



Leonard Barolli *Editor*

Advances on P2P, Parallel, Grid, Cloud and Internet Computing

The 19th International Conference
on P2P, Parallel, Grid, Cloud and
Internet Computing (3PGCIC-2024)

Lecture Notes on Data Engineering and Communications Technologies

232

Series Editor

Fatos Xhafa, *Technical University of Catalonia, Barcelona, Spain*

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

Leonard Barolli
Editor

Advances on P2P, Parallel, Grid, Cloud and Internet Computing

The 19th International Conference on P2P,
Parallel, Grid, Cloud and Internet Computing
(3PGCIC-2024)



Springer

Editor

Leonard Barolli
Department of Information and Communication
Engineering
Fukuoka Institute of Technology
Fukuoka, Japan

ISSN 2367-4512

ISSN 2367-4520 (electronic)

Lecture Notes on Data Engineering and Communications Technologies

ISBN 978-3-031-76461-5

ISBN 978-3-031-76462-2 (eBook)

<https://doi.org/10.1007/978-3-031-76462-2>

© 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.

Welcome Message from 3PGCIC-2024 Organizing Committee

Welcome to the 19th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2024), which will be held in conjunction with BWCCA-2024 International Conference from November 13–15, 2024, in San Benedetto del Tronto, Italy.

P2P, Parallel, Grid, Cloud and Internet computing technologies have been established as breakthrough paradigms for solving complex problems by enabling large-scale aggregation and sharing of computational, data and other geographically distributed computational resources.

Grid Computing originated as a paradigm for high-performance computing, as an alternative to expensive supercomputers. The Grid computing domain has been extended to embrace different forms of computing, including Semantic and Service-oriented Grid, Pervasive Grid, Data Grid, Enterprise Grid, Autonomic Grid, Knowledge and Economy Grid.

P2P Computing appeared as the new paradigm after client-server and web-based computing. These systems are evolving beyond file sharing toward a platform for large-scale distributed applications. P2P systems have as well inspired the emergence and development of social networking, B2B (Business to Business), B2C (Business to Consumer), B2G (Business to Government), B2E (Business to Employee) and so on.

Parallel Computing is an essential computational paradigm for solving complicated problems quickly. It divides a scientific computing problem into several small computing tasks and concurrently runs these tasks by utilizing parallel hardware and overcoming the memory constraint. Parallel computing is an important part of cloud environment. However, there are significant differences between cloud computing and parallel computing.

Cloud Computing has been defined as a “computing paradigm where the boundaries of computing are determined by economic rationale rather than technical limits”. Cloud computing is a multi-purpose paradigm that enables efficient management of data centers, timesharing and virtualization of resources with a special emphasis on business model. Cloud computing has fast become the computing paradigm with applications in all application domains and providing utility computing at large scale.

Finally, *Internet Computing* is the basis of any large-scale distributed computing paradigms; it has very fast developed into a vast area of flourishing field with enormous impact on today’s information societies. Internet-based computing serves thus as a universal platform comprising a large variety of computing forms.

The aim of the 3PGCIC-2024 conference is to provide a research forum for presenting innovative research results, methods and development techniques from both theoretical and practical perspectives related to P2P, Parallel, Grid, Cloud and Internet computing.

Many people have helped and worked hard to produce a successful 3PGCIC-2024 technical program and conference proceedings. First, we would like to thank all the

authors for submitting their papers, the PC members and the reviewers who carried out the most difficult work by carefully evaluating the submitted papers.

We thank Web Administrators for their excellent work and support with the Web Submission and Management System of conference. We would like to express our gratitude to Prof. Makoto Takizawa, Hosei University, Japan, and Prof. Flavio Corradini, Camerino University, Italy, as Honorary Co-Chairs of 3PGCIC-2024 for their support and help. We give special thanks to Keynote Speakers of 3PGCIC-2024 and local arrangement team.

We hope you will enjoy the conference proceedings.

3PGCIC-2024 Organizing Committee

Honorary Co-chairs

Makoto Takizawa
Flavio Corradini

Hosei University, Japan
Camerino University, Italy

General Co-chairs

Leonardo Mostarda
Tomoyuki Ishida
Mario Dantas

University of Perugia, Italy
Fukuoka Institute of Technology, Japan
Federal University of Juiz de Fora (UFJF), Brazil

Program Committee Co-chairs

Flora Amato
Juggapong Natwichai
Yusuke Gotoh

University of Naples “Federico I”, Italy
Chiang Mai University, Thailand
Okayama University, Japan

International Advisory Committee

Beniamino Di Martino
Chuan-Yu Chang
Wenny Rahayu
Kin Fun Li

University of Campania “Luigi Vanvitelli”, Italy
National Yunlin University of Science and
Technology, Taiwan
La Trobe University, Australia
University of Victoria, Canada

Publicity Co-chairs

Gianmarco Mazzante
Tomoki Yoshihisa
Marek Ogiela
Admir Barolli

Camerino University, Italy
Shiga University, Japan
AGH University of Krakow, Poland
Aleksander Moisiu University of Durres, Albania

Finance Chair

Makoto Ikeda

Fukuoka Institute of Technology, Japan

Web Administrator Chairs

Phudit Ampririt

Shunya Higashi

Fukuoka Institute of Technology, Japan

Fukuoka Institute of Technology, Japan

Local Organizing Co-chairs

Rosario Culmone

Diletta Cacciagrano

Camerino University, Italy

Camerino University, Italy

Steering Committee Chair

Leonard Barolli

Fukuoka Institute of Technology, Japan

Track Areas

1. Data Mining, Semantic Web and Information Retrieval

Co-chairs

Bowonsak Srisungsittisunti

Francesco Piccialli

Agnes Haryanto

University of Phayao, Thailand

University of Naples “Federico II”, Italy

Monash University, Australia

PC Members

De-Nian Yang

Nicola Cuomo

Marco Cesarano

Giuseppe Cicotti

Academia Sinica, Taiwan

ESET, Slovakia

Marvell Semiconductor, Santa Clara, California,
USA

Definiens, The Tissue Phenomics Company,
Munich, Germany

Marco Giacalone	Vrije Universiteit Brussel, Belgium
Seyedeh Sajedeh Saleh	Vrije Universiteit Brussel, Belgium
Luca Sorrentino	Brightstep AB, Stockholm, Sweden
Antonino Vespoli	Centre for Intelligent Power at Eaton, Dublin, Ireland
Wenny Rahayu	La Trobe University, Australia
David Taniar	Monash University, Australia
Eric Pardede	La Trobe University, Australia
Kiki Adhinugraha	La Trobe University, Australia

2. Cloud and Service-Oriented Computing

Co-chairs

Mario Dantas	Federal University of Juiz de Fora (UFJF), Brazil
Francesco Orciuoli	University of Salerno, Italy

PC Members

Douglas D. J. de Macedo	University of Santa Catarina, Brazil
Edelberto Franco Silva	University of Juiz de Fora, Brazil
Massimo Villari	University of Messina, Italy
Stefano Chessa	University of Pisa, Italy
Miriam Capretz	University of Western Ontario, Canada
Jean-Francois Mehaut	University of Grenoble Alpes, France
Giuseppe Fenza	University of Salerno, Italy
Carmen De Maio	University of Salerno, Italy
Angelo Gaeta	University of Salerno, Italy
Sergio Miranda	University of Salerno, Italy

3. Security and Privacy for Distributed Systems

Co-chairs

Aniello Castiglione	University of Naples Parthenope, Italy
Michal Choras	University of Bydgoszcz, Poland
Giovanni Mazzeo	University of Naples Parthenope, Italy

PC Members

Silvio Barra	University of Cagliari, Italy
Carmen Bisogni	University of Salerno, Italy
Javier Garcia Blas	Charles III University of Madrid, Spain
Han Jinguang	University of Surrey, UK
Sokol Kosta	University of Aalborg, Denmark
Gloria Ortega López	University of Malaga, Spain
Raffaele Montella	University of Naples Parthenope, Italy
Fabio Narducci	University of Naples Parthenope, Italy
Rafal Kozik	UTP Bydgoszcz, Poland
Joerg Keller	FUH Hagen, Germany
Rafal Renk	UAM Poznan, Poland
Salvatore D'Antonio	University of Naples Parthenope, Italy
Lukasz Apiecionek	UKW Bydgoszcz, Poland
Joao Campos	University of Coimbra, Portugal
Gerhard Habiger	Ulm University, Germany
Luigi Sgaglione	University of Naples Parthenope, Italy
Valerio Formicola	University of Naples Parthenope, Italy
Davinder Kaur	IUPUI, USA

4. P2P, Grid and Scalable Computing

Co-chairs

Nadeem Javaid	COMSATS University Islamabad, Pakistan
Keita Matsuo	Fukuoka Institute of Technology, Japan

PC Members

Joan Arnedo Moreno	Open University of Catalonia, Spain
Santi Caballe	Open University of Catalonia, Spain
Evjola Spaho	Polytechnic University of Tirana, Albania
Yi Liu	Oita National College of Technology, Japan
Yusuke Gotoh	Okayama University, Japan
Akihiro Fujimoto	Wakayama University, Japan
Kamran Munir	University of the West England, UK
Safdar Hussain Bouk	Daegu Gyeongbuk Institute of Science and Technology (DGIST), Korea
Muhammad Imran	King Saud University, Saudi Arabia

Syed Hassan Ahmed	Georgia Southern University, USA
Hina Nasir	Air University Islamabad, Pakistan
Sakeena Javaid	COMSATS University Islamabad, Pakistan
Rasool Bakhsh	COMSATS University Islamabad, Pakistan
Asif Khan	COMSATS University Islamabad, Pakistan
Adia Khalid	COMSATS University Islamabad, Pakistan
Sana Mujeeb	COMSATS University Islamabad, Pakistan

5. Bio-Inspired Computing and Pattern Recognition

Co-chairs

Francesco Mercaldo	Institute of Informatics and Telematics (IIT), CNR, Italy
Salvatore Vitabile	University of Palermo, Italy

PC Members

Andrea Saracino	Institute of Informatics and Telematics (IIT), CNR, Italy
Andrea De Lorenzo	University of Trieste, Italy
Fabio Di Troia	San Jose State University, USA
Jelena Milosevic	TU Wien, Austria
Martina Lindorfer	University of California, Santa Barbara, USA
Mauro Migliardi	University of Padua, Italy
Vincenzo Conti	University of Enna Kore, Italy
Minoru Uehara	Toyo University, Japan
Philip Moore	Lanzhou University, China

6. Intelligent and Cognitive Systems

Co-chairs

Serena Pelosi	University of Salerno, Italy
Alessandro Maisto	University of Salerno, Italy
Nico Surantha	Tokyo City University, Japan

PC Members

Lorenza Melillo	University of Salerno, Italy
Francesca Esposito	University of Salerno, Italy
Pierluigi Vitale	University of Salerno, Italy
Chiara Galdi	EURECOM, Sophia Antipolis, France
Marica Catone	University of Salerno, Italy
Annibale Elia	University of Salerno, Italy
Raffaele Guarasci	Institute for High Performance Computing and Networking (ICAR), CNR, Italy
Mario Monteleone	University of Salerno, Italy
Azzurra Mancuso	University of Salerno, Italy
Daniela Trotta	University of Salerno, Italy

7. Web Application, Multimedia and Internet Computing

Co-chairs

Flora Amato	University of Naples “Federico II”, Italy
Tomoyuki Ishida	Fukuoka Institute of Technology, Japan

PC Members

Vincenzo Moscato	University of Naples “Federico II”, Italy
Walter Balzano	University of Naples “Federico II”, Italy
Francesco Moscato	University of Campania “Luigi Vanvitelli”, Italy
Francesco Mercaldo	National Research Council of Italy (CNR), Italy
Tetsuro Ogi	Keio University, Japan
Hideo Miyachi	Tokyo City University, Japan
Kaoru Sugita	Fukuoka Institute of Technology, Japan
Akio Doi	Iwate Prefectural University, Japan

8. Distributed Systems and Social Networks

Co-chairs

Masaki Kohana	Chuo University, Japan
Jana Nowakova	VSB-Technical University of Ostrava, Czech Republic

PC Members

Jun Iio	Chuo University, Japan
Shusuke Okamoto	Seikei University, Japan
Hiroki Sakaji	Hokkaido University, Japan
Shinji Sakamoto	Kanazawa Institute of Technology, Japan
Masaru Kamada	Ibaraki University, Japan
Martin Hasal	VSB-Technical University of Ostrava, Czech Republic
Jakub Safarik	VSB-Technical University of Ostrava, Czech Republic
Michal Pluhacek	Tomas Bata University in Zlin, Czech Republic

9. IoT Computing Systems**Co-chairs**

Paskorn Champrasert	Chiang Mai University, Thailand
Lei Shu	Nanjing Agricultural University, China

PC Members

Chonho Lee	Cybermedia Center, Osaka University, Japan
Yuthapong Somchit	Chiang Mai University, Thailand
Pruet Boonma	Chiang Mai University, Thailand
Somrawee Aramkul	Chiang Mai Rajabhat University, Thailand
Roselin Petagon	Chiang Mai Rajabhat University, Thailand
Guisong Yang	University of Shanghai for Science and Technology, China
Baohua Zhang	College of Engineering, Nanjing Agricultural University, China
Ye Liu	College of Engineering, Nanjing Agricultural University, China
Kai Huang	College of Engineering, Nanjing Agricultural University, China
Jun Liu	Guangdong Polytechnic Normal University, China
Feng Wang	Hubei University of Arts and Science, China
Alba Amato	National Research Council of Italy (CNR), Italy
Salvatore Venticinque	University of Campania “Luigi Vanvitelli”, Italy
Flora Amato	University of Naples “Federico II”, Italy

10. Wireless Networks and Mobile Computing

Co-chairs

Akimitsu Kanzaki
Shinji Sakamoto

Shimane University, Japan
Kanazawa Institute of Technology, Japan

PC Members

Teruaki Kitasuka	Hiroshima University, Japan
Hiroyasu Obata	Hiroshima City University, Japan
Tetsuya Shigeyasu	Prefectural University of Hiroshima, Japan
Chisa Takano	Hiroshima City University, Japan
Shigeru Tomisato	Okayama University, Japan
Makoto Ikeda	Fukuoka Institute of Technology, Japan
Keita Matsuo	Fukuoka Institute of Technology, Japan
Admir Barolli	Aleksander Moisiu University of Durres, Albania
Evjola Spaho	Polytechnic University of Tirana, Albania
Tetsuya Oda	Okayama University of Science, Japan

3PGCIC-2024 Reviewers

Amato Flora	Ikeda Makoto
Barolli Admir	Ishida Tomoyuki
Barolli Leonard	Kamada Masaru
Bhed Bista	Kanzaki Akimitsu
Boonma Pruet	Kohana Masaki
Caballe Santi	Leung Carson
Cui Baojiang	Liu Yi
Dantas Mario	Lu Wei
El Madhoun Nour	Matsuo Keita
Enokido Tomoya	Maisto Alessandro
Fenza Giuseppe	Mizera-Pietraszko Jolanta
Fujihara Akihiro	Mostarda Leonardo
Fujisaki Kiyotaka	Natwichai Juggapong
Fun Li Kin	Nowakova Jana
Funabiki Nobuo	Oda Tetsuya
Gaeta Angelo	Ogiela Lidia
Gotoh Yusuke	Ogiela Marek
Hayashibara Naohiro	Okada Yoshihiro
Iio Jun	Okamoto Shusuke

Orciuoli Francesco
Pardede Eric
Rahayu Wenny
Rodriguez Jorge Ricardo
Sakaji Hiroki
Shinji Sakamoto
Spaho Eviola
Sugita Kaoru
Surantha Nico
Taniar David

Uchiya Takahiro
Uehara Minoru
Ullah Zia
Venticinque Salvatore
Vitabile Salvatore
Wang Xu An
Wei Shi
Woungang Isaac
Yoshihisa Tomoki
Xhafa Fatos

3PGCIC-2024 Keynote Talks

CPSs Modeling Challenge: From Real (Possibly Chaotic, Continuum and Nondeterministic) Systems to Computational Artifacts

Dr. Diletta Romana Cacciagrano

University of Camerino, Camerino, Italy

Abstract. Cyber-physical system (CPS) is a new generation of digital systems, composed of computational and physical capability that engages with humans like never before. It is designed to act like a network of multiple variables with both physical input and output—rather than standalone technology. This talk examines the role of modeling in the engineering of CPSs. Through several examples, it investigates how chaotic behavior, i.e., the inability of computers to numerically handle a continuum and the incompleteness of determinism, can limit the possibility to build a faithful model-driven approach for engineering CPSs.

Offloading in Cloud-to-Thing Continuum

Prof. Fatos Xhafa

Technical University of Catalonia, Barcelona, Spain

Abstract. With the fast widespread and adoption of Internet technologies, cloud computing has become a digital ecosystem, referred to as cloud-to-thing continuum computing, embracing an array of computing paradigms and infrastructures, from large servers and data centers to tiny sensors and actuators at the edges of the Internet. Thereby, the intelligent edge aims at placing intelligence to the end devices, at the edges of the Internet. The premise is that collective intelligence from the IoT data deluge can be achieved and used at the edges of the Internet by offloading the computation burden from the cloud systems and leveraging real-time intelligence. While (parallel) task offloading is a well-known problem from traditional distributed computing, it is more challenging in cloud-to-thing continuum. In this talk, we will discuss some offloading computing models in cloud-to-thing continuum, its challenges and opportunities for the intelligent edge. In particular, we will discuss the challenges of processing and analyzing the IoT data streams in real time and how offloading and agile optimization can be useful to harnessing the power of the intelligent edge. We will exemplify the discussion by a real-life scenario from augmented workspace based on affective computing and federated learning.

Advanced Diagnostic Techniques and Self-healing Approaches for Enhancing Resilience of Smart Manufacturing Systems

Dr. Inès Chihi

University of Luxembourg, Luxembourg

Abstract. Smart manufacturing systems, while transformative, are inherently vulnerable to various faults and failures in hardware, software, or communication networks. These vulnerabilities not only disrupt the operational efficiency of manufacturing systems but also have far-reaching implications on sustainability, including increased machine runtime, higher energy consumption, elevated maintenance costs, reduced equipment lifespan and have greater economic and environmental waste. This talk focuses on the presentation of a new holistic conceptual model designed to address these challenges. The focus will be on advanced diagnostic techniques and self-healing approaches that can significantly enhance the resilience of smart manufacturing systems. By integrating these strategies, we can improve system reliability, reduce downtime and contribute to more sustainable manufacturing practices.

Contents

A Chatbot for Specialized Domain	1
<i>Egidia Cirillo, Mattia Fonisto, Marco Giacalone, and Alberto Moccardi</i>	
Some Bibliometric Considerations for Computer Science Conferences	13
<i>Teodor-Florin Fortiș and Alexandra-Emilia Fortiș</i>	
Time Series Analysis and Modeling with Federated Learning Techniques in Cloud Edge Scenario: A Case Study on Environmental Air Quality in Homes	25
<i>Gennaro Junior Pezzullo, Beniamino Di Martino, Oguz Mulayim, and Eva Armengol</i>	
Cloud Framework for Data Practitioners for Research and Higher Education Community	35
<i>Shruthi Sreenivasa Murthy, Krishna Chaitanya Rao Kathala, and Guangli Zhang</i>	
P2FL: Privacy-Preserving Federated Learning Approach for Healthcare Informatics at the Edge	47
<i>Farhan Ullah, Leonardo Mostarda, Diletta Cacciagrano, Hamad Naeem, Shamsher Ullah, Pradeep Chaudhary, and Yue Zhao</i>	
Enhancing Customer-Perceived Value Through Personal Data Utilization in CRM Platforms: A Data Science Perspective	59
<i>Sutipong Sutinaphan and Juggapong Natwichai</i>	
Connecting AI and Blockchain to Improve Security of Financial Services	67
<i>Ramiz Salama, Diletta Cacciagrano, and Fadi Al-Turjman</i>	
A Comprehensive State-of-the-Art Review for Digital Twin: Cybersecurity Perspectives and Open Challenges	78
<i>Aws Jaber, Ioannis Koufos, and Maria Christopoulou</i>	
A Detour Route Selection Method Based on Node Density in Skip Graph	99
<i>Riku Kamiya and Tomoya Kawakami</i>	
EDoViT-Alz: Alzheimer's Disease Identification with Vision Transformer Using Extremely Downscaled MRI Data	109
<i>Diogen Babuc and Alexandra-Emilia Fortiș</i>	

A Comparison Study Between Cuckoo Search and Particle Swarm Optimization Based Intelligent Systems for Optimization of Mesh Routers in a Small-Scale WMN	121
<i>Shinji Sakamoto, Shigenari Nakamura, Leonard Barolli, and Makoto Takizawa</i>	
A Fuzzy-Based System for Assessment of Tie Strength in Online Social Networks	133
<i>Shunya Higashi, Phudit Ampririt, Ermioni Qafzezi, Makoto Ikeda, Keita Matsuo, and Leonard Barolli</i>	
An Efficient Algorithm to Prevent Procrastination in Spatial Crowdsourcing ...	142
<i>Naren Debnath, Sajal Mukhopadhyay, and Fatos Xhafa</i>	
A Learning Web System for Website Development	154
<i>Aino Nakamura and Masaki Kohana</i>	
A Community Web System for LGBTQ+ Students with Identification	165
<i>Miyu Sato and Masaki Kohana</i>	
Single Sign-on System with Local Personal Information Store	176
<i>Yoshiki Hosoda and Masaki Kohana</i>	
A Data Platform for the Integration of Smart City Subsystems	187
<i>Stefano Silvestri, Giuseppe Tricomi, Emanuele Damiano, Mario Sicuranza, and Mario Ciampi</i>	
Minimization of Transfer Time for User Files Through Read Control for Backup with Deadline Time	199
<i>Futa Takahashi and Takayuki Kushida</i>	
Distributing Energy Consumption in Multi-interface Networks: Dimension of Cycle Space	209
<i>Alessandro Aloisio and Diletta Cacciagranò</i>	
Min-Max Coverage in Multi-interface Networks: Pathwidth	221
<i>Alessandro Aloisio</i>	
A Scalable State Channel for IoT Using Interactive Consistency Protocols	233
<i>Gianmarco Mazzante, Leonardo Mostarda, Alfredo Navarra, and Davide Sestili</i>	
Digital Twins for Improving Buildings Performances: A Literature Review Methodology Use Case	245
<i>Ionica-Larisa Puiu and Teodor-Florin Fortis</i>	

Blockchain and Digital Twin Integration for Remote Control of Cyber-Physical Systems	258
<i>Alessandro Bigiotti, Purav Shah, and Ramona Trestian</i>	
Optimising Sea Rescue Missions by UAVs	270
<i>Sajjad Ghobadi and Francesco Piselli</i>	
Virtual Hazard Map for Disaster Prevention Education	282
<i>Yumemi Fukushima and Tomoyuki Ishida</i>	
Earthquake Virtual Reality Simulation System for Appropriate Evacuation Actions	292
<i>Koichi Nishino and Tomoyuki Ishida</i>	
Mixed Reality-Based Japanese Calligraphy Learning System: Development and Evaluation	302
<i>Riko Oohashi and Tomoyuki Ishida</i>	
Finding Representative Frames from Surveillance Video for Visualizing Viewer Behavior	312
<i>Kaoru Sugita</i>	
A Comparative Sensitivity Analysis of Loss Functions in Machine Learning-Based Weather Forecasting	318
<i>Aaron Van Poecke, Lukas Meuris, Matteo Cisneros, Michiel Van Ginderachter, Peter Hellinckx, and Hossein Tabari</i>	
Autonomous Shipping in Complex Situations	327
<i>Matteo Cisneros, Oliver Rommens, Renzo Massobrio, and Peter Hellinckx</i>	
Transfer Learning for Traffic State Predictions in Small and Medium-Sized Cities	336
<i>Mohammadmahdi Rahimiasl, Ynte Vanderhoydonc, Siegfried Mercelis, Laure De Cock, Thomas Kusmirczak, and Tamara De Swert</i>	
Evaluating the Impact of Suboptimal HVAC Systems on Control Strategies	347
<i>Pieter Jan Houben, Stef Jacobs, Renzo Massobrio, Hossein Tabari, Ivan Verhaert, and Peter Hellinckx</i>	
AI for Anticipating Human Behavior	356
<i>Jeoffrey Canters, Pieter Jan Houben, Renzo Massobrio, and Peter Hellinckx</i>	

Mamdani Type-1 Non-singleton Fuzzy Logic System (T1 NSFLS) for a Quality Control Process Based on Industrial Image Processing	364
<i>Pascual Noradino Montes-Dorantes, Adriana Mexicano-Santoyo, Jesús C. Carmona-Frausto, and Gerardo Maximiliano Mendez</i>	
Weed Detection in a Sunflower Field Using Supervised Learning Techniques	374
<i>A. Mexicano, J. C. Carmona, S. Cervantes, K. Bee, and P. N. Montes</i>	
Behavior Tree as a Decision Planning Algorithm for Industrial Robot	385
<i>Martina Hutter-Mironovova, Benjamin Blumhofer, Christopher Schneider, and Achim Wagner</i>	
A Grey-Box Model for Real-Time Control and Monitoring	395
<i>Ricardo Rodriguez-Jorge</i>	
Author Index	407



A Chatbot for Specialized Domain

Egidia Cirillo¹, Mattia Fonisto¹, Marco Giacalone², and Alberto Moccardi¹ (✉)

¹ Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy
{egidia.cirillo,mattia.fonisto,alberto.moccardi}@unina.it

² Private and Economic Law Digitalisation and access to justice, University of Bruxelles, Pleinlaan 2, 1050 Brussels, Belgium
Marco.Giacalone@vub.be

Abstract. Navigating the complexities of data within the legal framework presents a formidable challenge, as the corpora are often dense, verbose, and syntactically intricate. These characteristics necessitate sophisticated tools capable of not only accessing and interpreting the content but also effectively organizing it for practical use, particularly in applications such as chatbots or Retrieval Augmented Generation (RAG) systems. The existing legal databases, vital for the computationally intensive applications mentioned earlier, typically lack the systematic structure and indexing required for straightforward application, further complicating their integration into advanced conversational systems. The proposed work underscores the pivotal role of Generative AI, emerging as a dual-purpose tool by facilitating the structuring of vast, complex datasets and leveraging the optimized knowledge base. This dual functionality has been proven to significantly enhance the quality and interpretability of the outputs, fostering human validation, thereby mitigating the computational load and reducing the occurrence of inaccuracies commonly associated with Language Models (LMs).

1 Introduction

In the dynamic interplay of data and AI, integrating of data engineering with Generative AI (GAI) marks a significant milestone. As data complexity and volume escalate, alongside the computational demands of applications, addressing the challenges in analytical data management becomes paramount. This is even more crucial for the Large Language models (LLMs) powered conversational systems [11] which have reaped substantial benefits from generative AI wave, illustrating the technology's potential to enhance speed and accuracy in generating sophisticated textual interactions. Central to these advancements are the Retrieval Augmented Generation (RAG) algorithms, which blend traditional generative models with dynamic data retrieval, infusing the generative process with highly relevant contextual information, thereby enhancing the functionality and reliability of language models with granted external knowledge bases. In this context, the seminal works established within the classic research framework of

Natural Language Processing (NLP) [6, 7, 13, 24, 27] lay the foundational principles for this study, which is even more strongly founded on the noteworthy works by Lewis et al. [15] which introduced the mixed generation and retrieval model amalgamating parametric and non-parametric models to significantly improve the responsiveness and relevance of conversational systems. Building on this basement, the thorough survey by Gao et al. [12] in 2024 explores the evolution of RAG paradigms, offering a detailed examination of the progress and delineating both the extant challenges and prospective avenues for further research and development. The legal sector, moreover, represents a focal point and one of the most prominent research and development GAI application field due to its relevance and its straightforward applicability. A notable development in this area is the introduction of an open-source legal LLM, ChatLaw, by Cui et al. [8] in 2023, which addresses the challenge of hallucinations often encountered in generative models by implementing a RAG-fashioned system that improves the model’s accuracy by mitigating errors in reference data. Further enriching this landscape, the work of Pipitone et al. [23] which critically evaluates and benchmarks RAG pipelines within the legal domain.

Through the experience of these domain-based seminal papers, it emerged that the legal field often grapples with managing vast and poorly organized document archives complicating the retrieval of pertinent information from complex and sensitive documentation making the process of understanding and validating the result by the user even more complex. In this perspective, the deployment of a semantically driven GAI tool is here proposed and applied to unclassified and anonymized court cases corpus, specifically regarding disputes among UE civilians over assets in a divorce, inheritance, and the division of a company assets, provided by the CREA2 project. This LLM-based tool has been carefully designed to extract vital information and metadata from lengthy and complex datasets, processing each document individually and avoiding exceeding the computational burdens set by the LLM’s maximum window length while ensuring response quality and coherence. We also considered the use of federated learning [17–19]; it is important to account for both the privacy benefits of decentralized data processing and the potential challenges associated with communication overhead and model convergence, especially in scenarios involving heterogeneous data sources.

2 Methodology

The methodology depicted in this chapter introduces the emerging framework designed to harness synergistic potentials through the adept use of advanced computational tools for processing, analyzing, and semantically interpreting large and complex legal data. The framework architecture, as presented in Fig. 1, is delineated into two primary operational categories: “back-end operations” and “exposed operations”. Back-end operations encompass the initial stages of data processing till the embedding vector generation, laying the groundwork for advanced textual analysis and semantic processing [2–4]. Exposed operations,

conversely, comprise the retrieval system setup of the computationally intensive RAG equipped with the query expansion mechanism, directly interfacing with the user to deliver nuanced, contextually informed answers.

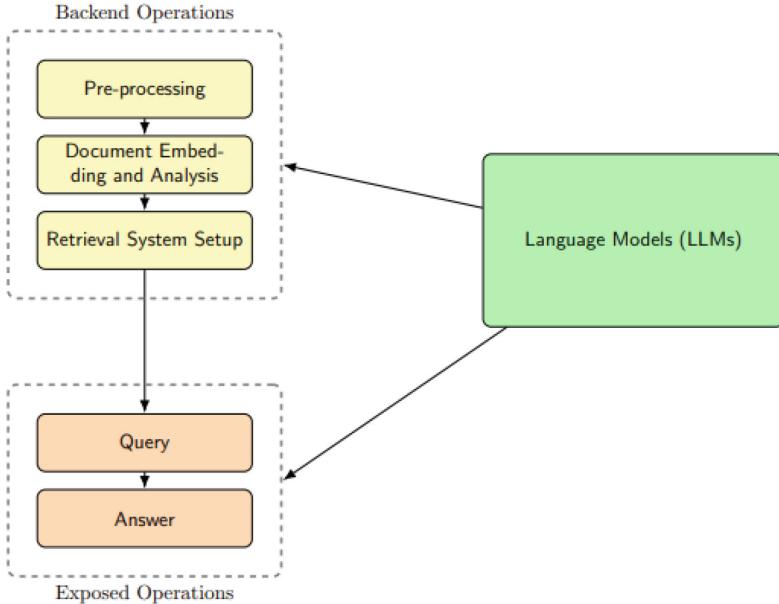


Fig. 1. General Methodology

The global methodological building aims to leverage a strategic use of Language Models (LMs) foundational for LLMs, from the Generative Pretrained Transformers (GPTs) series, to cover the areas of processing raw data and generating structured answers by covering the aforementioned strategic dual aspect of the proposed GAI application. The LLMs will navigate the tortuous legal lexicon and distill verbose documents into pithy summaries, capturing the essence of the documents. Further, the summaries generated shall be tagged with rich metadata derived from a carelessly crafted schema reflecting the domain-specific legislative content using custom NLP techniques. This encapsulation of metadata will enrich the corpus with critical case-specific information pertinent to dimensions of legal principles and procedural contexts, deepening the comprehension and analysis of jurisprudential domains and inserting the user into the validation loop. Secondly, an in-depth analytical exploration of the document's nature is conducted using cutting-edge embedding technologies. The corpus, filtered through a specific metadata analysis that connects the user's query to the corpus content, serves as the basis for constructing the retrieval system. This system utilizes a specialized vector store, which is the central component leveraged by the Retrieval Augmented Generation (RAG), to process the transformed

and filtered knowledge bases. Finally, through multiquery RAG, is possible to exploit the collective knowledge represented in the corpus through the real-time integration of multiple sets of semantically relevant documents into the answer generation process, leveraging, in this way, the embedded information through the retrieval mechanism. This all-embracing approach, therefore, has within it an elegant orchestration of cutting-edge computational tools and techniques meticulously integrated to cultivate innovative dynamic interactions with a variety of legal documentation.

Generative Data Structuring

In detail, the aforementioned methodology consists of an integrated preprocessing pipeline that transforms raw legal documents into structured, summarized, and metadata-enriched data suitable for retrieval-augmented processing tasks. The pipeline is divided into two principal stages: data cleansing and initial structuring, content summarization, and metadata enhancement with the consequent serialization of the processed output. A clear, in depth, visualization is proposed by the Fig. 2.

Data Cleansing and Initial Structuring

The commencement of our pipeline is marked by rigorous data cleansing, which involves normalization and error correction to produce a clean and homogeneous dataset. Concurrently, initial metadata tagging with GAI annotates the text with crucial categorical information that serves both as an index and guide for the further processing stages. This step is followed by the division and structuring of the text, systematically breaking it down into manageable and logically coherent segments for more focused analysis.

Content Summarization and Metadata Enhancement

Upon establishing a structured foundation, each segmented text undergoes a summarization process, distilling the essence of the content while preserving critical information. This is accomplished through the application of advanced NLP techniques that yield concise and meaningful summaries. To complement this, a secondary metadata tagging phase enriches the existing base, enhancing the overall data quality and depth. The final integration of metadata is executed through a left join operation, ensuring the enrichment of the initial metadata with the additional layers gathered in the last stage, thereby creating a comprehensive metadata profile for each document. The culmination of our preprocessing pipeline is the serialization stage. Here, the summarized content and fully integrated metadata are encoded into a JSON format, chosen for its widespread compatibility and ease of use in data interchange. A comparative analysis of this extraction process is further performed, depicted in Fig. 3, between the original and summarized documents, showing a significant reduction in word count per document (and memory), potentially offering computational advantages in the subsequent RAG workflow when interfacing with the computational constraints of LLMs, thereby facilitating a more efficient generation phase.

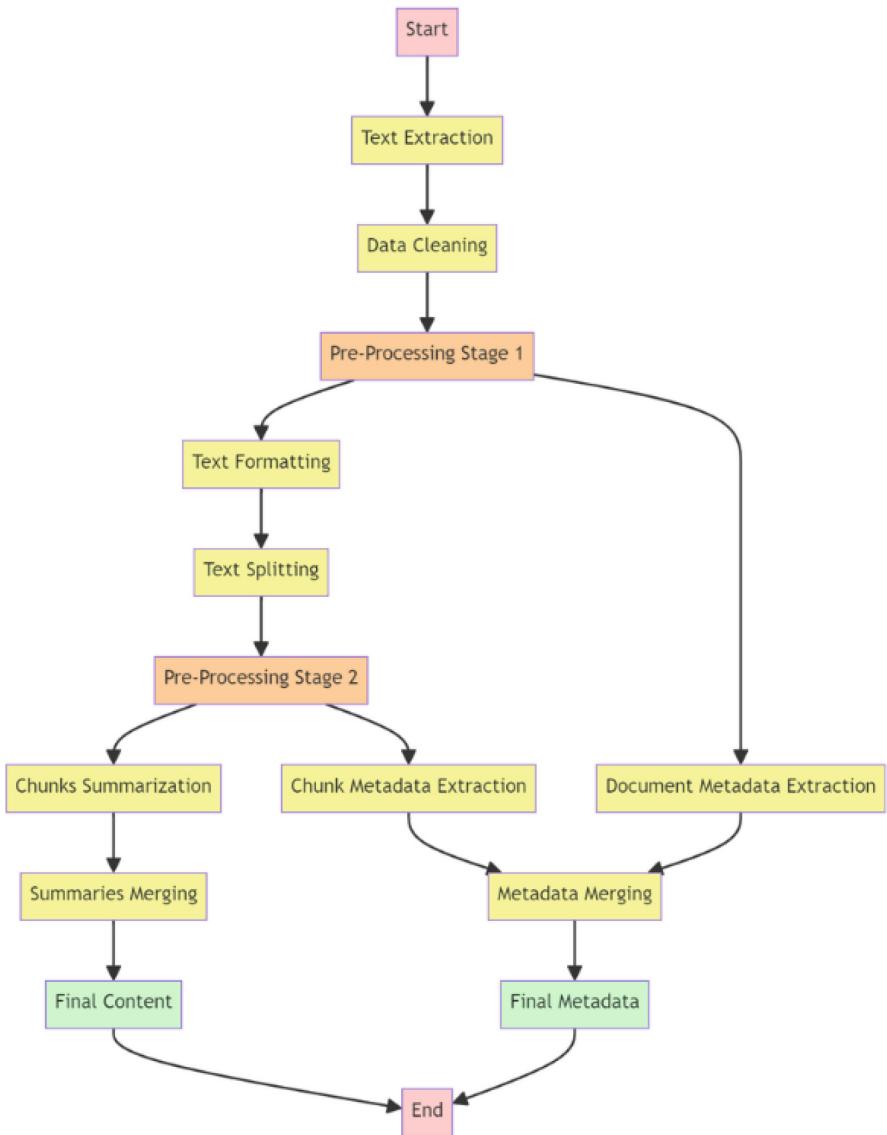


Fig. 2. Big Data Processing Workflow for Data & Computationally Intensive RAG Applications

Metadata Analysis and Knowledge Base Refinement

The integrity and utility of the knowledge base are further augmented through a detailed metadata analysis Fig. 4, concentrating on documents categorized under specific law-types, such as divorce and inheritance. Documents that possess clear

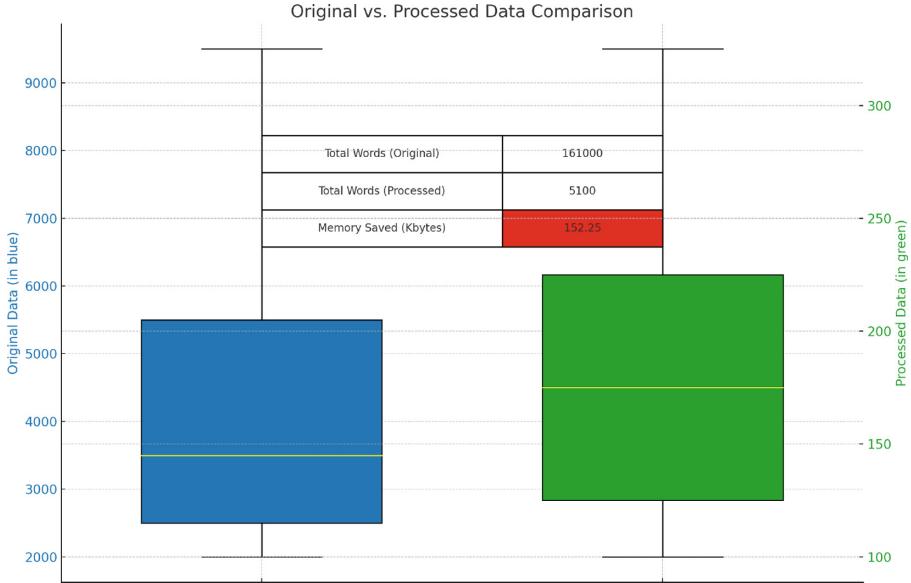


Fig. 3. Data distribution and memory before and after compression

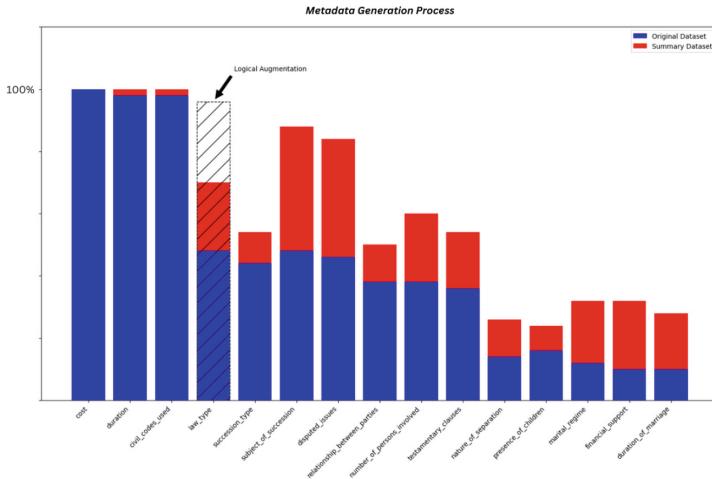


Fig. 4. Metadata extraction from the two-phases data processing with logical augmentation

and comprehensive law-type metadata undergo logical metadata augmentation and are retained within the database. Conversely, documents characterized by incomplete or ambiguous metadata are identified and marked for exclusion from the knowledge base.

