSMR Transmitter Only | Spektrum (2023). <https://www.spektrumrc.com/product/dx5-pro-5-channel-dsmr-transmitter-only/SPMR5025.html>. Accessed 29 August 2023
20. Leica Geosystems AG: TPS Getting Started Guide. https://www.gefos-leica.cz/ftp/GPS/Navody/EN_Originally/Viva_GNSS/Viva_TPS_Getting_started.pdf. Accessed 30 August 2023
21. AXIAL ADVENTURE: Axial Customer Support | Axial Adventure (2023). <https://www.axialadventure.com/rc-rig-support/>. Accessed 29 August 2023
22. van den Burg, G.J., Nazábal, A., Sutton, C.: Wrangling messy CSV files by detecting row and type patterns. *Data Min. Knowl. Disc.* **33**(6), 1799–1820 (2019)
23. Mijić, N.: Application of the airborne LIDAR technology on the quarry using AutoCAD Civil 3D software. In: Advanced Technologies, Systems, and Applications III: Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (IAT), vol. 2, pp. 43–51. Springer International Publishing (2019)



Exploring Critical Success Factors for Cloud Data Management in the South African Construction Industry

Wanda Buhle Mpingana¹(✉), Opeoluwa Akinradewo², and Clinton Aigbavboa¹

¹ Cidb Centre for Excellence & Sustainable Human Settlement and Construction Research Centre, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa

wanda101buhle@gmail.com, caigbavboa@uj.ac.za

² Department of Quantity Surveying and Construction Management, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa

Abstract. The construction industry is a project-based sector primarily driven by construction projects that typically generate diverse and enormous data. Therefore, due to the industry's intensive data feature, there is a need for a better data management tool that will ensure efficiency in the sector. Effective data management plays a pivotal role in the achievement of construction project goals, guaranteeing timely access to accurate information by the appropriate construction stakeholders. Hence, the paper seeks to investigate the critical success factors (CSFs) that support the effective implementation of cloud data management (CDM) solutions within the South African construction industry (SACI). The study employed a quantitative methodology, utilizing a carefully designed close-ended questionnaire that was disseminated among construction professionals and IT experts with construction backgrounds based in the Gauteng province, in South Africa (SA), and 104 individual responses were received. The data were analysed using descriptive statistical procedures and principal component analysis (PCA) as part of the exploratory factor analysis via the computer software SPSS. The study disclosed the multifaceted set of CSFs that substantially impact the efficient implementation of CDM, which include awareness of cloud services, top management commitment, and assurance of privacy. With factor analysis, these CSFs were lastly grouped into four clusters, namely "Efficiency and flexibility", "Organizational and external", "Usability," and "Organizational scale" factors. This shows the level of influence that construction companies can have in driving the adoption of CDM services in the sector. Thus, by understanding and implementing these CSFs, construction companies can enhance their data management capabilities, leading to improved project efficiency and overall competitiveness.

Keywords: Cloud Data Management · Critical Success Factors · Data Management · Efficiency and South African Construction Industry

1 Introduction

The construction industry is a project-based sector primarily driven by construction projects that typically generate diverse and enormous data. As construction projects progress, diverse data is regularly generated, making the construction sector a data-intensive industry (Bello et al. 2021). Likewise, Anumba et al. (2008: 78) concurred that construction projects are very information-intensive and involve a wide range of stakeholders. As a result, the data in the industry is ordinarily decentralized into compartmentalized storage locations such as personal computers, smartphones, team servers, desktops, and so on (Bello et al. 2021). Therefore, due to the built environment's intensive data feature, there is a need for a better data management tool that will ensure efficiency in the sector. Cloud services can be the data management tool that can be utilized instead of the expensive installation, operating, and maintenance costs that come with traditional IT computing facilities. With these current challenges of data management in the construction industry, Almaatouk et al. (2016: 5) found that cloud computing brings about adequate security and lower costs to data storage.

The concept of cloud computing began to mature in the late 1990s with the introduction of remotely provisioned services by Salesforce.com, followed by the introduction of Amazon Web Services (AWS) platform in 2002, and commercially emerged in 2006 with the launch of Amazon's Elastic Compute Cloud (EC2) and Google's provision of cloud of Google Apps (Alami 2016: 3). A well-known definition of cloud computing by the National Institute of Standard and Technology (NIST) describes the term as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Mell and Grance 2011: 6). Whereas Yogitha Lakshmi et al. (2019: 61) perceive cloud computing as a utility service that provides computing services which are software, servers, networking, database storage and virtualization over the internet on the cloud. Furthermore, Gangwar et al. (2015: 107) define cloud computing as digitized services provided via the internet in the form of infrastructure, platforms, and software to its end-users. The collaborative services offered by cloud computing enable the processing and sharing of large amounts of data, inter-application connectivity, and modular interaction (Chong et al., 2014: 153). The exchange of project information in the cloud can notably enhance a project's quality, cost, and performance, which, therefore, fulfills the client's requirements (Redmond et al. 2012: 177). This results in improved project delivery through the efficient information flow management system provided in cloud computing (Chong et al. 2014: 153).

There have been a lot of studies focusing on how cloud computing can bring about sustainable development in the construction sector (Chween et al. 2018; Chen et al. 2021). While some studies have looked at cloud computing advantages, quite a small number of studies have looked at success factors for the usage of cloud computing in developing countries (Waqar et al. 2023). There has been little research on the usage of cloud computing in the construction industry generally (Merschbrock and Munkvold 2015; Mandičák et al. 2016). Hence, this study intends to fill this gap by uncovering and appraising the CSFs that adopt cloud computing usage for the data management of construction projects in South Africa (SA). Thus, the study's research objective is:

RO1: To determine the critical success factors of cloud data management in the South African construction industry.

This study aims to create awareness among construction professionals and contractors about the benefits of using cloud computing for data management and the required factors they need to have in place to successfully implement cloud data management usage in SA. By uncovering and appraising the CSFs of adopting cloud computing usage, the study will provide a roadmap for improving the efficiency of construction projects, resulting in an enhanced project delivery through an efficient information flow management system.