Multiquery Retrieval Augmented Generation System

The Retrieval-Augmented Generation (RAG) system presents a sophisticated architecture designed for advanced legal document processing and knowledge-based question answering, with the main aim of testing the successful data engineering process. This system exemplifies a strategic integration of filtering mechanisms, query expansion [14, 26], and answer synthesis to enhance the efficiency, the interpretability, and the validity of legal information retrieval.

Embedding Generation and Analysis

The proposed embedding generation process is solidly grounded in the ongoing advancement of embedding technologies in NLP, commenced with the seminal contributions of Bengio et al. [5] who introduced the use of neural networks for generating linguistic embedding vectors, laying the groundwork for subsequent advancements. This foundational approach was further developed through the introduction of Word2Vec by Mikolov et al. [21], which pioneered the technique of embedding words into a vector space based on contextual similarities. Concurrently, Pennington et al. [22] proposed GloVe, an innovative method for aggregating global word-word co-occurrence statistics to enhance the representational capabilities of embeddings. The transformative potential of these techniques was fully realized with the introduction of the Transformer model by Vaswani et al. [28], which employs a self-attention mechanism to dynamically weigh the significance of different words in a sentence. Building upon this framework, Devlin et al. [9] introduced BERT, which leverages bidirectional training to significantly improve the contextuality of embedding vectors. The most recent synthesis and evaluation of these methodologies can be found in the comprehensive survey by Almeida et al. [1], which critically examines the evolution and impact of embedding technologies in natural language processing.

Retrieval and Filtering Mechanism

At its core, the proposed RAG system employs a robust filtering system that intelligently categorizes incoming queries into specific legal contexts. This categorization is facilitated through a preliminary extraction of metadata from user queries, identifying key legal dimensions that are critical for contextual relevance. Subsequent to metadata extraction, the RAG system utilizes a dynamic filtering approach to precisely target the relevant area of law. This specificity in document retrieval, supported by a vector database, ensures the relevance and precision of the documents fetched, substantially improving the outcome and transparency of the retrieval process.

Question Expansion Chain

A distinctive feature of the outlined RAG system is its ability to generate multiple rephrased versions of the initial user query. This process is conducted through

a dedicated question generation chain guided by a careful prompt engineering phase [25], which crafts up to three alternative queries that encapsulate varying legal perspectives and subtleties pertaining to the original question. This expansion of the query set is pivotal for mitigating the inherent limitations of distance-based similarity searches within vector databases, ensuring a thorough and comprehensive document retrieval process.

Answer Generation Chain

Following the retrieval phase, the system engages an answer generation chain to synthesize and formulate responses based on the information contained within the retrieved documents.

The answer generation template is fine-tuned to enhance the human comprehension and transparency of the responses provided by the RAG system, incorporating a structured approach that includes:

- **Contextualized Answers:** After retrieving relevant documents, the system synthesizes the content to construct answers that are not only direct but also richly contextualized within the legal framework discussed in the query.
- **Reasoning Process:** The template is structured to facilitate a reasoning process where the system answers the generated sub-questions. This method helps in breaking down complex legal information into understandable segments, making it easier for users without a legal background to comprehend the implications and jargons of their cases.
- **Incorporation of Similar Cases & Extracted Metadata:** To further aid understanding and provide comparative insights, the template includes a section that cites similar cases from the knowledge base and metadata extracted from the retrieval. This comparative analysis helps users grasp how their situation might unfold based on precedents.
- **Overall Evaluation:** Finally, the template provides an overall evaluation of the situation based on the answers and the cited cases.

3 Results and Evaluation

The generative data engineering process was finally tested with the previously presented RAG system using a small set of 10 queries, evenly split between succession and divorce-related questions. The performance of the system has been evaluated using the RAGAS (Retrieval-Augmented Generation Assessment Schema) [10] metrics, which are designed to provide a comprehensive assessment of both the retrieval and generation aspects of the system. For each query, the following components were analyzed:

- **User Query:** The initial input into the RAG pipeline represents real-world questions that users might pose regarding succession or divorce.
- **Generated Answer:** The output produced by the RAG system in response to the user query.

- **Contexts:** Documents used to generate answers retrieved by the system from an external knowledge source.
- **Ground Truths:** Human-annotated answers to the queries, serving as a benchmark for assessing the accuracy and relevance of the system's generated answers.

Four principal key metrics were employed to evaluate the performance of the RAG system:

1. **Context Precision:** This metric evaluates the signal-to-noise ratio within the retrieved contexts. It measures how many of the retrieved documents are actually relevant to the user's query.
2. **Context Recall:** Context recall assesses whether all necessary information required to answer the query has been retrieved. It relies on a comparison between the retrieved contexts and the ground truths.
3. **Faithfulness:** Faithfulness quantifies the factual accuracy of the answers generated by the RAG system. This metric ensures that the generated answers are not only relevant but also factually correct.
4. **Answer Relevancy:** This metric measures how well the generated answers address the user's queries. For example, if a query asks for multiple pieces of information, the relevancy score reflects how completely the response addresses all elements of the query.

Assessment of the Multiquery RAG System

The rigorous assessment of the system through RAGAS metrics has elucidated notable strengths, Figs. 5 and 6, particularly in the realm of Context Precision and Answer Relevancy. With mean values soaring at 87.50% and 89.22%, respectively, demonstrating a superior capacity for sourcing pertinent documents and generating answers with high relevance to user queries.

In areas where the performance did not reach the high marks seen in other domains, specifically in the context retrieval for divorce-related queries, a discerning analysis offers substantive insights. The less extensive document base for divorce cases has been a determinant in the modest recall rates. However, this is less a reflection of system inadequacy and more an indicator of the specialized and often nuanced nature of divorce proceedings which, historically, have been less documented and standardized in comparison to other legal domains.

This particular aspect provides a vital opportunity for targeted system enhancements. By expanding the external knowledge source and enriching the corpora with a broader spectrum of divorce-related documents, we anticipate measurable improvements in future performance [16, 20] evaluations. The system's overall efficacy, so, despite the challenge posed by the limited availability of comprehensive divorce case documents, is commendable. It underscores the system's robustness and the potential for amplified proficiency as the database grows.

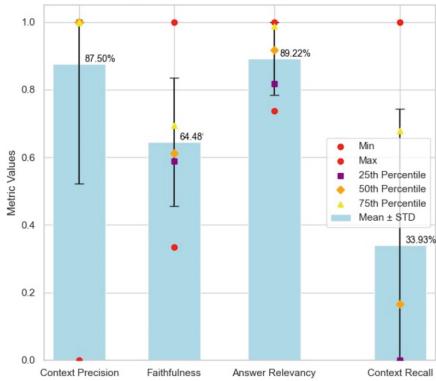


Fig. 5. RAGAS evaluation of Multiquery Retrieval Augmentation System

4 Conclusion

In conclusion, the proposed solution represents a significant leap forward in the application of generative Artificial Intelligence within the legal domain. It offers a systematic approach to tackling the prevalent challenge of disorganized legal knowledge bases, harnessing the power of GAI for data cleansing, text summarization, and metadata organization. The performances of the multiquery RAG system, demonstrate a keen understanding of the legal field's complexities, not only refining the query process with tailored filters but also employing sophisticated prompt engineering to generate nuanced and informative responses that are both precise and contextually relevant. The incorporation of an extensive metadata schema, coupled with the functionality to reference relevant articles and prior cases, significantly elevates the system's value as an indispensable resource for legal research and decision-making. This enhancement affords users the opportunity to critically evaluate the accuracy of responses by examining the extracted metadata, thereby increasing the accessibility and understanding of the information embedded within the original documents.

Generated Answer

Case Overview: In the case of John's contested will, the most important civil articles used were Article 190, Article 275, and Article 2964. These articles were utilized to address issues related to the reduction of donations, division of assets, injury of legitimacy, and the relationship between heirs.

Similar Cases Analysis: Similar cases involving the succession of deceased individuals have shown that the duration of such legal procedures can range from 1 year to 11 years, with an estimated medium cost of Euro 81,869.42. These cases also involved disputes over testamentary clauses, division of assets, and the relationship between parties, resulting in the court's intervention to ensure a fair distribution of the inheritance.

Conclusion: The legal implications of contested clauses in John's will can lead to prolonged legal battles, significant costs, and complex relationships between the parties involved. The distribution of assets among multiple heirs and legatees may be impacted by differing opinions, leading to the need for legal intervention to resolve disputes and ensure a fair division of the estate.

Subquestions

1. What are the legal implications of the contested clauses in John's will?
2. How might the complex relationships between the parties impact the distribution of assets among the multiple heirs and legatees?
3. What civil articles were used to address the issues in similar cases involving contested wills?

Query Metadata

Subject of Succession:	real estate, bank accounts,
Testamentary Clauses:	trusts, inheritance exclusion
Disputed Issues:	validity of will
Relationship Between Parties:	spouse, children, relatives
Number of Persons Involved:	4
Law Type:	Succession

Fig. 6. Generated answer example & structure

Acknowledgments. This research was funded by the European Union grant number 101046629 - CREA2. However, the views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

References

1. Almeida, F., Xexéo, G.: Word Embeddings: A Survey (2023). URL <https://arxiv.org/abs/1901.09069>
2. Amato, F., Mazzeo, A., Moscato, V., Picariello, A.: Semantic management of multimedia documents for e-government activity. In: 2009 International Conference on Complex, Intelligent and Software Intensive Systems, pp. 1193–1198. IEEE (2009)
3. Amato, F., Mazzeo, A., Moscato, V., Picariello, A.: A system for semantic retrieval and long-term preservation of multimedia documents in the e-government domain. *Int. J. Web Grid Serv.* **5**(4), 323–338 (2009)
4. Amato, F., Castiglione, A., Cozzolino, G., Narducci, F.: A semantic-based methodology for digital forensics analysis. *J. Parallel Distrib. Comput.* **138**, 172–177 (2020)
5. Bengio, Y., Ducharme, R., Vincent, P., et al.: A neural probabilistic language model. *J. Mach. Learn. Res.* **3**, 1137–1155 (2003)
6. Bonetti, F., Leonardelli, E., Trotta, D., Guarasci, R., Tonelli, S.: Work hard, play hard: collecting acceptability annotations through a 3d game. In: Proceedings of the Language Resources and Evaluation Conference (LREC 2022), pp. 1740–1750 (2022)
7. Chandrasekaran, D., Mago, V.: Evolution of semantic similarity-a survey. *ACM Comput. Surv. (CSUR)* **54**(2), 1–37 (2021)
8. Cui, J., Li, Z., Yan, Y., Chen, B., Li, Y.: Open-Source Legal Large Language Model with Integrated External Knowledge Bases, Chatlaw (2023)
9. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding (2019). <https://arxiv.org/abs/1810.04805>
10. Shahul, E.S., James, J., Espinosa-Anke, L., Schockaert, S.: Ragas: Automated Evaluation of Retrieval Augmented Generation (2023). <https://arxiv.org/abs/2309.15217>
11. Friedman, L., et al.: Leveraging Large Language Models in Conversational Recommender Systems (2023). <https://arxiv.org/abs/2305.07961>
12. Gao, Y., et al.: Retrieval-Augmented Generation for Large Language Models: A Survey (2024)
13. Guarasci, R., Catelli, R., Esposito, M.: Classifying deceptive reviews for the cultural heritage domain: a lexicon-based approach for the Italian language. *Expert Syst. Appl.* **252**, 124131 (2024)
14. Jagerman, R., Zhuang, H., Qin, Z., Wang, X., Bendersky, M.: Query Expansion by Prompting Large Language Models (2023). <https://arxiv.org/abs/2305.03653>
15. Lewis, P., et al.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks (2021)
16. Mazzeo, G., Arnautov, S., Fetzer, C., Romano, L.: Sgxtuner: Performance enhancement of intel SGX applications via stochastic optimization. *IEEE Trans. Dependable Secure Comput.* **19**(4), 2595 (2021)

17. Mazzocca, C., Romandini, N., Colajanni, M., Montanari, R.: Framh: a federated learning risk-based authorization middleware for healthcare. *IEEE Trans. Comput. Soc. Syst.* **10**(4), 1679–1690 (2022)
18. Mazzocca, C., Romandini, N., Mendula, M., Montanari, R., Bellavista, P.: Truflaas: trustworthy federated learning as a service. *IEEE Internet Things J.* **10**(24), 21266–21281 (2023)
19. Mazzocca, C., Romandini, N., Montanari, R., Bellavista, P.: Enabling federated learning at the edge through the iota tangle. *Future Gener. Comput. Syst.* **152**, 17–29 (2024)
20. Ménétrey, J.: A comprehensive trusted runtime for webassembly with intel SGX. *IEEE Trans. Dependable Secure Comput.* (2023)
21. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013. URL <https://arxiv.org/abs/1301.3781>
22. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
23. Pipitone, N., Alami, G.H.: Legalbench-rag: a Benchmark for Retrieval-Augmented Generation in the Legal Domain (2024). <https://arxiv.org/abs/2408.10343>
24. Reimers, N., Gurevych, I.: Sentence-Bert: Sentence Embeddings Using Siamese Bert-Networks (2019). arXiv preprint [arXiv:1908.10084](https://arxiv.org/abs/1908.10084)
25. Sahoo, P., Singh, A.K., Saha, S., Jain, V., Mondal, S., Chadha, A.: A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications (2024). <https://arxiv.org/abs/2402.07927>
26. Tang, Y., Yang, Y.: Multihop-rag: Benchmarking Retrieval-Augmented Generation for Multi-hop Queries (2024). <https://arxiv.org/abs/2401.15391>
27. Trotta, D., Guarasci, R., Leonardelli, E., Tonelli, S.: Monolingual and cross-lingual acceptability judgments with the Italian cola corpus. In: Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 2929–2940 (2021)
28. Vaswani, A., et al.: Attention is All You Need (2023). <https://arxiv.org/abs/1706.03762>



Some Bibliometric Considerations for Computer Science Conferences

Teodor-Florin Fortis^(✉) and Alexandra-Emilia Fortis

West University of Timișoara, 300223 Timișoara, Romania
{florin.fortis,alexandra.fortis}@e-uvt.ro

Abstract. Computer science conferences are an important means of disseminating scientific information in this field, with an increasing contribution to the scientific landscape. The share of indexed conference papers has increased significantly in recent years, with a similar increase in their citation rate. While, in the case of journals, there are means of measuring the impact, such tools are difficult to be identified and used in the case of computer science conferences. Our intention was to offer an overview of existing options and explore some of the numerical criteria for conference rankings.

Keywords: Computing conferences · Bibliometric investigation · h-Index · Conference rankings

1 Introduction

The importance of computer science conferences, as a main channel of dissemination, has significantly increased in last decades. For example, by analyzing the data available in SCOPUS¹ for the 2018–2022 year interval, it was found that the number of publications of the type ‘conference papers’ (CP) exceeds the number of publications of the type ‘article’ (AP), counting 1,403,201 entries for the first category, whereas 1,070,321 were found for the second category. It was somehow surprising the fact that, among the first ten highly cited entries, for the analyzed time interval, 8 are classified as CP, while only two are classified in the AP category.

This preference for computer science conferences, as the main dissemination channel of scientific results in the field, is not new. A document of the Computing Research Association (CRA) from 1999 mentioned that “relying on journal publications as the sole demonstration of scholarly achievement, especially counting such publications to determine whether they exceed a prescribed threshold, ignores significant evidence of accomplishment in computer science and engineering. For example, conference publication is preferred in the field, and computational artifacts – software, chips, etc. – are a tangible means of conveying ideas and insight”².

¹ <https://www.scopus.com/>.

² <https://cra.org/resources/best-practice-memos/evaluating-computer-scientists-and-engineers-for-promotion-and-tenure/>.

The main target of this paper was to discover some numerical patterns to estimate the potential quality of a conference series. To reach the specified target, several scientific outlets from the ‘Distributed computing and systems software’ area were analyzed, in order to formulate a set of initial suggestions.

2 Background Information

2.1 The Culture of Computer Science Conferences

A point of view expressed in [6], states that “in the Computer Science publication culture, prestigious conferences are a favorite tool for presenting original research – unlike disciplines where the prestige goes to journals and conferences are for raw initial results”, whereas “journals have their role, often to publish deeper versions of papers already presented at conferences.”

Another analysis, on a set of around 300,000 entries, including 195,000 CP, reveals that while computer science “values conferences as a publication venue more highly than any other academic field of study”, in what regards citation behavior, “citation rates in conferences are no higher than in journals”, while some computer science conferences “have the highest average paper citation rate of any publication type [10].”

On the other hand, while there are well-established means to measure the quality of periodicals, few ranking efforts for conferences exist, often adapted to a regional interest. Such initiatives include, for example:

1. CORE Conference Rankings³, an initiative based on previous evaluation exercises, launched in 2006, with periodic updates (the most recent being the exercise from 2023).
2. QUALIS-CAPES⁴, an initiative sponsored by the Brazilian Federal Agency for the Improvement of Higher Education. The QUALIS methodology provides an aggregation of information from different sources, including Google Scholar, Scopus, and Clarivate, the most recent update being made in 2020, for the 2017-2020 years interval.
3. GII-GRIN-SCIE (GGS)⁵, an initiative under the source of Informatics-Europe, which tries to rank and unify the results from different sources, with the most recent incomplete list compiled in the 2021 edition.
4. CCF List⁶, a list of recommended international conferences and periodicals, maintained by the China Computer Federation (CCF), currently at 2019 edition.

It is worth mentioning that, starting from the CORE and GCS activities, an extended framework for conference rankings (ICORE) was defined, and this will

³ <http://portal.core.edu.au/conf-ranks/>.

⁴ <https://qualis.ic.ufmt.br/>.

⁵ <https://scie.lcc.uma.es:8443/gii-grin-scie-rating/>.

⁶ <https://www.ccf.org.cn/c/2019-05-13/663884.shtml>.

prepare its first updates for 2026. The ICORE initiative is based on a collaboration between CORE, GCS, GRIN (Group of Italian Professors of Computer Science) and SCIE (Spanish Computer-Science Society) representatives⁷.

Considering that the updates of these ranking exercises are rather rare, whilst there is a tendency to favor computer science conferences that are intended for an audience from certain geographical areas (such as Australasian, in the case of CORE2023, Brasilian, for QUALIS-CAPES), the extensive use of these results can often be problematic, as mentioned in [4].

There are a number of approaches related to the quality of computer science conferences, by using some bibliometric indicators, such the one identified in [5], where the Scopus *CiteScore* value is used to evaluate 154 computer science conferences, compared to journals located in the first quartile. The conclusions of this study show that “publishing in conference proceedings – especially top-rated ones – are as important and influential as publishing in top journals.”

On the other hand, the dynamics of research in Computer Science is special, and rarely reflected in the various ranking approaches of computer science conferences. As an example, it is worth mentioning that, despite of the huge research funding invested in Cloud Computing and its spin-offs, there was a relatively modest coverage in the CORE rankings, until 2020.

2.2 Bibliometric Evaluations

In a column from 2008, Thomas E. Nisonger examines the 80/20 rule (Pareto distribution), in various contexts, including citation behavior. While this is an approximate pattern, it is still a valid method when applied to citation data [2,8].

The 80%–20% pattern is sometimes discussed in relation with the Bradford rule, which states that “when the journals are divided into groups, each containing the same number of articles on a given subject, then the number of journals in the succeeding groups form a geometrical progression” [1,7,8].

Citation patterns and citation behavior were also examined in [3,9]. The influence of the publication outlet on the citations they received was analyzed by N.Harwood in [3], while the work of Tahamtan et al. is offering an overview of the factors that are affecting the number of citations [9]. Finally, it is worth mentioning the title essay from 2011: “In scientific publishing at the article level, effort matters more than journal impact factors: hard work and co-authors overshadow journal venue in acquiring citations” [11].

3 Methodology

Our goal was to analyze a series of scientific outlets from the ‘Distributed computing and systems software’ area, and to explore specific numerical criteria for conference ranking. In order to get a better understanding of these numerical mechanisms, a set of eight scientific outlets was selected, based on their

⁷ <https://www.core.edu.au/icore-portal>.

CORE2023 rankings. For each of these events, we first collected data from Google Scholar metrics⁸. The collected information include the h5-index and h5-median values for each of these events.

Once the initial set of metrics were collected, a secondary search was executed on the Scopus database, as the main data source, to collect information about the full set of papers and their citation counts, for each of the initial set of conferences. Scopus was selected as the main data source, as it offers excellent coverage for this type of scientific outlets.

Starting from the aforementioned data sources, we derived some additional data to support the generation of conclusions for a potential ranking exercise of the conferences. These findings are presented in the concluding parts of the relevant sections.

3.1 Identification. The Initial Set of Conferences

As we aimed to analyze a variety of computer science conferences related to the field of research of *Distributed computing and systems software*⁹ (FoR 4606), we selected eight conferences with different CORE2023 rankings, ranging from A to C. For each of these conferences, initial data was collected from Google Scholar metrics. Once these metrics were identified, Scopus was used as the main source of information, for collecting the full collection of indexed entries, together with their associated meta-data.

The eight conferences included in our selection were¹⁰:

Ranked A (CORE2023)

1. **EuroSYS** – *ACM European Conference on Computer Systems*;
2. **ICSO** – *International Conference on Service Oriented Computing*;

Ranked B (CORE2023)

1. **AINA** – *International Conference on Advanced Information Networking and Applications*;
2. **CCGRID** – *IEEE International Symposium on Cluster Computing and the Grid*;
3. **Euro-PAR** – *European Conference on Parallel Processing*;

Ranked C (CORE2023)

1. **CISIS** – *International Conference on Complex, Intelligent and Software Intensive Systems*;
2. **CLOSER** – *International Conference on Cloud Computing and Services Science*;
3. **DS-RT** – *IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications*;

⁸ https://scholar.google.com/citations?view_op=metrics_intro&hl=en, 2023 edition.

⁹ The Field of Research (FoR) codes are available on the ICORE website.

¹⁰ In italics, the name used in GS.

3.2 Screening. Google Scholar Metrics

To initiate a comparison of the quality and impact of the selected conferences, we used the information which are available via Google Scholar (GS), and derived some additional values.

Venue Name: The official name of the conference series, as recorded in GS (their full names, as specified in the list from Sect. 3.1);

Metrics: The h5-index (h_5) and h5-median (h_{5m}) values, which measure the number and the distribution of citations of the most cited papers in the last five years;

Estimates: Two estimated values, b_{h5} and l_{h5} , which are approximations for the base value for the total number of citations of the top h_5 papers, $b_{h5} = h_5 \times h_{5m}$, and a linear approximation of the citation count, assuming an arithmetic progression of citations for the same group of papers, $l_{h5} = \frac{h_5 \times (h_5 + 1)}{2}$.

While the lower value for the number of citations for the first h_5 papers is h_5^2 , based on the the 80%–20% rule [8], and Bradford's law [7], it is unlikely that the most cited papers have equal citations. As the citation counts of the highly cited papers may vary, we decided to use the l_{h5} estimate as a lower value, even if this is an underestimation, compared with the values suggested by the two bibliometric rules.

The ratio $p_{h5} = \frac{b_{h5}}{l_{h5}}$, was used as an indicator for an expected performance of the conference compared to the linear assumption. A higher value suggests a better performance.

Also, calculations of the average expected performance for each conference were included, as the ratio of the estimated citations to the linear assumption. It was found that this ratio decreases as the number of accepted papers increases, meaning that having more accepted papers leads to lower impact, in line with some observations from literature, such as those from [3, 9].

3.3 Findings. Google Scholar Metrics

The findings for each of the analyzed computer science conferences, based on GS metrics, are as follows.

EuroSYS: The main metrics extracted are $h_5 = 47$, $h_{5m} = 76$, which suggest about $b_{h5} = 3,572$ citations for the group of entries used for h_5 computations. Thus, we have $l_{h5} = 3,337$, and the expected performance will be $p_{h5} = 1.07$.

In the case of EuroSYS, there are 11 highly cited papers (papers receiving at least 100 citations), of which one received less than 300 citations, other two less than 400 citations, and the top paper received more than 3,800 citations.

ICSO: We extracted the following values: $h_5 = 22$, $h_{5m} = 36$, with the expected values $b_{h5} = 792$ citations, and $l_{h5} = 737$, generating an expected performance of $p_{h5} = 1.07$. There are 2 highly cited papers for ICSO.

AINA: The GS values identified for AINA are: $h5 = 22$, $h5_m = 30$, which suggest about $b_{h5} = 660$ citations for the $h5$ group of entries.

The expected performance of this conference is $p_{h5} = 0.89$, against the value $l_{h5} = 737$, while only one paper received more than 100 citations;

CCGRID: In the case of CCGRID, the identified values included $h5 = 24$, $h5_m = 33$, combined with the expected values of $b_{h5} = 792$ and $l_{h5} = 876$.

The ratio $p_{h5} = 0.9$ indicates the expected performance of the conference, which is not very high as none of the papers is highly cited.

Euro-PAR: For this outlet, GS offers the following information $h5 = 19$, $h5_m = 31$, generating the set of basic values $b_{h5} = 589$, $l_{h5} = 551$, and $p_{h5} = 1.06$.

Even if Euro-PAR was, a decade ago, a conference classified as CORE A, the interest for this event seems in a slight decline, reflected also in the lack of highly cited papers.

CISIS: In the case of CISIS, the values are: $h5 = 15$ and $h5_m = 21$, based on which we generated the estimated values $b_{h5} = 315$ and $l_{h5} = 345$.

However, for this conference there is one highly cited paper, and the estimated performance value is $p_{h5} = 0.91$.

CLOSER: Based on GS, for the case of CLOSER we identified an $h5$ value of 19, $h5$ median of 27, and the associated values $b_{h5} = 513$, $l_{h5} = 551$, and $p_{h5} = 513$. Notice that there are no highly cited papers in relation with this conference series.

DS-RT: Finally, DS-RT, even if a smaller event, considering the number of accepted papers, managed to attract some quality papers, with $h5 = 11$, $h5_m = 15$, and $b_{h5} = 165$, $b_{h5} = 187$. Consequently, the estimated performance is $p_{h5} = 0.88$, for this event without highly cited entries.

A synthetic view of these initial findings is included in Table 1, where the performance value is related to the total number of entries for each event.

Table 1. GS metrics and main data

Conference	CORE2023	h5-Index	h5-Median	Papers	Est. Perf.
EuroSYS	A	47	76	287	1.07
ICSOC	A	22	36	473	1.07
AINA	B	24	32	875	0.85
CCGRID	B	24	33	476	0.9
Euro-PAR	B	19	31	455	1.06
CISIS	C	15	21	586	0.91
CLOSER	C	19	27	322	0.93
DS-RT	C	11	15	172	0.88

The information included in Table 1 also allows an initial observation for the conferences under analysis. Based on their estimated performance, we can divide the conferences into two groups.

Potentially High-Performing Group: This group includes, in a first iteration, the two CORE A conferences (EuroSYS and ICSOC) and Euro-PAR, a CORE B event, which has a lower h_5 value but a higher h_{5m} value.

These conferences have an estimated performance of at least 1.00, meaning that their citation distribution is skewed towards the top papers. However, further investigations for Euro-PAR are required, as the initial GS metrics are slightly below the other two CORE B entries.

Potentially Lower-Performing Group: This group includes the rest of the conferences, which have an estimated performance between 0.85 and 1.00.

However, this metric is not very reliable to distinguish the conferences in this group, as slight changes in h_5 or h_{5m} can affect the ranking significantly. Also, these conferences have a more balanced citation distribution, with fewer highly cited papers.

Based on these findings, we can infer that conferences with: a) an h_5 -index of at least 20 and an estimated performance of at least 1 are likely to be CORE A or A* candidates; b) an h_5 -index below 16 and a subunit estimated performance are likely to be classified as CORE C or lower. No initial conclusion can be drawn, based solely on GS data, for the case of CORE B entries.

3.4 Selection. Scopus Data

In this case, we selected the full set of papers from 2018 to 2022 for each of the selected conferences, which matches the GS coverage for computing h_5 and h_{5m} . We applied the following filters to the papers:

- Year range: 2018–2022 (`PUBYEAR > 2017 AND PUBYEAR $<$ 2023`), with a small variation for one conference.
- Document type: Conference paper (`((LIMIT-TO (DOCTYPE , "cp")))`).
- Source title: Manually refined to exclude incorrect entries.
- Subject area: Computer Science (`LIMIT-TO (SUBJAREA , "COMP")`).

We also considered the following semi-automatic analyses to measure the core performance of each conference:

- Counting the number of papers included in Scopus;
- Computing the h-index and i10-index values for each set of papers;
- Estimating an impact factor based on the previous values, with intermediate values, for 2, 3, and 4 years, respectively;
- Exporting the full list of papers and their citations to support further investigations.

The impact factor of each conference was estimated based on the collected Scopus data, following a mechanism which is similar with the original definition of this indicator: the ratio between the number of citations received in a specific year for papers published in the two preceding years, and the total number of indexed papers for the same period of time. For example, the impact factor of ICSOC for 2020 was calculated as the result of the ratio 417/200, where 417 is the number of citations received for the period 2018-2020 and 200 is the number of indexed papers, presented in 2018 and 2019.

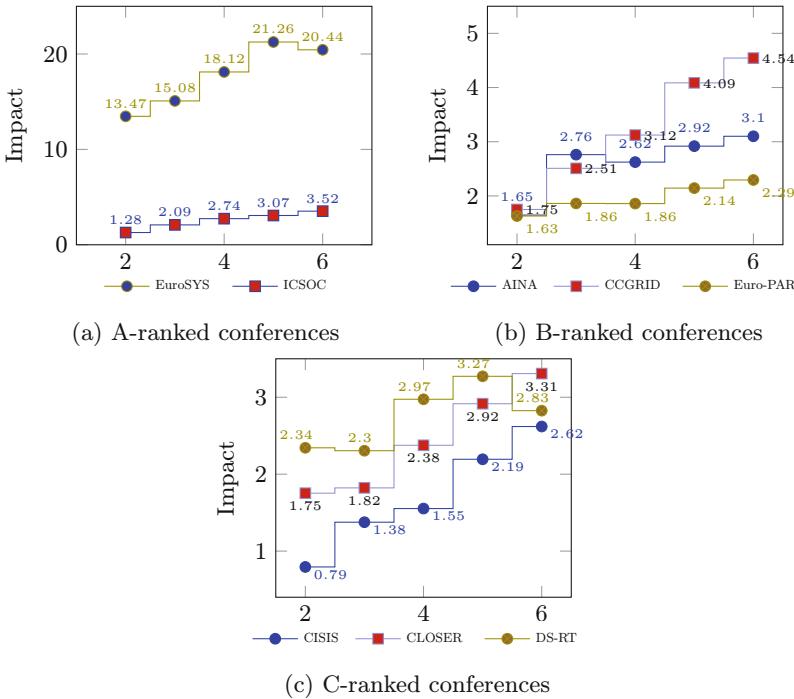


Fig. 1. Impact evolution for the set of conferences

Figure 1 shows the estimated impact of the conferences based on their citation counts, with some unexpected findings. As Fig. 1 illustrates, the first CORE A conference, EuroSYS, has a much higher impact than the other conferences. However, the other seven conferences have similar impact values after five years of analysis.

Moreover, one of the CORE B conferences, CCGrid, has a higher impact than the second CORE A conference, ICSOC. This might not be surprising, considering that CCGrid was, up to the 2023 edition of CORE rankings, a rank A conference. On the other hand, Euro-PAR is offering yet another surprising result: despite being a CORE B event, it has a lower impact than all of the

CORE C conferences considered for our analysis, thus confirming its decline, as indicated in the GS section.

The raw data that was collected from Scopus is synthesized in Table 2. This data include:

1. The total number of indexed conference papers and their yearly distribution from 2018 to 2022;
2. The total number of citations and their yearly distribution for the same period;
3. The computed h-index, which measures the impact and productivity of indexed conference papers. It is expected that this value is close to the h_5 value reported by GS;
4. The i10-index, which counts the number of papers with at least 10 citations. This value was presented as a percent of the total papers for a conference series.

Table 2. Scopus-based findings

Conference	Papers	Citing documents	h-Index	i10-Index (%)	Impact (estimate)
EuroSYS	287	5,866	39	47.39%	20.44
ICSOC	473	1,667	18	9.51%	3.52
AINA	875	2,713	20	9.03%	3.1
CCGrid	476	2,163	20	14.29%	4.54
Euro-PAR	413	1,149	13	5.57%	2.78
CISIS	586	1,551	15	5.97%	2.65
CLOSER	322	1,065	14	8.7%	3.31
DS-RT	172	486	9	4.65%	2.83

Using the Scopus data, one can notice that conferences with an impact factor of 3.00 or higher and an h-index of 18 or higher are likely to be ranked A. Conversely, conferences with low h-index and i10-values are likely to be ranked C or lower. This suggests that EuroSYS is a strong top-tier conference, while ICSOC is a borderline case for rank A outlets. CCGrid could also be considered for rank A. AINA is a solid rank B conference, with a large number of papers. The suggested rank for the rest of the conferences is C or lower.

3.5 Aggregation. Google Scholar and Scopus

Next, by using the GS-based b_{h5} -value, as estimated in Sect. 3.2, and the number of citations (#citations) and the number of indexed conference papers (#papers), as extracted from Scopus database, one can compute the expected impact of the set of conference papers which are outside the set of h_5 papers, by using the following estimate: $I_{h5} = \frac{\#citations - h_5 \times h_5}{\#papers - h_5}$. The resulting values are, in descending order

9.55 (EuroSYS, A), 3.03 (CCGrid, B), 2.4 (AINA, B),
 2.16 (CISIS, C), 1.99 (DS-RT, C), 1.98 (ICSOC, A),
 1.82 (CLOSER, C), and 1.42 (Euro-PAR, B).

To confirm these values, we collected the actual number of citations for the set of h_5 papers, and repeated the calculations, based on the actual data. The main findings are included in Table 3. Based on these information, we can notice that the impact of the group of top h_5 papers is quite important for all conferences.

Table 3. Impact for top h_5 conference papers vs. impact for other conference papers

Conference name	Impact/ h_5 (top h_5)	Impact	
		(w/o h_5)	(estimation, w/o h_5)
EuroSYS	66.25	7.07	(9.55)
ICSOC	17.72	2.25	(1.98)
AINA	13.7	2.38	(2.40)
CCGrid	9.55	3.44	(3.03)
Euro-PAR	8.00	2.19	(1.42)
CISIS	20.66	1.77	(2.16)
CLOSER	16.35	2.07	(1.82)
DS-RT	4.11	2.25	(1.99)

The results for EuroSYS are outstanding, again, showing an impressive impact for the top h_5 papers, under the influence of a group of hot papers, while keeping a large enough impact for the other papers. On the other hand, Euro-PAR is again with lower values, with an impact for the top h_5 papers lower than the GS h_5 -median value, of 31 (for the actual impact value, one have to add the h-index value to the information from Table 3). All other events, except DS-RT, have a top impact value close to the GS h_5 -median and, consequently, there are small differences between the estimated impact for the remainder of papers, and the real impact of the same group of papers.

Based on these findings, one can notice that a conference can qualify for a superior ranking when the computed impact of its top h_5 papers, relative to its h_5 value, and the overall impact of the ‘other papers’ are high enough. Thus, EuroSYS qualify as an excellent outlet, while CCGrid provide a larger base of citable papers, even if the top h_5 papers receive, on average, almost half of the h_5 value, in terms of citations. Also, as the top h_5 papers represent the most cited papers, this computed value loses its relevance and must be considered only as an additional criterion to support impact analysis for the group of ‘other papers’.

On the other hand, conferences with modest impact factors for their top h_5 papers, or for the ‘other papers’, are potential candidates for the third category. A top limit of 10 for the h_5 impact value will put Euro-PAR under the threshold,

together with DS-RT, doubled by a modest impact for the ‘other papers’. On the other hand, despite of its high impact value for the top h_5 papers, CISIS is more likely a superior third rank conference, given its modest impact, of only 1.77, for the ‘other papers’.

4 Discussion

Conference classification is a rather complex process, which is based on various inputs, including event metrics. For the case of CORE conference rankings, we can find that “Conference rankings are determined by a mix of indicators, including citation rates, paper submission and acceptance rates, and the visibility and research track record of the key people hosting the conference and managing its technical program”¹¹.

As our study is limited to the simplest set of meta-information available: the list of published and indexed papers, citations and citation rates in GS and Scopus, h_5 -index, h_5 -median, $i10$ -index from both databases, as well as some derived metrics, it will be able to offer only a tentative classification of these outlets. However, the combination of information that we can gather by using the two data sources have the ability to put a conference event in the right context.

Google Scholar metrics can offer an initial estimate:

- A conference with $h_5 > 20$ and $h_{5m} > 1.5 \times h_5$ has a potential classification similar with CORE A, provided that its estimated performance is above 1.00;
- For a conference with $h_5 < 16$, the potential classification is similar with at most CORE C;

Notice that some events will escape this initial filter. Also the size of an event (total number of published papers) will, most probably, affect its classification.

When an estimation of an impact factor is possible, the performance of all analyzed events is good enough, where the most performing events will receive an impact factor above 3.5, while the lower values are below 3.00. It is expected that events with an estimated impact below 2.0 are unclassified or, at most, CORE C.

However, when we combine the different information that can be computed, we can assess that:

- A CORE A event will have $h - index > 20$, $i10 - index > 10\% \#papers$, and $IF > 4$;
- An entry with $i10 - index < 6\% \#papers$ and $h - index < 15$ or $IF < 3$ will receive at most a CORE C rank.

On the other hand, the suggested category must be confirmed by a good performance for the group of h_5 top papers, as well as a good performance of the papers outside of h_5 area.

¹¹ <https://www.core.edu.au/conference-portal>.