The rest of the study is organized as follows. Section 2 outlines the cloud data management and CSFs of cloud data management usage in the body of literature. Section 3 presents the research methodology process. Section 4 displays the results of the data analysis. Section 5 presents the discussion and implications, and Sect. 6 addresses the conclusions and recommendations of this study.

2 Literature Review

2.1 Cloud Data Management

Cloud data management is all about the management of data in the cloud, which involves processing, analyzing, and storing data, making it an effective method for long-term archiving since resources can be purchased on demand (Sikwela 2020). Therefore, cloud data management refers to the use of cloud computing technology to manage, store, and process data. So, it involves the use of cloud-based platforms and services to store and manage data, applications, and infrastructure. However, in the construction industry, data management is critical to the success of construction projects. The efficacy of such data management guarantees that the right information is available to the right people at the right time. Cloud data management is a modern approach to data management that offers many benefits over traditional on-premises data management. The built environment sector is one of the world's largest and most complex industries. It involves many stakeholders, including architects, engineers, contractors, subcontractors, suppliers, and clients. The industry generates a vast amount of data, including project plans, designs, specifications, contracts, schedules, budgets, and invoices (Shash and Almufadhi 2021: 170). Managing this data is critical to the success of construction projects. Additionally, the prominent problems of cloud data management are the analytical processing of data in the cloud where database information is extracted for decision making, online query execution, and failure of parallel processing of data execution, which involves multiple parallel processing of data failing to divide the workload (Yogitha Lakshmi et al. 2019: 61). Oke et al. (2021) argue that construction stakeholders can store and retrieve development data in real-time by collaborating on cloud data management technologies since it has online storage with self-service capabilities.

2.2 Critical Success Factors of Cloud Data Management Usage in the SACI

This section highlights and examines the various CFSs of cloud data management usage in the South African construction industry based on the existing literature. These various CSFs are coded as shown in Table 1.

Table 1. Critical success factors and their codes

| Code | Critical success factors | References |
|-------|---|---|
| CSF1 | Top management commitment | Raut et al. (2017), Adendorff and Smuts (2019: 8); Okour (2022: 365) |
| CSF2 | Government support | Alharthi et al. (2017: 669); Prasad et al. (2018: 439); AL-Shboul (2019: 891); Aghimien et al. (2022:5) |
| CSF3 | Availability of regulations for cloud usage | Alreshidi et al. (2018); Sahanto et al. (2019) |
| CSF4 | Organizational competency | Tongsuksai et al. (2019: 5); Hentschel et al. (2019: 7345); Eze et al. (2022: 30); Okour (2022: 365) |
| CSF5 | Organizational readiness | Gangwar et al. (2015: 113); Raut et al. (2017: 132); Prasad et al. (2018: 441); Aghimien et al. (2022: 4) |
| CSF6 | Size of organization | Saedi & Iahad (2013: 54); Mohd Asfahani (2020); Raut et al. (2017: 131) |
| CSF7 | Reliability of cloud services | Alharthi et al. (2017: 667); Raut et al. (2017: 130); Tongsuksai et al. (2019: 5) |
| CSF8 | Assurance of data security | Raut et al. (2017: 132); Hentschel et al. (2019: 7344); Oke et al. (2021) |
| CSF9 | Assurance of privacy | Raut et al. (2017: 132); Hentschel et al. (2019: 7344); Oke et al. (2021) |
| CSF10 | Awareness of cloud services | Alharthi et al. (2017: 669); Menta (2022: 830) |
| CSF11 | Comprehension of cloud data management services | Alharthi et al. (2017: 669); Menta (2022: 830) |
| CSF12 | Perceived ease of use | Senk (2013); Gangwar et al. (2015); (Raut et al., 2017: 132); Tongsuksai et al. (2019:6) |
| CSF13 | Perceived usefulness | Senk (2013); Gangwar et al. (2015); Raut et al. (2017: 132); Tongsuksai et al. (2019: 6) |
| CSF14 | Customizable service level agreement | Otuka et al. (2014: 8); Alharthi et al. (2017: 669); Raut et al. (2017: 132) |
| CSF15 | Sufficient internet bandwidth and performance | Ramagoffu (2012); Alharthi et al. (2017: 668) |
| CSF16 | Low latency in the cloud | Alharthi et al. (2017: 668); Adendorff and Smuts (2019: 8); Gillis (2020); Yasar (2022) |
| CSF17 | Elasticity of cloud services | Tao et al. (2011: 2050); Chong et al. (2014: 24); Raut et al. (2017: 131); Adendorff and Smuts (2019: 3) |

(continued)

Table 1. (*continued*)

| Code | Critical success factors | References |
|-------|---|--|
| CSF18 | Transparency in data management of the cloud | Hentschel et al. (2019: 7345); Adendorff and Smuts (2019: 7); Aghimien et al. (2022: 4) |
| CSF19 | Realtime access of the cloud | Tongsuksai et al. (2019: 5); Adendorff and Smuts (2019) |
| CSF20 | Interoperability of cloud data management services | Gangwar et al. (2015: 113); Alharthi et al. (2017: 667); Hentschel et al. (2019: 7345); Mohd Asfahani (2020) |
| CSF21 | Compatibility with organization's IT infrastructure | Gangwar et al. (2015: 113); Alharthi et al. (2017: 667); Hentschel et al. (2019: 7345); Mohd Asfahani (2020) |

CSF1, CSF2 & CSF3: Adendorff and Smuts (2019: 8), Raut et al. (2017), and Okour (2022: 365) highlight that top management support attributed to the management's commitment through allocating resources is crucial for the successful adoption of cloud computing technology in organizations, influencing organizational vision and improving employee morale and motivation. Moreover, government support has a significant role in facilitating the adoption of cloud data management services in the construction industry via policies and regulations that will give regulatory clarity, user protection and foster the industry's growth, especially in developing countries (Alharthi et al. 2017: 669; Prasad et al. 2018: 439; AL-Shboul 2019: 891; Aghimien et al. 2022: 5). Furthermore, Alreshidi et al. (2018) and Sahanto (2019) highlight the presence of regulations for cloud computing usage as necessary for safeguarding the sensitive data and ensuring compliance in the interaction between cloud service providers (CSPs) and construction companies, thereby making sure that cloud data management usage is sustainable and it addresses issues like collaboration and data access in the cloud.

CF4, CSF5, CSF6: Studies by Tongsuksai et al. (2019: 5), Hentschel et al. (2019: 7345), Eze et al. (2022: 30) and Okour (2022: 365) postulate that organizational competency attributed by corporate culture, IT expertise, and resource availability is critical for the effective adoption of cloud data management services in organizations. Whereas organizational readiness is characterized by management commitment, organizational governance, technological and financial preparedness, and executive understanding of cloud computing, which drive the readiness for cloud computing implementation (Gangwar et al. 2015: 113; Raut et al. 2017: 132; Prasad et al. 2018: 441; Aghimien et al. 2022: 4). Moreover, the size and the nature of an organization play a crucial role in the implementation of cloud data management services, as larger firms with more resources like skilled IT staff and advanced technology have an advantage in managing associated risks (Mohd Asfahani 2020). However, Raut et al. (2017: 131) indicate that smaller firms have the advantage of flexibility and adaptability, enabling them to implement innovative practices in response to competitive changes while achieving effective coordination.

CSF7, CSF8, CSF9, CSF10 & CSF11: Cloud data management services must be reliable for construction companies to trust the CSPs, and this is characterized by consistent delivery of promised results, real-time data provision, continuous availability, redundant systems for failure protection, and 24/7 accessibility, as exemplified by services like Gmail, which are significantly more dependable than traditional email systems (Alharthi et al. 2017: 667; Raut et al. 2017: 130; Tongsuksai et al. 2019: 5). CSPs need to assure data security and privacy by ensuring that they have data provision and robust security measures which are also governed by service-level agreements between the user and the CSP and the country-specific regulations to facilitate the cloud computing adoption (Raut et al. 2017: 132; Hentschel et al. 2019: 7344; Oke et al. 2021). Lastly, Alharthi et al. (2017: 669) and Menta (2022: 830) emphasize that for construction companies to adopt cloud data management services effectively, it is essential to improve organizational awareness and understanding of cloud computing through education, training the construction professionals and hiring knowledgeable IT personnel, integrating awareness as a key part of the implementation process.

CSF12, CSF13, CSF14 & CSF15: Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) are the vital factors influencing the implementation of cloud computing in organizations, with PEOU relating to users' perception of the accessibility, learnability, and usability of the technology while PU relates to users' perception of how cloud computing can improve job efficiency, performance, and productivity, both crucial factors for successful integration of cloud data management services in construction companies (Senk 2013; Gangwar et al. 2015; Raut et al. 2017: 132; Tongsuksai et al. 2019: 6). Furthermore, Otuka et al. (2014: 8) and Raut et al. (2017: 132) postulate that customizable service level agreements (SLAs) between CSPs and end users can improve user control over data and its availability, with government regulations potentially standardizing SLA requirements in cloud data management services. Alharthi et al. (2017: 669) describe SLAs as crucial contractual documents between cloud service users and providers, outlining uptime, performance, and user experience expectations and offering remedies for service failures. Likewise, sufficient internet bandwidth and performance are crucial for the vital implementation of CDM services since they ensure that construction stakeholders experience optimal cloud-based services without disruptions, as indicated by Ramagoffu (2012) and Alharthi et al. (2017: 668).

CSF16, CSF17 & CSF18: Yasar (2022) argues that low latency has a positive user experience in the adoption of cloud computing since there is high internet bandwidth compared to high latency. According to the studies by Adendorff and Smuts (2019: 8), the adoption of cloud computing in South Africa is hindered by high latency caused by last-mile connectivity and expensive broadband. Also, Raut et al. (2017: 131) and Adendorff and Smuts (2019: 3) indicate that the rapid elasticity of cloud services is a crucial factor in adopting cloud data management services in the construction industry, as companies will adopt this technology due to this feature. According to Adendorff and Smuts (2019: 7), transparency of data management in the cloud is vital as it shows how the security is managed in the cloud to the client so that users can have trust in the CSP.

CSF19, CSF20, CSF21: Tongsuksai et al. (2019: 5) and Adendorff and Smuts (2019) emphasize the importance of realtime access to the cloud as a driving factor for cloud implementation since users safely and remotely access data anytime and share

it among construction stakeholders with optimized data retrieval process reducing time and resource costs. To sum up, Alharthi et al. (2017: 667), Hentschel et al. (2019: 7345), and Mohd Asfahani (2020) postulate that to avoid vendor lock-in and ensure successful CDM adoption in construction, CSPs must provide interoperable solutions across various providers, facilitate easy migration of services, and ensure compatibility with existing IT infrastructures, addressing potential errors for seamless integration.

3 Research Methodology

To investigate the CFSs for cloud data management usage in the SACI, this study adopted a quantitative research methodology. This was achieved via a closed, well-structured questionnaire as they were conveniently coded for analysis in SPSS (Arowooya et al. 2020). The quantitative research approach was chosen due to its capacity to make credible inferences with an unbiased examination of the statistical data (Creswell and Creswell 2017). The questionnaire was sent out to all registered construction professionals in the SACI, which consisted of the Architects, Quantity Surveyors, Construction Project Managers, Construction Managers, Civil, Mechanical, and Electrical Engineers, as well as the IT experts with construction background either via a qualification or experience in the field. This study was conducted in the Gauteng province of South Africa. Gauteng was chosen because the researcher was familiar with the province, and it was easy to access the target area. The samples were selected from the five municipalities in the province: the City of Ekurhuleni Metropolitan, the City of Johannesburg Metropolitan, the City of Tshwane Metropolitan, Sedibeng, and West Rand district. For this research, simple random sampling was used to get a good sample size. A sample size of 104 research participants was achieved out of approximately 33 206 registered construction professionals in the Gauteng province in the SACI according to the 2022–2023 councils report (ECSA 2023; SACAP 2023; SACPCMP 2023; SACQS 2023). The first section of the questionnaire obtained the biographical information of the research participants, while the second section got the ranking of the CSFs from the respondents' view using the agreement likert scale. The data analysis was performed using the Statistical Package for the Social Sciences (SPSS) version 29 to give the mean item score, standard deviation and exploratory factor analysis of the variables.