5 Conclusions

Although there is a preference for computer science conferences, as an important dissemination channel, few information on the evaluation of these scientific outlets exist, compared to the available metrics used to evaluate journals. Using similar principles with those established for journal impact factor, or similar metrics, we investigate the usability of conference-related data, which are available via the annual edition of Google Scholar (GS) metrics, or via the Scopus database.

This initial investigation revealed that GS-computed values, like h_5 and h_5 -median, can offer an estimate of the CORE category of a conference, with better results for events from top ranking categories.

At the same time, when there is full coverage in the Scopus database, an estimate of the performance of conference impact factor becomes possible, and the performance of conferences can be thus compared.

References

1. Brookes, B.: “Sources of information on specific subjects” by S. C. Bradford. *J. Inf. Sci.* **10**(4), 173–175 (1985). <https://doi.org/10.1177/016555158501000406>
2. Garfield, E.: The history and meaning of the journal impact factor. *JAMA* **295**(1), 90 (2006). <https://doi.org/10.1001/jama.295.1.90>
3. Harwood, N.: Publication outlets and their effect on academic writers’ citations. *Scientometrics* **77**(2), 253–265 (2008). <https://doi.org/10.1007/s11192-007-1955-x>
4. Li, X., Rong, W., Shi, H., Tang, J., Xiong, Z.: The impact of conference ranking systems in computer science: a comparative regression analysis. *Scientometrics* **116**(2), 879–907 (2018). <https://doi.org/10.1007/s11192-018-2763-1>
5. Meho, L.I.: Using Scopus’s CiteScore for assessing the quality of computer science conferences. *J. Informet.* **13**(1), 419–433 (2019). <https://doi.org/10.1016/j.joi.2019.02.006>
6. Meyer, B., Choppy, C., Staunstrup, J., van Leeuwen, J.: Viewpoint research evaluation for computer science. *Commun. ACM* **52**(4), 31–34 (2009). <https://doi.org/10.1145/1498765.1498780>
7. Naranan, S.: Bradford’s law of bibliography of science: an interpretation. *Nature* **227**(5258), 631–632 (1970). <https://doi.org/10.1038/227631a0>
8. Nisonger, T.E.: The “80/20 rule” and core journals. *Ser. Libr.* **55**(1–2), 62–84 (2008). <https://doi.org/10.1080/03615260801970774>
9. Tahamtan, I., Safipour Afshar, A., Ahamdzadeh, K.: Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics* **107**(3), 1195–1225 (2016). <https://doi.org/10.1007/s11192-016-1889-2>
10. Vrettas, G., Sanderson, M.: Conferences versus journals in computer science. *J. Am. Soc. Inf. Sci.* **66**(12), 2674–2684 (2015). <https://doi.org/10.1002/asi.23349>
11. Winker, K.: In scientific publishing at the article level, effort matters more than journal impact factors: Hard work and co-authors overshadow journal venue in acquiring citations. *BioEssays* **33**(6), 400–402 (2011). <https://doi.org/10.1002/bies.201100020>



Time Series Analysis and Modeling with Federated Learning Techniques in Cloud Edge Scenario: A Case Study on Environmental Air Quality in Homes

Gennaro Junior Pezzullo^{1,2(✉)}, Beniamino Di Martino^{1,3,4}, Oguz Mulayim⁵, and Eva Armengol⁵

¹ Department of Engineering, University of Campania “Luigi Vanvitelli”, Caserta, Italy

beniamino.dimartino@unicampania.it

² Department of Engineering, University of Rome “Campus Bio-Medico”, Rome, Italy

gennaro.pezzullo@unicampus.it

³ Department of Computer Science, University of Vienna, Vienna, Austria

⁴ Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan

⁵ Artificial Intelligence Research Institute (IIIA-CSIC), Barcellona, Spain
{oguz,eva}@iiia.csic.es

Abstract. This study addresses the problem of air quality in homes, focusing on the analysis and research of environmental patterns linked to different domestic activities. Based on a real-world dataset composed of sensor data collected in real time from six homes over a 14-day period, this research aims to identify how daily activities impact indoor air quality using two different approaches: the first in which each household uses its own data to train its own models. The second scenario in which there is a federated collaboration without sharing the data. In particular, through the definition of a detailed methodology, a pipeline is described that starts from data preprocessing, followed by the application of machine and federated learning techniques including the use of K-Means with Dynamic Time Warping (DWT) to detect and associate patterns of air quality variation, leading up to the validation of the results.

1 Introduction

According to the 2021 report from the Institute of Neurological Sciences of Bologna, exposure to air pollution causes around 4.2 million deaths every year worldwide [1]. The study of such data can be mainly divided into two macro areas: causes and effects. More in detail, on the one hand it is necessary to understand what are the causes that positively or negatively modify the quality of the air; on the other hand, it is interesting to understand what are the effects

associated with these changes on human health. The work introduced in the present paper focuses on the first issue. To achieve this objective, starting from a partially dataset of real air quality data of houses, associated with a report of the activities developed in them, we propose two different approaches, single or collaborative methodology, to be able to understand changes in air quality associated to activities. The goal is to better understand the causes of variations in air quality and identify specific patterns associated with different daily activities, in order to comprehend the causes of low air quality and their negative effects on human health [2]. Through the use of pattern recognition, machine and federated learning techniques, we intend to identify the variations in air quality related to the different activities carried out in the monitored houses. The structure of this paper is as follows: in the next chapter we will discuss how the various real data were collected and preprocessed. After this we define a methodology that we follow during the implementation in order to achieve the objective of associating the different patterns to different activities, and at the end we define two different algorithms that will be implemented and validated in a future work.

2 Data Collection

The data from which the development of the project started was collected through the use of sensors deployed in the kitchen and in the living room of six houses. Sensors are capable of measuring the following variables: Humidity, CO₂ (ppm), Formaldehyde ($\mu\text{g}/\text{m}^3$), TVOC (ppb), TVOC Index, PM 1.0 ($\mu\text{g}/\text{m}^3$), PM 4 ($\mu\text{g}/\text{m}^3$), PM 10 ($\mu\text{g}/\text{m}^3$), PM 2.5 ($\mu\text{g}/\text{m}^3$). Data from sensors are taken in regular intervals of 10 min over a 14 day period. In the current paper we will focus on the kitchen's sensors. This choice is motivated by the fact that the kitchen is the room where more activities take place. Regarding these data, it is important to underline that these were treated as time series. This type of approach was found to be fundamental for the following reasons:

- Continuous measurement: in many monitoring contexts, data are intrinsically continuous over time. Using time series therefore allows us to capture trends that develop over time. Furthermore, all parameters can change significantly over time and in this context the use of time series can help us understand their future implications.
- Time intervals: Time series analysis can be performed in different time intervals. Through this data representation we can easily model the time interval of both the single step and the single series. This flexibility guarantees us greater accuracy on tests.
- Temporal Correlation: Often, data collected at successive times are correlated with each other. Recognizing and exploiting these correlations can significantly improve the accuracy.

In time series analysis, handling data errors is critical to maintaining the integrity of the results. Sensors, susceptible to malfunctions such as power failures or

errors, often register zero values. These issues can negatively distort the analysis. To resolve this, the first manipulation we did was to replace the null values with the global averages of the single series. This replacement ensures the consistency of the entire dataset. Furthermore, in our data analysis, we were frequently faced with two main challenges: the former in the presence of short-term fluctuations: these can often obscure more relevant, long-term trends that are critical for analysis. The latter regards noise in the data: random variations that can distort results and make it difficult to identify clear, meaningful patterns. These issues require an effective solution to improve the clarity and reliability of our analyses. The application of a rolling average [3] proves fundamental in this context. By smoothing out short-term fluctuations, the rolling average makes it easier to spot long-term trends. Additionally, by averaging data over a specified period, this technique filters out random variations, exposing clearer, more meaningful patterns. In addition to time series captured, each family living in each one of the homes was asked to compile reports detailing the activities they carried out throughout the period of operation of the sensors. Thus, they report in natural language the activity they carry out and also the time of this (for instance, cooking from 6 pm to 7 pm, watching tv from 9 pm to 11 pm). This type of report is simple to understand for a human, but, for a computer the analysis of such data would lead to many errors. There are two possible paths: the first uses AI techniques such as natural language processing to map the possible scenarios at keywords [4]. The second way, the one we have used, is to do this job manually. The reason why the second approach was chosen was to try to limit the errors caused by data formatting as much as possible with the aim of having measurements as precise as possible. Although more onerous in terms of time, this approach was considered more suitable.

3 Methodology

After having illustrated the preprocessing process of the data at our disposal, in this chapter we define the methodology adopted for identifying patterns associated with the different activities. A crucial aspect to consider is that, since we are dealing with different people in different contexts, we do not know precisely the duration of each activity. The activities themselves, in fact, were reported in a textual file compiled by the owners of the house, which does not allow us to rely completely on the time intervals indicated. Furthermore, we do not have exact knowledge of the total number of activities carried out in the house, but only an estimate provided by the owners themselves through the activity sheet. Therefore, it will be necessary to select an appropriate model for the classification of the different actions, and subsequently define a methodology capable of partially mitigating the inaccuracies of the data mentioned above. Once the methodology is established, it will be essential to translate it into an algorithm to implement and test.

3.1 Choice of Model or Algorithm for Classification

For the first phase of methodology it is necessary to define the model of the algorithm to be used and then verify for each simulation which are the best. Different approaches were tested with the aim of identifying the most suitable one. All the approaches adopted are listed below:

- 1. Euclidean Distance:** A metric that measures the distance between two data points. It is calculated by extracting the square root of the sum of the squared differences between the corresponding coordinates of each point [5]. The formula is as follows:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

where p and q represent the data points in n-dimensional space and n is the number of dimensions. Despite its usefulness, Euclidean distance is remarkably sensitive to noise and outliers. These sensitivities can limit its effectiveness, especially in environments such as homes, where data are influenced by constant external events. Although we have implemented techniques to significantly reduce these noises, Euclidean distance may not always be the most reliable method for assessing similarity between data in these contexts.

- 2. Dynamic Time Warping (DTW):** A technique used to measure the similarity between two temporal sequences that may vary in time or speed. Unlike methods such as Euclidean distance [6], DTW dynamically aligns sequences to minimize temporal differences and calculate an optimal path that highlights the most similar alignment. The formula is as follows:

$$dtw_matrix[i, j] = \min \begin{cases} dtw_matrix[i - 1, j - 1] + d(ts_A[i], ts_B[j]), \\ dtw_matrix[i - 1, j] + d(ts_A[i], ts_B[j]), \\ dtw_matrix[i, j - 1] + d(ts_A[i], ts_B[j]) \end{cases}$$

- 3. Symbolic Aggregate approXimation (SAX):** A time series analysis method that transforms complex data into simple symbols [7]. The series are divided into segments of equal length, the average value of each segment is calculated and they are associated with symbols based on predefined quantiles of a normal distribution. This method is widely used for pattern recognition, clustering, and anomaly identification in time series.

$$\alpha[i] = j \text{ if } q_{j-1} \leq v_i < q_j$$

- α : represents the string of symbols resulting from the SAX transformation.
- i : index that scrolls through the segments of the time series.
- j : symbol index, determined by the position of the segment average value relative to the breakpoints.

- q_{j-1} and q_j : are the breakpoints that define the intervals for the symbolic transformation; these are the quantiles of the normal distribution that divides the value axis into a equally probable intervals.
- v_i : the average value of the i -th segment of the time series, used to determine the corresponding symbol.

4. **Pearson correlation:** A technique that quantifies the linear relationship between two variables [5]. It is used to identify the strength and direction of a linear relationship between two time series. In formulas:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

5. **Levenshtein distance:** Also known as edit distance, it is a metric for measuring the difference between two strings [8]. It represents the minimum number of single modification operations (insertions, deletions or replacements) necessary to transform one string into another.

$$d(a, b) = \begin{cases} |a| & \text{se } |b| = 0 \\ |b| & \text{se } |a| = 0 \\ d(\text{tail}(a), \text{tail}(b)) & \text{se head}(a) = \text{head}(b) \\ 1 + \min \left\{ d(\text{tail}(a), b), d(a, \text{tail}(b)), (\text{tail}(a), \text{tail}(b)) \right\} & \text{altrimenti} \end{cases}$$

Among this, the choice was to use Dynamic Time Warping (DTW) in combination with the K-Means algorithm. This decision was driven by promising preliminary results, which indicate that this combination is particularly effective in specific time series analysis contexts. From here on it is necessary to understand for each series which is the best set of hyperparameters to use for searching for patterns within the series.

3.2 Description of Pattern Identification Algorithm

For the second part of the pipeline, as explained at the beginning of the chapter, it is necessary to structure a logic to mitigate the intrinsic inaccuracies of the data. To do this, it is necessary to define a methodology for evaluating the hyperparameters which is the following: let n_i be the number of activities reported in the report for the house i . We define the set of possible clusters as $K_i = \{k : k \in [n_i, n_i + 5]\}$. For each time interval $T \in \{24, 12, 6, 3\}$ hours, we perform the following procedure:

1. Divide the time series into segments of length T .
2. For each $k \in K_i$, apply clustering to split the segments into k clusters.
3. Identifies for each cluster j , the activity \hat{a}_{ij} that maximizes the frequency of appearance in the cluster.

The comparison between the identified \hat{a}_{ij} activities and those reported in the report determines the validity of the association for each cluster. We define the success rate for house i as:

$$\text{Success Rate}_i = \frac{\text{Number of matches between } \hat{a}_{ij} \text{ and the report}}{\text{Total number of activities in the report}} \times 100$$

3.3 Motif Search

In order to identify the pattern associated with an activity after having clustered the various time series according to the best parameters, it is necessary to identify the pattern. Therefore, through the use of specific techniques, the associated motif will be identified for each best result [9]. These will then be compared between the different homes to verify whether, given 2 identical activities, they have the same motif. The idea behind this process is to take all the cluster centers associated with the same activities and, on them, recalculate the clusters to further isolate the most representative motif. This second phase of clustering is crucial because it allows refining the search for the pattern, focusing exclusively on the data that represents the core of similar activities. In practice, this means that the initially grouped data undergoes further selection to more precisely identify common patterns that characterize a specific activity in different environments.

4 Detail of the Algorithm

For implementation the first thing to define was the development environment. In this case, due to the type of project and algorithm, the choice fell on Google Coolab [10] for easy of use of libraries. Python was used exclusively as the programming language. The flexibility of this language, combined with the vast number of libraries for manipulating data frames and the use of artificial intelligence techniques, has certainly made this language the best choice. Among the main libraries to use we certainly have: pandas for the entire dataframe manipulation part and in general for developing what was mentioned in the methodology in the "Preprocessing" section, matplotlib for the visualization of the various graphs and tslearn for what concerns the use of KMeans. In particular, in this specific case, a function specifically specialized for TimeSeries called TimeSeriesKMeans was used. Regarding the algorithm described in the methodology two different versions were created using the machine and federated leaning respectively.

4.1 Machine Learning Approach

In the first scenario, the single user algorithm allows all the data from a single houses to be analyzed. Therefore, the pseudo code associated with this algorithm is shown below.

Algorithm 1. Analysis of activities for home

```

1: for each house  $i$  do
2:    $n_i \leftarrow$  number of activities in the report
3:   for  $K_i = n_i - 2$  to  $n_i + 2$  do
4:     for  $T \in \{24, 12, 6, 3\}$  do
5:       Divide the time series into intervals of  $T$  hours
6:       Perform clustering with  $K_i$  cluster
7:       for each cluster  $j$  do
8:         Counts the number of activities  $r_{ij}$ 
9:         Finds the most frequent activity in cluster  $j$ 
10:        Associates the most frequent activity with cluster  $j$ 
11:      end for
12:      Compares the cluster activities with those in the report
13:      Calculates the success rate
14:    end for
15:  end for
16: end for

```

The single iteration of this pseudo code gives the following output as a result Fig. 1. The figure shown represents one of the 900 simulations carried out to test the algorithm. It illustrates the clustering process, where each line within a cluster represents an activity pattern over time. The red line in each cluster identifies the centroid [11], which is the most representative pattern of that cluster's activities. This visualization is crucial for understanding the variations and commonalities within the activities of each house, providing a clear depiction of how activities are grouped according to similarity in different time intervals. Overall, however, while this approach is advantageous in terms of privacy, considering different bitrates within the model could further increase the accuracy rate.

4.2 Federated Learning Approach

The use of a federated approach in a context of this type represents a key resource for two main reasons: it allows maintaining privacy on devices by not directly sharing the data, but rather the weights of the associated models and also move on server side a part of computation. In this case the algorithm changes slightly as the activities are first identified on the individual houses and subsequently aggregated. The underlying idea is to send only the models with the related cluster centroids and then be able to carry out the aggregation and average of the centroids [12]. Specifically, the algorithm is the following:

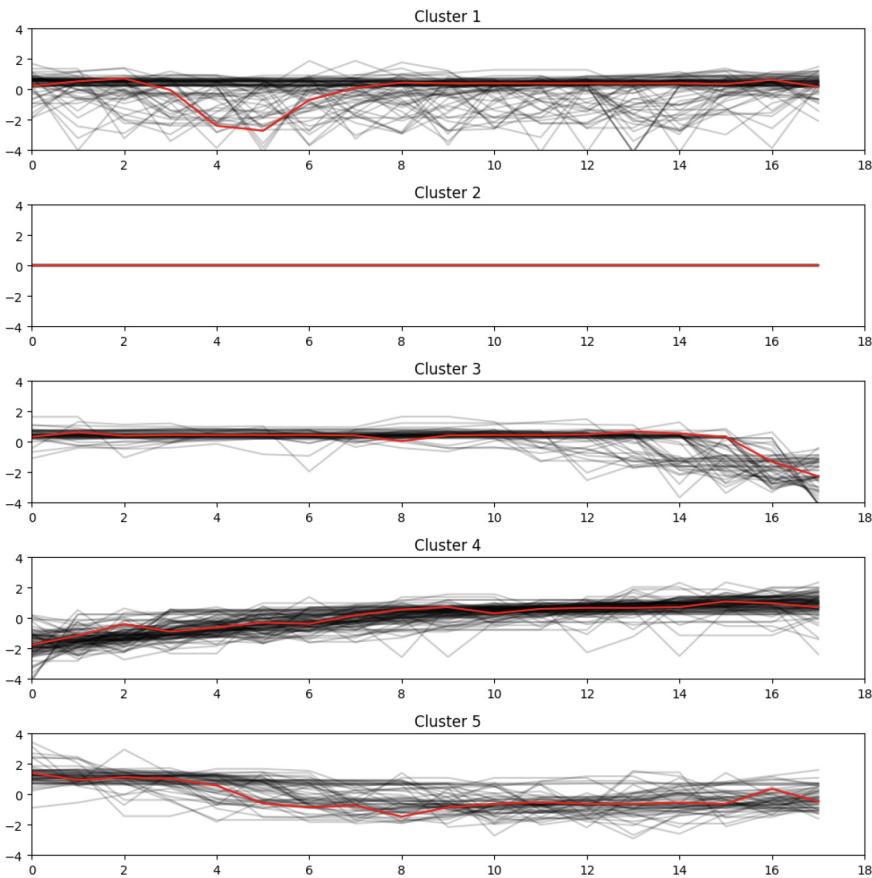


Fig. 1. Single iteration of defined code

Algorithm 2. Client-Side Algorithm (for each house)

```

1: Initialize:  $n_i \leftarrow$  Number of tasks in the report
2: for  $K_i = n_i - 2$  to  $n_i + 2$  do
3:   for  $T \in \{24, 12, 6, 3\}$  do
4:     Divide the time series into intervals of  $T$  hours
5:     Cluster the data locally into  $K_i$  clusters
6:     for each cluster  $j$  do
7:       Count the number of activities  $r_{ij}$  in cluster  $j$ 
8:       Identify the most frequent activity in cluster  $j$ 
9:       Associate the most frequent activity with cluster  $j$ 
10:    end for
11:    Send cluster centroids to the central server
12:  end for
13: end for
14: Wait for the global model from the central server

```

Algorithm 3. Server-Side Algorithm

```

1: Initialize: Receive centroids from all clients
2: for each aggregation cycle do
3:   Aggregate centroids received from all clients to form a global model
4:   Send the global model to all clients
5: end for
6: Repeat aggregation cycle as needed

```

5 Conclusion and Future Developments

In conclusion, this study proposes a detailed methodology to analyze the impact of domestic activities on air quality in homes. This methodology represents only a starting point and requires further validation. The next phase will include the definition of several case studies and the application of various measures to verify the effectiveness and accuracy of the results obtained. Only through this validation process will it be possible to confirm the robustness of the proposed approach and its applicability in real contexts, thus contributing to a better understanding and management of indoor air quality. Furthermore fine-grained approach, for example differentiating the types of cleaning or household appliances used, could guarantee more precise results. In addition, the use of more edge components [13] to be able to better divide the tasks and optimize the entire process could be another important factor to consider [14].

References

1. Della Qualità, D.: Valutazione Sanitaria Della Qualità Dell'aria a Bologna 2020 (2021)
2. Jiang, S.-Y., Ma, A., Ramachandran, S.: Negative air ions and their effects on human health and air quality improvement. *Int. J. Mol. Sci.* **19**(10), 2966 (2018)
3. Eshragh, A., Livingston, G., McCann, T.M., Yerbury, L.: Rollage: Efficient Rolling Average Algorithm to Estimate ARMA Models for Big Time Series Data (2021). arXiv preprint [arXiv:2103.09175](https://arxiv.org/abs/2103.09175)
4. Di Martino, B., Pezzullo, G.J., Grassia, E.: Support for automated story telling using natural language processing techniques aimed at recognizing narrative elements. In: International Conference on Emerging Internet, Data & Web Technologies, pp. 607–616. Springer (2024)
5. Berthold, M.R., Höppner, F.: On Clustering Time Series Using Euclidean Distance and Pearson Correlation (2016). arXiv preprint [arXiv:1601.02213](https://arxiv.org/abs/1601.02213)
6. Senin, P.: Dynamic time warping algorithm review. *Inf. Comput. Sci. Dept. Univ. Hawaii Manoa Honolulu USA* **855**(1–23), 40 (2008)
7. Yu, Y., Zhu, Y., Wan, D., Liu, H., Zhao, Q.: A Novel Symbolic Aggregate Approximation for Time Series. In: Lee, S., Ismail, R., Choo, H. (eds.) IMCOM 2019. AISC, vol. 935, pp. 805–822. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-19063-7_65
8. Behara, K.N., Bhaskar, A., Chung, E.: A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies* **111**, 513–530 (2020)

9. Huo, H., Zhao, Z., Stojkovic, V., Liu, L.: Optimizing genetic algorithm for motif discovery. *Math. Comput. Model.* **52**(11–12), 2011–2020 (2010)
10. Bisong, E., Bisong, E.: Google colaboratory. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform: a Comprehensive Guide for Beginners, pp. 59–64 (2019)
11. Zhang, Z., Zhang, J., Xue, H.: Improved k-means clustering algorithm. In: Congress on Image and Signal Processing, vol. 5, pp. 169–172. IEEE (2008)
12. Cui, L., Su, X., Zhou, Y., Zhang, L.: Clustergrad: adaptive gradient compression by clustering in federated learning. In: GLOBECOM 2020–2020 IEEE Global Communications Conference, pp. 1–7. IEEE (2020)
13. Di Martino, B., Pezzullo, G.J., Low, W., Ljungberg, P., Saha, S.: Survey on reference architecture for cloud continuum and multi-access edge computing (MEC) in 5g networks. In: International Conference on Advanced Information Networking and Applications, pp. 141–150. Springer (2024)
14. Pezzullo, G.J., Di Martino, B.: Artificial intelligence techniques for dynamic offloading in cloud continuum environment: a review. In: International Conference on Complex, Intelligent, and Software Intensive Systems, pp. 405–412. Springer (2024)



Cloud Framework for Data Practitioners for Research and Higher Education Community

Shruthi Sreenivasa Murthy¹(✉), Krishna Chaitanya Rao Kathala²,
and Guangli Zhang³

¹ Saint Louis University, 219 ISE Building, 240 North Grand, St. Louis, MO 63103,
USA

shruthi.sreenivasamurthy@health.slu.edu

² University of Massachusetts Amherst, Office of Vice Provost for Global Affairs, 70
Butterfield Terrace, Amherst, MA 01003, USA

kkathala@umass.edu

³ Saint Louis University, Fusz Hall 366, 3700 W Pine Mall Blvd,
St. Louis, MO 63108, USA

guangli.zhang@slu.edu

Abstract. This paper presents a state-of-the-art cloud-based system for academic research GPS data processing on a big scale. The framework tackles the major issues of cost-effectiveness, processing efficiency, and data volume, which are frequently obstacles to utilizing location-based insights. The suggested method makes use of a thorough Geo Data Warehouse strategy and makes use of Apache Spark, AWS services, and Apache Sedona. This integrated solution significantly reduces computing time and resource needs by streamlining the whole data lifecycle, from ingestion to analysis. This framework's main contributions are its automated ETL orchestration, powerful post-processing phase designed for geographic analysis, and effective data-splitting algorithms. These developments have made large-scale location data available to researchers in a variety of fields and allowed for significant gains in cost savings, analytical power, and data handling efficiency. Best practices for putting such a framework into place and keeping it up to date in academic settings are also covered in the article. To guarantee the long-term acceptance and application of this framework entails attending to important issues like security, data governance, and knowledge transfer, improving research capacities, encouraging interdisciplinary cooperation, and democratizing access to sophisticated data analytic methods in higher education are the main objectives of this endeavor. This framework enables researchers to fully utilize location-based insights, spurring creativity and discovery in a variety of academic domains by tackling operational and technical obstacles.

1 Introduction

The exponential rise of available data, combined with the increasing complexity of analytical tasks, has revolutionized the landscape of academic research. This transition is especially visible in the field of economic research, where the availability of massive volumes of raw location data has created new pathways for insight and discovery. GPS signals, which number in the billions, provide unparalleled chances to research human movement patterns, economic behaviors, and geographical relationships at a granular scale. However, this quantity of data poses substantial hurdles for researchers as well as institutions alike.

The importance of cloud frameworks in higher education research cannot be overstated in this context. According to Berisha et al. [3]: “Cloud computing has emerged as a critical infrastructure for managing and analysing big data in academic settings.” These frameworks offer the scalability, flexibility, and processing capacity required to handle the massive datasets that characterize modern research projects. They enable researchers to bypass traditional hardware limitations and gain access to cutting-edge tools and technologies that were previously only available to huge firms or specialized research institutes [7].

Despite these advantages, data practitioners in academia face a distinct set of obstacles. The massive amount of data, frequently measured in terabytes, necessitates sophisticated pre-processing and storage solutions. Shah [20] states that “the complexity of big data in academic research often outpaces the technical infrastructure and skills available within many institutions”. This disparity between data availability and institutional preparation is a substantial challenge for researchers attempting to use large-scale datasets.

These issues are especially significant in the context of GPS and position data. To be ready for analysis, the data, which is often given in formats such as Parquet files, must go through considerable pre-processing. Storage of such massive volumes of data provides yet another challenge, with cost-effectiveness and long-term accessibility coming into play. Furthermore, the need for effective data retrieval using specified geocodes and date ranges complicates the study process [11].

The purpose of this study is to solve these issues by presenting a comprehensive cloud-based data pipeline tailored to the demands of researchers and data practitioners in higher education. This framework seeks to simplify the processing, storage, and retrieval of large-scale location data, making it more accessible and controllable for academic study. According to Hussain et al. [12], “Efficient data pipelines are crucial for unlocking the full potential of big data in academic research.”

Our proposed solution takes a holistic approach to data management, from initial ingestion to final analysis. It uses cloud technology to deliver scalable computer resources, automated processing workflows, and low-cost storage solutions. By using this methodology, we want to dramatically minimize the time and resources necessary for data preparation, allowing researchers to concentrate on analysis and insight production. The scope of this work goes beyond the technical implementation. We address the framework’s broader implications

for the higher education community, such as its ability to stimulate collaboration, improve research skills, and democratize access to advanced data analysis tools. According to Mahony [15], “cloud frameworks in academia have the potential to level the playing field, allowing smaller institutions to compete with larger, better-funded research centers.”

In the following sections, we will look at the specifics of our proposed cloud framework, including its design, implementation considerations, and potential benefits to the research community. We will also examine the problems of integrating such systems into existing academic infrastructures and provide solutions to overcome these obstacles. This framework intends to speed research, improve data analysis quality, and contribute to more rigorous and insightful economic studies by addressing the specific demands of data practitioners in higher education. As we traverse the complexity of big data in academia, such solutions will become increasingly important for maintaining the competitiveness and relevance of higher education research in a quickly changing digital context.

2 Literature Review

The difficulties of dealing with enormous datasets in academic research have grown more apparent as data volume and complexity have increased. Researchers from various disciplines, particularly economics and social sciences, are dealing with datasets that far beyond standard processing capabilities. In the words of Philip Chen et al [16], “The era of big data has transformed the landscape of academic research, presenting both unprecedented opportunities and formidable challenges.” One of the most significant challenges that researchers encounter is the enormous amount of processing time required for large-scale data analysis. Traditional computing infrastructure frequently struggles to manage terabytes of data efficiently. This constraint can cause major delays in research initiatives and prevent the timely publication of findings [2]. Another important consideration is cost, as high-performance computing resources are typically too expensive for many academic institutions [22]. Infrastructure restrictions exacerbate these challenges since many universities lack the appropriate gear and software to adequately analyze and store enormous datasets [14]. Existing methods for managing huge amounts of GPS data in academic contexts have limits in terms of scalability, cost-effectiveness, and dependability. G.S. Bhunia [5] made the observation that “current methods for processing geospatial big data often fail to meet the demands of real-time analysis and suffer from high computational costs.” Traditional GIS methods, while effective for smaller datasets, frequently struggle with the volume and velocity of data produced by current GPS systems.

The introduction of cloud computing has created new opportunities to address these difficulties. Cloud-based solutions provide scalability and flexibility, which are ideal for the changing demands of academic research. However, as Darvish Dina (2024) notes, “while cloud computing provides a promising platform for big data analytics in academia, issues of data security, privacy, and cost management remain significant concerns” [6]. Data pipelines have developed as

an important tool for handling large-scale data processing. These automated workflows improve the transportation and transformation of data, allowing for more efficient analysis [17]. However, the adoption of effective data pipelines in academic contexts has been hampered by a lack of standardized methodologies and the requirement for specific technical knowledge.

Several technologies and tools have emerged as effective solutions to these difficulties. AWS S3 has grown in popularity as a data storage option because to its scalability and cost-effectiveness. The Parquet file format, with its effective compression and encoding algorithms, has proven useful for storing huge datasets. Apache Spark, a unified analytics engine, has been widely used due to its ability to process enormous amounts of data quickly. In the words of Sham-sinejad: “Spark’s in-memory processing capabilities have significantly reduced computation times for complex data analysis tasks” [21]. AWS Elastic MapReduce (EMR) has emerged as an effective tool for executing large data frameworks such as Hadoop and Spark in the cloud. Its capacity to flexibly scale resources makes it ideal for the fluctuating workloads seen in academic research. Apache Sedona, a cluster computing framework for processing large-scale spatial data, has shown promise in handling geospatial big data efficiently [19].

The cost of storing and processing large GPS datasets can be financially challenging for the research community. Traditional HPC clusters are often inadequate when it comes to handling large datasets, specifically with frequent and periodic data ingestion. The storage and retrieval also is limited to the availability of on-prem resources.

Data security and compliance are also a very important consideration when handling GPS data. We are proposing a framework on the cloud that could address all these challenges. We aim to provide scalable and robust infrastructure, provide secure and compliant data management, optimize cost and processing efficiency, and enable researchers to analyze large-scale GPS data effectively.

Despite these advances, there is still a considerable vacuum in systems designed expressly for the needs of university researchers dealing with large-scale GPS data. Existing frameworks can need significant technical skills to build and maintain, posing challenges for many researchers [10]. Furthermore, integrating these tools with current academic IT infrastructures poses continual hurdles. The suggested architecture in this research seeks to fill these gaps by providing a comprehensive, user-friendly solution that takes advantage of cloud computing and modern data processing technologies. By integrating the strengths of solutions like as AWS S3, EMR, and Apache Sedona, the framework aims to provide a scalable, cost-effective, and efficient solution for processing and analyzing huge amounts of GPS data in academic contexts.

3 Proposed Framework

We propose a comprehensive Geo Data Warehouse solution to address the challenges of processing large-scale GPS data in academic settings. Our framework is designed to streamline data handling, improve efficiency, and reduce costs for

economic research and other disciplines requiring extensive location data analysis, as shown in Fig. 1.

We begin by eliminating intermediate storage, which greatly lowers costs and processing overhead. By implementing a partitioning method in the format of ($\text{geoShort} = \text{xxx/year} = \text{xxxx/month} = \text{xx/day} = \text{xx}$), we minimize storage and increase data retrieval efficiency [18]. For implementation, we use AWS Elastic Map Reduce (EMR) in conjunction with Apache Spark, which provides the scalable data processing capabilities required to manage large datasets. We also use Infrastructure as Code (IaaC) principles, using AWS Step Functions to automate EMR cluster construction and configuration [8]. This automation extends to incremental data processing, ensuring that our system remains efficient even as new data is introduced.

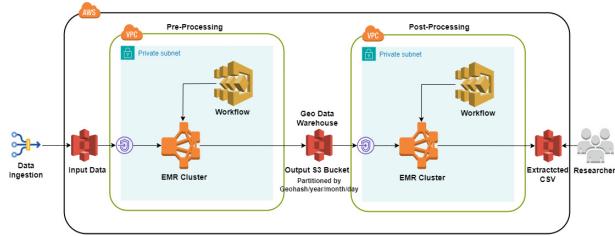


Fig. 1. Framework for Data Practitioners

Our framework incorporates a number of critical optimizations. We boost calculation speed significantly by using Spark DataFrames and adding data broadcasting to executors. An efficient data repartitioning procedure eliminates small-sized outputs, which improves S3 write speeds and cost-effectiveness. The usage of Gp3 volumes improves I/O operations, hence increasing overall performance [23]. Data is uploaded and saved in an S3 bucket, categorized by year, month, and day. For long-term storage, we preserve the data in S3 Glacier and keep it in Parquet format for easy retrieval when needed [9].

Time zone mapping and cleansing is a critical stage in our approach. We assign a time zone to each record using a geohash (6-character) mapping from a time JSON file, which includes Date and Time columns. We've developed a method to prevent duplication that can arise when a geohash is mapped to different time zones, ensuring data accuracy. Our framework's core is based on data processing and partitioning. We read raw Parquet data, convert it to Spark DataFrames, and remove duplicates by mapping data to the appropriate time zone using geohash mappings [4]. To enhance this process, we broadcast the time zone DataFrame to all worker nodes, allowing for more efficient joining with geohash codes. Unix time is then transformed into local time for each record, and data is repartitioned based on geohash, year, month, and day. For storage in our Data Warehouse, we save processed data in an S3 bucket, divided by a 3-character geohash, year, month, and date [1]. This structure allows quick

retrieval of specific files from Glacier, accelerating the data input process during post-processing.

To improve the usefulness of our system, we included a comprehensive post-processing phase that uses Apache Sedona for geographical data analysis. This step uses single and double-pass processing to extract useful information based on the researchers' individual needs. Apache Sedona, a distributed computing engine for large-scale geographical data, provides numerous advantages. We lowered processing time dramatically by dividing polygon operations across executor nodes. Our innovative methodology reduces processing time by more than half when compared to traditional methods, allowing for lower instance capacities and significant cost reductions.

Our post-processing approach begins with researchers submitting input shapefiles containing polygon features relevant to their research, as shown in Fig. 2. We then retrieve the necessary data from our Geo Data Warehouse. In the single-pass processing stage, we use the polygon information from the shapefiles to filter warehouse data, reducing the dataset to only records that meet the specified geographical requirements. The single-pass approach enables researchers to analyze foot traffic data at the point-of-interest (POI) level. To achieve more precise results, we use a double-pass processing stage, which refines the filtered data by matching it with Cell phone ID (CAID), yielding a CAID-level panel dataset. This results in a highly focused dataset for researchers. The last step is geographic analysis, which combines the filtered and extracted data into a single CSV file for further analysis with popular tools like GeoDa, STATA, python, and ArcGIS.

By integrating this Geo Data Warehouse and post-processing methodology, we can greatly improve performance and scalability while lowering expenses. Our framework offers academics an effective and dependable tool for economic data analysis, addressing present and future data processing needs in the academic community. We invite collaboration and comments to improve and adapt this solution to the changing needs of data-intensive academic research.

Because of its modular architecture, the framework can easily process datasets that go beyond GPS data, including data on education, health, the economy, and the environment. Its key components, which include the usage of Spark DataFrames and automated ETL orchestration via AWS Step Functions, make it flexible and scalable for a variety of data sources, including unstructured and structured data. For more complex data processing requirements, more domain-specific libraries can be used in place of tools like Apache Sedona. The framework's adaptability is demonstrated, for example, by substituting PySpark or Pandas for Sedona in non-geospatial analysis scenarios, or by incorporating various file formats like CSV, JSON, or Avro. With these modifications, the system can support several academic disciplines while maintaining efficiency and cost-effectiveness. According to Johnson, E. et al. [13], cloud frameworks can handle any large-scale dataset with the right setups, not just GPS data. This is because they provide scalability and flexibility, which are essential for solving a variety of data analysis difficulties.

We chose AWS Step Functions to manage our ETL process because we needed a flexible solution that could respond to changing research needs. Step Functions give us fine-grained control over the ETL process, allowing us to use alternative services for specific pipeline stages while maintaining the overall design. The versatility of Step Functions has been invaluable in our research setting. For example, when we wanted to add a new data cleaning step to accommodate an unexpected format in certain GPS data, we smoothly integrated it into our pipeline without interrupting any existing operations. This amount of adaptability means that our ETL orchestration is strong and can evolve in tandem with our research demands.

We used a rigorous code review process with several team members to verify code quality and detect any issues early in the development cycle. Our testing technique goes beyond simple success scenarios, integrating edge cases and stress tests to ensure the pipeline's functionality under varying conditions. This comprehensive approach has been critical in identifying and fixing possible bottlenecks or failure areas before they affect our study data. By recognizing these issues early on, we've been able to consistently improve our pipeline's dependability and performance.

Given the various sources and formats of GPS data we encounter, data validation has emerged as an important stage in our workflow. Prior to processing, we conducted rigorous validation checks on both metadata and data volumes. This phase has proven useful in detecting user input problems and finding non-conforming files early in the pipeline. Any file that does not satisfy our preset criteria is instantly flagged for additional analysis. This approach prevents contamination of our core dataset and protects the integrity of our research data, which is critical for sustaining the validity of our findings.

To keep track of the costs related to our big data processing, we set up a budget and cost alert system. These alerts are calibrated to match our predicted workload costs and are configured to notify us of any deviations at both the individual workload and AWS account levels. This proactive approach to cost management has enabled us to optimize resource utilization while avoiding unexpected charges, which is critical when managing research budgets. It has also given us significant insights into the cost implications of various processing methodologies, which have helped us make decisions on pipeline optimization.

We have enabled thorough monitoring for all services engaged in our workload, recording important parameters that reflect the pipeline's health and performance. This ongoing attention enables us to immediately detect and address any difficulties that develop, maintaining the seamless operation of our data processing processes. Our monitoring system has proven especially useful for large-scale data processing operations, allowing us to identify and address any bottlenecks in real-time. This feature has substantially increased the efficiency of our research processes.