4 Results

4.1 Respondents' Demographic Information

The demographic data obtained was related to the respondent's educational qualifications, profession, years of experience, number of construction projects involved in, and the nature of the construction organizations they belong to. In terms of educational qualifications, 47.1% (49) had an honours degree, 23.1% (24) had a bachelor's degree and 22.1% (23) with a master's degree. The respondents with diplomas sat at 4.8% (5), and those with doctoral degrees sat at 2.9% (3). This means the construction professionals who participated in this research are educated enough to assist on the subject matter. In terms of their respective professions, 39.4% (41) were quantity surveyors, 23.1% (14)

were construction project managers, 13.5% (14) were construction managers, architects were sitting at 7.7% (8), while Structural / Civil Engineer and Mechanical Engineer were both sitting at 5.8% (6), 3.8% (4) were electrical engineers and 1% (1) was consisting of an IT expert. In terms of their years of experience, 45.2% were respondents with years of experience of 1–5 years, 22.1% were research participants with years of experience of 6–10 years, 12.5% were research participants with years of experience of 11–15 years, 8.7% were respondents with years of experience of less than 12 months and respondents with about 16–20 years and more than 20 years of experience were both 5.8% respectively.

4.2 Critical Success Factors of Cloud Data Management Usage in the SACI

A summary of the findings from the CSFs' mean ranking analysis is also shown in Table 2. Following a similar approach by Aliu et al. (2023), CSFs were deemed significant if they reached a threshold of 3.5. Nevertheless, if the mean score is less than 3.5, it is deemed negligible. All 21 CSFs possessed mean values above 3.5, thus, they were considered critical to the research. Based on mean rankings, the top five CSFs were CSF10 & CSF1 (Awareness of cloud services with $SD = 0.673$ and Top management commitment with $SD = 0.713$, both sharing a $MIS = 4.38$), CSF9 (Assurance of privacy with $MIS = 4.34$; $SD = 0.677$), CSF5 (Organizational readiness with $MIS = 4.30$; $SD = 0.774$), CSF4 (Organizational competency with $MIS = 4.28$; $SD = 0.830$), CSF21 (Compatibility with organization's IT infrastructure was ranked sixth with $MIS = 4.23$; $SD = 0.815$). Due to the sample size being less than 2000, the Shapiro-Wilk normality test was used to assess the data's normality and ascertain if the data was parametric or nonparametric (Aliu et al. 2023). All of the evaluated CSFs had a significant value of 0.001, meaning that the data did not follow a normal distribution as the p-value is less than 0.05.

4.3 Results of the Factor Analysis

Before reorganizing the identified CSFs into more logical sections that reflect relationships among multiple sets of related factors, factor analysis was used to determine the links between the variables. The Bartlett's test of sphericity (BTS) and the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy were used to assess the suitability and appropriateness for factor analysis of the collected data. For this study, a KMO value of 0.895 was recorded, meaning that 89.5% of the data obtained was acceptable for factor analysis. As indicated in Table 3, the BTS in this study achieved statistical significance and produced a high chi-squared value of 1416.418 at 210 degrees of freedom. This proves that the correlation matrix is not an identity matrix, and factor analysis is appropriate. Factor analysis was conducted using the principal component analysis (PCA) with direct oblimin with Kaiser Normalization. Table 3 also indicates that all factor loadings exceeded 0.30, meaning that all variables are considered for this study. The PCA results show four rotated components with eigenvalues above 1, with a cumulative variance of 66.11%. The four components were named to reflect the variables' traits for simplicity and interpretability purposes (Aliu et al. 2023).

Table 2. Mean rankings of critical success factors

| Critical success factors | Mean (MIS) | Std. Deviation (SD) | Rank | Normality (Shapiro-Wilk) | | |
|---|------------|---------------------|------|--------------------------|-----|-------|
| | | | | Statistic | Df | Sig |
| Awareness of cloud services (CSF10) | 4,38 | 0,673 | 1 | 0,757 | 104 | 0,001 |
| Top management commitment (CSF1) | 4,38 | 0,713 | 1 | 0,737 | 104 | 0,001 |
| Assurance of Privacy (CSF9) | 4,34 | 0,677 | 3 | 0,754 | 104 | 0,001 |
| Organizational readiness (CSF5) | 4,30 | 0,774 | 4 | 0,756 | 104 | 0,001 |
| Organizational competency (CSF4) | 4,28 | 0,830 | 5 | 0,776 | 104 | 0,001 |
| Compatibility with organization's IT infrastructure (CSF21) | 4,23 | 0,815 | 6 | 0,796 | 104 | 0,001 |
| Assurance of data security (CSF8) | 4,22 | 0,775 | 7 | 0,778 | 104 | 0,001 |
| Perceived usefulness (CSF13) | 4,22 | 0,788 | 7 | 0,806 | 104 | 0,001 |
| Realtime access of the cloud (CSF19) | 4,18 | 0,773 | 9 | 0,810 | 104 | 0,001 |
| Reliability of cloud services (CSF7) | 4,17 | 0,769 | 10 | 0,811 | 104 | 0,001 |
| Perceived ease of use (CSF12) | 4,15 | 0,879 | 11 | 0,810 | 104 | 0,001 |
| Sufficient internet bandwidth and performance (CSF15) | 4,15 | 0,901 | 11 | 0,806 | 104 | 0,001 |
| Comprehension of cloud data management services (CSF11) | 4,13 | 0,878 | 13 | 0,796 | 104 | 0,001 |
| Availability of regulations for cloud usage (CSF3) | 4,11 | 0,775 | 14 | 0,815 | 104 | 0,001 |

(continued)

Table 2. (*continued*)

| Critical success factors | Mean (MIS) | Std. Deviation (SD) | Rank | Normality (Shapiro-Wilk) | | |
|--|------------|---------------------|------|--------------------------|-----|-------|
| | | | | Statistic | Df | Sig |
| Interoperability of cloud data management services (CSF20) | 4,08 | 0,797 | 15 | 0,834 | 104 | 0,001 |
| Customizable Service Level Agreements (CSF14) | 4,06 | 0,933 | 16 | 0,826 | 104 | 0,001 |
| Transparency in data management of the cloud (CSF18) | 4,05 | 0,840 | 17 | 0,829 | 104 | 0,001 |
| Government support (CSF2) | 3,95 | 1,109 | 18 | 0,826 | 104 | 0,001 |
| Size of the organization (CSF6) | 3,89 | 1,051 | 19 | 0,840 | 104 | 0,001 |
| Low latency in the cloud (CSF16) | 3,86 | 0,908 | 20 | 0,868 | 104 | 0,001 |
| Elasticity of cloud services (CSF17) | 3,85 | 0,890 | 21 | 0,866 | 104 | 0,001 |

5 Discussion

5.1 Cluster 1: Efficiency and Flexibility Factors

A sum of seven critical success factors loaded onto this cluster and they are ‘Low latency in the cloud’ (89.2%), ‘Elasticity of cloud services’ (82%), ‘Transparency in data management of the cloud’ (74.7%), ‘Realtime access of the cloud’ (65%), ‘Interoperability of cloud data management services’ (56.9%), ‘Customizable Service Level Agreements’ (48.9%) and ‘Awareness of cloud services’ (41%). This cluster accounted for 48.5% of the total variance, and these factors relate to the efficiency and flexibility that cloud data management services offer construction companies. Prominent CFSs that bring out a flexibility feature in cloud data management services are low latency in the cloud services, which results in adequate internet bandwidth coupled with the elasticity of the cloud services, allowing construction companies to scale up or down on the resources they need and the real-time accessibility of the stored data in the cloud (Alharthi et al. 2017: 668; Raut et al. 2017: 131; Adendorff and Smuts 2019: 8). The openness of CSPs about the management of the data in the cloud plays a substantial role in building trust with users, which in turn can bring about efficiency in the data management of construction projects (Adendorff and Smuts 2019: 7). Since construction projects normally a short span, interoperability of cloud data management (CDM) services will result in efficient management of data in construction organizations (Alharthi et al. 2017: 667).

Table 3. Summary of factor analysis

| Groupings | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|---|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 10,175 | 48,454 | 48,454 | 10,175 | 48,454 | 48,454 |
| 2 | 1,354 | 6,447 | 54,900 | 1,354 | 6,447 | 54,900 |
| 3 | 1,241 | 5,907 | 60,808 | 1,241 | 5,907 | 60,808 |
| 4 | 1,113 | 5,298 | 66,106 | 1,113 | 5,298 | 66,106 |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | | | | | 0,895 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | | | 1416,418 | | |
| | | | | Df | 210 | |
| | | | | Sig | <,001 | |
| Critical Success factors of Cloud Data Management Usage | Component | | | | | |
| | 1 | 2 | 3 | 4 | | |
| <i>Cluster 1: Efficiency and flexibility factors</i> | - | | | | | |
| Low latency in the cloud | 0,892 | | | | | |
| Elasticity of cloud services | 0,820 | | | | | |
| Transparency in data management of the cloud | 0,747 | | | | | |
| Realtime access of the cloud | 0,650 | | | | | |
| Interoperability of cloud data management services | 0,569 | | | | | |
| Customizable Service Level Agreements | 0,489 | | | | | |
| Awareness of cloud services | 0,410 | | | | | |
| <i>Cluster 2: Organizational and external factors</i> | - | | | | | |
| Top management commitment | | 0,747 | | | | |
| Organizational competency | | 0,706 | | | | |
| Government support | | 0,703 | | | | |
| Availability of regulations for cloud usage | | 0,587 | | | | |
| Organizational readiness | | 0,563 | | | | |
| <i>Cluster 3: Usability factors</i> | - | | | | | |

(continued)

Table 3. (*continued*)

| Groupings | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|---|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| Sufficient internet bandwidth and performance | | | | 0,779 | | |
| Perceived usefulness | | | | 0,756 | | |
| Assurance of Privacy | | | | 0,688 | | |
| Reliability of cloud services | | | | 0,555 | | |
| Compatibility with organization's IT infrastructure | | | | 0,512 | | |
| Perceived ease of use | | | | 0,488 | | |
| Assurance of data security | | | | 0,475 | | |
| Comprehension of cloud data management services | | | | 0,437 | | |
| <i>Cluster 4: Organizational scale factor</i> | - | | | | | |
| Size of the organization | | | | | 0,880 | |

The awareness of construction professionals about the CDM services will result in fruitful customizable SLAs between CSPs and users (Alharthi et al. 2017: 669; Raut et al. 2017: 132). Promoting the factors in this cluster for the widespread use of CDM services would involve awareness campaigns through seminars between industry professionals and research institutions. Moreover, Case studies showing the efficiency brought by cloud computing would encourage widespread use of the technology.

5.2 Cluster 2: Organizational and External Factors

This cluster accounted for 6.4% of the total variance and contains five extracted critical success factors. These include ‘Top management commitment’ (74.7%), ‘Organizational competency’ (70.6%), ‘Government support’ (70.3%), ‘Availability of regulations for cloud usage’ (58.7%) and ‘Organizational readiness’ (56.3%). These findings relate to the organizational factors that construction companies would need to ensure that they possess and external (government) factors for the effective usage of cloud computing in the construction industry. According to Adendorff and Smuts (2019: 8), top management support is a substantial CSF for the adoption of CDM services in the construction sector. The commitment is influenced by the presence of positive relationships between organizational structures which may come in the form of resources being allocated for the implementation process and motivation for increasing employee morale (Raut et al. 2017: 132; Adendorff and Smuts 2019: 8; Okour 2022: 365). Furthermore, Prasad et al. (2018: 441) opined that organizational executive readiness is a significant factor driven by the organizations understanding of CDM services, which shows the competency of

the companies to adopt the services. External factors such as government commitment and regulations that govern the usage of CDM services also play a crucial factor in the adoption of CDM services. Moreover, Prasad et al. (2018: 439) identified regulatory clarity as the biggest player in expediting the adoption of blockchain-based CDM technology by reducing uncertainty and ensuring user protection. Government support also encompasses regulations, initiatives, policies, and instructions designed to provide backing and assistance to the organization in using CDM services (Alharthi et al. 2017: 669; Alreshidi et al. 2018; Sahanto 2019; AL-Shboul 2019: 891). Therefore, having developed sector standards with clear guidelines for implementing the technology from the built environment companies' and government's point of view would encourage the widespread adoption of CDM in SA.