To tie all of these monitoring efforts together, we created a single CloudWatch dashboard for each task. This dashboard gives us a consolidated view of all major indicators across our services, providing invaluable insights into the overall per-

formance of our pipeline. During test runs, this dashboard has been especially valuable in validating that our process is performing as intended and rapidly highlighting any anomalies that require attention. It has also facilitated communication among our research team members, offering a common understanding of pipeline performance that influences our conversations and decision-making processes.

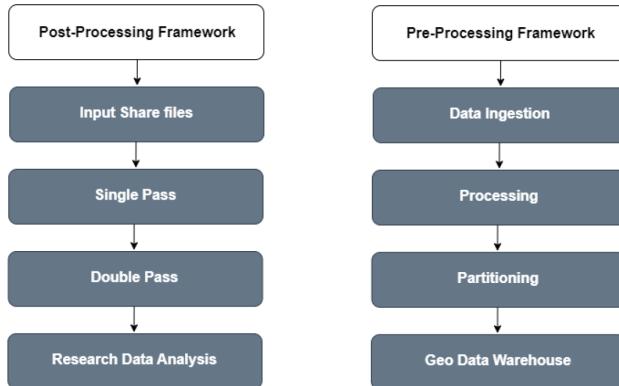


Fig. 2. Pre-Post Processing Framework

By implementing these additional steps, we've significantly enhanced the robustness and reliability of our big data pipeline. This proactive approach to pipeline management has not only met our current research needs but has also positioned us well to adapt to future challenges in GPS data processing and analysis. As we continue to refine our approach, we remain committed to sharing our experiences with the broader research community, contributing to the development of more efficient and reliable data processing frameworks across various fields of study.

4 Benefits of the Enhanced Post-processing Approach

Our proposed framework has shown significant advantages by leveraging Apache Sedona and distributed processing techniques, improving our handling and analysis of large-scale geospatial data.

Integrating Apache Sedona has transformed our data processing. We reduced processing times by half by distributing difficult polygon operations across numerous executor nodes, as compared to earlier methods. This is especially useful for huge datasets, as it enables us to process terabytes of GPS data quickly. The distributed architecture has also decreased computational resource requirements, allowing us to manage larger datasets with lower-capacity instances while optimizing computing resources.

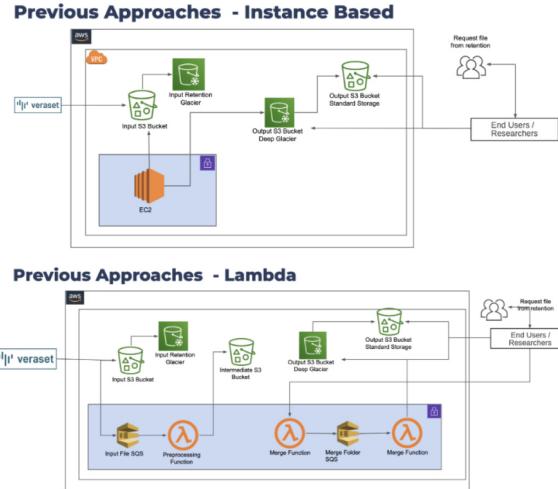


Fig. 3. Previous Approaches

Improved processing efficiency has resulted in significant cost reductions. Our operational expenditures have lowered due to reduced processing time and necessary instance capacity. This cost-effectiveness is critical in academic research, where funding is sometimes limited. Our improved technique allocates more resources to research tasks rather than computational overhead. In Fig. 3, we show two previous approaches. Compared to the previous lambda-based approach, we reduced costs by over 55% for processing 5 years of data. Compared to the instance-based approach, we reduced costs by over 70%. Our robust performance improvement plan has significantly reduced costs and processing time, and enhanced workflow efficiency.

Double-pass processing has enhanced our datasets to satisfy specific study requirements. We ensure extremely relevant and personalized datasets by filtering using geographical parameters and linking them with Census Area ID data. This tailored method increases the accuracy of our analysis while reducing dataset noise, laying a solid platform for economic and geospatial research.

Our post-processing output, a thorough CSV file, greatly improved our analytical skills. This standard format works perfectly with complex analytical tools like GeoDa, STATA, and ArcGIS. Researchers may now execute complex spatial econometrics, develop precise visualizations, and conduct advanced statistical analyses with more efficiency, resulting in more insightful research outputs.

5 Best Practices for Data Practitioners

In developing and implementing our Geo Data Warehouse and post-processing framework, we've identified several best practices that are crucial for success in academic research environments.

We discovered that having a secure cloud infrastructure is critical to the success of our research endeavors. This includes installing strong access controls, encryption techniques, and conducting frequent security audits. We utilize AWS Identity and Access Management (IAM) to manage user permissions in granular detail, ensuring that researchers only have access to the resources they require. In addition, we've used virtual private clouds (VPCs) to isolate our research environments, providing an extra layer of protection. Regular training sessions on cloud security best practices have been critical to ensuring the integrity of our research data.

Effective data governance has been critical for handling our large-scale geospatial datasets. We've created explicit procedures for data access, usage, and sharing to ensure compliance with institutional and funding agency standards. Our data management strategy includes extensive methods for data storage, backup, and archiving, with an emphasis on long-term accessibility and reproducibility of research findings.

We have put in place rigorous monitoring systems to keep our data pipeline healthy and efficient. We use AWS CloudWatch to track important metrics and set up automated notifications for anomalies or performance issues. This proactive strategy enables us to address possible issues before they affect our research work.

Giving researchers proper access to our cloud infrastructure has been critical in encouraging collaboration and innovation. We created a tiered access system that strikes a balance between security concerns and the researchers' desire for flexibility. This provides pre-configured environments for simple analysis activities, as well as sophisticated access choices for expert users.

Perhaps most crucially, we have acknowledged the importance of continuous education and knowledge transmission within our research community. We propose holding regular workshops and training sessions on topics ranging from fundamental cloud computing ideas to advanced data analysis techniques on our infrastructure. These seminars not only improve our researchers' technical skills but also foster a culture of data-driven research and collaborative problem-solving. By creating a knowledge-sharing atmosphere, we ensure that our infrastructure and techniques evolve and improve over time.

6 Conclusion and Future Scope

The suggested cloud-based data pipeline offers a comprehensive answer to the issues that academics confront when working with enormous amounts of GPS data. The framework uses scalable cloud technologies and automated procedures to drastically minimize the time and resources necessary for data preparation, allowing researchers to focus on analysis and insight development. The integration of tools like as AWS S3, EMR, and Apache Sedona provides effective data management, storage, and retrieval, meeting the specific requirements of academic research. The framework's capacity to promote cooperation and democratize access to advanced data analysis tools holds great promise for improving

research skills in higher education. This method not only enhances the efficiency and quality of present research, but also serves as a solid platform for future advancements in data-intensive academic studies. Future work will focus on improving the data pipeline's ability to handle larger datasets and more complicated analytical jobs. Enhancements in data security, privacy, and cost management will be investigated to ensure the framework's strength and sustainability. The application of advanced machine learning algorithms for predictive analysis and real-time data processing will be examined. Furthermore, the framework's relevance to domains other than economic study, such as environmental studies and public health, will be explored. Continuous engagement with the research community will be required to adapt the framework to changing needs and integrate feedback for future enhancements.

Acknowledgements. We would like to thank Michael Podgursky, SCAER's Director, Saint Louis University, for supporting this work. We definitely couldn't have achieved this without all the knowledge and resources that SLU has offered.

References

1. Al-Mamun, A., Wu, H., Aref, W.G.: A tutorial on learned multi-dimensional indexes. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, pp. 1–4 (2020)
2. Asch, M., et al.: Big data and extreme-scale computing: pathways to convergence—toward a shaping strategy for a future software and data ecosystem for scientific inquiry. *Int. J. High Perform. Comput. Appl.* **32**(4), 435–479 (2018)
3. Berisha, B., Mëziu, E., Shabani, I.: Big data analytics in cloud computing: an overview. *J. Cloud Comput.* **11**(1), 24 (2022)
4. Bhatlawande, S., Rajandekar, R., Shilaskar, S.: Implementing middleware architecture for automated data pipeline over cloud technologies. In: 2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT), pp. 506–513. IEEE (2024)
5. Bhunia, G.S., Shit, P.K.: Big data analysis for sustainable land management on geospatial cloud framework. In: Geospatial Practices in Natural Resources Management, pp. 3–17. Springer (2024)
6. Darwish, D.: Emerging Trends in Cloud Computing Analytics, Scalability, and Service Models (2024)
7. Das, D., Nayak, M.: Big data analytics: an overview. In: Applications of Machine Learning in Big-Data Analytics and Cloud Computing, pp. 271–287 (2022)
8. Das, P., Begum, S.A., Buyya, R.: Advanced Computing, Machine Learning, Robotics and Internet Technologies: First International Conference, AMRIT 2023, Silchar, India, March 10–11, 2023, Revised Selected Papers. Springer Nature (2024)
9. Diamantini, C., Mircoli, A., Potena, D., Tempera, V., Moretti, M.: Workload-driven database optimization for cloud applications. In: 2017 International Conference on High Performance Computing & Simulation (HPCS), pp. 595–602. IEEE (2017)
10. Dwivedi, Y.K., et al.: “real impact”: Challenges and Opportunities in Bridging the Gap Between Research and Practice—Making a Difference in Industry, Policy, and Society (2024)

11. Hamed, N., Rana, O., Orozco-terWengel, P., Goossens, B., Perera, C.: A Comparison of Open Data Observatories (2024)
12. Hussain, M., Zhang, T., Seema, M.: Adoption of big data analytics for energy pipeline condition assessment. *Int. J. Press. Vessels Pip.* **206**, 105061 (2023)
13. Johnson, E., Seyi-Lande, O.B., Adeleke, G.S., Amajuoyi, C.P., Simpson, B.D.: Developing scalable data solutions for small and medium enterprises: challenges and best practices. *Int. J. Manag. Entrep. Res.* **6**(6), 1910–1935 (2024)
14. Mafukidze, H.D., Nchibvute, A., Yahya, A., Badruddin, I.A., Kamangar, S., Hussien, M.: Development of a modularized undergraduate data science and big data curricular using no-code software development tools. *IEEE Access* (2024)
15. Mahony, S.: Toward openness and transparency to better facilitate knowledge creation. *J. Am. Soc. Inf. Sci.* **73**(10), 1474–1488 (2022)
16. Philip Chen, C., Zhang, C.Y.: Data-Intensive Applications, Challenges, Techniques and Technologies: a Survey on Big Data (2014)
17. Polimetla, K., Jenny, F.: Spearheading big data solutions: optimizing data pipelines for enhanced efficiency and performance. *Educ. Adm. Theory Pract.* **30**(6), 4106–4116 (2024)
18. Ponnusamy, S., Gupta, P.: Scalable data partitioning techniques for distributed data processing in cloud environments: a review. *IEEE Access* (2024)
19. Sanjay, R., Pulakhandam, D., Nirmalrani, V.: Real-time dashboarding using big data tools. In: 2024 International Conference on Inventive Computation Technologies (ICICT), pp. 629–635. IEEE (2024)
20. Shah, T.H.: Big data analytics in higher education. Research Anthology on Big Data Analytics, Architectures, and Applications pp. 1275–1293 (2022)
21. Shamsinejad, E., Banirostam, T., Pedram, M.M., Rahmani, A.M.: Representing a model for the anonymization of big data stream using in-memory processing. *Annals of Data Science* pp. 1–30 (2024)
22. Silva, C., Vilaça, R., Pereira, A., Bessa, R.: A review on the decarbonization of high-performance computing centers. *Renew. Sustain. Energy Rev.* **189**, 114019 (2024)
23. Zhou, Y., Zhou, J., Lu, K., Zhan, L., Xu, P., Wu, P., Chen, S., Liu, X., Wan, J.: A contract-aware and cost-effective LSM store for cloud storage with low latency spikes. *ACM Trans. Storage* **20**(2), 1–27 (2024)



P2FL: Privacy-Preserving Federated Learning Approach for Healthcare Informatics at the Edge

Farhan Ullah¹(✉), Leonardo Mostarda², Diletta Cacciagrano³, Hamad Naeem⁴, Shamsher Ullah⁵, Pradeep Chaudhary⁶, and Yue Zhao⁷

¹ Cybersecurity Center, Prince Mohammad Bin Fahd University, Khobar, Dhahran 34754, Saudi Arabia

fullah@pmu.edu.sa

² Department of Mathematics and Computer Science, University of Perugia, 06121 Perugia, Italy

leonardo.mostarda@unipg.it

³ Division of Computer Science, University of Camerino, 62032 Camerino, Italy
diletta.cacciagrano@unicam.it

⁴ Department of Computer Science, King Faisal University, Hofuf, Al-Hassa 31982, Saudi Arabia

⁵ National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen 518060, China

shamsher@szu.edu.cn

⁶ Chaudhary Charan Singh University, Department of Statistics, Meerut, Uttar Pradesh 250004, India

⁷ Department of Computer Science, College of Science, Mathematics and Technology at Wenzhou-Kean University, Wenzhou, China

yuezhao@kean.edu

Abstract. The healthcare industry is at significant risk of cybercrime and privacy violations due to the extensive distribution and sensitivity of health data. Recent trends in confidentiality breaches across multiple industries highlight the critical need for improved data security technologies that ensure privacy, accuracy, and reliability. Furthermore, decentralized healthcare systems suffer from intermittent remote clients and imbalanced data. This paper proposes a privacy-preserving federated learning (P2FL) approach for improving medical data privacy in health care informatics on edge devices. The remote clients in FL may have imbalanced datasets for local training, leading to lower results. Data augmentation procedures are employed in local model training to address this issue and balance datasets. Dynamic customer interactions with the global server via remote healthcare facilities are challenging. In real-world scenarios, technical or connectivity issues may cause clients to attend or exit the training session. The proposed methodology is tested on various clients and image sizes to evaluate its efficacy under various conditions. The suggested method achieved a 99.04% classification accuracy using two standard datasets. These findings enable collaborative efforts among medical institutions to leverage private data efficiently, further developing strong patient diagnostic models.

Keywords: Federated Learning · Healthcare · Medical Imaging · Deep Learning · Privacy · Security

1 Introduction

The advent of the digital revolution and globalization has led to extensive data gathering by various businesses and organizations. This data is obtained from various sources, including the economic, commercial, and healthcare sectors. Advancements in Machine Learning (ML) and Deep Learning technologies are driven by the need to analyze massive amounts of data gathered daily. However, because of its immense value, maintaining the confidentiality and accuracy of this data is crucial. Preserving confidential information frequently requires compliance with regulations such as the General Data Protection Regulation (GDPR) [1].

The traditional approach for using machine learning to decentralize data is a centralized method where multiple parties control the data, as illustrated in Figure 1. This method requires clients to send their data to a central server to train a model that generates predictions [2]. After training, each client receives their results. This technique has several drawbacks, mainly concerning data privacy and preserving unique data characteristics. Transmitting data from clients to a central server poses a risk of interception during transit. Real-time predictions require fast, low-latency communication links for large datasets. An alternative is to train a model with client data and share the model with all clients, allowing each data owner to keep the model and avoid relocating data when new data is obtained. This approach offers two main benefits over the traditional method: reduced latency by enabling individual predictions for each client, and lower communication costs, decreasing network dependence. However, during early training, clients must send data to the central server. This raises security concerns since sensitive data can be intercepted, especially over slow networks like those used by IoT devices or in cloud computing, potentially leading to data privacy breaches.

FL is gaining popularity because it allows data to be stored on the servers of individual data owners, even during training [3]. This approach secures sensitive information by removing the necessity of collecting data from clients for centralized training. FL guarantees that user data is not transmitted to central servers by conducting data analysis in a distributed manner. FL-based ML algorithms provide secure communication among edge devices. This technology decreases data transmission between edge devices and centralized systems, providing new potential for consumers and businesses in automotive, security, surveillance, and healthcare [4]. Edge-enabled ML is especially useful in healthcare, where patient privacy and ethical data use are critical. Furthermore, edge computing can potentially influence healthcare services, particularly remote healthcare delivery. Patients living in rural places sometimes have restricted access to healthcare due to a lack of modern medical equipment and logistics. Subsequently, integrating intelligent data management capabilities into peripheral devices can enhance the efficiency and capability of healthcare services while maintaining data privacy [5]. The main contributions of this study are the following:

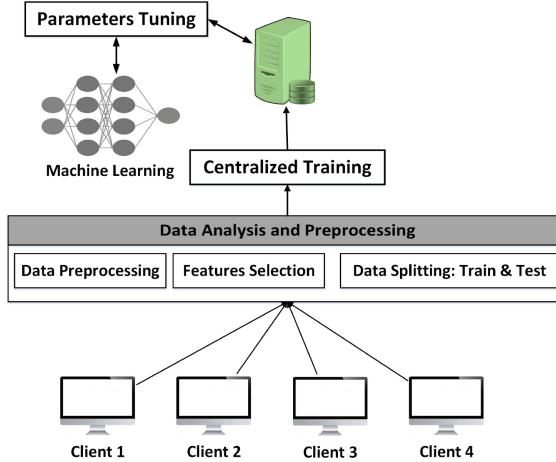


Fig. 1. Traditional ML with centralized training process.

- We propose a privacy-preserving FL technique for intelligent medial image classification at the edges. This FL paradigm enables remote healthcare organizations to benefit from remote data without exchanging critical information, protecting privacy and security.
- Remote clients may have imbalanced sets of medical images for local training purposes. Thus, the minority class may not provide enough data for the training model. We utilize various data augmentation techniques to significantly increase the dataset and investigate this impact on the proposed method. We further demonstrate that the proposed approach ensures uniform training data distribution across all training data centers.
- Dynamic client connectivity across several hospitals may influence classification performance. Our solution uses the FL architecture to get superior classification results despite challenges such as imbalanced data and unpredictable distribution from intermittent clients.

The remaining part of the paper is organized as follows: Part 2 explains the related work, Part 3 elaborates on the proposed method, Part 4 shows the experimental analysis, and finally, Part 5 concludes the work.

2 Related Work

FL is commonly used to merge data from specific clients while prioritizing security and privacy. The FedAvg approach, developed by Google [6], was used to train machine learning models by collecting data from several mobile devices without exchanging any of that data. This method focuses on transferring model parameters rather than direct data transfers, successfully addressing concerns about isolated data islands. Haya et al. [7] created FL technologies that combine the Internet of Things (IoT) with medical facilities. They provided a data cooperation architecture for remote patient monitoring utilizing IoT, which excluded FL from the tracking process in certain scenarios.

Their examination, which used ECG data from several approaches, revealed that deep learning outperformed other systems. They combined FL with an IoT digital infrastructure to safeguard personal privacy. Yu and Liu [8] investigated the application of FL in object detection. They devised a federated average technique incorporating aberrant weight clipping to enhance the model's performance on non-independent and identically distributed (IID) data. However, their object detection dataset is concerned more with general items than medical-specific objects.

Lee et al. [9] suggested a privacy-preserving platform in a federated setting focusing on cross-institutional patient similarity learning. Their method can identify similar patients across several institutions without disclosing individual patient information. Similarly, Huang et al. [10] addressed the issue of non-IID (differently distributed) ICU patient records, hampered decentralized training. They could accurately estimate fatality rates and hospitalization durations by categorizing individuals into clinically important groups. Baheti et al. [11] employed FL techniques to detect respiratory lung nodules using CT imaging. Rahman et al. [12] suggested an FL-based approach for digital health that uses a model incorporating a deep learning edge element and blockchain to improve dependability and safety. In addition, they devised a mechanism for transmitting industrial IoT data utilizing FL and blockchain. Lee et al. [9] prioritized patient privacy by designing a framework for detecting patient similarities in a federated system. Their method may detect patients with similar characteristics across several health facilities without sharing medical data. Liu et al. [13] present an effective and privacy-focused incentive model for Mobile Edge Computing (MEC)-assisted FL healthcare that takes into consideration dynamic interactions between Base Stations (BS), MEC servers, and MEC users. They analyze transmit power allocation, Differential Privacy (DP) budgets for MEC users and incentive strategies.

McMahan et al. [14] introduced FL, a distributed technique for constructing data-driven models that do not require centralized storage. FL divides training assignments across specialized clients, each working with its own set of data, and updates a centralized global model regularly. This decentralized technique improves privacy by keeping client data on their separate devices while decreasing latency. Figure 2 depicted FL process, clients use their data to make Local Model Updates (LMUs), which are then sent to a central server. These weights are used to create Global Model Updates (GMUs). This iterative technique optimizes the global model while keeping data private.

3 Our Proposed P2FL Method

Figure 3 shows the privacy-preserving FL for edge-based healthcare informatics using medical imaging. FL for intelligent medical image categorization at the edges protects privacy, allowing remote healthcare institutions to use data without exposing sensitive information. Our technique also adjusts to dynamic client connectivity across hospitals, maintaining good classification performance despite unbalanced data and sporadic client involvement.

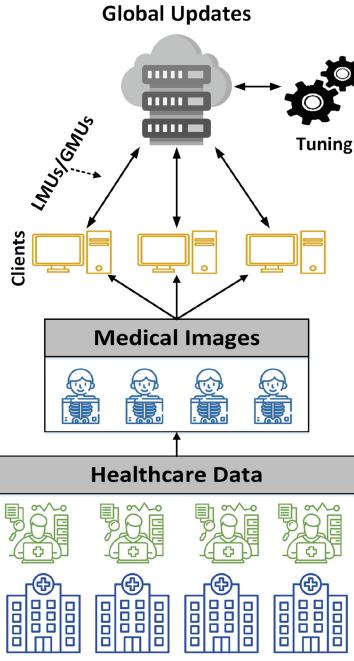


Fig. 2. Collaborative FL for distributed healthcare data

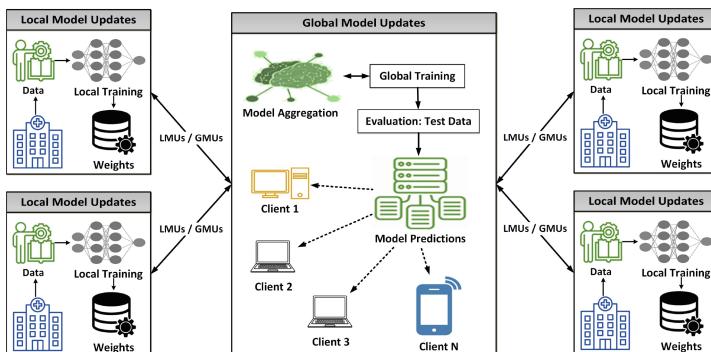


Fig. 3. Privacy-preserving FL for edge-based healthcare informatics using medical imaging

3.1 Dataset Preparation

The approach is evaluated by conducting tests on two large datasets from the Kaggle repository. The first dataset is chest X-ray images (pneumonia)¹ [15], and the second dataset is the COVID-19 radiography database² [16] [17]. The first dataset contains 5,863 pediatric chest X-ray scans (ages 1-5) from Guangzhou Women and Children

¹ <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>

² <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>

Medical Centre, focusing on anterior-posterior views. These images, used in clinical treatment, are classified into normal and pneumonia (further divided into bacterial and viral). A normal X-ray shows clear lungs, while viral pneumonia has a diffuse interstitial pattern, and bacterial pneumonia shows localized lobar consolidation (Figure 4). The second dataset, created by Qatar University, the University of Dhaka, and collaborators from Pakistan and Malaysia, includes chest X-rays of COVID-19 patients, healthy lungs, and viral pneumonia cases. It comprises 3,616 COVID-19 images, 10,192 normal lung images, 6,012 lung opacity images (non-COVID infections), and 1,345 viral pneumonia images.

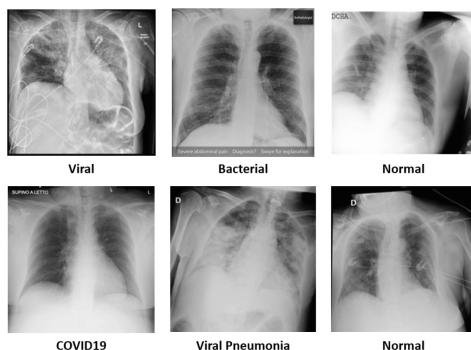


Fig. 4. Medical image classes from two standard datasets

3.2 Data Augmentation

Data augmentation techniques enhance image quality and balance datasets, improving classification efficiency. These methods increase training data diversity without needing new images, crucial in medical imaging where labeled data is hard to gather. We used the following techniques in medical imaging:

Rotation: Rotating images by ± 10 degrees to achieve model invariance to orientation, useful for varying patient positions.

Translation: Shifting images horizontally or vertically to help the model identify objects off-center, replicating patient placement variations.

Scaling: Adjusting magnification to account for different zoom levels, allowing the model to adapt to varied focus areas.

Contrast Adjustment: Modifying image contrast based on patient characteristics and machine settings for better performance across different scenarios.

Cropping and Padding: Mimicking different zoom levels and focus areas to improve detection skills by learning from various image portions.

Histogram Equalisation: Enhancing contrast in X-rays to reveal details for better feature extraction and classification.

Data augmentation is essential for improving the performance and resilience of models trained on X-ray images from balanced datasets, as shown in Figure 5. These

techniques help overcome dataset limitations, resulting in more precise and efficient classification.

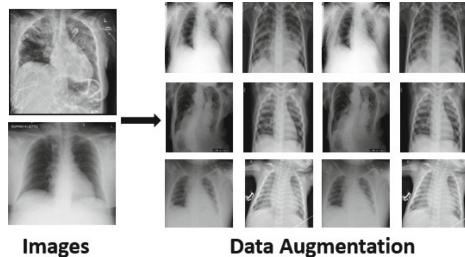


Fig. 5. Medical images using data augmentation techniques

3.3 Deep Convolutional Neural Network (DCNN): LMU

Federated clients locally train DCNN models using mini-batch Stochastic Gradient Descent (SGD) with their datasets and resources. Each client uses the same DCNN structure and loss functions, training with local data and global weights. After local training, updates are sent to the global server, which updates the global model. No data is shared between clients during training. DCNNs excel in categorizing medical images in intelligent medical facilities, extracting important features from large datasets, and reducing computational overhead. Figure 6 shows the DCNN architecture used for client-side training. The model uses a two-dimensional CNN network with convolutional layers, pooling layers, dropout layers, a fully connected layer, and a softmax layer. Convolutional layers generate feature maps with filters optimized by hyperparameters to enhance feature extraction.

One crucial component in our CNN architecture is the Batch Normalization layer. Batch normalization helps stabilize the training process and accelerates convergence by ensuring that the mean and standard deviation of inputs to each layer are close to 0 and 1, respectively:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (1)$$

where \hat{x} is the normalized input, x is the input, μ is the mean, σ^2 is the variance, and ϵ is a small constant to prevent division by zero.

The Batch Normalization process includes scaling and shifting the normalized input:

$$\begin{aligned} BN(x) &= \gamma \hat{x} + \beta \\ &= \gamma \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \end{aligned}$$

where γ is the scaling factor and β is the shifting factor, both learnable parameters during training.

This normalization reduces internal covariate shifts, making the model less sensitive to weight initialization and hyperparameters. Consequently, batch normalization enables higher learning rates and faster training, improving CNN model performance and robustness. Our network includes three convolutional layers with 128, 256, and 512 filters. Max-pooling layers reduce the dimensions of the feature maps, focusing on the most critical features and reducing computational costs.

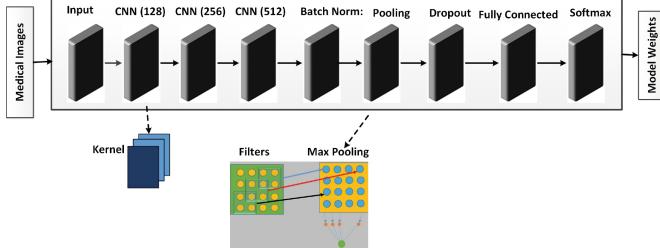


Fig. 6. Local training process using DCNN with batch normalization

3.4 Privacy-Preserving FL: GMU

The process begins by training the model locally on intermittent clients, producing weights known as LMU (w_i). Each client then transmits its LMU to the central server, where Federated Learning (FL) operates. The server aggregates these LMUs to update the global model (GMU), a process commonly achieved through federated averaging. For each client i , the weight w_i is determined as:

$$w_i = \frac{n_i}{\sum_{i=1}^N n_i}$$

where N represents the total number of clients, and n_i signifies the data volume of client i . This weighting mechanism ensures that clients with larger datasets contribute proportionally more to the aggregation process.

The aggregation of model weights is expressed as:

$$x_{\text{aggregated}} = \sum_{i=1}^N n_i w_i$$

Here, $x_{\text{aggregated}}$ is the combined weighted average of the model weights from all clients, reflecting the overall model update after aggregation.

The FL server's process involves the following steps:

- Initially, the server maintains a global model with preliminary weights, shared selectively with a subset of clients denoted as C_t .

2. Clients within C_i train on their local data and send their updated weights (LMU) back to the server.
3. The server aggregates these updated weights using the aggregation equation, resulting in the GMU.
4. This iterative process continues for multiple rounds, with the server sending updated weights to clients in each round.

4 Experiments and Discussions

4.1 Results Analysis

The proposed method is analyzed using two standard datasets with medical images. Figure 7 shows the epoch curves for accuracy and loss values using five and eight clients. The five clients start at 0.20, while the global server starts at 0.18. These curves gradually increase against each epoch, which shows the continual growth in local and global accuracy. However, some drops in accuracy are experienced in some epochs. For instance, the four clients have high accuracy on Epoch 8th, but Client 1 has an extensive drop in accuracy on the same epoch; the four clients, excluding client 1, have 0.48 while client 1 has 0.28, respectively. There is another sudden drop in local and global accuracy on epoch 15th, but it grows again. Overall, these curves are growing continuously and become more or less constant. The maximum accuracy achieved is 0.92. Part b shows the local and global loss for individual clients and global servers. These loss values display the values for each epoch, contrasting with the accuracy, demonstrating that our method performs better. For instance, the five clients, including the global server, start from 2.0 and then gradually drop in loss values, which shows their efficiency increase. The lowest loss is approximately 0.18.

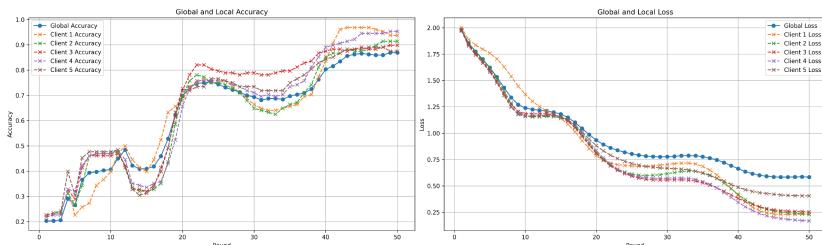


Fig. 7. Local and dynamic accuracy/loss curves using 5 clients

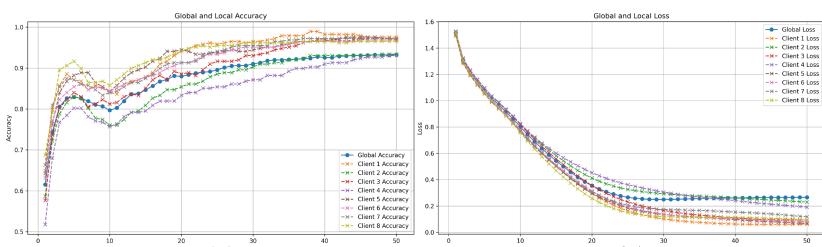


Fig. 8. Local and dynamic accuracy/loss curves using 8 clients

We conducted experiments with more clients to better show the performance of the proposed approach. Figure 8 shows the local and global epoch curves for eight clients with the global server. Now, these accuracy curves perform better with eight clients than Figure 7, using five clients. For instance, the eight clients and global server start from 0.5 and gradually increase their values on each epoch. There is a sudden drop in accuracy on Epoch 8th, but it is a continuously increasing trend. The eight clients and global server show a growing trend; the maximum achieved accuracy is 0.96, which shows an excellent performance. It indicates the local and global loss values to epochs for eight clients and server. It shows an excellent decreasing trend against each epoch, which indicates better performance of our proposed method. The minimum achieved loss is 0.04. Tables 1, 2, 3 show the performance measures using several number clients with two datasets.

Table 1. Dataset 1 using 64x64 images

	Clients	Precision	Recall	F-measure	Accuracy
Bacterial	3	72.41	86.5	80.31	79.5
Normal		85.48	88.21	85.78	
Virul		78.61	72.46	75.51	
Bacterial	5	85.67	97.01	92.45	91.21
Normal		94.16	96.12	95.04	
Virul		94.67	85.36	86.13	

Table 2. Dataset 1 using 128x128 images

	Clients	Precision	Recall	F-measure	Accuracy
Bacterial	8	93.36	94.52	94.82	95.87
Normal		100	96.61	99.14	
Virul		93.12	93.25	94.68	
Bacterial	10	98.54	98.45	98.24	98.61
Normal		100	100	98.13	
Virul		98.13	99.21	98.41	

Table 3. Dataset 2 using 128x128 images

	Clients	Precision	Recall	F-measure	Accuracy
COVID	5	96.86	96.18	93.66	92.76
Normal		87.72	94.42	88.72	
Virul Pneumonia		93.18	89.68	89.52	
COVID	10	99.1	100	100	99.04
Normal		100	98.42	96.72	
Virul Pneumonia		97.68	96.78	97.14	

5 Conclusion

The healthcare sector faces heightened risks from cyber threats and privacy breaches due to the widespread dissemination and sensitivity of health data. This study underscores the pressing need for advanced data security measures prioritising privacy, accuracy, and reliability. Decentralized healthcare systems encounter challenges such as intermittent remote client access and data imbalance. We propose a P2FL approach to mitigate these issues and enhance medical data privacy in edge device-based healthcare informatics. Our methodology incorporates data augmentation techniques to address imbalanced datasets during local training. Testing across various clients and image sizes demonstrated its efficacy under diverse conditions, achieving a notable 99.04% classification accuracy using two standard datasets. These results highlight opportunities for collaborative efforts among medical institutions to leverage private data effectively, advancing the development of robust patient diagnostic models. Future directions include exploring enhanced security protocols and scalability for broader implementation in healthcare settings.

References

1. Voigt, P., Von dem Bussche, A.: The EU General Data Protection Regulation (GDPR). A Practical Guide, 1st Ed. Springer International Publishing, Cham 10(3152676), 10–5555 (2017)
2. Callahan, A., Shah, N.H.: Machine learning in healthcare. In: Key advances in clinical informatics, pp. 279–291. Publisher, Elsevier (2017)
3. Nguyen, D. C., Pham, Q.-V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., & Hwang, W.-J.: Federated learning for smart healthcare: A survey. ACM Comput. Surv. **55**(3), 1–37 (2022)
4. Ullah, F., Srivastava, G., Xiao, H., Ullah, S., Lin, J.C.-W., Zhao, Y.: A scalable federated learning approach for collaborative smart healthcare systems with intermittent clients using medical imaging. IEEE J. Biomed. Health Inform. (2023)
5. Han, Y., Li, D., Qi, H., Ren, J., & Wang, X.: Federated learning-based computation offloading optimization in edge computing-supported Internet of Things. In: Proceedings of the ACM Turing Celebration Conference-China, pp. 1–5 (2019)

6. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282 (2017)
7. Elayan, H., Aloqaily, M., Guizani, M.: Sustainability of healthcare data analysis IoT-based systems using deep federated learning. *IEEE Internet of Things J.* **9**(10), 7338–7346 (2021)
8. Yu, P., Liu, Y.: Federated object detection: Optimizing object detection model with federated learning. In: Proceedings of the 3rd International Conference on Vision, Image and Signal Processing, pp. 1–6 (2019)
9. Lee, J., Sun, J., Wang, F., Wang, S., Jun, C.-H., Jiang, X., et al.: Privacy-preserving patient similarity learning in a federated environment: development and analysis. *JMIR Med. Inform.* **6**(2), e7744 (2018)
10. Huang, L., Shea, A.L., Qian, H., Masurkar, A., Deng, H., Liu, D.: Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *J. Biomed. Inform.* **99**, 103291 (2019)
11. Baheti, P., Sikka, M., Arya, K.V., Rajesh, R.: Federated Learning on Distributed Medical Records for Detection of Lung Nodules. In: Proceedings of VISIGRAPP (4: VISAPP), pp. 445–451 (2020)
12. Rahman, M.A., Hossain, M.S., Islam, M.S., Alrajeh, N.A., Muhammad, G.: Secure and provenance enhanced internet of health things framework: A blockchain managed federated learning approach. *IEEE Access* **8**, 205071–205087 (2020)
13. Liu, J., Chang, Z., Wang, K., Zhao, Z., Hämäläinen, T.: Energy Efficient and Privacy-Preserved Incentive Mechanism for Mobile Edge Computing – Assisted Federated Learning in Healthcare System. In: IEEE Transactions on Network and Service Management (2024)
14. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282 (2017)
15. Kermany, D., Zhang, K., Goldbaum, M., et al.: Labeled optical coherence tomography (OCT) and chest X-ray images for classification. *Mendeley Data* **2**(2), 651 (2018)
16. Chowdhury, M.E.H. et al.: Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* **8**, 132665–132676 (2020)
17. Rahman, T., et al.: Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Comput. Biol. Med.* **132**, 104319 (2021)



Enhancing Customer-Perceived Value Through Personal Data Utilization in CRM Platforms: A Data Science Perspective

Sutipong Sutinaraphan^{1(✉)} and Juggapong Natwichai²

¹ Data Science Consortium, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

Sutipong_s@cmu.ac.th

² Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

juggapong.n@cmu.ac.th

Abstract. Customer data or so called personal data has play an important role in nowadays business. Tremendous business strive to gather and analyze such data for their business growth which make customer privacy at risk in many perspectives. In order to regulate such risk, government of many countries begin to impose personal data privacy protection law which on the other hand caused difficulties for business to enlarge an insightful data. Studying about customer perceived value to voluntary give away their data is one of the key for gaining competitive advantage and to success in understanding customer and market insight in the real world. By simulate membership platform and let our sample engage and giving value through membership reward for their personal data, hence, it will convince business including managerial decision to convince and balance between personal privacy and business growth which also enhance the economic in the big picture.

Keywords: customer perceived-value · personal data privacy · data marketing · customer relationship management

1 Introduction

Economic and commercial world has evolve faster than before as a result of digital revolution and data-driven trend and culture. Businesses especially SMEs has also massively risen through out the world to capture revenue from limited and slightly growth number of population [9]. With an imbalance between seller and buyer in the economic system digital and data-driven tools has become an important factor to compete for customer [2].

Personalization [6] is one of the most focused strategies that business will give priority to compete with rivals in the same industry. That is why many businesses begin to develop their own strategy to win customer data and importantly

to overcome their privacy concern and prolong them with marketing loyalty program.

Especially after the COVID-19 pandemic, consumers are mostly penetrating to digital solution such as delivery and e-commerce platform [8]. Those digital solution aim to gather more personal data including indirect data from device usage. Let assume, a company is among competitively cosmetic shop market such as Sephora globally or Cosme in Asian market, knowing more on customer preference and able to communicate to them can be super weapon to win the market share.

Whereas business would like together as much consumer data and information as much as they can [7], but consumers may think in the other ways. How to persuade consumer to allow and give correct and recent data to business is one of the main challenges [10].