5.3 Cluster 3: Usability Factors

A total of eight extracted critical success factors loaded onto this cluster and they are 'Sufficient internet bandwidth and performance' (77.9%), 'Perceived usefulness' (75.6%), 'Assurance of Privacy' (68.8%), 'Reliability of cloud services' (55.5%), 'Compatibility with organization's IT infrastructure' (51.2%), 'Perceived ease of use' (48.8%), 'Assurance of data security' (47.5%) and 'Comprehension of cloud data management services' (43.7%). This cluster accounted for 5.9% of the total variance, and these factors relate to the usability characteristics needed for the effective adoption of cloud computing in the built environment. In a study conducted by Ramagoffu (2012), sufficient internet bandwidth is a crucial usability characteristic of CDM services as cloud services will not be interrupted by any issues or glitches. Furthermore, the TAM framework consisting of PEOU and PU is a crucial CSF as construction companies will understand the ease of use and useful of CDM services (Gangwar et al. 2015: 112; Raut et al. 2017: 132). Likewise, the assurance of data security and privacy by CSPs also increases the usability and trust of CDM services by construction companies as their crucial data will be protected in the cloud (Raut et al. 2017: 132; Hentschel et al. 2019: 7344). Studies by Alharthi et al. (2017: 667), Hentschel et al. (2019: 7345), and Mohd Asfahani (2020), outline the reliability in the cloud services and the compatibility of these cloud services with an organization's IT infrastructure as crucial usability factors that bring about seamless integration with existing IT systems and a continuous availability of data in the cloud with redundant systems in case of failure.

5.4 Cluster 4: Organizational Scale Factor

This cluster accounted for 5.3% of the total variance and contains only one extracted critical success factor. The factor is 'Size of the organization' (88%). This variable does not correlate substantially with other variables, and it qualifies to be a single variable factor. It relates to the organizational scale that will influence the usage of cloud data management services in the construction industry. The size of the organization determines the available resources supporting the effective implementation of CDM services in the construction sector, where larger firms typically have IT support staff with experience and skills in cloud-based technology that will assist the organization with the implementation (Saedi and Iahad 2013: 54; Raut et al. 2017: 131; Mohd Asfahani 2020). The use

of pilot projects showcasing the benefits of using CDM by larger construction companies can help boost the widespread use of CDM in the SA built environment sector.

6 Conclusion and Recommendations

Cloud data management services have the potential to revolutionize the data management of construction projects, increasing the efficiency of the built environment sector. A holistic adoption approach for cloud computing technology that includes technological, regulatory, and human capacity development factors would ensure the widespread use of the technology. Thus, this study investigated the CSFs of cloud data management usage in the South African built environment for the efficient data management of construction projects by construction professionals in a developing economy like South Africa. In order to accomplish this goal, this study used quantitative research methodology and a survey questionnaire to gather data from construction industry professionals on 21 CSFs that were taken from the body of knowledge. From the data analysis, this paper revealed four main clusters of CSFs: efficiency and flexibility factors, organizational and external factors, usability factors, and organizational scale factors, which can drive the adoption of cloud data management services in the SACI. Therefore, it is evident that CDM's technical aspects are crucial for the operational efficiency of cloud services, which are driven by the leadership and strategic direction provided by construction companies and CSPs. The user experience and trust are crucial for successfully implementing CDM services. The research limitation of this paper is the primary focus on the CSFs of cloud data management usage in the context of the South African construction industry. Further studies can be carried out on the role of training and education about CDM services to increase tech-savviness among construction professionals and the inclusion of cloud service developers in identifying the driving factors of implementing CDM services in the construction industry.

References

- Adendorff, R., Smuts, H.: Critical success factors for cloud computing adoption in South Africa. Twenty-Fifth Am. Conf. Inf. Syst., Cancun **2019**, 1–10 (2019)
- Aghimien, D., Aigbavboa, C.O., Chan, D.W., Aghimien, E.I.: Determinants of cloud computing deployment in South African construction organisations using structural equation modelling and machine learning technique. Eng., Constr. Arch. Manag. (2022)
- Alani, M.M.: What is the Cloud? In Elements of Cloud Computing Security, pp. 1–14. Springer International Publishing (2016)
- Alharthi, A., Allassafi, M.O., Walters, R.J., Wills, G.B.: An exploratory study for investigating the critical success factors for cloud migration in the saudi arabian higher education context. Telematics Inform. **34**(2), 664–678 (2017)
- Aliu, J., Oke, A.E., Abayomi, T., Aigbavboa, C., Makanjuola, S.: Exploring the critical success factors for adopting gamification in the Nigerian construction sector. Built Environ. Proj. Asset Manag. (2023)
- Almaatouk, Q., Othman, M.S.B., Al-khazraji, A.: A review on the potential of cloud-based collaboration in construction industry. In: 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), pp. 1–5. IEEE (2016)

- Alreshidi, E., Mourshed, M., Rezgui, Y.: Requirements for cloud-based BIM governance solutions to facilitate team collaboration in construction projects. *Requirements Eng.* **23**, 1–31 (2018)
- AL-Shboul, M.A.: Towards better understanding of determinants logistical factors in SMEs for cloud ERP adoption in developing economies. *Bus. Process. Manag. J.* **25**(5), 887–907 (2019)
- Anumba, C.J., Pan, J., Issa, R.R.A., Mutis, I.: Collaborative project information management in a semantic web environment. *Eng. Constr. Archit. Manag.* **15**(1), 78–94 (2008)
- Arowooya, V.A., Oke, A.E., Aigbagboea, C.O., Aliu, J.: An appraisal of the adoption internet of things (IoT) elements for sustainable construction. *J. Eng., Des. Technol.* **18**(5), 1193–1208 (2020). <https://doi.org/10.1108/JEDT-10-2019-0270>
- Bello, S.A., et al.: Cloud computing in construction industry: Use cases, benefits and challenges. *Autom. Constr.* **122**, 103441 (2021)
- Chen, Y., Yao, D., Duan, Y.: Complexity of the analysis of financial cloud based on fuzzy theory in the wisdom of sustainable urban development. *Complexity* **2021**, 1–15 (2021)
- Chong, H.Y., Wong, J.S., Wang, X.: An explanatory case study on cloud computing applications in the built environment. *Autom. Constr.* **44**, 152–162 (2014)
- Chween, L.X., Yahya, M.Y., Yassin, A.M.: Challenges of implementation of cloud computing technology among construction participants. *Adv. Sci. Lett.* **24**(5), 3300–3305 (2018)
- Creswell, J.W., Creswell, J.D.: Research design: qualitative, quantitative, and mixed methods approaches. Sage publications (2017)
- Engineering Council of South Africa.: ECSA Annual Report 2022–2023 (2023). https://www.ecsa.co.za/about/pdfs/2022%202023%20ECSA_Annual%20Report%20Fin.pdf [Accessed: 21 Jan. 2024]
- Eze, F.O., Mbah, P.C., Oboko, W.C.: Organisational competency and the performance of academic staff of state tertiary institutions in enugu state. *Adv. J. Curr. Res.* **7**(7), 27–48 (2022)
- Gangwar, H., Date, H., Ramaswamy, R.: Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. *J. Enterp. Inf. Manag.* **28**(1), 107–130 (2015)
- Gillis, A.S.: What is Latency? (2020). <https://www.techtarget.com/whatis/definition/latency> [Accessed: 17 Jun. 2023]
- Hentschel, R., Leyh, C., Baumhauer, T.: Critical success factors for the implementation and adoption of cloud services in SMEs. Proceedings of the 52nd Hawaii International Conference on System Sciences, pp. 7342–7351 (2019)
- Huzsvai, L., et al.: Analysis of sweet corn nutritional values using multivariate statistical methods. *Acta Agrar. Debr. Iensis* **1**, 103–108 (2021)
- Mandičák, T., Mesároš, P., Kozlovská, M.: Exploitation of cloud computing in management of construction projects in Slovakia. *Organ., Technol. & Manag. Constr.: Int. J.* **8**(1), 1456–1463 (2016)
- Mell, P., Grance, T.: The NIST definition of cloud computing, pp. 1–7 (2011)
- Menta, A.J.C.: Adopting technology environment organization framework in virtual learning through cloud computing of higher education institution. *Int. J. Comput. Sci. Res.* **6**, 822–841 (2022)
- Merschbrock, C., Munkvold, B.E.: Effective digital collaboration in the construction industry—a case study of BIM deployment in a hospital construction project. *Comput. Ind.* **73**, 1–7 (2015)
- Mohd Asfahani, S.: The critical success factors of cloud based application implementation in construction management. Dissertation: University of Malaya (2020)
- Oke, A.E., Farouk Kineber, A., Abdel-Tawab, M., Abubakar, A.S., AlbuKhari, I., Kingsley, C.: Barriers to the implementation of cloud computing for sustainable construction in a developing economy. *Int. J. Build. Pathol. Adapt.* 2398–4708 (2021)
- Okour, S.M.: Critical success factors of cloud enterprise resource planning systems and financial performance: evidence from emerging markets. *J. Governance Regul.* **11**(1), 361–375 (2022)

- Otuka, R., Preston, D., Pimenidis, E.: The use and challenges of cloud computing services in SMEs in Nigeria. In: Proceedings of the European Conference on Information Management, p. 325 (2014)
- Prasad, S., Shankar, R., Gupta, R., Roy, S.: A TISM modeling of critical success factors of blockchain based cloud services. *J. Adv. Manag. Res.* **15**(4), 434–456 (2018)
- Ramagoffu, M.: The impact of network related factors on internet based technology in South Africa: a cloud computing perspective. Dissertation: Gordon Institute of Business Science, University of Pretoria (2012)
- Raut, R.D., Gardas, B.B., Jha, M.K., Priyadarshinee, P.: Examining the critical success factors of cloud computing adoption in the MSMEs by using ISM model. *J. High Technol. Managem. Res.* **28**(2), 125–141 (2017)
- Redmond, A., Hore, A., Alshawi, M., West, R.: Exploring how information exchanges can be enhanced through cloud BIM. *Autom. Constr.* **24**, 175–183 (2012)
- Saedi, A., Iahad, N.A.: An integrated theoretical framework for cloud computing adoption by small and medium-sized enterprises. Pacific Asia Conference on Information Systems 2013 proceedings, pp. 48–60 (2013)
- Senk, C.: Adoption of security as a service. *J. Internet Serv. Appl.* **4**(1), 1–16 (2013)
- Shash, A.A., Almufadhi, S.: Constructability: owners, designers, and contractors practices in industrial projects. *Journal of Eng., Proj. & Prod. Manag.* **11**(3), 169–180 (2021)
- Sikwela, J.G.: Data management considerations in the design of internet of things applications. Dissertations: University of Johannesburg (South Africa) (2020)
- South African Council for the Project and Construction Management Profession: SACPCMP Annual Report 2022–2023 (2023). https://static.pmg.org.za/SACPCMP_Annual_Report_2023.pdf [Accessed: 21 Jan 2024]
- South African Council for the Quantity Surveying Profession: SACQS Annual Report 2022–2023 (2023). https://static.pmg.org.za/SACQSP_Annual_Report_2022-23.pdf [Accessed: 21 Jan 2024]
- South African Council for the Architectural Profession: SACAP Annual Report 2022–2023 (2023). https://www.sacapsa.com/html/SACAP_Annual_Report_2022_2023/index_70.html#page=60 [Accessed: 21 Jan 2024]
- Suhanto, A., Hidayanto, A.N., Naisutu, M., Bowo, W.A., Budi, N.F.A., Phusavat, K.: Hybrid cloud data integration critical success factors: A case study at PT Pos Indonesia. In: 2019 Fourth International Conference on Informatics and Computing (ICIC), pp. 1–6: IEEE (2019)
- Tao, J., Marten, H., Kramer, D., Karl, W.: (2011). An intuitive framework for accessing computing clouds. *Procedia Comput. Sci.* 42049–2057
- Tongsuksai, S., Mathrani, S., Taskin, N.: Cloud enterprise resource planning implementation: a systematic literature review of critical success factors. In: 2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1–8: IEEE (2019)
- Yasar, K.: What are network packets and how do they work? (2022). <https://www.techtarget.com/searchnetworking/definition/packet> [Accessed: 17 Jun. 2023]
- Yogitha Lakshmi, K., Dhanalakshmi, S., Obula Reddy, B.G.: An overview of data management in cloud computing. *Int. J. Recent. Technol. Eng.* **7**(5C), 61–64 (2019)