In the rapidly evolving digital economy, businesses, particularly Small and Medium-sized Enterprises (SMEs), are increasingly reliant on Customer Relationship Management (CRM) platforms to understand and enhance customer value. This paper aims to investigate how SMEs can leverage personal data within CRM systems to maximize customer-perceived value while maintaining data privacy and trust. For instance, personalized customer experiences can be achieved by analyzing CRM data to offer tailored recommendations and targeted communications. Moreover, anonymizing and aggregating data protect individual identities, respecting privacy regulations.

In [4], a reverse auction is proposed within a real-life experimental field setting to control consumer decision-making process for customer-perceived value. Utilizing a reverse-auction methodology, such work inspected consumers' valuation of data, personalized as their willingness to sell such information. The selection of data items for the auction was informed by prior research findings, aiming to provide a comprehensive representation of various data types. The study introduced a measurement model to observe consumers' privacy calculus in real-world decisions via search engine platform, with a specific focus on evaluating the significance of the company's data usage in consumers' data valuation. The findings substantiated that consumers incorporate perceived value into their data valuation, thereby confirming the presence of a discernible cost-benefit trade-off when disclosing personal data.

However, reverse auctions have proven successful in various global markets, their implementation in certain Asian countries encounters distinctive challenges. Issues such as cultural nuances, varying negotiation styles, and preferences for traditional procurement methods may pose hurdles to the seamless adoption of reverse auction strategies in these regions. For instances, in some Asian countries, there may be a higher sensitivity to costs due to varying economic conditions and standards of living. Companies and consumers may be more focused on cost-effectiveness and may prefer negotiation methods that directly address pricing concerns. On the other hand, Asian cultures often prioritize relationship-building in business dealings. Negotiation styles may be more indirect, with an emphasis on building trust over time. Reverse auctions, which are transactional and

competitive, might be perceived as impersonal and contradicting relationship-building values.

Therefore, the context of our work in this paper will diverge from the reverse auction method. We propose a preliminary analysis to evaluate whether implementing a membership system can effectively overcome the challenges associated with the reverse auction method within the CRM framework. This assessment will consider key factors such as customer engagement, satisfaction, and the overall alignment with our CRM and marketing objectives.

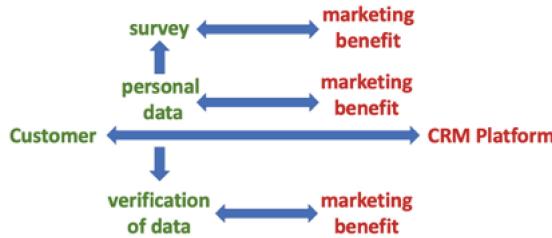


Fig. 1. A potential conceptual framework

In Fig. 1, customers will be offered to register for membership platform so called CRM and will be asked for input of personal data to trade for specific marketing benefit specified for each field which reflect level of sensitivity of personal data. Moreover, each customer will be asked if they would like to verify their personal data for correctness and to prevent faking input into CRM platform due to marketing incentive which after verified, they will receive more marketing benefit for verification. In addition, the online survey will be sent directly to the registered and verified user to gather more insight on how and what influence them to make decision to trade off their personal data which result will later analyze, as an initial idea.

In addition, in [11], the study focuses on a supermarket with a membership program, utilizing data mining techniques to scrutinize customer relationships, derive insights from customer information, and offer assistance for enterprise management and operations. The investigation concentrates on the analysis of customer background details and specific data variables, aiming to categorize customers according to external attributes, intrinsic characteristics, and consumer behavior. The outcomes of the study empower the enterprise to tailor marketing strategies for diverse customer segments, leading to reduced marketing expenditures and the cultivation of customer resources. Thus, our work entails a meticulous examination of how customers perceive the value associated with the implemented membership system.

Commencing with an exploration of whether individual customer traits, such as shopping habits, exert influence on their opinions, deviations from anticipated patterns in our collected data can be observed. Consequently, adjustments are made to enhance the precision of our understanding. By employing analytical

tools, we identified a positive correlation between increased membership usage and heightened customer satisfaction. Furthermore, additional exploration will focus on identifying specific membership features that played a key role in generation positive influence. Then, the numerical insights with existing literature on similar membership systems to substantiate our findings can be made, aligning them with broader perspectives on customer perceptions. In summary this comprehensive analytical approach can ensure the nuanced exploration of intricate dynamics characterizing the relationship between membership system utilization and customer-perceived value.

The rest of the paper is organized as follows: In Sect. 2, we present relevant literature by reviewing and discussing some customer perceived-value concepts utilized in our work. In Sect. 3, the problem statement and progress are presented. Thereafter, the conclusion remarks on the preparation of this work is presented in Sect. 4.

2 Related Literatures

In this section, we review the concept of blocking and matching strategies within the context of the entity matching process.

2.1 The Evolution of CRM in the Digital Era

The evolution of CRM in the digital era is a crucial and dynamic exploration of how organizations navigate technological advancements to enhance customer relationships, offering practical insights for staying competitive in the digital landscape. In [1], a comprehensive examination of customer relationship management (CRM) in the banking sector within the digital age is proposed. Primarily a literature review, the study synthesizes existing research to elucidate the impact of CRM on banks. Additionally, a survey involving 60 bank employees and 30 clients to gather quantitative data was conducted, which was subsequently analyzed. The findings underscore the advantageous effects of electronic customer relationship management (e-CRM) on banks, encompassing reduced routine tasks, streamlined access to customer information, and prompt customer service. Notably, the study establishes a significant correlation between the success of the banking industry and effective CRM strategies, influencing customer satisfaction and retention. Despite the absence of specific survey results, the paper emphasizes the critical role of information and communication technologies (ICT) in enhancing relationship management, fostering customer loyalty, and ultimately leading to heightened revenues and profits.

2.2 Theoretical Frameworks on Customer-Perceived Value

Recently, businesses worldwide benefit from understanding theoretical frameworks on customer-perceived value to optimize strategies, enhance customer satisfaction, and stay competitive in dynamic markets. In [3], the intricate interplay

between customer satisfaction and customer value within the realm of business marketing is proposed, seeking to dispel uncertainties surrounding both constructs. Through the development and testing of two alternative models, the study explores the relationship between customer-perceived value and behavioral outcomes. The first model posits a direct influence of perceived value on purchasing managers' intentions, while the second suggests that satisfaction mediates this relationship. Employing structural equation modeling (SEM) and a cross-sectional survey with purchasing managers in Germany, the work convincingly establishes that customer-perceived value and customer satisfaction are distinct yet complementary constructs. This distinction not only enhances the conceptual clarity of both concepts but also provides valuable insights into their individual and interrelated roles in influencing behavioral outcomes in the business marketing context.

2.3 The Interplay Between Personal Data Usage and Privacy Concerns in Marketing

The real-world dynamics of personal data usage and privacy concerns in marketing underscore the intricate balance businesses must achieve between leveraging customer data for personalized strategies and upholding ethical and regulatory standards to foster trust in the digital age. In [5], a telephone survey involving 480 consumers in a highly populated three-county area of a large southern state was conducted using computer-assisted telephone interviewing (CATI) software. This method ensured direct data entry to minimize errors. The surveys were carried out on November 15 and 17, 2000, between 4 p.m. and 8 p.m. The results highlight consumers' substantial privacy concerns related to the collection and use of personal information, encompassing issues such as the release of medical records to research organizations without patient consent, referencing individual medical histories by politicians without patient clearance, and internet purchasing. Additionally, the study reveals a limited awareness among consumers regarding the utilization of discount (loyalty) cards for personal purchase data collection. The findings indicate variations in privacy concerns across demographic market segments and establish a significant relationship between privacy concerns and consumers' internet purchasing behaviors.

3 Problem Statement

3.1 Problem Statement

Harmony between Data Utilization and Trust Preservation. The work aims to navigate the delicate balance between utilizing individual data for marketing advantages and the critical responsibility of safeguarding customer trust and privacy in the realm of digital marketing.

Innovative Approach with Membership Systems. Unlike traditional methods such as reverse auctions, the primary focus is on pioneering new strategies,

specifically through the implementation of a membership system, to enhance how customers perceive value. The goal is to address a significant gap in existing work by investigating the positive impact of such a system on customer perceptions, offering a more nuanced and ethical perspective on utilizing customer data for marketing purposes. The anticipated outcomes include insights that can contribute to establishing a mutually beneficial relationship harmonizing marketing advantages, customer trust, and privacy in the evolving landscape of the digital age.

In order to address such problem, the foundation of our work lies in the collection of customer data through our membership system. The system captures a comprehensive set of identifiers, including phone numbers, date of birth, email addresses, other demographic data, and customer addresses. This wealth of information provides a detailed snapshot of our customer base, allowing for a nuanced exploration of customer behavior and preferences. As we progress in our analysis, the inclusion of diverse demographic data enhances the depth and richness of our dataset, enabling a more thorough investigation into the factors shaping customer-perceived value.

Our membership system operates as a dynamic platform, intricately designed to meet the diverse needs and desires of our customers. Within this framework, customers are not only recognized individually but are also incentivized through a points and coupon system. These incentives serve as a flexible currency, enabling customers to trade-off accumulated points and coupons for exclusive discounts and tailored offerings aligned with their desires. This unique approach sets our membership system apart, and our ongoing investigation aims to unravel the impact of these incentives on customer satisfaction and perceived value. By delving into the interplay between demographic factors and the utilization of incentives, we seek to understand the varying dynamics that contribute to customer engagement and loyalty.

Beyond the conventional loyalty programs, the system offers a personalized experience by aligning incentives with individual preferences. This progress marks the initial stages of our exploration into the effectiveness of this model, seeking to understand how it influences customer behavior and shapes their perception of value. By continually examining the interconnections between demographic attributes, customer desires, and the utilization of incentives.

From the system, we aim at collecting such data as the following characters.

1. The data should be directly related to customer interactions within the membership system, focusing on information that influences perceived value.
2. The data must be covered various aspects of interactions, including transaction history and engagement patterns.
3. The data must be validated and cleaned to eliminate errors and inconsistencies.
4. The data must be acquired data ethically, obtaining customer consent and complying with privacy regulations.

This work employs a mixed-method approach, combining quantitative surveys and qualitative interviews to investigate the impact of a retail loyalty mem-

bership system on customer-perceived value. Methodologies include applying the Hausman test to choose between fixed and random effects models, using the Shapiro-Wilk test to assess normality, employing Pearson's correlation coefficient to evaluate linear relationships, and conducting regression analyses with fixed effects to model the intricate dynamics of the membership system's influence on perceived value.

4 Conclusion Remarks

To summary, our objectives are firstly to analyze customer-perceived value on marketing benefit for their personal data through CRM platforms. We aim at studying how customers see the value in a business, especially when using their information through membership systems in CRM platforms, also at aiding companies create personalized experiences, and thus building trust by examine the balance between data utilization and trust preservation and provide a nuanced and ethical perspective on data utilization. Secondly, to identify strategies that SMEs can adopt to enhancing CRM Platform and balance data utilization with privacy concerns and evaluate the impact of membership systems on customer perception which lead to exploration of innovation strategies for value perception and generate insights for establishing a mutually beneficial relationship.

We believe that such work can give significant evidence-based on how consumer value their personal data for marketing benefit. In addtion, it can suggest effective strategy to adopt and use personal data for CRM platform with consumer privacy concern.

Last, we elaborate on the approaches to improve and accelerate the value of our study as follows.

1. Diverse Incentives: Instead of just offering discounts or points, consider using a range of incentives such as personalized product recommendations, early access to new products, and special recognition for loyalty. This can provide a deeper understanding of what types of incentives are most effective in different demographics.
2. Gamification: Integrate gamification elements into the membership system, such as rewards for completing certain activities or levels, to increase engagement and perceived value.
3. Feedback Loops: Implement a system where customers can provide feedback on the incentives and offers they receive, allowing for real-time adjustment of strategies to better match customer preferences.
4. Cultural Sensitivity: Tailor the CRM experience to reflect cultural values and preferences, which can be particularly relevant for SMEs operating in diverse markets.
5. Demographic Segmentation: Analyze how different demographic segments respond to CRM strategies to identify patterns and preferences unique to each group.

References

1. Bachir, S.: The evolution of customer relationship management in the digital age and its impact on banks. EURASEANs J. Glob. Socio Econ. Dyn. **3**, 50–63 (2021). [https://doi.org/10.35678/2539-5645.3\(28\).2021.50-63](https://doi.org/10.35678/2539-5645.3(28).2021.50-63)
2. Coreyenen, W., MatthysSENS, P., Van Bockhaven, W.: Boosting servitization through digitization: pathways and dynamic resource configurations for manufacturers. Ind. Mark. Manag. **60**, 42 (2016). <https://doi.org/10.1016/j.indmarman.2016.04.012>
3. Eggert, A., Ulaga, W.: Customer perceived value: A substitute for satisfaction in business markets? J. Bus. Ind. Mark. **17**, 107–11 (2002). <https://doi.org/10.1108/08858620210419754>
4. Fehrenbach, D., Herrando, C.: The effect of customer-perceived value when paying for a product with personal data: a real-life experimental study. J. Bus. Res. **137**, 222–232 (2021). <https://doi.org/10.1016/j.jbusres.2021.08.029>
5. Graeff, T., Harmon, S.: Collecting and using personal data: consumers' awareness and concerns. J. Consum. Mark. **19**, 302–318 (2002). <https://doi.org/10.1108/07363760210433627>
6. Kumar, V., Rahman, Z., Kazmi, A.: Sustainability marketing strategy: an analysis of recent literature. Glob. Bus. Rev. **14**, 601–625 (2013). <https://doi.org/10.1177/0972150913501598>
7. Line, N., Dogru-Dr True, T., El-Manstrly, D., Buoye, A., Malthouse, E., Kandampully, J.: Control, use and ownership of big data: a reciprocal view of customer big data value in the hospitality and tourism industry. Tour. Manag. **80**(104), 10 (2020). <https://doi.org/10.1016/j.tourman.2020.104106>
8. Priyono, A., Moin, A., Putri, V.N.A.O.: Identifying digital transformation paths in the business model of SMEs during the covid-19 pandemic. J. Open Innov. Technol. Mark. Complex. **6**(4), 104 (2020). <https://doi.org/10.3390/joitmc6040104>
9. Surya, B., Menne, F., Sabhan, H., Suriani, S., Abubakar, H., Idris, M.: Economic growth, increasing productivity of SMEs, and open innovation. J. Open Innov. Technol. Mark. Complex. **7**(1), 20 (2021). <https://doi.org/10.3390/joitmc7010020>
10. Yaghtin, S., Safarzadeh, H., Zand, M.: Planning a goal-oriented b2b content marketing strategy. Mark. Intell. Plan. (2020) (Ahead of print). <https://doi.org/10.1108/MIP-11-2019-0559>
11. Zhou, S., Lei, G.: Application of data mining technology in membership supermarket's customer segmentation. In: 2011 International Conference on Business Computing and Global Informatization, pp. 181–183 (2011). <https://doi.org/10.1109/BCGI.2011.53>



Connecting AI and Blockchain to Improve Security of Financial Services

Ramiz Salama¹(✉), Diletta Cacciagrano², and Fadi Al-Turjman³

¹ Research Center for AI and IoT, Near East University, Department of Computer Engineering, AI and Robotics Institute, Nicosia, Mersin 10, Turkey
ramiz.salama@neu.edu.tr

² University of Camerino, Computer Science, Camerino, Italy

³ AI and Robotics Institute, Near East University, Artificial Intelligence, Software, and Information Systems Engineering Departments, Nicosia, Mersin10, Turkey

Abstract. Blockchain technology, paired with Artificial Intelligence (AI), provides a revolutionary solution to increase the security of financial transactions. When blockchain technology and AI are coupled, fraud, identity theft, and data breaches are reduced, while financial transactions are seen as more trustworthy and legal. This abstract examines the potential synergies between blockchain technology and AI, as well as the uses, benefits, and combined strengths of the two in the financial services sector. AI technologies, which include machine learning and natural language processing, now allow financial institutions to examine large volumes of data in real time and identify trends, irregularities, and suspicious activities that may indicate fraudulent behavior. By combining blockchain-based transactional networks with AI-powered fraud detection systems, organizations may increase overall security and transparency across the financial ecosystem. Blockchain technology's immutable record ensures transaction integrity and traceability, while AI algorithms provide powerful analytics and predictive insights to effectively detect and prevent fraudulent activity. Furthermore, AI-driven identity verification and authentication systems enhance the security of digital transactions by precisely confirming user identities and detecting unauthorized access attempts. Financial institutions may increase security, speed up customer onboarding, and prevent unauthorized access to sensitive data by merging AI-based biometric authentication with blockchain-based identity management systems. Furthermore, by automating and enforcing the execution of contractual agreements, AI-powered smart contracts reduce the risk of financial transaction fraud, errors, and disputes. Organizations may use blockchain technology and AI-powered smart contract platforms to provide secure, transparent, and immutable transactions. As a result, middlemen and transaction expenses will decrease. Blockchain and AI integration creates new opportunities for risk management and regulatory compliance in the financial services industry. AI-powered regulatory compliance solutions examine massive amounts of regulatory data, identify compliance issues, and ensure that the law is followed. Combining these solutions with blockchain-based regulatory reporting platforms can help financial organizations increase regulatory oversight, transparency, and auditability. This will increase loyalty and trust in the banking business. Finally, the combination of blockchain technology and AI has enormous promise for increasing financial service security. By combining blockchain technology with AI, organizations may boost regulatory compliance, detect and stop

fraudulent activity, speed up transaction processes, and improve identity verification and authentication. This abstract describes how blockchain technology and AI are transforming financial service security and creating a more stable, efficient, and secure financial environment.

Keywords: Financial · Services Security · AI · Blockchain · Enhanced

1 Introduction

Security in financial services is critical for securing sensitive data, transactions, and assets in today's rapidly changing digital environment. Financial institutions are constantly seeking for new and innovative methods to decrease risks and improve security as cyber threats develop and digital transactions become more common. Combining blockchain technology and Artificial Intelligence (AI) is one such approach that offers a revolutionary method of improving financial services security.

AI is the ability of machines to simulate human intelligence. Because of this, machines can perform cognitive tasks such as learning and problem solving that would normally require human intelligence. AI encompasses a wide range of techniques, including computer vision, machine learning, and natural language processing, that enable machines to analyze data, draw conclusions, and make intelligent decisions on their own. Blockchain, on the other hand, is a computer network-based, decentralized, secure distributed ledger system for recording transactions. A chain of blocks is formed when each transaction, or "block," is cryptographically linked to the one before it; hence the term "blockchain." Because blockchain technology is decentralized, transparent, and immutable, it may be used to conduct secure, trustless transactions. Security is critical in the financial services business because it ensures transaction integrity and protects the interests of stakeholders and customers [1–3]. Among other security issues, financial organizations may face fraud, identity theft, data breaches, and cyberattacks. These instances may result in monetary losses, reputational damage, and legal consequences. As a result, it is critical to implement strong security measures to retain trust, protect sensitive financial data, and ensure the financial system's resilience and stability.

Fintech security could be enhanced by merging blockchain technology and AI. By combining the benefits of blockchain technology and AI (see Fig. 1), financial institutions may offer cutting-edge solutions to a variety of security concerns such as fraud detection, identity verification, risk management, and regulatory compliance. Computers can now immediately review massive amounts of data using AI algorithms, which can detect suspicious actions, patterns, and anomalies that may indicate fraud. In the meantime, blockchain technology protects the integrity and immutability of financial data by providing a transparent and secure platform for transaction recording and verification. The distribution and consumption of financial services are transformed when blockchain technology and AI are coupled because the strong combination improves financial transaction security, transparency, and trust [4].

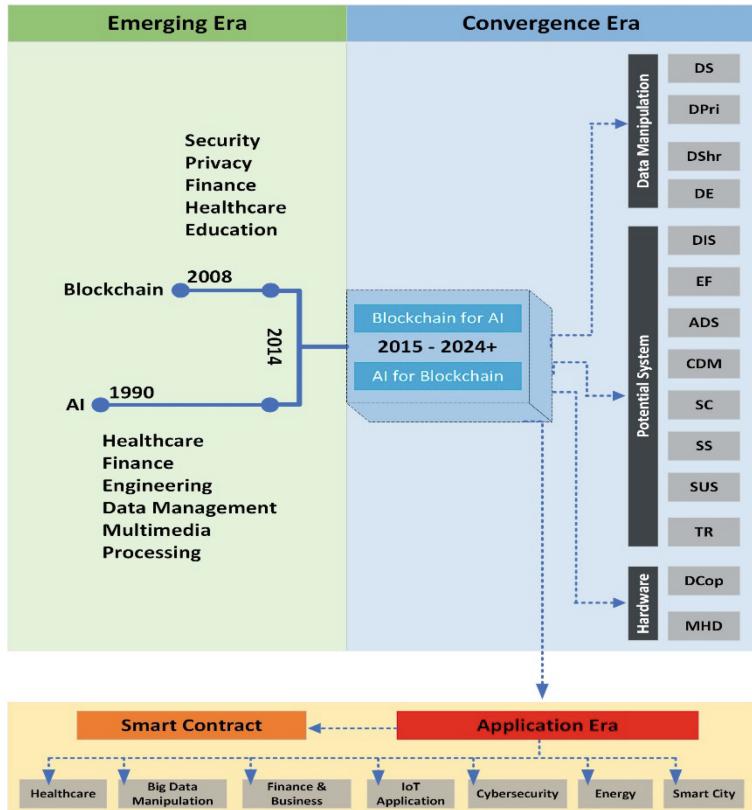


Fig. 1. Taxonomy for the timeline of blockchain and AI integration.

2 AI and Blockchain Synergy

In the financial services business, blockchain technology and AI work together to generate a powerful synergy that has the potential to totally change security procedures. This combination leverages the unique characteristics of blockchain technology and AI to deliver cutting-edge solutions that increase the security, trust, and transparency of financial transactions. To fully realize AI and blockchain's promise to improve security in the financial services industry, it is critical to understand how they complement one other (see Fig. 2). AI encompasses a diverse set of talents, including machine learning, computer vision, and natural language processing. AI technologies have enabled robots to do activities such as data processing, insight extraction, and independent intelligent decision-making that would normally need human cognitive capacities. AI is critical to financial sector security because machine learning techniques enable computers to spot patterns, learn from previous data, and predict future events. In the field of fraud detection, machine learning algorithms analyze transactional data to identify trends, highlight suspicious behavior, and quickly escalate the risk of fraud. Natural language processing encourages human-machine communication and interaction by enabling machines to

understand and interpret human language. Natural Language Processing (NLP) techniques are used in the financial services industry to detect fraud, analyze sentiment, and provide customer support. These solutions enable enterprises to examine textual data and derive actionable insights that will help them better serve their consumers. Three basic aspects of blockchain technology distinguish it from conventional databases: decentralization, transparency, and immutability. Because of these features, blockchain is the most secure and impenetrable option for logging and confirming transactions [5–7]. Blockchain offers the following benefits for financial service security:

Once saved on the blockchain, information cannot be modified or tampered with, ensuring the transaction records' integrity and immutability. Blockchain is the most secure method for storing sensitive financial data, such as identity certificates, asset ownership records, and transaction logs. The blockchain promotes transparency by allowing all network users to access transaction records for verification. Transparency boosts trust and accountability by allowing stakeholders to authenticate transactions, audit financial activities in real time, and monitor money movement. The cornerstone of blockchain technology is a decentralized network of computers known as nodes, which collaborate to record and confirm transactions. Decentralization prohibits any single party from controlling the network, reducing the likelihood of fraud, censorship, and single points of failure [8].

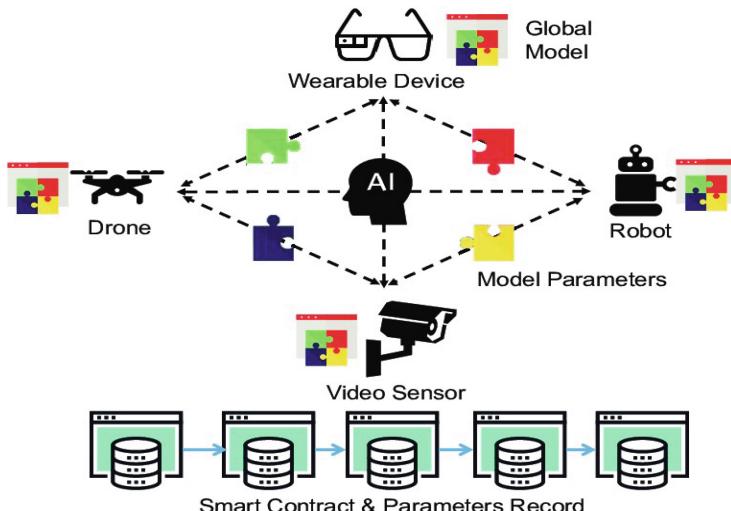


Fig. 2. Blockchain-powered AI for the Internet of Things (IoT).

The most effective method for the two technologies to collaborate is to combine blockchain's transparent and secure platform with AI's analytical capabilities. Combining AI and blockchain technology could enable financial institutions to provide cutting-edge solutions for a variety of security challenges, including fraud detection, identity verification, risk management, and regulatory compliance. Fraud detection systems powered by AI employ machine learning algorithms to examine transactional data for trends,

anomalies, and potential fraud. Businesses may increase the security and transparency of the whole financial ecosystem by integrating these systems with blockchain-based transactional networks. AI-powered identity verification systems use biometric authentication and natural language processing approaches to certify user identities and detect unauthorized access attempts. Integrating these solutions with blockchain-based identity management platforms allows financial institutions to strengthen security, prevent unwanted access to key data, and speed up the customer onboarding process. AI-powered solutions for risk management and regulatory compliance scan massive amounts of data, identify compliance issues, and ensure the law is obeyed. Combining these solutions with blockchain-based regulatory reporting platforms can help financial institutions increase regulatory oversight, transparency, and auditability. This will increase loyalty and trust in the banking business.

Finally, merging blockchain technology with AI has enormous promise for improving financial service security. By integrating blockchain's transparent and secure platform with AI's analytical talents, organizations may speed up procedures, provide creative solutions to a variety of security concerns, and increase trust in financial transactions. The integration of AI with blockchain technology is poised to usher in a new era of innovation, efficiency, and resilience in the financial services security landscape.

3 AI-Powered Security Apps for Financial Services

Security is critical in the financial services industry for protecting confidential information, preventing fraud, and ensuring regulatory compliance. AI enables financial companies to more efficiently reinforce their security protocols. AI provides proactive threat mitigation by utilizing powerful algorithms and analytics to detect patterns, anomalies, and potential threats in real time. Financial institutions are particularly concerned about fraud because hackers are continuously looking for new methods to exploit weaknesses in procedures and policies. AI fraud prevention and detection systems employ machine learning algorithms to swiftly identify fraudulent activity by evaluating vast volumes of transactional data. These algorithms can detect trends such as unusual transaction amounts, peculiar spending habits, or questionable login attempts that indicate fraud. AI systems can detect potential fraud signals by continuously monitoring user behavior and transaction trends. This enables financial institutions to respond quickly to mitigate risks and protect their consumers. Furthermore, as new fraud patterns and hazards arise, AI-powered fraud detection systems may adapt and learn from them, increasing efficacy and accuracy over time. This dynamic method allows financial firms to cut losses while staying ahead of evolving fraud attempts [9–11]. In the digital age, identity theft and account takeover are prevalent issues that endanger both individuals and financial organizations. Businesses can use AI algorithms to properly authenticate their customers' identities and detect unauthorized access attempts. AI algorithms provide extra possibilities for identification and identity verification. AI-powered identity verification systems use biometric authentication, natural language processing, and machine learning algorithms to examine a wide range of identification documents, including passports, government-issued identification cards, and biometric data (such as fingerprints and facial recognition). These programs can detect attempts by con artists to validate identifying documents and pose as legitimate individuals.

AI systems can also analyze user interactions and behaviors to establish a standard for normal conduct. When there is a deviation from the planned baseline, financial institutions will be notified about potential fraud, providing them the opportunity to intervene and prevent unauthorized access to accounts or important data. Financial institutions must follow stringent standards and be closely watched by regulators in a highly regulated environment. AI analytics provide practical capabilities for regulatory compliance and risk management, allowing firms to easily discover, assess, and mitigate risks while adhering to legal standards [12]. AI-powered risk management systems examine massive amounts of data, including as transactional data, market patterns, and macroeconomic indicators, to identify potential dangers and weaknesses. These systems may do extensive risk modeling, scenario analysis, and stress testing to assess the impact of various risk factors on the financial stability and health of the firm. Financial organizations can also employ AI analytics to automate regulatory compliance operations such as reporting suspicious activity, monitoring for money laundering, and conducting Know-Your-Customer (KYC) due diligence. These systems can identify potential compliance issues and ensure that regulatory standards are met by evaluating consumer data, transactional trends, and other data sources. Using AI analytics, financial institutions can successfully decrease regulatory risks, accelerate compliance procedures, and increase risk management skills. Finally, there are numerous sophisticated applications for AI in financial services security. AI algorithms offer cutting-edge capabilities for improving security procedures and mitigating emerging risks. These skills include anything from identity verification and regulatory compliance to fraud prevention and detection. By continuing to employ AI-driven technology, financial institutions may enhance their defenses, decrease risks, and maintain the integrity of their operations in an increasingly global and digitally interconnected world.

4 Blockchain's Impact on Financial Services Security

The financial services industry is experiencing a dramatic shift in security measures as a result of the introduction of blockchain technology. Blockchain's unique characteristics, such as immutability, decentralization, and smart contracts, make it an efficient platform for protecting financial transactions and data. In this article, we look at how blockchain technology improves the security of financial services, focusing on the essential components of its immutable ledger, which ensures the correctness and transparency of transactions recorded on the network. When a transaction is confirmed and recorded on the blockchain, it creates an immutable and impenetrable chain of blocks that are cryptographically linked to one another. Because of its immutability, blockchain technology eliminates the possibility of fraud, manipulation, or unauthorized changes, making it ideal for recording and certifying financial transactions. Blockchain technology enables financial institutions to maintain unhackable and secure transaction records, reducing the risk of fraud, disputes, and data breaches. Furthermore, blockchain's transparency promotes accountability and trust by enabling all network users to review and verify transaction data. Blockchain improves the integrity of the financial system and increases trust in financial transactions by providing an auditable and transparent record of transactions.

Blockchain technology enables decentralized identity management solutions, which allow for the trustworthy and secure authentication of individuals and things in financial transactions. Traditional identity management systems are prone to data breaches, identity theft, and single points of failure since they rely on a centralized authority to validate and verify identities. These authorities could include national governments or financial institutions. On the other hand, blockchain-based identity management solutions leverage decentralized networks to ensure secure identity verification and identification. Every user on the network has their own digital identity, which can be independently confirmed by other users, encrypted, and stored on the blockchain. Because these decentralized identity management solutions do not require centralized authority or middlemen, they reduce the risk of identity theft and data breaches. Financial transactions can be made safer and more private by giving users control over their identities and personal information. Furthermore, blockchain-based identity management solutions enable secure and easy access to financial services by providing seamless and compatible authentication across a variety of platforms and apps [13–15]. Clear contract criteria are included into the code of smart contracts, which are self-executing. When certain conditions are met, these contracts automatically execute and enforce their terms; no middlemen or human intervention are required. Blockchain technology offers an auditable and tamper-proof foundation for automated transactions, allowing for the creation and execution of smart contracts on a decentralized network. Smart contracts automate the execution of contractual agreements based on established criteria, which can speed up a variety of financial processes such as payments, settlements, and asset transfers. Transaction security and reliability are enhanced by the tamper-proof nature of blockchain smart contracts, which cannot be modified or interfered with once deployed. Furthermore, by reducing the risk of censorship or single points of failure, blockchain's decentralized structure ensures the availability and integrity of transactions executed via smart contracts.

To summarize, blockchain technology is critical for enhancing security protocols in the financial services industry. Blockchain provides a solid framework for protecting financial transactions and data, with its immutable ledger for transaction integrity, decentralized identity management for secure authentication, and smart contracts for automated and impermeable transactions. Financial institutions that embrace blockchain technology can improve security standards, reduce risks, and increase trust in the financial ecosystem.

5 Integrating Blockchain and AI to Boost Security

Blockchain technology and AI combine to provide a powerful synergy that improves the security of the financial services sector. Organizations can find innovative solutions to a variety of security issues by combining the transparent and secure blockchain platform with AI's analytical capabilities, as shown in Fig. 3. AI-powered fraud detection systems scan massive volumes of transactional data and apply machine learning algorithms to detect anomalies that may indicate fraudulent behavior. Integrating these technologies with blockchain transaction networks allows organizations to improve security and transparency throughout the financial ecosystem. Blockchain protects the integrity and immutability of transaction data by providing a secure, unhackable environment for

transaction recording and validation. Blockchain technology and AI combine to provide a powerful synergy that improves the security of the financial services sector. Organizations can find innovative solutions to a variety of security issues by combining the transparent and secure blockchain platform with AI's analytical capabilities. AI-powered fraud detection systems scan massive volumes of transactional data and apply machine learning algorithms to detect anomalies that may indicate fraudulent behavior. Integrating these technologies with blockchain transaction networks allows organizations to improve security and transparency throughout the financial ecosystem. Blockchain protects the integrity and immutability of transaction data by providing a secure, unhackable environment for transaction recording and validation. Identity theft and account takeover are prevalent issues in the digital era, and identity verification is an important part of financial transaction security. AI-driven identity verification systems use a variety of ways to confirm user identities, including machine learning, biometric authentication, and natural language processing. Integrating these technologies with blockchain-based identity management solutions allows organizations to increase the security and privacy of identity verification procedures. Blockchain protects the security and validity of identity data by providing a decentralized, unbreakable infrastructure for digital identity verification and storage [16].

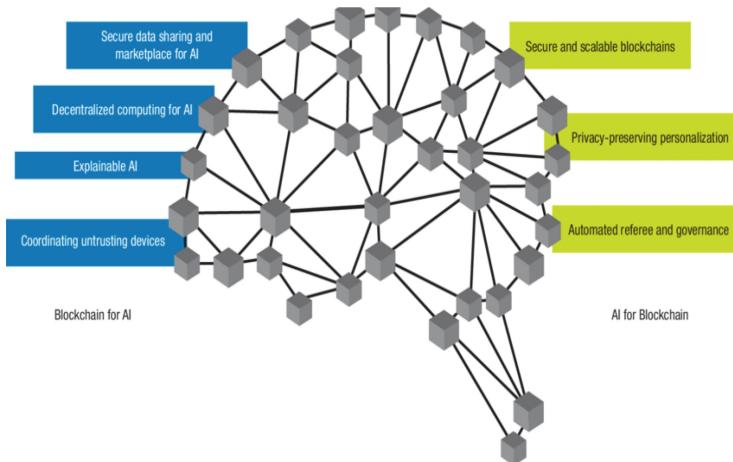


Fig. 3. Integration of AI and blockchain (a) blockchain for AI and (b) AI for blockchain.

AI systems confirm user identities by examining a variety of identity documents such as official identification cards, passports, and biometric data. After validation, digital identities are securely stored on the blockchain, where other network users can view and confirm them. Blockchain-based identity management services improve financial transaction security and privacy by granting consumers control over their identities and personal data. Enterprises that integrate AI-powered identity verification with blockchain-based identity management can improve security, speed up client onboarding, and protect private information from unauthorised access.

Clear contract criteria are included into the code of smart contracts, which are self-executing. When certain conditions are met, these contracts automatically execute and enforce their terms; no middlemen or human intervention are required. Blockchain technology protects smart contracts' immutability and resistance to manipulation while ensuring security and transparency during deployment and execution. Using AI-driven smart contracts on blockchain networks, organizations can automate and accelerate a number of financial transactions such as asset transfers, payments, and settlements.

Algorithms using AI can analyze transactional data and external variables to determine when smart contracts should be applied. Once implemented on the blockchain, smart contracts eliminate the need for human interaction and reduce the chance of errors or conflicts by automatically enforcing the terms of the agreement when pre-determined conditions are met. Furthermore, because of the openness and immutability of blockchain technology, smart contracts cannot be modified or tampered with once implemented, hence improving transaction security and reliability. Fusing blockchain platforms with AI-driven smart contracts allows organizations to expedite procedures and decrease risks while creating a more transparent, safe, and efficient financial eco-system [17–19]. Finally, the marriage of blockchain technology and AI has the potential to significantly improve security standards in the financial services business. Organizations can strengthen and secure their financial ecosystems by combining AI-powered fraud detection with blockchain transactional networks, implementing AI-driven smart contracts on blockchain platforms, and integrating AI-powered identity verification with blockchain-based identity management [20]. Financial institutions that integrate blockchain technology and AI can improve security standards, reduce risks, and increase trust in the financial ecosystem.

6 Conclusion

To summarize, the combination of blockchain technology and AI represents a revolutionary opportunity to improve security procedures in the financial services industry. Enterprises may provide novel solutions to a variety of security challenges by combining AI algorithms with blockchain technology's immutable ledger. The combination of blockchain technology and AI has significant potential for improving transaction security and trust. The transparent and secure platform of blockchain, when integrated with AI-driven fraud detection, identity verification, and smart contract automation, improves security procedures, reduces risks, and increases public trust in the financial sector.

Throughout this study, we demonstrated how firms may improve fraud detection, speed up procedures, and comply with regulations by merging blockchain technology and AI. We've also discussed the challenges and considerations that must be considered in order to fully benefit from the combination of blockchain technology and AI, such as algorithmic fairness, data privacy, and regulatory compliance. Stakeholders in the financial services sector will need to collaborate, come up with new ideas, and allocate resources to blockchain and AI integration as we move forward. Together, we can overcome hurdles, establish norms, and stimulate use so that blockchain technology and AI can fully realize their promise of improving financial services security.

To summarize, merging AI and blockchain technology is a practical solution to build a financial ecosystem that is more transparent, efficient, and safe. Working together and

putting in consistent effort, we can employ blockchain technology and AI to improve security standards, reduce risks, and increase consumer and enterprise confidence in financial transactions.