Author Index

A

- Abellanosa, Abbey Dale 455
AbouRizk, Simaan 455
Adekunle, Samuel 509
Adimi, Neda 252
Aghimien, Douglas 509
Aigbavboa, Clinton 429, 509, 652
Akcay, Emre Caner 361
Akinradewo, Opeoluwa 429, 509, 652
Al Nama, Entesar 294
Ali, Sherief 159
Aliu, John 509
Aminbakhsh, Saman 361
Anugwo, Iruka C. 585
Anugwo, Jachike C. 585
Aquino, Enrique A. Juarez 571
Arcolezi, Karina Hwang 489
Asgari, Sadegh 205, 473
Auflič, Mateja Jemec 498
Azar, Elie 169

B

- Bagasi, Omar 562
Bai, Ruihan 325
Basson, Jean-Pierre 520, 533
Béland, Laurent Karim 275
Berger, Markus 15
Block, Marlena 398
Boya, Ayanda 429
Budiman, Jonathan Aloysius 217
Buttgereit, Alexander 398

C

- Caron, Jean-François 421
Cauti, Luis L. Espinoza 571
Chang, Gary 238
Chang, Luh-Maan 264
Chen, Daoxiang 464
Chen, Po-Han 264, 283

- Chen, Yufan 325

D

- Díaz, Joaquín 181, 443
Disser, Michael 92
Divakaran, Anagha 169
Dragos, Kosmas 337, 347
Dujc, Jaka 498
Duski, Husan 107

E

- Eder, Friedrich 54
Elsafty, Abdullah 408

F

- FatehiJananloo, Marjan 311
Filardo, Martina Mellenthin 148
Flood, Ian 133
Fonsati, Arianna 64

G

- García de Soto, Borja 76
Gehring, Maximilian 92
Georg, Olm 191
Glunz, Matthias W. 148
Golabchi, Hamidreza 455
Golazad, Seyedeh Zahra 473
Gu, Yi 227
Guan, Hong 227
Gudmundsson, Kjartan 64

H

- Hague, Stephen 455
Hartmann, Timo 408
Hatami, Mohsen 133
Häusler, Felix S. 337
Heiß, Marcel 29, 119
Hoeng, Simon K. 54

J

- Jia, Fanglai 64
 Jin, Willy 421
 Juraboev, Abduaiziz 443

K

- Kandil, Mohamed S. 238
 Kandler, Benedikt 119
 Kern, Rüdiger 443
 Köhle, Jan-Friedrich 181
 König, Markus 107, 159, 398
 Kotzé, Riëtte 533
 Krall, Jenna 44
 Krishnan, Shivram 169

L

- Latifilkhechi, Leva 361
 Lee, Bogyeong 374
 Lee, Seulbi 374
 Li, Guo 464
 Liao, Wen-jie 227
 Lin, Peng 464
 Lin, Yijun 325
 Lu, Ming 640
 Lu, Xin-zheng 227
 Lu, Yujie 325

M

- Macvicar, Bruce 626
 Mahmud, Maqsood 294
 Maibaum, Jonas 107
 Manga, Ashvin 520
 Markarian, Elin 169
 Masaki, Munechika 383
 McArthur, J. J. 238, 252, 311
 Mchunu, Johannes S. 585
 Melzner, Jürgen 148
 Mevissen, Martin 15
 Mohamed, Yasser 455
 Mohammadi, Abbas 473
 Mohammadi, Pouria 205
 Motamedi, Ali 613
 Moyo, Thembani 599
 Mpingana, Wanda Buhle 652
 Musonda, Innocent 599

N

- Najafi, Fazil 562

Nawari, Nawari 562

Neumann, Patrick P. 337

O

- Obergriesser, Mathias 54
 Onososen, Adetayo 599
 Ordubadi, Farzad 473
 Othman, Abdulrahman 443
 Ouellet-Plamondon, Claudiane 275, 421, 489

P

- Paek, Irene 44
 Park, Samuel 44
 Ploennigg, Joern 15

Q

- Qiblawi, Seif 169
 Qin, Si-zhong 227
 Qu, Tan 546

R

- Rafatian, Benjamin 613
 Rashidi, Abbas 205, 473
 Robinson, Derek T. 626
 Rocha, Devon 509
 Rostami, Ghodsiyeh 283
 Rüppel, Uwe 29, 92, 119, 443

S

- Schäfer, Nils 181
 Schmailzl, Marc 54
 Schönfelder, Phillip 107
 Shaban, Marwan 546
 Shahinuzzaman, 264
 Sharafdin, Pezhman 613
 Shinoda, Naoya 383
 Smarsly, Kay 15, 337, 347
 Steinbach, Florian 148
 Stopps, Helen 311
 Stührenberg, Jan 337, 347

T

- Tajalli, Armin 473
 Takeuchi, Kazuyuki 383
 Tandon, Aditya 347
 Thiele, Christian-Dominik 29
 Thomas, Albert 169

U

Ung, Bora 489

W

Wala, Jens 3
Wang, Chen 640
Wang, Sida 640
Wang, Yang 283
Wedekind, Lina 148
Wei, Wei 325
Weng, Huaiyuan 626
Wu, Qingying 640

X

Xue, Hong-jing 227

Y

Yamada, Tsuyoshi 383
Yang, I.-Tung 217
Yao, Dongchi 76
Yeum, Chul Min 626

Z

Zhang, Ruichuan 44
Zhang, Zhiwei 464
Zheng, Zhe 227
Zhong, Lijian 325
Zongo, Karim 275