References

1. Odeyemi, O., Okoye, C.C., Ofodile, O.C., Adeoye, O.B., Addy, W.A., Ajayi-Nifise, A.O.: Integrating AI with blockchain for enhanced financial services security. *Finance Account. Res. J.* **6**(3), 271–287 (2024)
2. Alenizi, A., Mishra, S., Baihan, A.: Enhancing secure financial transactions through the synergy of blockchain and artificial intelligence. *Ain Shams Eng. J.* **15**(6), 102733 (2024)
3. Dhar Dwivedi, A., Singh, R., Kaushik, K., Rao Mukkamala, R., Alnumay, W.S.: Blockchain and artificial intelligence for 5G-enabled Internet of Things: Challenges, opportunities, and solutions. *Trans. Emerg. Telecommun. Technol.* **35**(4), e4329 (2024)
4. Nguyen Thanh, B., Son, H.X., Vo, D.T.H.: Blockchain: the economic and financial institution for autonomous AI? *J. Risk Finan. Manage.* **17**(2), 54 (2024)
5. Tyagi, P., Shrivastava, N., Sakshi, Jain, V.: Synergizing Artificial Intelligence and Blockchain. In: *Next-Generation Cybersecurity: AI, ML, and Blockchain* (pp. 83–97). Springer Nature Singapore, Singapore (2024)
6. Alhalafi, A., Veeraraghavan, P., Hanna, D.: Artificial Intelligence (AI) and Blockchain-based Online Payments in the Global World. *IJCSNS* **24**(3), 1 (2024)
7. Kanaparthi, V.: Exploring the Impact of Blockchain, AI, and ML on Financial Accounting Efficiency and Transformation. In: *International Conference on Multi-Strategy Learning Environment*, pp. 353–370. Springer Nature Singapore, Singapore (2024)
8. Kuznetsov, A., Sernani, P., Romeo, L., Frontoni, E., Mancini, A.: On the integration of artificial intelligence and blockchain technology: a perspective about security. *IEEE Access* (2024)
9. Hemamalini, V., Mishra, A.K., Tyagi, A.K., Kakulapati, V.: Artificial intelligence-blockchain-enabled–internet of things-based cloud applications for next-generation society. *Automated Secure Computing for Next-Generation Systems*, 65–82 (2024)
10. Tun, V.H.Z., Jahankhani, H.: Using Artificial Intelligence (AI) and Blockchain to Secure Smart Cities' Services and Applications. In: *Cybersecurity and Artificial Intelligence: Transformational Strategies and Disruptive Innovation*, Springer Nature Switzerland, Cham pp. 163–184 (2024)
11. Salama, R., Al-Turjman, F.: Future uses of AI and blockchain technology in the global value chain and cybersecurity. In: *Smart Global Value Chain*, CRC Press, pp. 257–269 (2024)
12. Tyagi, A.K.: Blockchain and Artificial Intelligence for Cyber Security in the Era of Internet of Things and Industrial Internet of Things Applications. In: *AI and Blockchain Applications in Industrial Robotics*, IGI Global, pp. 171–199 (2024)
13. Ruzbahani, A.M.: AI-Protected Blockchain-based IoT environments: Harnessing the Future of Network Security and Privacy. *arXiv preprint arXiv:2405.13847* (2024)
14. Bhumichai, D., Smiliotopoulos, C., Benton, R., Kambourakis, G., Damopoulos, D.: The convergence of artificial intelligence and blockchain: the state of play and the road ahead. *Information* **15**(5), 268 (2024)
15. Li, Z., Liang, X., Wen, Q., Wan, E.: The Analysis of Financial Network Transaction Risk Control based on Blockchain and Edge Computing Technology. *IEEE Transactions on Engineering Management* (2024)
16. Sharma, A., Gupta, N.: Developing An Automated And Transparent Wealth Framework With AI And Blockchain Interoperability. *Educ. Administ.: Theor. Pract.* **30**(5), 10030–10041 (2024)

17. Ressi, D., Romanello, R., Piazza, C., Rossi, S.: AI-enhanced blockchain technology: A review of advancements and opportunities. *J. Netw. Comput. Appl.* **103858** (2024)
18. Hu, B.: Digital transformation of retail financial services marketing in the information era: opportunities, risks and future. *Highlights Bus. Econ. Manage.* **24**, 2587–2594 (2024)
19. Yoganandham, G.: Transformative impact: the role of modern and innovative banking technologies in driving global economic growth. *J. Propul. Technol.* **45**(1) (2024)
20. Mahajan, S., Nanda, M.: Revolutionizing banking with blockchain: opportunities and challenges ahead. *Next-Generation Cybersecurity: AI, ML, and Blockchain*, 287–304 (2024)



A Comprehensive State-of-the-Art Review for Digital Twin: Cybersecurity Perspectives and Open Challenges

Aws Jaber¹ , Ioannis Koufos² , and Maria Christopoulou²

¹ Division of Network and Systems Engineering, KTH Royal Institute of Technology,
Stockholm, Sweden
awsj@kth.se

² Institute of Informatics and Telecommunications, National Centre for Scientific
Research “Demokritos”, Athens, Greece
{ikoufos,maria.christopoulou}@iit.demokritos.gr
<https://www.kth.se>

Abstract. The growing complexity and interconnectivity of modern systems have made the proactive identification and mitigation of vulnerabilities increasingly challenging. DTs, which offer dynamic and real-time virtual representations of physical systems, have emerged as a potential solution to enhance cybersecurity. However, several significant challenges exist, including the integration of DTs with existing cybersecurity frameworks and the effective use of DTs to detect and mitigate zero-day vulnerabilities. In this study, we conduct a systematic literature review to explore these challenges by addressing four key research questions. Our findings reveal critical challenges and open research problems associated with the implementation of DTs for cybersecurity. These insights are essential for advancing the security of complex systems, paving the way toward building resilient and secure cyber-infrastructures, enabling real-time assessment, effective remediation actions and cyber perception on emerging threats.

Keywords: Digital Twin · Cybersecurity · Attack Surface · Threat modeling · Game theory

1 Introduction

The concept of the Digital Twin (DT) was first introduced by Michael Grieves in 2002, who defined it as the digital representation of a physical entity, process, or system [19]. A DT mirrors its physical counterpart, continuously updating through real-time data, simulating behavior, analyzing outcomes, and providing optimized action commands. DTs are extensively used in various sectors, including industrial applications, lifecycle management platforms, predictive maintenance, and the automotive industry. However, their role in cybersecurity is an emerging area of research, where DT technology is being leveraged to enhance

security measures and enable comprehensive security analysis. Recent research has investigated the use of DTs in several cybersecurity applications, such as intrusion detection, anomaly detection, predictive analytics, and monitoring. Despite the promise of DTs in bolstering cybersecurity, numerous challenges and solutions remain. The increasing academic and industry interest in the intersection of DT and cybersecurity is reflected in the growing number of publications on this topic over recent years. As shown in Fig. 1, the annual publications related to Digital Transformation and Cybersecurity have significantly increased from 2018 through August 2024. This trend underscores the critical importance of understanding and addressing the security challenges associated with DTs.

Contribution: This state-of-the-art review paper, as of August 2024, offers a comprehensive examination of the interplay between DT and cybersecurity. It makes significant contributions to the academic field in the following ways:

1. **Comprehensive Analysis of DT Security Challenges:** The paper thoroughly examines the primary security challenges facing DT by analyzing various cyber attack models. It explores the effectiveness of tools specifically designed to counter a wide range of cybersecurity threats targeting DTs.
2. **Systematic Review of DT Attack Modeling Techniques:** An extensive and methodical review is conducted on DT attack modeling techniques, involving three snowball sampling rounds. This review provides readers with a deep understanding of the tools used for attack modeling and their respective functionalities within the DT context.
3. **Insights into Enhancing DT Security with AI:** The authors present valuable insights on how to enhance DT security by integrating AI-driven

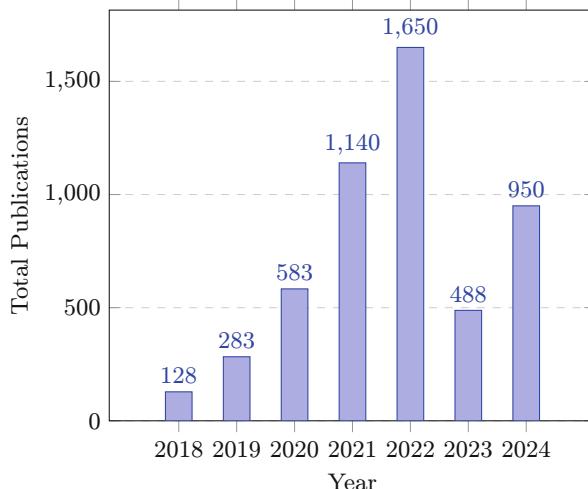


Fig. 1. Annual Publications on Digital Transformation and Cybersecurity through August 2024

attack mechanisms into existing platforms. This approach lays the groundwork for more robust and adaptive cybersecurity solutions tailored to the unique needs of DT environments.

Paper Structure: Section 2 outlines the systematic literature review (SLR) process adopted and describes the research questions that guide the study, the search strategy used to gather relevant research papers, and the phases of the literature review. Section 3 explains how DTs can be integrated into cybersecurity frameworks, enhance predictive security measures, enable real-time assessment and support remediation strategies. Section 4 summarizes key takeaways from the research and stress the importance of balancing robustness, stability and security in DTs and concludes with the challenges in implementing DTs in complex and dynamic environments. However, Table 1 provides a comparison of recent surveys and reviews on DT and cybersecurity, highlighting the focus areas, methodologies, and key findings of each work. This comparison helps to contextualize the distinct approaches taken by various researchers and identifies gaps that this paper aims to address.

Table 1. Comparison of Various Surveys and Reviews on DT and Cybersecurity

Reference	Year	Focus Area	Methodology	Key Findings
[3]	2022	Security Threats in DT	Survey	Comprehensive overview of security threats in DT.
[9]	2021	AR and DT in Cybersecurity	Review	Explores the integration of AR and DT with cybersecurity, highlighting future perspectives.
[45]	2022	Human DT	Ontological and Network Science	Proposes the use of human DTs for cybersecurity simulations in the metaverse.

2 Methodology

In the context of cybersecurity for DTs, a systematic approach to examining the literature involves employing a method called a Systematic Literature Review (SLR). This process, adhering to the foundational guidelines proposed by Kitchenham [34], consists of three primary phases: planning, executing, and reporting. This approach ensures a comprehensive understanding of the existing research and developments in cybersecurity pertaining to digital twins.

2.1 Research Questions

The research questions form the cornerstone of this academic inquiry, guiding the direction and focus of the literature review. These questions are not only

foundational to structuring the investigation but also crucial for identifying and addressing the key challenges in the domain of cybersecurity for DTs.

In this literature review, the research questions presented in Table 2 are carefully crafted to explore the critical aspects of DT security. These questions aim to bridge gaps in existing knowledge by investigating emerging trends, potential vulnerabilities, and the effectiveness of current security measures. By posing these questions, the review seeks to contribute meaningful insights that can inform both academic research and practical implementations in the field.

Table 2. Research Questions

No.	Type	Question
RQ1	DT Integration	How can DT technology be effectively integrated into existing cybersecurity frameworks and solutions?
RQ2	Predictive Cyber Security	How can DT technology be utilized to enhance the accuracy and reliability of predictive cybersecurity measures?
RQ3	Real-Time Assessment and Remediation	What role does DT play in enabling real-time assessment and remediation of cybersecurity threats?
RQ4	Remediation and Countermeasures	How can DT technology support the development and deployment of effective remediation strategies and countermeasures for various cybersecurity threats?
RQ5	Zero-day Simulation and Defense	How can DT technology contribute to the simulation and defense of zero-day vulnerabilities and attacks in cybersecurity?

2.2 Search Strategy

- **Search Term and Literature Resources:** To gather information on DT technology, a thorough literature review was conducted using Scopus, Elsevier, and ScienceDirect databases. The following search string was used: “DT” AND (“cybersecurity” OR “digital twin cybersecurity” OR “digital twin intrusion detection” OR “intrusion prevention” OR “reinforcement learning”). This systematic search encompassed a broad range of peer-reviewed interdisciplinary research papers.
- **Search Process:** Conducting a systematic literature review is a crucial component of academic research in cybersecurity, demanding the highest level of rigor and completeness. In the context of our study on cybersecurity in DTs, an exhaustive search of academic databases and relevant journals yielded a total of 235 papers. Each of these papers was subjected to a meticulous examination of the full text to ensure that only studies directly related to our research questions were included. Consequently, after a rigorous selection process, **68 papers** were deemed pertinent and included in our review. To facilitate replication and further research, we have included a comprehensive list of the selected studies. We believe that our study contributes significantly

to the academic literature on cybersecurity in DTs by providing a rigorous and thorough analysis of the available research. Through our approach to literature selection, we aim to advance the field's understanding of this critical topic and lay a foundation for future research in this area.

2.3 Phases of the Literature Review

- **Phase 2: Forward and Backward Snowballing**

- In Phase 2, we employed forward and backward snowballing techniques to expand our research base. Starting from key studies such as [6, 25, 59], we explored both the citations these studies made (backward snowballing) and the papers that cited these studies (forward snowballing). This allowed us to identify additional relevant articles that may not have been captured in the initial search.
- We particularly focused on articles that discussed the use of DT in the context of cybersecurity, such as [47, 54, 66].

- **Phase 3: Full-Text Review and Inclusion**

- In Phase 3, we conducted a full-text review of the studies identified during Phase 2. The criteria for inclusion were based on the relevance of the study to our research questions and its methodological rigor.
- Articles like [3, 28] were selected for their comprehensive coverage of DT in cybersecurity contexts, while others like [48, 70] were included for their focused discussions on specific aspects of DT and security.

- **Phase 4: Final Study Selection**

- The final phase involved synthesizing the selected studies to build a comprehensive understanding of the current state of research on DT in cybersecurity. Studies like [14, 21] were instrumental in providing insights into advanced applications of DT for security purposes.
- This phase also included the identification of gaps in the literature, guiding future research directions.

2.4 Study Selection

To be considered for inclusion in this study, prospective articles must satisfy a set of rigorous criteria. Firstly, a strong emphasis on DT within the field of cybersecurity is essential. Additionally, the articles must have been published recently, between 2018 and 2024. Furthermore, they must have undergone a peer-review process and be published in a reputable scholarly journal or conference proceedings. Lastly, authors are required to have conducted a comprehensive comparative or empirical analysis of DT in the context of cybersecurity, highlighting its potential and relevance. These stringent inclusion criteria are implemented to ensure that only the most impactful and innovative studies are included in our analysis.

2.5 Data Extraction

To ensure a high level of quality and relevance in our analysis of DT in cybersecurity, certain exclusion criteria were implemented. Duplicate studies were eliminated from consideration, as were studies written in languages other than English. The exclusion of grey literature, including blogs and short notes, was deemed necessary due to a lack of sufficient technical details. Additionally, to maintain focus on search-based approaches and experimental studies related to DT, studies lacking in these areas were not included in our analysis. Literature written in languages other than English often presents challenges due to unavailable abstracts and full texts. This lack of accessibility makes it difficult to assess the relevance of such literature to our research.

These exclusion criteria were put in place to ensure that our analysis is of the highest quality and provides valuable insights into the utilization of DT technology in cybersecurity.

Game-Theoretical Approaches in Classifying Attack and Defense Strategies in Digital Twins. Game theory, a mathematical framework for analyzing strategic interactions among rational players, offers valuable insights into cybersecurity. Key concepts such as Nash equilibrium and mixed strategies are crucial in understanding the dynamics of cyber threats and defenses (see Fig. 2) [20, 23]. Various attack strategies like data tampering and identity spoofing can be analyzed through game theory. Defense mechanisms such as encryption and anomaly detection can be modeled within a game-theoretical framework. We discuss real-world scenarios where game theory has been applied to digital twin security [22].

Within the realm of digital twin security, several types of attacks are commonly encountered. These include:

- **Data Tampering:** Manipulating the data sent to or from a digital twin to misrepresent the state or control the physical system.
- **Identity Spoofing:** Impersonating a legitimate entity in the digital twin ecosystem to gain unauthorized access or privileges.
- **Denial of Service (DoS):** Overwhelming the digital twin’s network or resources, rendering it unable to perform its functions.
- **Man-in-the-Middle Attacks:** Intercepting communication between the digital twin and other entities to steal or manipulate data.

Graph Theory in DT: Attackers and Defenders. In the realm of DT, graph theory emerges as an indispensable tool. This mathematical discipline is utilized to intricately map the relationships and dynamics within these systems, enhancing the understanding of their structural and functional complexities. However, this integration of graph theory also introduces a spectrum of vulnerabilities, particularly in the context of graph-based attacks [31, 37]. These attacks, which target the graph models underpinning digital twins, can lead to

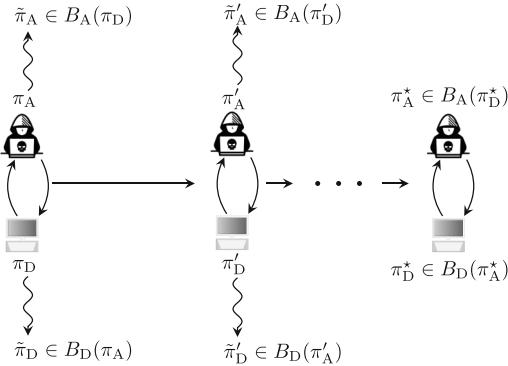


Fig. 2. The fictitious self-play process, which is a method for approximating an equilibrium of the game [23].

significant distortions in the representation of the physical systems. Such manipulations can vary in complexity and impact, ranging from alterations in node and edge attributes to more sophisticated forms of attacks that involve restructuring the graph’s topology. This poses a significant challenge in maintaining the accuracy and integrity of digital twins, especially given their growing application across diverse sectors such as manufacturing, healthcare, and urban infrastructure. While, in Table 3 provides a comprehensive classification of potential attacks and corresponding defense mechanisms in digital twins, particularly focusing on learning fields that utilize game and graph theory concepts.

Consequently, the development of robust security measures to safeguard against graph attacks is becoming a pivotal concern. This is crucial not only for the reliability of digital twin models but also for ensuring the operational safety of the systems they emulate. As such, research in this domain is increasingly focusing on identifying potential vulnerabilities and devising strategies to fortify DT against these graph-based security threats, as seen in Fig. 3.

Attack graphs in the context of DTs are a crucial application of graph theory in cybersecurity, aimed at understanding and mitigating security threats. They model potential attack paths within a system, helping to identify vulnerabilities and predict exploitation methods. We discuss various types of attack graphs, each providing unique insights into system security:

- **State Transition Graphs:** These graphs model system state changes due to potential attacks. Nodes represent system states, and edges denote transitions triggered by attack actions.
- **Attack Trees:** Hierarchical structures that decompose attacks into steps or actions. The root symbolizes the attack’s ultimate goal, with branches detailing methods to achieve this goal.
- **Exploit Dependency Graphs:** Focus on the dependencies between exploits. Nodes represent vulnerabilities, and edges illustrate dependency relations.

- **Attack-Defense Trees (ADTs):** Extend attack trees by incorporating defense mechanisms, providing a comprehensive view of attack paths and countermeasures.
- **Bayesian Attack Graphs:** Use Bayesian networks to model the probability of successful exploits and attack paths, aiding in risk assessment.
- **Kill Chain Graphs:** Represent the stages of an attack from initial reconnaissance to final damage, highlighting the attacker's objectives and tactics.

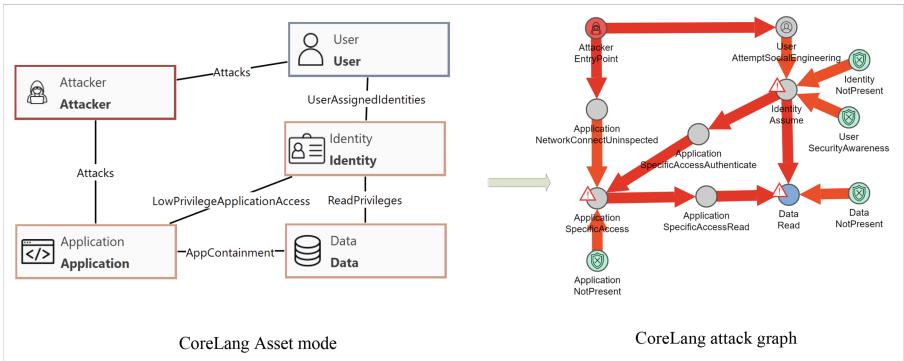


Fig. 3. The left one is the DL in coreLang asset model, and the one on the right is a selection of the coreLang attack graph [32].

2.6 Cybersecurity in Digital Twin Applications

As shown in Table 4, the cybersecurity aspect of DT applications is critical, as these systems often deal with sensitive data, control critical operations, and are interconnected with other digital infrastructures [26]. The convergence of IT (Information Technology) and OT (Operational Technology) in DTs creates a complex landscape where traditional cybersecurity strategies may not be sufficient. Ensuring the integrity, confidentiality, and availability of data within a Digital Twin is paramount, not only to protect the digital replica but also to safeguard the physical entity it represents. There are four types of the DT as shown below:

- a. **Component/Part Twins:**
 - Simulate behavior of individual system components under cyber-attacks.
 - Predict effects of known vulnerabilities, enabling proactive mitigation.
- b. **Asset Twins:**
 - Monitor real-time health and security status of larger entities like servers or routers.
 - Provide advanced warning of cyber threats through behavioral analysis.

Table 3. Deeply Advanced Classification of DTs Attacks and Defenses Across Learning Fields with Game and Graph Theory Concepts

Theory Method and Type	Potential Attacks	Defense Mechanisms
Game Theory [24]		
Zero-sum Games [41]	Nash equilibrium exploitation, Best-response attacks	Mixed strategy, Minimax theorem, Stackelberg leadership
Non-zero-sum Games [55]	Coalition attacks, Strategy manipulation	Correlated equilibrium, Mechanism design, Reputation systems
Evolutionary Games [68]	Fitness landscape distortion, Replicator dynamics manipulation	Stability analysis, Adaptive strategies, Population control
Graph Theory [46]		
Graph-based ML [71]	Edge attacks, Node influence attacks, Graph topology manipulation	Subgraph analysis, Node anonymization, Robust graph neural networks
Graph embeddings [61]	Embedding distortion, Homophily attacks, Structural hole attacks	Embedding regularization, Structure-preserving embeddings, Graph partitioning
Graph-based RL [46]	Transition graph manipulation, State graph poisoning, Reward graph tampering	Graph verification, Topological sorting, Graph isomorphism checks

c. **System/Unit Twins:**

- Model complex entities like data centers or entire network infrastructures.
- Visualize effects of an attack on one system part and its ripple effects.
- Simulate large-scale cyber-attacks to predict outcomes and devise defensive strategies.

d. **Process Twins:**

- Model data flows and organizational processes to identify vulnerable points.
- Understand and mitigate the human factor in cybersecurity risks.

DTs are being utilized across a wide range of industries, including manufacturing, healthcare, and space exploration. Despite their enormous potential, many people still lack a complete understanding of what DTs can accomplish [10]. These applications are not limited to manufacturing alone; DTs are increasingly finding applications in information technology infrastructure and space exploration as well.

Effective signaling mechanisms in 'Wireless for Twins' facilitate various services for twin-designed wireless systems [43]. The primary considerations involve air interface configuration, interactions with twin entities, and the assurance of security and privacy. The air interface is designated for signaling between twins. The connection to twin objects pertains to the link between twins and their corresponding physical devices. Moreover, it's imperative to have robust security and privacy measures in place for wireless systems based on twins [58].

The security challenges in these systems can be grouped into two main categories: the security of the physical device and the security of the interface [15]. To prevent unauthorized access to devices, edge/cloud servers, and blockchains,

reliable authentication methods are essential. In a network structured around SDN, potential risks encompass vulnerabilities in network controllers, the introduction of counterfeit control packets, improperly implemented policy directives, and insufficient device authentication [8]. Poorly configured interfaces, such as those between twins, and protocols for routing twin packets, can create multiple security loopholes. Inadequate authentication procedures and unencrypted communication channels can pave the way for security breaches [42]. The integration of DTs with 5G technologies has seen a surge of academic interest [44]. In [1] have been instrumental in framing this discourse, offering a holistic architecture tailored for an IoT-focused manufacturing setting [33, 52]. Their approach underscores the synergies of DTs and burgeoning 5G wireless modalities. Qi et al. (2021) have noted the pervasive ambiguity in the engineering sector regarding the selection of appropriate technological tools [50]. Their scholarship meticulously cataloged the foundational technologies propelling DTs and provided a roadmap for overarching research trajectories.

By predicting maintenance needs and optimizing designs, DTs help save time, money, and resources. As we move towards the future, the emergence of machine learning and AI will enable us to build even more precise and sophisticated digital twins. These advanced technologies will allow us to anticipate and prevent potential issues before they occur, leading to enhanced safety and efficiency.

Evaluation of Digital Twin Attacks Over Years. Creating a DT of a system has the potential to increase the attack surface, as adversaries can target both the physical systems and their digital counterparts [57]. However, this is particularly significant when the underlying systems are not easily accessible from external sources, as DTs can expose previously hidden aspects of an enterprise [64]. For example, in the past, accessing a data center's power supply required a technician to be physically present at a nearby control terminal. However, with the introduction of a digital twin for that infrastructure, the technician can now monitor the device remotely [40]. Unfortunately, this also opens up the possibility for hackers to exploit the digital twin if they manage to gain access.

Digital twins not only pose security risks but also offer opportunities to enhance cybersecurity within organizations [56]. Many companies are utilizing DTs as valuable tools to strengthen their defenses [39]. As shown in Fig. 4, they employ them as early-warning systems for detecting attacks, as honey traps to lure and capture malicious actors, and as testing sandboxes. By generating virtual replicas of systems, DTs enable organizations to identify and eliminate vulnerabilities through comprehensive security testing. They contribute to cybersecurity efforts by simulating the response of an actual system when faced with cyber threats. However, the main challenge lies in ensuring that the digital twin accurately represents its physical counterpart. This accuracy is crucial for the effectiveness of simulations in cybersecurity. The process of creating DTs can be straightforward for physical objects, as specific sensors can create detailed representations. However, the complexity increases when replicating systems and

Table 4. Classification of DTs in Terms of Attacks and Defenses Across Various Fields

DT Application	Potential Attacks	Defense Mechanisms
IoT Applications [63]	Device spoofing Eavesdropping Data tampering	Strong authentication Encryption Regular software updates
Smart Homes [62]	Unauthorized access Privacy breaches Malware	Network security Anomaly detection Access controls
Smart Cities [18]	Service disruption Surveillance Data manipulation	Resilient infrastructure Cyber-incident response teams Privacy-aware data handling
Energy Management [67]	Grid manipulation Energy theft Supply chain attacks	Real-time monitoring Intrusion detection systems Secure communication protocols
Intelligent Transportation Systems [38, 53, 60]	GPS spoofing Signal jamming Vehicle hacking	Redundant systems Secure V2X communication Continuous system evaluation
Healthcare [2]	Patient data compromise Ransomware Service denial	HIPAA compliance Endpoint security Disaster recovery plans
Agriculture [49]	Data falsification Equipment theft Supply chain attacks	Access management Equipment tracking Data integrity checks
Education [30]	Academic data breaches Identity theft Cheating and fraud	Data privacy policies User education Secure testing environments
Sports [36]	Performance data theft Illegal betting Doping data manipulation	Anonymization of data Fair play technology Security awareness training
Government [27]	Espionage Election interference Public record tampering	Secure civic engagement platforms Transparent processes Audit trails
Retail [17]	Payment fraud Customer data breaches Supply chain disruption	PCI DSS compliance Anti-fraud algorithms Secure logistics
Infrastructure Security [4]	Critical infrastructure attacks Terrorist threats Natural disaster exploitation	NIST framework adherence Red team exercises Emergency response protocols

processes, necessitating manual revisions to ensure emulation or simulation accuracy. Additionally, there are concerns about the affordability of digital twin technology, especially for small businesses, which poses a challenge for widespread adoption across various industries.

Simulation of Cyber Threats with Digital Twins. In the increasingly complex world of cybersecurity, the use of DTs for the emulation or simulation of cyber threats presents a cutting-edge approach. DTs - virtual replicas of physical systems - can be effectively utilized to understand, analyze, and mitigate a wide array of cyber threats [13]. By creating detailed simulations, these digital models allow for a safer and more controlled environment to study malware behavior, botnet structures, and various cyber attack strategies. From testing the resilience of systems against sophisticated phishing attacks to exploring the nuances of AI-driven password cracking techniques, this approach provides invaluable insights into the dynamics of cyber threats. Additionally, it aids in the development of

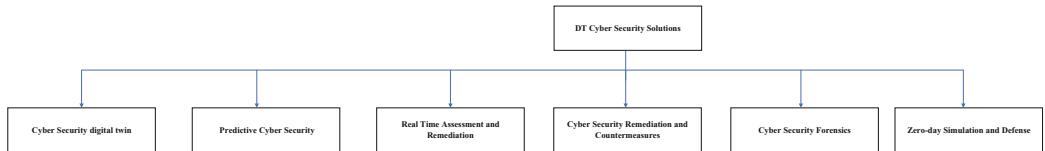


Fig. 4. DT in Cyber Security Solutions

robust defense mechanisms, predictive models for potential attacks, and strategies for real-time threat detection and response.

Several use cases for Adverse AI-Digital Twin have been summarized as shown in Fig. 5, and itemized as below:

1. Rural Movements: This could imply the simulation of population movements or behaviors in rural areas, which may have implications for understanding and predicting social or political events.
2. Behavior analysis: This could refer to using AI to analyze patterns of behavior which could be replicated in a digital twin environment.
3. System mapping: This involves creating a digital representation of a physical system, capturing all its components and interactions, which is a fundamental part of building a digital twin.
4. Natural language manipulations: This may involve using AI to understand and generate natural language, which can be used to simulate human-like interactions within a digital twin environment.
5. Command and control of botnets: Refers to using DTs to simulate the command and control structures of botnets for the purpose of understanding how they operate or to train defensive AI systems.
6. Self-learning malware: This could imply creating a digital twin of a system to train malware that can adapt and learn from its environment to improve its effectiveness.
7. Domain generation: This typically refers to algorithms used by malware to generate new domain names for command and control servers, which can be tested and refined within a digital twin.
8. Discovery obfuscation: This might involve using a digital twin to test methods of hiding from discovery tools, making it harder for defensive systems to detect the digital twin's actions.
9. Low-and-slow exfiltration: This use case could involve simulating data exfiltration techniques that aim to be slow and stealthy to avoid detection, which could be tested against a digital twin of a network.
10. Captcha attack: Testing AI systems that can bypass CAPTCHA protections within a simulated digital environment.
11. Password attack: Simulating various password attack techniques like brute force or dictionary attacks to understand how they can be executed and mitigated.
12. Classifier manipulation: This involves using a digital twin to test ways to fool machine learning classifiers, which can be used to bypass security systems.

13. Attack code generator: Developing AI that can autonomously create new attack vectors or malware, tested within the safety of a digital twin environment.
14. Phishing: Simulating phishing attacks within a digital twin to understand how they work and how to defend against them.
15. Attack planning: Using a digital twin to plan cyber attacks, possibly to understand their potential impact and to develop countermeasures.
16. Outcome Prediction: Using AI to predict the outcomes of actions taken within the digital twin, which can inform decision-making in the real world.
17. Vulnerability detection: Scanning the digital twin for vulnerabilities as a means to anticipate and mitigate potential security threats.
18. Target profiling: Collecting data on potential targets within a digital twin to understand how to better tailor attacks.
19. Cyber Threat Intelligence: Using the digital twin to gather intelligence that could be used in the planning and execution of adversarial actions.

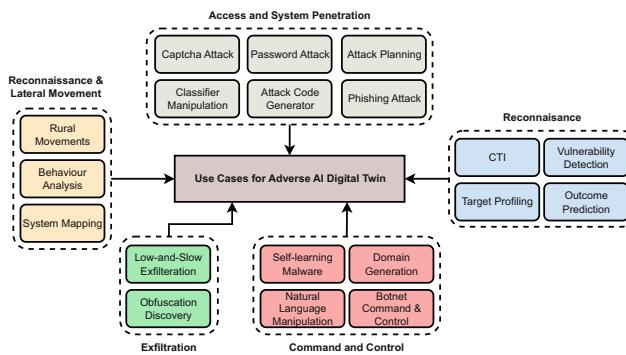


Fig. 5. Use cases for Adverse AI-Digital Twin

Intrusion Response Mechanism in Digital Twin. While , the Table 5 illustrates various intrusion response mechanisms implemented in Digital Twin systems, categorized into active and passive responses. Active responses include measures that directly interact with the system or user, such as generating reports, enabling additional logging, locking user accounts, terminating sessions, or even shutting down hosts to mitigate the impact of an attack [5]. Passive responses focus on alerting and preparing the system to better handle potential intrusions by generating alarms, creating backups, blocking IP addresses, and restricting user activities [5]. These mechanisms collectively enhance the resilience of Digital Twin systems against security threats.

Table 5. Intrusion Response Mechanism in Digital Twin Systems

Category	Response Mechanisms
Active Responses [5]	Generate a Report Enable Additional Logging Enable Additional IDS Lock User Account Terminate User Session Shutdown Host Disable the attacked port of Service Trace the Connection Employ Temporary Shadow Files
Passive Responses [5]	Generate an Alarm Enable Remote Logging Create Backup Suspend User Jobs Block IP address Disconnect from the Network Warn the Intruder Force Additional Authentication Restrict User Activity

Advanced Classification of DTs Attacks and Defenses Across ML, DL, RL Learning Fields. Table 6 provides an advanced classification of various attacks and defense mechanisms applicable across different learning methods, specifically within the fields of Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL). Each row in the table details specific learning types, the potential attacks they are susceptible to, and the corresponding defense mechanisms that can be employed to mitigate these threats.

– Machine Learning (ML)

- *Supervised Learning:* Supervised learning methods face a range of attacks including data snooping, evasion attacks, and model inversion [65]. Defense mechanisms to counter these include feature selection, regularization techniques, and differential privacy to protect sensitive information during the learning process.
- *Unsupervised Learning:* This category is vulnerable to attacks like data poisoning, clustering attacks, and adversarial manipulations [7]. Defenses such as noise injection, robust clustering techniques, and secure multi-party computation help in safeguarding the integrity of the models.

– Deep Learning (DL)

- *Convolutional Neural Networks (CNN):* CNNs are particularly susceptible to adversarial image attacks, model inversion, and transfer attacks [39]. To defend against these, techniques like feature squeezing, adversarial training, and spatial smoothing are commonly used to enhance model robustness.
- *Recurrent Neural Networks (RNN):* RNNs face unique threats such as sequence attacks, state inference, and adversarial sequence generation

- [35]. Defense mechanisms like gradient masking, regularization, and memory protection are employed to secure RNNs against these vulnerabilities.
- *Transformers*: The powerful attention mechanisms in transformers make them targets for attention attacks, embedding attacks, and model extraction threats [51]. Defenses include attention masking, knowledge distillation, and the use of secure embeddings to reduce the impact of these attacks.

– Reinforcement Learning (RL)

- *Model-free RL*: In model-free RL, exploration exploitation attacks, policy poisoning, and reward shaping attacks are prevalent [29]. Defensive strategies such as safe exploration, regularized policies, and reward clipping are critical to maintaining the integrity of learning processes in dynamic environments.
- *Model-based RL*: Model-based RL methods are vulnerable to model biasing, transition dynamics attacks, and reward inference attacks [69]. To combat these, model ensembling, transition verification, and robust reward estimation techniques are employed to ensure reliable performance.

This classification highlights the diverse range of attack vectors in the various fields of learning and underscores the importance of tailored defense mechanisms to ensure the security and reliability of DTs in different AI applications.

Table 6. Advanced Classification of DTs Attacks and Defenses Across ML, DL, RL Learning Fields

Learning Method and Type	Potential Attacks	Defense Mechanisms
Machine Learning (ML)		
Supervised [65]	Data snooping, Evasion attacks, Model inversion	Feature selection, Regularization, Differential privacy
Unsupervised [7]	Data poisoning, Clustering attacks, Adversarial attacks	Noise injection, Robust clustering, Secure multi-party computation
Deep Learning (DL)		
Convolutional Neural Networks (CNN) [39]	Adversarial image attacks, Model inversion, Transfer attacks	Feature squeezing, Adversarial training, Spatial smoothing
Recurrent Neural Networks (RNN) [35]	Sequence attacks, State inference, Adversarial sequence generation	Gradient masking, Regularization, Memory protection
Transformers [51]	Attention attacks, Embedding attacks, Model extraction	Attention masking, Knowledge distillation, Secure embeddings
Reinforcement Learning (RL)		
Model-free RL [29]	Exploration exploitation, Policy poisoning, Reward shaping attacks	Safe exploration, Regularized policies, Reward clipping
Model-based RL [69]	Model biasing, Transition dynamics attacks, Reward inference attacks	Model ensembling, Transition verification, Robust reward estimation

Cyber Threat Intelligence in DTs. As sensors and enterprise networks become more integrated into industrial environments, security strategies struggle to keep up with the expanding attack surface resulting from the convergence of IT infrastructure and industrial systems [11]. Security Operations Centers (SOCs) are overwhelmed with the task of integrating these systems, and DTs offer a valuable concept for effectively monitoring industrial assets [12]. DTs can enhance enterprise security by simulating attacks and analyzing their impact on virtual assets. However, the integration of security simulations for DTs into enterprise security strategies, primarily managed by SOCs, is currently being overlooked[16].

3 Search Results

This section presents the findings from our systematic literature review, which focused on identifying how DT can be integrated into cybersecurity frameworks. Our review was structured around five key research questions (RQs) aimed at exploring the integration of DTs, their predictive capabilities, real-time assessment potential, remediation strategies, and their role in simulating and defending against zero-day vulnerabilities.

Below, we summarize the insights gathered from the literature concerning each research question.

– RQ1: DT Integration

The integration of Digital Twin (DT) technology into existing cybersecurity frameworks has shown promise in enhancing threat detection and mitigation capabilities. Studies indicate that DTs can provide real-time, dynamic insights into system vulnerabilities, allowing for more proactive security measures. However, challenges remain in ensuring seamless integration with legacy systems and maintaining data integrity during synchronization processes.

– RQ2: Predictive Cybersecurity

DT technology can significantly improve the accuracy and reliability of predictive cybersecurity measures by simulating potential attack scenarios and assessing system vulnerabilities before they can be exploited. Machine learning algorithms integrated within DTs help in forecasting threats, enabling early intervention. Despite these advancements, there is a need for more robust models that can handle the complexity of real-world cyber threats.

– RQ3: Real-Time Assessment and Remediation

DTs play a critical role in enabling real-time assessment and remediation of cybersecurity threats. By continuously mirroring the physical system, DTs provide immediate feedback on the state of the system, allowing for prompt detection of anomalies and quick deployment of countermeasures. This real-time capability is crucial for minimizing the impact of cyberattacks. However, the effectiveness of this approach is contingent on the quality of the data fed into the DT and the speed of the system's response.

– **RQ4: Remediation and Countermeasures**

DT technology supports the development and deployment of effective remediation strategies and countermeasures by modeling the impact of various defensive actions in a simulated environment. This capability allows organizations to test and refine their cybersecurity strategies without risking the actual system. The primary challenge lies in ensuring that the DT accurately reflects the real system, as any discrepancies could lead to ineffective or even counterproductive responses.

– **RQ5: Zero-Day Simulation and Defense**

The simulation capabilities of DTs are particularly valuable for defending against zero-day vulnerabilities. By replicating potential zero-day attacks in a controlled environment, DTs enable security teams to understand and prepare for these unknown threats. The ability to simulate and rehearse responses to zero-day attacks before they occur is a significant advantage, although it requires continuous updates to the DT to remain relevant as new vulnerabilities emerge.

4 Conclusion

In the context of dynamical systems, balance is a necessary requirement, under which different purposes, like robustness, stability, and security, can be required of DT in different ways. Production review, including construction for DT, is still at an introductory stage, especially at the theoretical level, because of challenges from changes in the situation, modeling mistakes of physical and software services, undesired network-induced phenomena, and various attacks and disturbances. Immobility problems on multi-agent methods and clever grids are investigated using the recently developed modeling.

Acknowledgement. This work was funded by the European Union as part of the European Defence Fund (EDF) project AIInception (GA No. 101103385). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

References

1. Al Ali, A., Cabibihan, J.J., Meskin, N., Rossi, S., Jiang, W., He, H., Ge, S.S.: Social Robotics: 15th International Conference, ICSR 2023, Doha, Qatar, December 3–7, 2023, Proceedings, Part I, vol. 14453. Springer Nature (2023)
2. Al-Dalati, I.: Digital twins and cybersecurity in healthcare systems. In: Digital Twin for Healthcare, pp. 195–221. Elsevier (2023)
3. Alcaraz, C., Lopez, J.: Digital twin: A comprehensive survey of security threats. IEEE Commun. Surv. Tutori. (2022)
4. Allison, D., Smith, P., McLaughlin, K.: Digital twin-enhanced incident response for cyber-physical systems. In: Proceedings of the 18th International Conference on Availability, Reliability and Security, pp. 1–10 (2023)

5. Anwar, S., et al.: From intrusion detection to an intrusion response system: fundamentals, requirements, and future directions. *Algorithms* **10**(2), 39 (2017)
6. Attaran, M., Celik, B.G.: Digital twin: Benefits, use cases, challenges, and opportunities. *Decision Anal.* J., 100165 (2023)
7. Bernieri, G., Conti, M., Turrin, F.: Evaluation of machine learning algorithms for anomaly detection in industrial networks. In: 2019 IEEE International Symposium on Measurements & Networking (M&N). pp. 1–6. IEEE (2019)
8. Bhuiyan, Z.A., Islam, S., Islam, M.M., Ullah, A.A., Naz, F., Rahman, M.S.: On the (in) security of the control plane of SDN architecture: a survey. *IEEE Access* (2023)
9. Böhm, F., Dietz, M., Preindl, T., Pernul, G.: Augmented reality and the digital twin: state-of-the-art and perspectives for cybersecurity. *J. Cybersecur. Privacy* **1**(3), 519–538 (2021)
10. Böttjer, T., et al.: A review of unit level digital twin applications in the manufacturing industry. *CIRP J. Manuf. Sci. Technol.* **45**, 162–189 (2023)
11. Dietz, M., Schlette, D., Pernul, G.: Harnessing digital twin security simulations for systematic cyber threat intelligence. In: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 789–797. IEEE (2022)
12. Dietz, M., Vielberth, M., Pernul, G.: Integrating digital twin security simulations in the security operations center. In: Proceedings of the 15th International Conference on Availability, Reliability and Security, pp. 1–9 (2020)
13. Eckhart, M., Ekelhart, A.: Towards security-aware virtual environments for digital twins. In: Proceedings of the 4th ACM Workshop on Cyber-Physical System Security, pp. 61–72 (2018)
14. Eckhart, M., Ekelhart, A.: Digital twins for cyber-physical systems security: State of the art and outlook. *Security and Quality in Cyber-Physical Systems Engineering: With Forewords by Robert M. Lee and Tom Gilb*, pp. 383–412 (2019)
15. El-Kady, A.H., Halim, S., El-Halwagi, M.M., Khan, F.: Analysis of safety and security challenges and opportunities related to cyber-physical systems. *Process Safety and Environmental Protection* (2023)
16. Faleiro, R., Pan, L., Pokhrel, S.R., Doss, R.: Digital twin for cybersecurity: Towards enhancing cyber resilience. In: Broadband Communications, Networks, and Systems: 12th EAI International Conference, BROADNETS 2021, Virtual Event, October 28–29, 2021, Proceedings 12, pp. 57–76. Springer (2022)
17. Far, S.B., Rad, A.I.: Applying digital twins in metaverse: user interface, security and privacy challenges. *J. Metaverse* **2**(1), 8–15 (2022)
18. Farsi, M., et al.: Digital Twin Technologies and Smart Cities. Springer (2020)
19. Grieves, M.: Digital twin: manufacturing excellence through virtual factory replication. *Cybersecur. J.* **12**(3), 123–134 (2017). <https://doi.org/10.1007/springer12345>
20. Hammar, K., Stadler, R.: Learning security strategies through game play and optimal stopping (2022). arXiv preprint [arXiv:2205.14694](https://arxiv.org/abs/2205.14694)
21. Hammar, K., Stadler, R.: Digital twins for security automation. In: NOMS 2023–2023 IEEE/IFIP Network Operations and Management Symposium, pp. 1–6. IEEE (2023)
22. Hammar, K., Stadler, R.: Digital twins for security automation. In: NOMS 2023–2023 IEEE/IFIP Network Operations and Management Symposium, pp. 1–6 (2023). [10.1109/NOMS56928.2023.10154288](https://doi.org/10.1109/NOMS56928.2023.10154288)
23. Hammar, K., Stadler, R.: Learning near-optimal intrusion responses against dynamic attackers (2023). arXiv preprint [arXiv:2301.06085](https://arxiv.org/abs/2301.06085)
24. Hammar, K., Stadler, R.: Scalable learning of intrusion responses through recursive decomposition (2023). arXiv preprint [arXiv:2309.03292](https://arxiv.org/abs/2309.03292)

25. Holmes, D., Papathanasaki, M., Maglaras, L., Ferrag, M.A., Nepal, S., Janicke, H.: Digital twins and cyber security—solution or challenge? In: 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), pp. 1–8. IEEE (2021)
26. Homaei, M., Mogollón-Gutiérrez, Ó., Sancho, J.C., Ávila, M., Caro, A.: A review of digital twins and their application in cybersecurity based on artificial intelligence. *Artif. Intell. Rev.* **57**(8), 1–65 (2024)
27. Hossain, S.T., Yigitcanlar, T., Nguyen, K., Xu, Y.: Local government cybersecurity landscape: a systematic review and conceptual framework. *Appl. Sci.* **14**(13), 5501 (2024)
28. Hu, W., Chang, C.H., Sengupta, A., Bhunia, S., Kastner, R., Li, H.: An overview of hardware security and trust: threats, countermeasures, and design tools. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **40**(6), 1010–1038 (2020)
29. Jaber, A.: Transforming cybersecurity dynamics: Enhanced self-play reinforcement learning in intrusion detection and prevention system. In: 2024 IEEE International Systems Conference (SysCon), pp. 1–8. IEEE (2024)
30. Kandasamy, N.K., Venugopalan, S., Wong, T.K., Nicholas, L.J.: Epictwin: an electric power digital twin for cyber security testing, research and education (2021). arXiv preprint [arXiv:2105.04260](https://arxiv.org/abs/2105.04260)
31. Katsikeas, S., Hacks, S., Johnson, P., Ekstedt, M., Lagerström, R., Jacobsson, J., Wällstedt, M., Eliasson, P.: An attack simulation language for the it domain. In: International Workshop on Graphical Models for Security, pp. 67–86. Springer (2020)
32. Katsikeas, S., Ling, E.R., Johnsson, P., Ekstedt, M.: Empirical evaluation of a threat modeling language as a cybersecurity assessment tool. *Comput. Secur.* **140**, 103743 (2024)
33. Kirchhof, J.C., Malcher, L., Rumpe, B.: Understanding and improving model-driven iot systems through accompanying digital twins. In: Proceedings of the 20th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences, pp. 197–209 (2021)
34. Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering—a systematic literature review. *Inf. Softw. Technol.* **51**(1), 7–15 (2009)
35. KP, S., et al.: Rnmsecuronet: Recurrent neural networks for cyber security use-cases (2019). arXiv preprint [arXiv:1901.04281](https://arxiv.org/abs/1901.04281)
36. Laamarti, F.: Towards Standardized Digital Twins for Health, Sport, and Well-being. Ph.D. thesis, Université d’Ottawa/University of Ottawa (2019)
37. Lagerstrom, R., et al.: Probabilistic model for graph-based security analysis. *J. Cybersecur.* **10**(3), 200–213 (2023). <https://doi.org/10.1093/cybsec/tyab009>
38. Liu, J., Li, C., Bai, J., Luo, Y., Lv, H., Lv, Z.: Security in IoT-enabled digital twins of maritime transportation systems. *IEEE Trans. Intell. Transp. Syst.* (2021)
39. Lv, Z., Chen, D., Cao, B., Song, H., Lv, H.: Secure deep learning in defense in deep-learning-as-a-service computing systems in digital twins. *IEEE Trans. Comput.* (2023)
40. Lv, Z., Fridenfalk, M.: Digital twins for building industrial metaverse. *J. Adv. Res.* (2023)
41. Meakins, J.: A zero-sum game: the zero-day market in 2018. *J. Cyber Policy* **4**(1), 60–71 (2019)
42. Neupane, S., et al.: Security considerations in AI-robotics: a survey of current methods, challenges, and opportunities (2023). arXiv preprint [arXiv:2310.08565](https://arxiv.org/abs/2310.08565)

43. Ngo, D.T., Aouedi, O., Piamrat, K., Hassan, T., Raipin-Parvédy, P.: Empowering digital twin for future networks with graph neural networks: overview, enabling technologies, challenges, and opportunities. *Future Internet* **15**(12), 377 (2023)
44. Nguyen, H.X., Trestian, R., To, D., Tatipamula, M.: Digital twin for 5G and beyond. *IEEE Commun. Mag.* **59**(2), 10–15 (2021)
45. Nguyen, T.N.: Toward human digital twins for cybersecurity simulations on the metaverse: ontological and network science approach. *JMIRx Med.* **3**(2), e33502 (2022)
46. Nyberg, J., Johnson, P., Méhes, A.: Cyber threat response using reinforcement learning in graph-based attack simulations. In: NOMS 2022–2022 IEEE/IFIP Network Operations and Management Symposium, pp. 1–4. IEEE (2022)
47. Parnianifard, A., Jearavongtakul, S., Sasithong, P., Simpan, N., Poomrittigul, S., Bajpai, A., Vanichchanunt, P., Wuttisittikulkij, L.: Digital-twins towards cyber-physical systems: a brief survey. *Eng. J.* **26**(9), 47–61 (2022)
48. Pinto, A., Herrera, L.C., Donoso, Y., Gutierrez, J.A.: Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors* **23**(5), 2415 (2023)
49. Purcell, W., Neubauer, T.: Digital twins in agriculture: a state-of-the-art review. *Smart Agric. Technol.* **3**, 100094 (2023)
50. Qi, Q., et al.: Enabling technologies and tools for digital twin. *J. Manuf. Syst.* **58**, 3–21 (2021)
51. Qin, B., Pan, H., Dai, Y., Si, X., Huang, X., Yuen, C., Zhang, Y.: Machine and deep learning for digital twin networks: A survey. *IEEE Internet Things J.* (2024)
52. Raymat, D., Chaker, M.: Analysis of performance parameters for service assurance in radio access networks (2023)
53. Sellitto, G.P., Masi, M., Pavleska, T., Aranha, H.: A cyber security digital twin for critical infrastructure protection: the intelligent transport system use case. In: IFIP Working Conference on The Practice of Enterprise Modeling, pp. 230–244 (2021)
54. Shaikh, E., Mohammad, N., Al-Ali, A., Muhammad, S.: A probabilistic model checking (PMC) approach to solve security issues in digital twin (DT). In: 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 192–197. IEEE (2023)
55. Soper, B.C.: Non-zero-Sum, Adversarial Detection Games in Network Security. University of California, Santa Cruz (2015)
56. Sorensen, A.: Risk Management in Digital Twin Systems. Ph.D. thesis, University of Technology (2023)
57. Stefanidou, A., et al.: Leveraging digital twin technologies for public space protection and vulnerability assessment (2024). arXiv preprint [arXiv:2408.17136](https://arxiv.org/abs/2408.17136)
58. Tao, Z., Xu, W., Huang, Y., Wang, X., You, X.: Wireless network digital twin for 6g: Generative AI as a key enabler (2023). arXiv preprint [arXiv:2311.17451](https://arxiv.org/abs/2311.17451)
59. Wanasinghe, T.R., et al.: Digital twin for the oil and gas industry: overview, research trends, opportunities, and challenges. *IEEE Access* **8**, 104175–104197 (2020)
60. Wang, Z., Lv, C., Wang, F.Y.: A new era of intelligent vehicles and intelligent transportation systems: digital twins and parallel intelligence. *IEEE Trans. Intell. Veh.* (2023)
61. Xiao, Q., Liu, J., Wang, Q., Jiang, Z., Wang, X., Yao, Y.: Towards network anomaly detection using graph embedding. In: Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part IV 20, pp. 156–169. Springer (2020)

62. Xiao, Y., Jia, Y., Hu, Q., Cheng, X., Gong, B., Yu, J.: Commandfence: a novel digital-twin-based preventive framework for securing smart home systems. *IEEE Trans. Dependable Secure Comput.* (2022)
63. Xu, H., Wu, J., Pan, Q., Guan, X., Guizani, M.: A survey on digital twin for industrial internet of things: applications, technologies and tools. *IEEE Commun. Surv. Tutor.* (2023)
64. Xu, J., et al.: Traversing digital twins in cybersecurity. *Cybersecur. J.* **12**(3), 123–134 (2023). <https://doi.org/10.1007/springer12345>
65. Xu, Q., Ali, S., Yue, T.: Digital twin-based anomaly detection in cyber-physical systems. In: 2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST), pp. 205–216. IEEE (2021)
66. Xu, Q., Ali, S., Yue, T.: Digital twin-based anomaly detection with curriculum learning in cyber-physical systems. *ACM Trans. Softw. Eng. Methodol.* (2023)
67. Yu, W., Patros, P., Young, B., Klinac, E., Walmsley, T.G.: Energy digital twin technology for industrial energy management: classification, challenges and future. *Renew. Sustain. Energy Rev.* **161**, 112407 (2022)
68. Zhang, H., Tan, J., Liu, X., Huang, S., Hu, H., Zhang, Y.: Cybersecurity threat assessment integrating qualitative differential and evolutionary games. *IEEE Trans. Netw. Serv. Manag.* **19**(3), 3425–3437 (2022)
69. Zhang, T., et al.: When moving target defense meets attack prediction in digital twins: a convolutional and hierarchical reinforcement learning approach. *IEEE J. Sel. Areas Commun.* (2023)
70. Zheng, T., Liu, M., Puthal, D., Yi, P., Wu, Y., He, X.: Smart grid: cyber attacks, critical defense approaches, and digital twin (2022). arXiv preprint [arXiv:2205.11783](https://arxiv.org/abs/2205.11783)
71. Zonneveld, G., Principi, L., Baldi, M.: Using graph theory for improving machine learning-based detection of cyber attacks (2024). arXiv preprint [arXiv:2402.07878](https://arxiv.org/abs/2402.07878)



A Detour Route Selection Method Based on Node Density in Skip Graph

Riku Kamiya and Tomoya Kawakami^(✉)

Graduate School of Engineering, University of Fukui, Fukui, Japan
tomoya-k@u-fukui.ac.jp

Abstract. Among overlay networks, those that form a certain data structure to realize efficient routing are called structured overlay. One type of range-searchable structured overlay is Skip Graph. However, the Skip Graph routing does not always result in the shortest path, and Detouring Skip Graph, a method that uses detouring paths, has been proposed to achieve more efficient routing. In this paper, based on the Detouring Skip Graph algorithm, we propose an algorithm that selects detouring routes based on node density to improve the routing efficiency by adjusting parameters such as thresholds. The results show that the proposed method is slightly more efficient than Detouring Skip Graph.

1 Introduction

In recent years, IoT devices have been increasing, and the amount of data handled has been growing along with it. The client-server model of managing data is a centralized data management system that uses a server, but this may place a heavy burden on the server. A solution to this problem is a P2P model. In the P2P model, computers (nodes) on the network communicate with each other on an equal basis, rather than through a server as described above.

P2P networks can be broadly classified into two categories: structured overlay networks and unstructured overlay networks. First, an overlay network is an application-level independent network built on an existing network such as the Internet. In particular, a structured overlay network is an overlay network in which the adjacencies between nodes are regular. The structured overlay network provides efficient routing to target nodes with high scalability. In addition, structured overlay networks such as Skip Graph [1] and Chord# [8] treat range search, and the related methods have been proposed in a range-searchable environment [6, 9].

There are studies of hierarchical Skip Graph in environments with a limited list of nodes [3], and studies of range-key Skip Graph (RKSG) [4] that show that the routing efficiency is comparable to that of Multi-Range Forwarding (MRF) [2] by allowing the registration of range keys, which are keys with a range, and that the routing table can be greatly simplified. Detouring Skip Graph [5], which is a modification of the routing algorithm to utilize detouring paths in order to achieve routing closer to the shortest path.

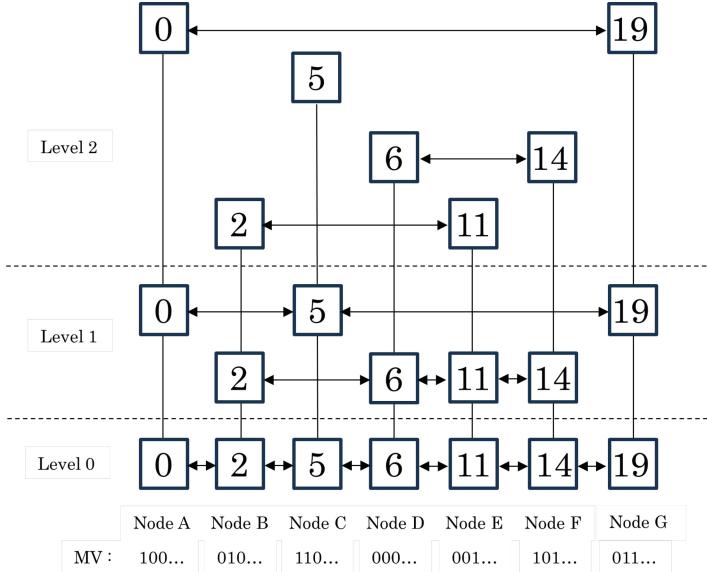


Fig. 1. Example of skip graph topology

In this paper, we focus on the Detouring Skip Graph algorithm and propose a selection algorithm of detouring routes based on different criteria to achieve more efficient routing in the viewpoint of the load on the entire network.

2 Related Work

2.1 Skip Graph

Skip Graph [1] is a type of structured overlay. The main feature is the topology shown in Fig. 1, which is based on the skip list [7]. Each node has a key that is total ordered and a randomly generated membership vector (MV) string. Nodes that match the first i digits of the MV value form a bidirectional link at level i . The topology is such that each node belongs to multiple linked lists. Thus, as the level value increases, the likelihood of having links to distant nodes in the key space increases. At level 0, all nodes belong to a single linked list.

In Skip Graph routing, when a search query is issued at a node, the search starts at the maximum level at that node. The condition for deciding whether or not to send the query to a node is that the key of the node with a link is less than or equal to the query's search key. At the maximum level, the system determines whether the node with the link satisfies the sending condition, and if so, sends the query with information on the starting node, the search key, and the level at which the query was sent. These operations are repeated, and if the key of the node that received the query matches the search key, the discovery query is sent to the start node, and routing is completed.

2.2 Detouring Skip Graph

Detouring Skip Graph [5] is a type of structured overlay proposed by Kaneko and others. It is a Skip Graph routing algorithm with two modifications described below.

2.2.1 Search from the Maximum Level

The first, search from the maximum level, means that every time a query starts routing from a new node, it searches from the maximum level. In Skip Graph routing, the level information is maintained when the query is forwarded, and the level decreases monotonically throughout the routing. On the other hand, by performing a search from the maximum level each time a query is moved, links formed at high levels can be utilized, and shorter route lengths can be achieved.

2.2.2 Use of Detour Routes

The second use of a bypass route is to go through a node with a key larger than the search key. In Skip Graph routing, a node with a key smaller than the search key is a condition for sending a query, so a node with a key larger than the search key is not routed. On the other hand, a short route length can be achieved by using a detour route.

To determine whether or not to utilize a detour route, when the key key_{next} of a node with a link at the current level is larger than the search key key_{target} , the median of the key key and the key key_{lower} of the node with a link at the lower level is used, and if the median If the search key is greater than the median, a detour route is selected. This is equivalent to satisfying the following equation.

$$\frac{key_{lower} + key_{next}}{2} < key_{target} \quad (1)$$

In other words, the decision method is such that the algorithm selects the node with the smaller distance to the search key in the key space when comparing nodes with links at the current level and nodes with links at lower levels.

3 Proposed Method

3.1 Issues

The objective of this research is to further improve the efficiency of routing based on the Skip Graph and Detouring Skip Graph presented in Sect. 2, thereby reducing the load on the overall network.

As described in Sect. 2.2, Detouring Skip Graph improves routing efficiency by modifying the algorithm of the skip graph. However, focusing on the algorithm, we can see that it is not always possible to select the shortest paths in routing.

The reason why the efficiency is worse than that of a Skip Graph is that the decision to select a detour route is based on the median of the keys, which

may not be able to handle environments where the distribution of nodes in the key space is biased. Therefore, to improve the efficiency of routing based on the Detouring Skip Graph, this paper proposes a method to select a detour route using the density of the distribution of nodes instead of the median of the keys of multiple nodes. However, in this study, it is assumed that the information on the distribution of nodes is available in advance.

3.2 Overview

As shown in Sect. 3.1, in order to realize an algorithm that can perform more appropriate routing in environments where the distribution of nodes in the key space is uneven, we propose to change the decision method for selecting a detour route to one that uses information on the density of nodes around the query. The specifics are described below. For the sake of explanation, we will refer to the larger key in the key space as the right side and the smaller key as the left side.

Figure 2 shows the model of the proposed method.

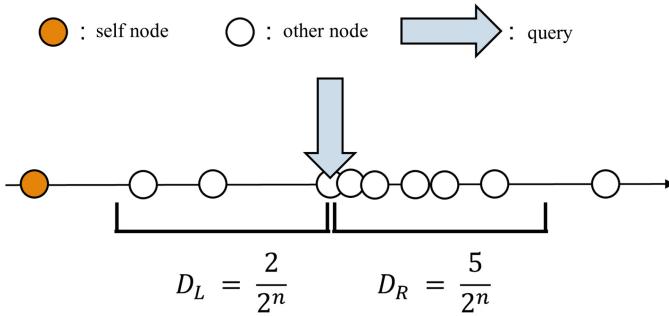


Fig. 2. Model of the proposed method

The overall flow of the proposed method is as follows.

1. Determine the range 2^n for measuring node density and threshold T
2. measure the node densities on the left and right sides of the query
3. Compare the difference between the left and right node densities and the threshold T
4. Based on the comparison results, select an algorithm that utilizes the detour route or does not utilize the detour route, and perform routing

First, 2^n , the range over which the density of nodes is measured in the key space, and the threshold value T are set in advance. The threshold T is the value expressed below, based on the percentage of nodes P and the number of nodes N_{all} in the entire network, where the number of nodes in the entire network is 1.

$$T = \frac{N_{\text{all}} \times P}{2^n} \quad (2)$$

When a query is generated, the number of nodes N_R and N_L on the left and right 2^n of the query are counted, respectively, and the density of nodes on the right side of the query D_R and the density of nodes on the left side of the query D_L for the key space 2^n are calculated by the following formula. $D_R = \frac{N_R}{2^n}$, $D_L = \frac{N_L}{2^n}$. Then, the difference between D_R and D_L is compared with the threshold T according to the following formula.

$$D_R - D_L > T \quad (3)$$

When Eq. (3) is true, the algorithm does not select a detour route, and when Eq. (3) is false, routing is performed using the Detouring Skip Graph algorithm.

4 Evaluation

4.1 Simulation Environments

The simulation environment is shown in Table 1. The queries are set to a Gaussian distribution with the center of the key space as the mean value, and the nodes are set to a distribution as shown in Figs. 3 and 4. This creates an environment where the distribution of the nodes is biased, which tends to deteriorate the results of the Skip Graph.

Table 1. Simulation environments

Key space	$65536 (= 2^{16})$
Number of queries per simulation	10000
Number of simulation trials	10
Number of nodes	250, 500, 750, 1000
Query distribution	Fig. 3 Gaussian distribution (ave.: median key space standard deviation: ave./5)
Node Distribution	Fig. 4
Evaluation index	Number of messages, worst hop count

In Sects. 4.2 and 4.3, we will examine appropriate values for the variable n and the threshold value T . Section 4.4 then uses the values obtained to make a final evaluation.

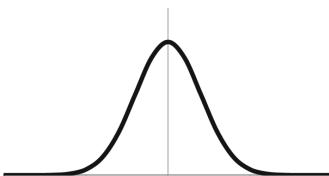


Fig. 3. Query distribution

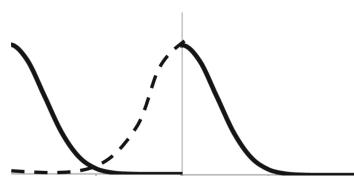


Fig. 4. Node distribution

4.2 Width of Used Node Density

The results of varying the range 2^n over which the density of nodes is measured are shown in Figs. 5 and 6. The horizontal axis is the threshold value, and the values of 2^n are compared at 12, 13, 14 and 15. Since n is the range over which the node density is measured, the percentage of the range 2^n in the total key space represented by 2^{16} is 6.25%, 12.5%, 25%, and 50%, respectively.

Figures 5 and 6 show the results for the number of messages for 500 nodes. Figure 5 shows the results for the number of messages for 500 nodes. The results show that the number of messages is reduced the most when $n = 14$. Figure 6 shows the results for the number of hops. The results show a slightly better result for $n = 13$, but not much difference for any value. From the above, $n = 14$ is considered a more suitable value.

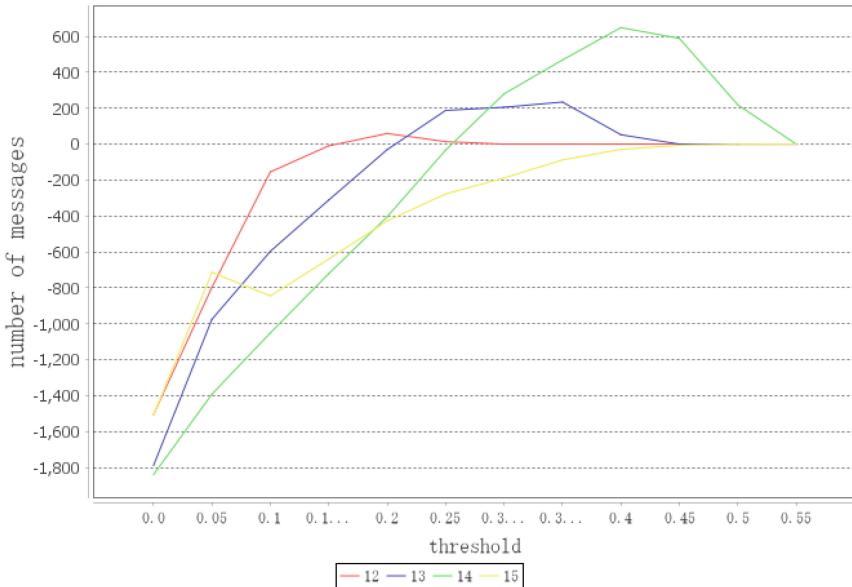


Fig. 5. Number of messages when the number of nodes is 500

Table 2. Optimal threshold in the number of messages

number of nodes	message	worst hop
250	0.44	0.44
500	0.40	0.43
750	0.48	0.49
1000	0.43	0.47

4.3 Comparison of Threshold Values

Figures 7 and 8 show the simulation results of the proposed method when the threshold T is varied in a narrow interval. $n = 14$ was set for the variable n , which determines the range 2^n , as described in Sect. 4.2. In each graph, multiple values for different number of nodes are shown for comparison.

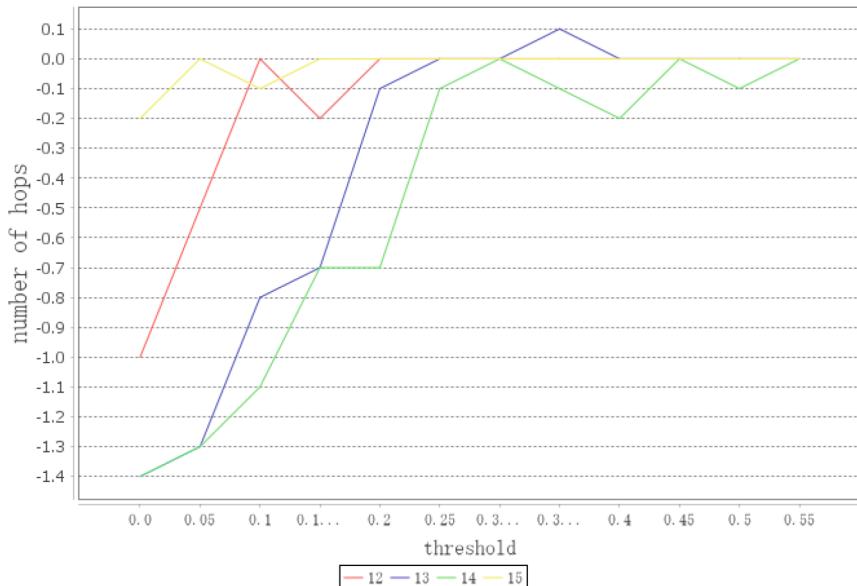


Fig. 6. Worst number of hops when the number of nodes is 500

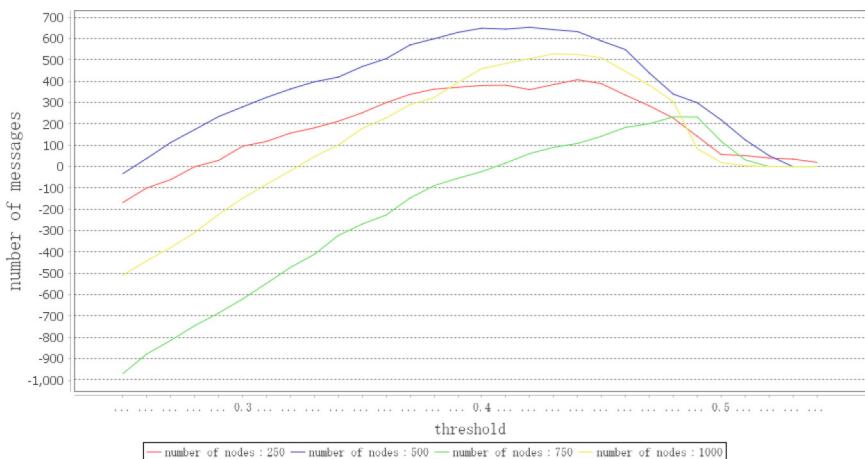
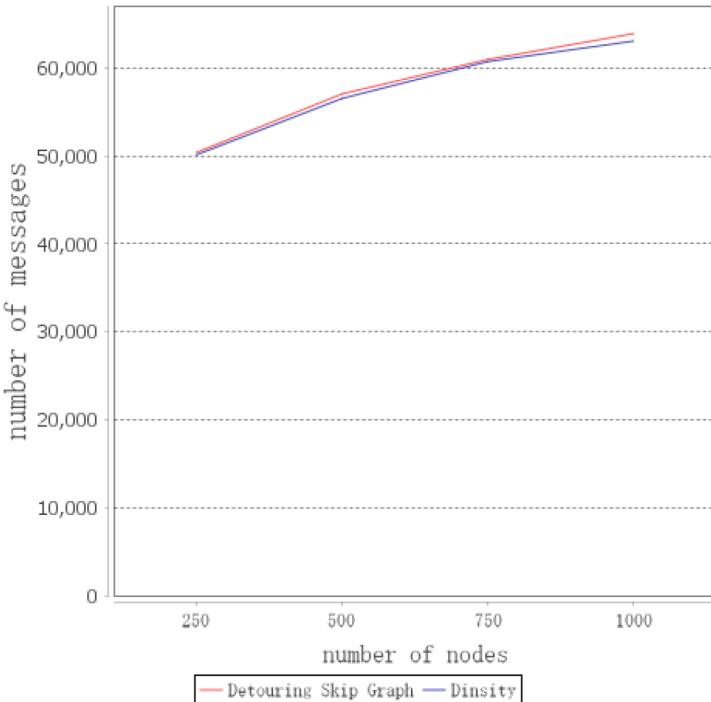


Fig. 7. Number of messages by the threshold value

**Fig. 8.** Number of hops by the threshold value

For each evaluation index and number of nodes, the values of the threshold T with the best results obtained from Figs. 7 and 8 are shown in Table 2. Throughout, good results are obtained around 0.45. Therefore, we assume that

**Fig. 9.** Number of messages by the number of nodes

0.4475, the average of all values in the table, is the optimal value for the final evaluation.

4.4 Final Evaluation

The final evaluation is performed using the values obtained in Sects. 4.2 and 4.3. From Figs. 9 and 10, the number of messages could be reduced by several percent. While some conditions showed slightly better results for the number of hops, some conditions were seen to deteriorate. When the distribution was unbiased, the results were comparable to Detouring Skip Graph.

The overall result is an algorithm that achieves the same level of routing as the Detouring Skip Graph in environments where the distribution of nodes is not so skewed, and may slightly outperform the Detouring Skip Graph in environments where the distribution of nodes is skewed. We were able to realize an algorithm that may yield slightly better results than Detouring Skip Graph in environments where the distribution is biased.

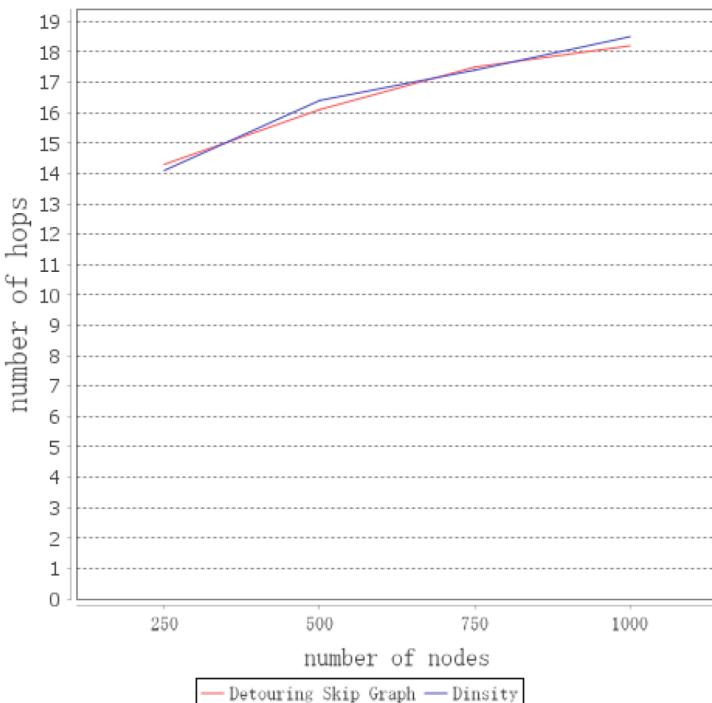


Fig. 10. Number of hops by the number of nodes

5 Conclusion

In this paper, we showed the challenges of the Detouring Skip Graph algorithm proposed in the past and study how to improve the efficiency of routing based on these challenges. The proposed method produced results with the same number of worst hops as Skip Graph in an environment with a skewed distribution of nodes. The number of messages was improved by several percent. In addition, in an environment where the nodes follow a uniform distribution, the routing is similar to that of Detouring Skip Graph, and the efficiency does not deteriorate.

Future issues include further study on how to select detour routes. In addition, methods for collecting and predicting node distribution information have not yet been determined in this paper. We would like to study specific methods for this purpose.

Acknowledgment. This work was partially supported by JSPS KAKENHI Grant Number JP22K12009, the Okawa Foundation for Information and Telecommunications, and Research Grants from the University of Fukui (FY 2024).

References

1. Aspnes, J., Shah, G.: Skip graphs. *ACM Trans. Algorithms* **3**(4), 37 (2007)
2. Banno, R., Shudo, K.: An efficient routing method for range queries in skip graph. *IEICE Trans. Inf. Syst.* **E103.D**(3), 516–525 (2020)
3. Fujita, Y., Fujimoto, A., Tode, H.: A study on hop count reduction of skip graph with arbitrary number of layers. *IEICE Techn. Rep.* **121**(433), 218–223 (2022). (in Japanese)
4. Ishi, Y., Teranishi, Y., Yoshida, M., Takeuchi, S., Shimojo, S., Nishio, S.: Range-key extension of the Skip Graph. In: Proceedings of the 2010 IEEE Global Telecommunications Conference (GLOBECOM 2010), pp. 1–6 (2010)
5. Kaneko, T., Banno, R., Shudo, K., Abe, K., Teranishi, Y.: Detouring skip graph: efficient routing via detour routes on skip graph topology. *IEEE Open J. Commun. Soc.* **1**, 1658–1673 (2020)
6. Kimata, T., Teranishi, Y., Hosokawa, T., Harai, H., Shimojo, S.: An electricity saving load balancing method for cloud storage and processing platforms. *IPSJ J.* **61**(2), 339–350 (2020). (in Japanese)
7. Pugh, W.: Skip lists: a probabilistic alternative to balanced trees. *Commun. ACM* **33**(6), 668–676 (1990)
8. Schutt, T., Schintke, F., Reinefeld, A.: Range queries on structured overlay networks. *Comput. Commun.* **31**(2), 280–291 (2008)
9. Shao, X., Jibiki, M., Teranishi, Y., Nishinaga, N.: An efficient load-balancing mechanism for heterogeneous range-queriable cloud storage. *Futur. Gener. Comput. Syst.* **78**, 920–930 (2018)



EDoViT-Alz: Alzheimer's Disease Identification with Vision Transformer Using Extremely Downscaled MRI Data

Diogen Babuc^(✉) and Alexandra-Emilia Fortiș

West University of Timișoara, Timișoara 300223, Romania
diogen.babuc@e-uvt.ro

Abstract. Alzheimer's disease is a neurological disorder that can be diagnosed by using PET or MRI scans. Reduction of images' resolution can be suitable to solve resource-limited tasks, such as Alzheimer's disease prediagnosis. This paper proposes the EDoViT-Alz model, an approach based on a vision transformer that utilizes MRI scans downsampled to an 8×8 resolution. This method significantly reduces computational demands while keeping other important features. EDoViT-Alz is evaluated using a comprehensive dataset that demonstrates high performance across various dementia stages. In addition to the patch embedding layer and positional encoding, the model has transformer encoders with multi-head self-attention mechanism on a feed-forward network to increase the efficacy of the approach.

Keywords: Alzheimer's disease · Downscaled MRI scans · Vision transformer · Pattern recognition · Machine learning

1 Introduction

With Alzheimer's disease (AD) and other forms of dementia included on a global top ten that ranks diseases causing death [4], there is huge interest in early diagnose, monitoring and support for all those affected. Various inter- and multidisciplinary approaches are carried out, amplified by the effects of digitalization in the health area for offering rapid responses to different issues generated by the growing number of patients and the cost pressure over medical healthcare systems all over the world. Among those approaches, promising results seem to be generated through the investigation on large sets of medical images via Vision Transformer models empowered with Machine Learning features immersed into Magnetic Resonance Imaging (MRI) [11].

In the monitoring of AD, magnetic resonance plays an important role [3]. MRI scans offer details about the brain anatomy. This allows medical personnel to observe changes that appear from the onset of Alzheimer. However, the use of MRI scans in the clinical sphere is often limited by the computational demands associated with processing high-resolution images [17]. Extreme downscaling of

MRI images presents a potential solution to these challenges. We decrease the data size by drastically reducing the resolution of MRI images. This also leads to reductions in storage [6]. However, this approach raises a critical question. Can essential diagnostic information be retained in such extremely downsampled images?

In recent years, deep learning demonstrated a potential in medical analysis. Convolutional Neural Network (CNN) models are used for tasks such as image classification and segmentation due to their ability to capture local spatial hierarchies in images. Despite their success, CNN models face challenges when dealing with extremely low-resolution images, as their reliance on local feature extraction becomes less effective with reduced detail. Vision Transformer (ViT) represents a modern approach to image analysis. ViT increases self-attention mechanisms to capture global dependencies and contextual information across an image [15]. Unlike CNN, ViT does not depend on local convolutions. This makes them more suitable for analyzing low-resolution images with fine details.

This paper introduces a proposed approach to vision transformers for the detection of Alzheimer's disease using extremely downsampled MRI images, a method we refer to as EDoViT-Alz (**E**xremely **D**ownscaled **V**ision **T**ransformer for **A**lzheimer's **D**etection). By reducing MRI images to a resolution 8×8 , we aim to significantly lower computational demands while still maintaining the critical features necessary for a performant prediagnosis.

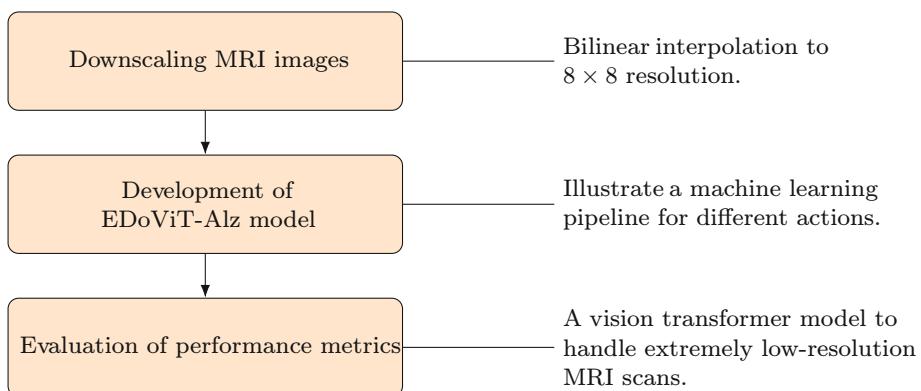


Fig. 1. Flowchart for the primary objectives and contributions of the study.

The objectives of this paper include (see also the Fig. 1):

1. **Implementing an effective downscaling method.** Use of bilinear interpolation to downscale MRI images to an 8×8 resolution.
2. **Developing the EDoViT-Alz vision transformer model** to prediagnose Alzheimer's disease on downsampled MRI scans.
3. **Evaluating the performance of the model.**

2 Background Information and Related Works

2.1 Magnetic Resonance Imaging

Magnetic resonance imaging is a diagnostic tool that uses magnetic fields and radio waves to generate detailed images of the organs and tissues [7]. In the sphere of neurology, MRI is invaluable due to its ability to produce high-resolution images of brain structures without exposure to ionizing radiation. This capability makes MRI essential for diagnosing various neurological disorders, including Alzheimer's disease. It allows clinicians to monitor changes in brain volume, to detect neurodegenerative markers that are critical in the early diagnosis and progression assessment [2].

2.2 Machine Learning in Medical Imaging

The integration of machine learning in medical imaging has transformed diagnostics. It allows more precise and automated analysis of medical images. Machine learning techniques such as image segmentation, anomaly detection, and pattern recognition are commonly applied here [9]. These technologies are capable to process spatial hierarchies in complex image data. This offers enhanced disease screening. Machine learning models are trained to identify subtle patterns that may not be evident to the human eye. However, the challenge lies in managing the high volumes of data produced by high-resolution imaging technologies, which require a lot of computational resources for processing and analysis.

2.3 Vision Transformers

Vision transformers represent a change from traditional convolution-based methods to models that use self-attention mechanisms to process images [8]. In computer vision, ViT models segment an image into multiple patches and then analyze these patches collectively in a global context. This allows the model to prioritize information from different parts of an image based on its relevance. This approach can be advantageous when we discuss the relationships between different anatomical structures. ViT models have shown promising results in various image recognition tasks [18]. They offered improvements in efficiency and accuracy over CNN when images are downscaled or when dealing with large datasets.

2.4 Related Works

The paper [18] introduces an approach for prediagnosing AD using ViT applied to a multi-modal Positron Emission Tomography (PET) images. The methodology involves converting 3D PET images into 2D images via a CNN to reduce computational costs. These 2D images are then processed by a ViT model, which uses self-attention mechanisms to capture global and local features.

The advantages of this approach include improved performance metrics compared to traditional CNN models. They also include efficient processing due to 3D-to-2D conversion and enhanced feature extraction through the ViT model. Specifically, the ADViT model achieved an accuracy of 0.9134 and an AUC of 0.9522. It outperformed other models (Table 1). Additionally, integrating PET-AV45 and PET-FDG images provides an analysis of brain abnormalities associated with AD.

However, the method has some disadvantages. Converting 3D images to 2D can result in the loss of critical spatial information. Also, the complexity of the ViT model can lead to overfitting, especially with small medical datasets. The fusion of 3D to 2D images also lacks clear medical interpretation.

Table 1. Performance metrics of different models on PET-AV45 and PET-FDG images.

Model	Accuracy	AUC	F1-Score	Precision	Sensitivity
3DCNN-AV45	0.79	0.8032	0.7531	0.7771	0.7305
3DCNN-FDG	0.8337	0.8537	0.75	0.7843	0.7186
3DCNN-Joint (multi-modality)	0.895	0.9066	0.8773	0.8994	0.8563
ViT-AV45	0.8635	0.8887	0.8395	0.8662	0.8144
ViT-FDG	0.8661	0.8863	0.8401	0.8816	0.8024
ADViT (multi-modality)	0.9134	0.9522	0.8972	0.9351	0.8623

The authors of the paper [13] present a novel method to increase the early detection of AD by combining PET and MRI imaging data. This approach involves several key steps. These are preprocessing the images to reduce noise, discrete wavelet transformation (DWT) to decompose the images into frequency bands, and extracting features from these bands using a pre-trained VGG-16 model. The features from images are then fused using inverse DWT to create compounded images. They integrate structural information from MRI and functional information from PET. Finally, these fused images are classified using a fine-tuned ViT. The advantages of this approach include improved accuracy by capturing comprehensive structural and functional information, efficient feature extraction through pre-trained models such VGG-16 and ViT, robustness with limited data due to ViT's efficiency, and high performance on PET data, highlighting functional abnormalities linked to AD.

However, the approach also has disadvantages such as computational complexity, high dependency on the quality and preprocessing of input data, and potential overfitting, particularly with small training datasets. The model achieved 81.2% accuracy using MRI data and 93.75% using PET data on both early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI), with validation accuracies of 98.50% for EMCI and 99.58% for late mild LMCI (see Table 2). Compared to existing models, the ViT-based approach

outperformed them, especially with PET data. In conclusion, this multimodal fusion approach with ViT shows significant improvements in early AD detection, demonstrating robustness and high accuracy, particularly with functional PET data, though its complexity and data dependency suggest areas for further optimization.

Table 2. Results of the proposed ViT-based model.

Metric	MRI Data	PET Data
Accuracy (EMCI)	0.8125	0.9375
Accuracy (LMCI)	0.8125	0.9375
Validation Accuracy (EMCI)		0.985
Validation Accuracy (LMCI)		0.9958

The study [16] evaluates the efficacy of ViT models in comparison to CNN models, specifically the VGG-19 model, for classifying AD through positron emission imaging (see Table 3). ViT models have shown promise in capturing direct relationships between different regions of images. This could be particularly beneficial for analyzing the intricate network of the brain. The research involved both binary classification (distinguishing between normal controls and those with AD or mild cognitive impairment (MCI)) and ternary classification (distinguishing among healthy controls, MCI patients, and AD patients).

Table 3. Ternary classification performance metrics for ViT and VGG-19 models with original and augmented datasets.

Model	Accuracy	Sensitivity	Precision	F1 Score
ViT (original data)	0.5667	0.5667	0.5278	0.5455
ViT (augmented data)	0.5333	0.5333	0.5056	0.5174
VGG-19 (original data)	0.6667	0.6667	0.6794	0.6660
VGG-19 (augmented data)	0.4667	0.4667	0.3286	0.3673

The dataset comprised 383 subjects, including 220 AD patients, 113 MCI patients, and 50 healthy controls. The images were preprocessed by converting 3D PET scans into 2D slices and applying data augmentation techniques such as image rotations to address data scarcity. Both the ViT and VGG-19 models were pretrained on the ImageNet dataset and subsequently fine-tuned using the PET image data. The key hyperparameters used in the study included a batch size of 16, 100 epochs, and an input size of 224×224 pixels.

3 Proposed Approach

The *EDoViT-Alz* (Extremely Downscaled Vision Transformer for Alzheimer’s Disease Detection) is a specialized application of vision transformers designed to prediagnose Alzheimer’s Disease using MRI images, downscaled ones. This approach uses the Vision Transformer architecture, which is adapted to analyze images on a reduced scale while aiming to retain the essential diagnostic features.

3.1 Layers’ Composition

The EDoViT-Alz receives an input of extremely downscaled MRI images, typically resized to a dimension of 8×8 pixels (see Fig. 2). This extreme downscaling significantly reduces the computational load and memory requirements [12]. Each input image is split into fixed-size patches. In the case of EDoViT-Alz, given the already small size of the image, each pixel can effectively be treated as a patch. These patches are then linearly embedded into a higher dimensional space to provide richer features for the subsequent layers. To retain information on the position of each patch within the original image, positional encodings are added to the patch embeddings. This step is crucial because the transformer architecture itself does not have any notion of order or position [14]. The core of EDoViT-Alz is composed of multiple layers of transformers. Each layer contains multi-head self-attention mechanisms and feed-forward neural networks [10]. These layers enable the model to focus on different parts of the image and integrate information across the entire field of view, which is essential for identifying patterns indicative of Alzheimer’s Disease. The final transformer output is passed through a classification head, typically consisting of a layer normalization followed by a linear layer, to predict the presence of Alzheimer’s Disease and afterwards the type of detected dementia. The output from the linear layer typically goes through a softmax function in cases where multi-class classification is required. The softmax function converts the logits (raw prediction values from the linear layer) into probabilities by dividing the exponential of each output logit by the sum of the exponentials of all logits [1]. This output is a probability distribution for the classes.

3.2 Properties

Using a transformer architecture, EDoViT-Alz becomes scalable. It can handle varying image sizes and resolutions effectively, although it is optimized for low-resolution images. Vision transformers are flexible in processing different parts of the image, thanks to the self-attention mechanism. This mechanism adapts to focus on the most relevant features for diagnosis. The model is designed to be computationally efficient, particularly in environments with limited resources, due to the extreme downscaling of input images.

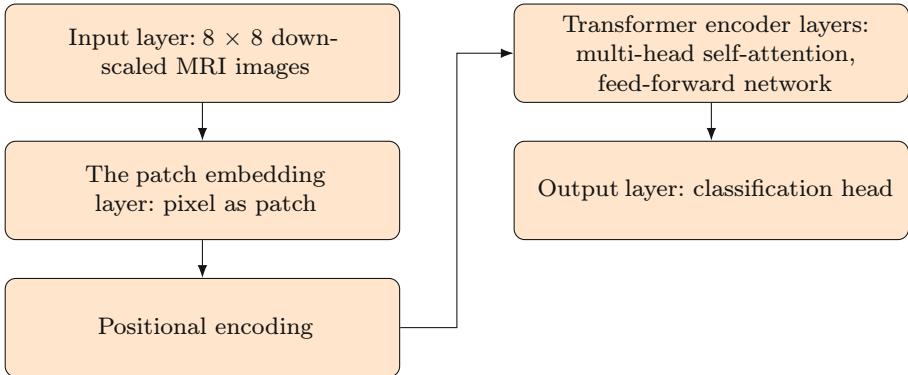


Fig. 2. Architecture of the EDoViT-Alz model considering the main actions.

3.3 Advantages and Disadvantages

The extreme downscaling of MRI images reduces the computational burden and data storage needs. This makes the system accessible and faster. Despite the low resolution, vision transformers can extract meaningful features from the global context of the image. In this way, they capture critical diagnostic information that traditional methods might miss.

Extreme downscaling can lead to a loss of crucial diagnostic details, which may affect the accuracy and reliability of the diagnosis. The effectiveness of the model could depend on the initial quality of the magnetic resonance scans and the consistency of the downscaling process.

3.4 Applicability in Alzheimer's Disease Detection

The EDoViT-Alz is particularly applicable in scenarios where large volumes of MRI data are required efficiently and cost-effectively, such as in regions with limited medical imaging resources. Its ability to work with low-resolution images opens up possibilities for using older or less powerful imaging systems without compromising the ability to perform preliminary diagnostics (Fig. 3). In addition, the adaptability of the model makes it suitable for screening applications, which can help in early detection.

4 Results and Discussions

The process begins with data collection and preparation (see Fig. 4). MRI scans relevant to Alzheimer's disease are collected and subsequently downsampled to an 8×8 resolution to reduce computational load. This step is crucial as it allows for quicker processing while aiming to retain essential diagnostic features. In addition to downscaling, the images are normalized and may undergo further preprocessing steps, such as noise reduction, to enhance image quality for better

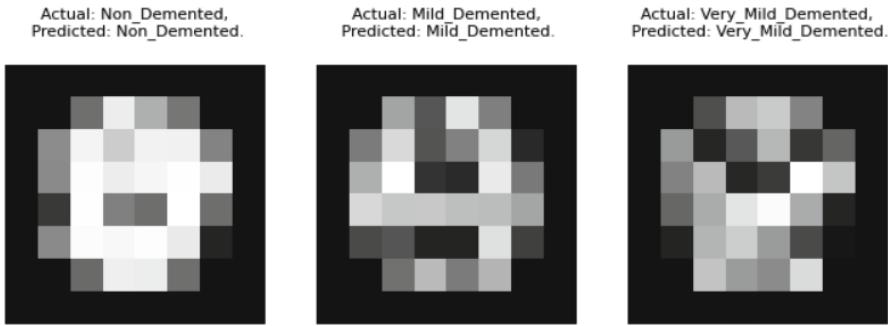


Fig. 3. Classification of dementia stages using extremely downscaled MRI images.

model training. Following data preparation, the model training phase involves several key steps. The dataset [5] with a total of 6400 MRI scans is divided into training, validation, and testing subsets (with a total of 5120 training images, 640 validation and 640 test samples) to support effective learning and evaluation. The EDoViT-Alz architecture is then configured. It includes the setup of patch embedding, positional encoding, and transformer encoder layers as previously described. Training is conducted on the dataset using appropriate loss functions and optimizers, with the validation set used to tune hyperparameters and prevent overfitting. The training process employed the cross-entropy loss as the primary loss function due to its effectiveness in handling multi-class classification problems. The Adam optimizer was selected for this study due to its ability to adapt the learning rate for each parameter individually. This accelerates convergence and improves the stability of the training process. The key hyperparameters, such as the learning rate, batch size, and the number of epochs, were tuned during the validation phase. A grid search was conducted to explore different combinations. The final values were selected based on the ability to minimize the validation loss while avoiding overfitting.

Model evaluation is the next critical phase. The model's performance is assessed on an unseen test set, which provides insights into its effectiveness in diagnosing Alzheimer's from extremely downscaled MRI images (Fig. 3).

This holistic approach ensures that the EDoViT-Alz system is scientifically and practically responsible in its application.

The model has a precision of 0.99 for the *Mild Demented* class, meaning that 99% of the instances it predicted as *Mild Demented* were correct (see Table 4). For the *Mild Demented* class, the sensitivity of 0.95 implies that the model correctly identified 95% of all actual *Mild Demented* cases. The *F1-score* is the harmonic mean of precision and sensitivity, providing a single metric to evaluate the balance between them. For instance, the *Mild Demented* class has an *F1-score* of 0.97, indicating a strong balance between precision and sensitivity. The *Support* column shows the number of actual occurrences of each class in the dataset used for evaluating the model. The *Mild Demented* class, for example, appears 78 times. The *Macro Avg* and *Weighted Avg* rows provide averages of

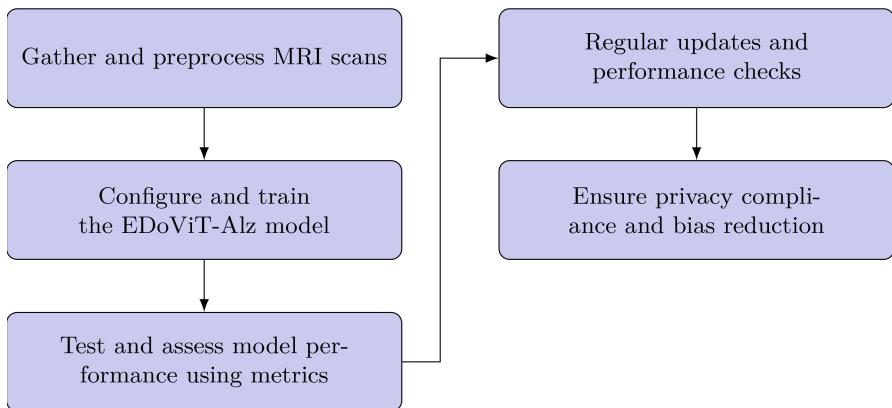


Fig. 4. Flowchart of the EDoViT-Alz machine learning pipeline.

Table 4. Classification results for Alzheimer's disease detection model for test data.

Class	Precision	Sensitivity	F1-Score	Support
Mild Demented	0.99	0.95	0.97	78
Moderate Demented	1.00	0.80	0.89	10
Non Demented	0.98	0.96	0.97	327
Very Mild Demented	0.94	0.99	0.96	225
Macro Avg	0.98	0.92	0.95	640
Weighted Avg	0.97	0.97	0.97	640

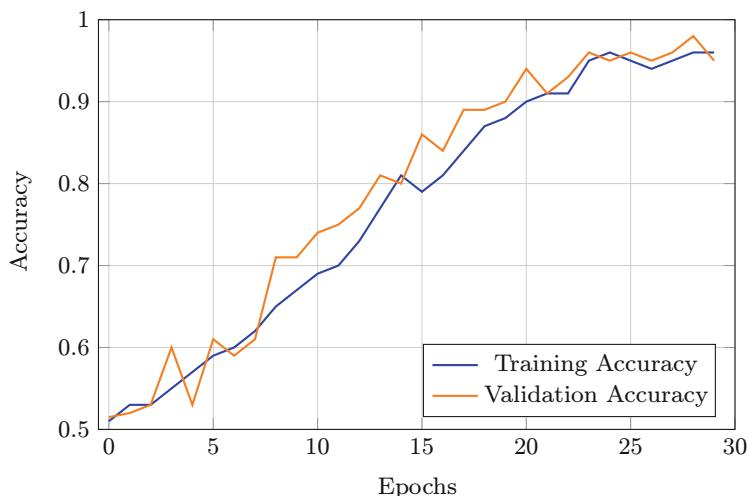


Fig. 5. Training and validation accuracies of the EDoViT-Alz model for each epoch.

precision, sensitivity, and F1-score across all classes. The Macro Avg calculates these metrics uniformly across classes, while the Weighted Avg accounts for the support of each class, providing insight into the model's effectiveness on a class-weighted basis.

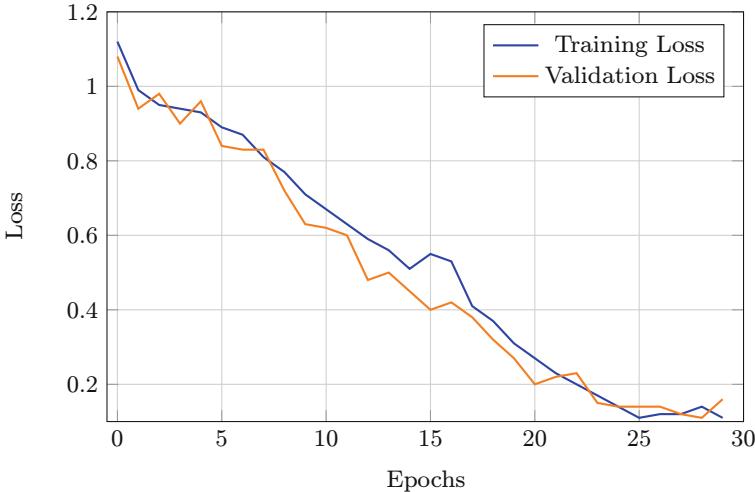


Fig. 6. Training and validation losses of the EDoViT-Alz model for each epoch.

The Figs. 5 and 6 present two graphs that show the training and validation dynamics (accuracies and loss values) of the EDoViT-Alz model over 30 epochs. The *training accuracy vs. validation accuracy* graph shows both accuracy metrics improving over time, with training accuracy slightly outperforming validation accuracy by the end of the training process. This suggests that the model is learning effectively, although the slight divergence might indicate minor overfitting as the model becomes too tailored to the training data. On the other hand, the *training loss vs. validation loss* graph depicts both loss metrics decreasing over the same period, converging towards the end of the training. The convergence of training and validation loss indicates that the model is generalizing well to new data, which is a good sign of its predictive reliability and robustness.

5 Conclusion

The development and evaluation of the EDoViT-Alz model represent significant strides in the application of advanced machine learning techniques, particularly Vision Transformers, for the detection of Alzheimer's Disease using extremely downsampled MRI images. Our experiments have demonstrated that the EDoViT-Alz model not only achieves high accuracy in diagnosing different stages of dementia but also maintains a balance between precision and sensitivity,

as evidenced by the performance metrics across various dementia categories. The model's ability to effectively learn and generalize from downsampled images was illustrated by the convergence of training and validation losses, alongside consistent improvements in accuracy over 30 training epochs. These results highlight the model's robustness and its potential utility in clinical settings, where rapid and reliable diagnostics are crucial.

Future work will focus on refining the model to handle larger and diverse datasets, exploring the impact of further optimizations on model performance. We will compare the results from our model with other existing models to see the reliability in terms of providing output. Another future direction is to develop a decision-support pipeline for deep learning architecture. This will provide a comprehensive flow for the established processes in Alzheimer's disease tasks. Ethical considerations and privacy compliance will remain at the forefront of our efforts. We seek to deploy this technology in a manner that is both responsible and beneficial to patients.

Acknowledgments. The authors thank the West University of Timișoara for the resources provided and Florin Fortiș and Gabrijel Babuc for their suggestions.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Carcagnì, P., Leo, M., Del Coco, M., Distante, C., De Salve, A.: Convolution neural networks and self-attention learners for Alzheimer dementia diagnosis from brain MRI. *Sensors* **23**(3), 1694 (2023). <https://doi.org/10.3390/s23031694>
2. Fleisher, A.S., Houston, W.S., Eyler, L.T., Frye, S., Jenkins, C., Thal, L.J., Bondi, M.W.: Identification of Alzheimer disease risk by functional magnetic resonance imaging. *Arch. Neurol.* **62**(12), 1881–1888 (2005). <https://doi.org/10.1001/archneur.62.12.1881>
3. Frisoni, G.B., Fox, N.C., Jack, C.R., Jr., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* **6**(2), 67–77 (2010). <https://doi.org/10.1038/nrneurol.2009.215>
4. Gustavsson, A., et al.: Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's Dementia* **19**(2), 658–670 (2022). <https://doi.org/10.1002/alz.12694>
5. Kumar, S., Shastri, S.: Alzheimer MRI preprocessed dataset (2023). <https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset>. Accessed 24 July 2024
6. Leng, Y., Cui, W., Peng, Y., Yan, C., Cao, Y., Yan, Z., Chen, S., Jiang, X., Zheng, J., Initiative, A.D.N., et al.: Multimodal cross enhanced fusion network for diagnosis of Alzheimer's disease and subjective memory complaints. *Comput. Biol. Med.* **157**, 106788 (2023). <https://doi.org/10.1016/j.compbiomed.2023.106788>
7. Liang, Z.P., Lauterbur, P.C.: Principles of magnetic resonance imaging. SPIE Opt. Eng. Press Bellingham (2000). <https://doi.org/10.1109/9780470545652>

8. Lyu, Y., Yu, X., Zhu, D., Zhang, L.: Classification of alzheimer's disease via vision transformer: Classification of Alzheimer's disease via vision transformer. In: Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 463–468 (2022). <https://doi.org/10.1145/3529190.3534754>
9. Mirzaei, G., Adeli, H.: Machine learning techniques for diagnosis of Alzheimer disease, mild cognitive disorder, and other types of dementia. *Biomed. Signal Process. Control* **72**, 103293 (2022). <https://doi.org/10.1016/j.bspc.2021.103293>
10. Mu, X., Zhang, J., Zhang, K., Jin, H., Zhou, X., Xie, Z., Toe, T.T.: Alzheimer classification based on convolutional neural network and vision transformer. In: 2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), pp. 329–334. IEEE (2023). <https://doi.org/10.1109/icicml60161.2023.10424819>
11. Nguyen, H.D.: Deep learning for the detection of neurological diseases. Theses, Université de Bordeaux (2023). <https://theses.hal.science/tel-04311995>
12. Nguyen, H.D., Clément, M., Planche, V., Mansencal, B., Coupé, P.: Deep grading for MRI-based differential diagnosis of Alzheimer's disease and frontotemporal dementia. *Artif. Intell. Med.* **144**, 102636 (2023). <https://doi.org/10.1016/j.artmed.2023.102636>
13. Odusami, M., Maskeliūnas, R., Damaševičius, R.: Pixel-level fusion approach with vision transformer for early detection of Alzheimer's disease. *Electronics* **12**(5), 1218 (2023). <https://doi.org/10.3390/electronics12051218>
14. Pranav, G., Varsha, K., Gayathri, K.: Early alzheimer detection through speech analysis and vision transformer approach. In: International Conference on Speech and Language Technologies for Low-resource Languages, pp. 265–276. Springer (2022). https://doi.org/10.1007/978-3-031-33231-9_19
15. Sarraf, S., Sarraf, A., DeSouza, D.D., Anderson, J.A., Kabia, M., Initiative, A.D.N.: Ovitad: optimized vision transformer to predict various stages of Alzheimer's disease using resting-state fMRI and structural MRI data. *Brain Sci.* **13**(2), 260 (2023). <https://doi.org/10.3390/brainsci13020260>
16. Shin, H., Jeon, S., Seol, Y., Kim, S., Kang, D.: Vision transformer approach for classification of Alzheimer's disease using 18f-florbetaben brain images. *Appl. Sci.* **13**(6), 3453 (2023). <https://doi.org/10.3390/app13063453>
17. Small, S.A., Nava, A.S., Perera, G.M., Delapaz, R., Stern, Y.: Evaluating the function of hippocampal subregions with high-resolution MRI in Alzheimer's disease and aging. *Microsc. Res. Tech.* **51**(1), 101–108 (2000). [https://doi.org/10.1002/1097-0029\(20001001\)51:1<101::aid-jemt11>3.0.co;2-h](https://doi.org/10.1002/1097-0029(20001001)51:1<101::aid-jemt11>3.0.co;2-h)
18. Xing, X., Liang, G., Zhang, Y., Khanal, S., Lin, A.L., Jacobs, N.: Advit: Vision transformer on multi-modality pet images for Alzheimer disease diagnosis. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–4. IEEE (2022). <https://doi.org/10.1109/isbi52829.2022.9761584>



A Comparison Study Between Cuckoo Search and Particle Swarm Optimization Based Intelligent Systems for Optimization of Mesh Routers in a Small-Scale WMN

Shinji Sakamoto^{1(✉)}, Shigenari Nakamura², Leonard Barolli³,
and Makoto Takizawa⁴

¹ Department of Information and Computer Science, Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi, Ishikawa 921-8501, Japan
shinji.sakamoto@ieee.org

² Department of Information System Engineering, Tokyo Denki University, 5 Senju Asahi-cho, Adachi-ku, Tokyo 120-8551, Japan
s.nakamura@mail.dendai.ac.jp

³ Department of Information and Communication Engineering, Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
barolli@fit.ac.jp

⁴ Department of Advanced Sciences, Faculty of Science and Engineering, Hosei University, Kajino-Machi, Koganei-Shi, Tokyo 184-8584, Japan
makoto.takizawa@computer.org

Abstract. Wireless Mesh Networks (WMNs) are efficient networks due to their high robustness and rapid deployment capabilities. But they face some challenges such as network congestion, interference, reduced data transfer rates, packet losses, and increased latency. The optimization of mesh router placement is a good approach to solving these problems. However, finding the best location of mesh routers in the considered area is classified as an NP-hard problem. To deal with this problem, we propose and develop two intelligent simulation systems based on Cuckoo Search (CS) and Particle Swarm Optimization (PSO), called WMN-CS and WMN-PSO, respectively. In this study, we compare the performance of WMN-CS and WMN-PSO systems for a small-scale WMN. The simulation results show that WMN-CS performs better and converges faster than WMN-PSO in the considered scenario.

1 Introduction

Wireless Mesh Networks (WMNs) are recognized for their ability to provide wireless Internet access. They are dynamically organized and configured automatically, and can establish and maintain connectivity of mesh topology. This capability brings many advantages [2]. However, they have several problems that

need to be solved, such as network congestion, interference, reduced data transmission speeds, packet loss, and increased latency [13, 17]. The optimization of mesh routers placement is very good approach for solving these problems. However, finding the exact optimal location of mesh routers in WMNs is difficult and classified as NP-hard [3].

Finding the exact optimal location of mesh routers is not an efficient approach [7, 12]. Therefore, some researchers are focusing on getting near optimal solutions in real time by using meta-heuristic approaches. In recent years, several studies have proposed systems based on intelligent and meta-heuristic algorithms [5, 6, 8, 16, 18, 20, 22].

In WMN research field, researchers are concentrated on resource allocation and routing techniques [1, 23]. A lot of studies try to improve WMN performance by using optimization strategies such as Mixed-Integer Linear Programming (MILP) and meta-heuristic approaches [5, 6, 10, 11, 14–16, 18, 20, 22]. Also, techniques and approaches using Hill Climbing (HC), Simulated Annealing (SA), and Genetic Algorithms (GA) have been proposed and implemented. It has been seen that Meta-heuristic techniques are often preferred over MILP methods for solving NP-hard problems due to their efficiency in real-time computation [25].

In our previous work, we developed intelligent systems based on the Cuckoo Search (CS) algorithm and Particle Swarm Optimization (PSO), referred to as WMN-CS and WMN-PSO. These systems try to optimize the locations of mesh routers using meta-heuristic approaches. In this paper, we compare the performance of WMN-CS and WMN-PSO systems for a small-scale WMN.

The rest of the paper is organized as follows. In Sect. 2, we explain the node placement challenges in WMNs. Section 3 introduces the CS algorithm and PSO. The intelligent simulation systems are presented in Sect. 4. In Sect. 5, we discuss the simulation results. Finally, Sect. 6 concludes the paper.

2 Node Placement Problem in WMNs

Some issues of WMNs related to wireless communication must be solved in order to use the full potential of WMNs. For instance, the signal attenuation, caused by distance or physical obstacles like walls, directly affects the communication between nodes and brings to the formation of dead zones [23]. Also, the number of available channels often limits network scalability and signal interference may decrease the performance, particularly in high-population areas where numerous mesh nodes compete for connectivity. Finding the optimal location of mesh routers in WMNs can solve the above problems. However, it is known to be an NP-hard problem [3]. We call this issue the node placement problem in WMNs.

For node placement problem in WMNs, we consider a two-dimensional continuous space designated for deploying both mesh client and router nodes. The objective is to find the optimal location of mesh routers to enhance network connectivity and client coverage. The network connectivity is quantified through a graph metric known as the Size of Giant Component (SGC). The client coverage is evaluated through the Number of Covered Mesh Clients (NCMC), which

counts the number of clients within the communication range of at least one router. The instance of the problem can be formulated as follows.

- N mesh router nodes, each identified by an ID and defined by its radio coverage, are considered as a vector of routers.
- There is a considered area with two dimensions $W \times H$ for the placement of N mesh routers and M mesh clients.
- M client mesh nodes, which are randomly distributed within the considered area, are depicted as a matrix of clients.
- The optimal positions of mesh routers need to be decided.

It is essential to note that both network connectivity and mesh client coverage are important because they directly influence the performance of WMNs.

The network is described as an adjacency matrix, with each mesh node being seen as a vector $v = \langle x, y, r \rangle$ allocated within the considered area. When node v is within the communication radius of node u , an edge is established in the WMN graph, showing that they are connected.

The effectiveness of WMNs is influenced by both environmental factors and the locations of mesh nodes. So, their performance is evaluated through computer simulations that explore a range of theoretical scenarios. Heuristic methods are particularly good approach for their ability to rapidly generate feasible solutions to the node placement problem in WMNs [28].

3 Intelligent Algorithms

3.1 CS Algorithm

The CS algorithm is inspired by the brood parasitism behaviour observed in some cuckoo species. The basic concepts of the CS algorithm are based on three key assumptions [29]:

1. Each cuckoo lays a single egg at a time.
2. Higher-quality eggs have a greater likelihood of advancing to future generations.
3. There is a certain probability p_a that a host bird will recognize a cuckoo egg. If the host bird is recognized, the host bird may either reject the egg or completely abandon the nest.

The CS algorithm has three hyperparameters: the number of nests, the host bird recognition rate and the scale parameter (σ) used in the Lévy flight distribution.

The pseudocode for the CS algorithm is presented in Algorithm 1. After generating the initial solution, the algorithm checks the fitness of each nest and identifies the one with the highest fitness value. Then, it compares the fitness value of the new solution with current solution. In the case when the fitnees value of the new solution is higher then the current solution, the new solution replaces

Algorithm 1. Pseudo code of CS algorithm.

```

1: Initialize parameters:
2:   Computation step  $t = 0$  and maximum steps  $T_{max}$ 
3:   Number of nests  $n$  (ensure  $n > 0$ )
4:   Scale parameter for Lévy flight  $\gamma$  (ensure  $\gamma > 0$ )
5:   Host bird recognition rate  $p_a$  (where  $0 < p_a < 1$ )
6: Define the fitness function to calculate fitness value as  $f$ 
7: Generate initial set of solutions  $S_0$ 
8: while  $t < T_{max}$  do
9:   while  $i < n$  do
10:     $j := i \% \text{len}(S_t)$  //  $j$  represents the modulo of  $i$  and the current number of
11:    Generate a new solution  $S_{t+1}^i$  from  $S_t^j$  using Lévy flights
12:    if  $(f(S_{t+1}^i) < f(S_t^j))$  and  $(\text{rand()} < p_a)$  then
13:      Remove solution  $S_{t+1}^i$ 
14:    end if
15:     $i = i + 1$ 
16:  end while
17:   $t = t + 1$ 
18: end while
19: return Best solution

```

the current solution. Otherwise, the new solution is discarded. This iterative process continues until the termination criterion is satisfied.

The CS algorithm utilizes the Lévy flights, which is a type of random walk characterized by a long-tailed probability distribution not limited by the size or complexity of solution spaces. Its adaptability allows it to efficiently explore a wide range of solution spaces, enhancing the likelihood of finding optimal solutions in diverse optimization scenarios.

Lévy flights generally have short movements, but occasionally, they result in long jumps due to the long-tailed characteristics of the Lévy distribution. The Probability Density Function (PDF) for the Lévy distribution is defined as:

$$P(x; \mu, \gamma) = \sqrt{\frac{\gamma}{2\pi}} \frac{e^{-\gamma/2(x-\mu)}}{(x-\mu)^{3/2}}, \quad (1)$$

where μ is the location parameter ($x \geq \mu$) and γ is the scale parameter ($\gamma > 0$).

The Cumulative Distribution Function (CDF) for the Lévy distribution is given by:

$$F(x; \mu, \gamma) = \text{erfc} \left(\sqrt{\frac{\gamma}{2(x-\mu)}} \right), \quad (2)$$

where erfc is the complementary error function.

To generate variables that follow the Lévy distribution, the inverse transformation method is used:

Algorithm 2. Pseudocode for PSO.

```

/* Initialize solutions and parameters */
Set maximum computation time:=  $T_{max}$ , and current time  $t = 0$ ;
Define the number of particle-patterns:=  $m$ ,  $2 \leq m \in \mathbf{R}^1$ ;
Set initial solutions for particle-patterns:=  $\mathbf{P}^0$ ;
Initialize the global best solution:=  $\mathbf{G}^0$ ;
Initialize particle positions:=  $\mathbf{x}_{ij}^0$ ;
Initialize particle velocities:=  $\mathbf{v}_{ij}^0$ ;
/* Determine the following PSO parameters: */
Set PSO parameters:=  $\omega$ ,  $0 < \omega \in \mathbf{R}^1$ ;
Set PSO parameters:=  $C_1$ ,  $0 < C_1 \in \mathbf{R}^1$ ;
Set PSO parameters:=  $C_2$ ,  $0 < C_2 \in \mathbf{R}^1$ ;
/* Begin the PSO process */
Compute fitness for each particle-patterns in ( $\mathbf{P}^0$ );
Compute the initial global best solution ( $\mathbf{G}^0$ )
while  $t < T_{max}$  do
    /* Update velocities and positions */
     $\mathbf{v}_{ij}^{t+1} = \omega \cdot \mathbf{v}_{ij}^t$ 
     $+ C_1 \cdot \text{rand}() \cdot (\text{best}(\mathbf{P}_{ij}^t) - \mathbf{x}_{ij}^t)$ 
     $+ C_2 \cdot \text{rand}() \cdot (\text{best}(\mathbf{G}^t) - \mathbf{x}_{ij}^t);$ 
     $\mathbf{x}_{ij}^{t+1} = \mathbf{x}_{ij}^t + \mathbf{v}_{ij}^{t+1};$ 
    Update_Best_Solutions( $\mathbf{G}^t$ ,  $\mathbf{P}^t$ );
    /* Update_Best_Solutions" compares and updates the best solutions for each particle and the global best solution if the new fitness values are better. */
    Recompute fitness for ( $\mathbf{G}^{(t+1)}$ ,  $\mathbf{P}^{(t+1)}$ );
    Increment time:  $t = t + 1$ ;
end while
Final update of best solutions: Update_Best_Solutions( $\mathbf{G}^t$ ,  $\mathbf{P}^t$ );
return The best solution found;

```

$$F^{-1}(x; \mu, \gamma) = \frac{\gamma}{2(\text{erfc}^{-1}(x))^2} + \mu, \quad (3)$$

where F^{-1} is the inverse function of the CDF.

3.2 PSO

In PSO, each individual in the particle swarm consists of three \mathcal{D} -dimensional vectors, where \mathcal{D} represents the dimensionality of the search space. These vectors are the current position \mathbf{x}_i , the previous best position \mathbf{p}_i , and the velocity \mathbf{v}_i .

The particle swarm is more than just a simple collection of particles. It interacts with other particles during the optimization progress. So, the optimization problem is based on the collective behaviour of the particles through their interactions.

Some communication structures or topologies organize populations, often conceptualized as social networks. This topology consists of bidirectional edges that connect pairs of particles, meaning that if particle j is in the neighbourhood of particle i , then i is also in j 's neighbourhood. Each particle exchanges

information with others and is influenced by the best position found by any member of its topological neighbourhood. This optimal position denoted as \mathbf{p}_g , represents the best position among the neighbours.

In PSO process, the velocity of each particle is iteratively adjusted, causing the particle to stochastically oscillate around the \mathbf{p}_i and \mathbf{p}_g positions. The pseudocode of PSO algorithm is shown in Algorithm 2.

4 Design and Implementation WMN-CS and WMN-PSO Systems

4.1 Common Design Issues

For both WMN-CS and WMN-PSO systems, the solution is evaluated by the fitness function, which considers two metrics: Size of Giant Component (SGC) and Number of Covered Mesh Clients (NCMC).

SGC:

This metric indicates the WMN connectivity. Establishing the interconnection among mesh routers is important for effective and robust data transmission in the network. A decrease in the SGC value suggests that there are some isolated mesh routers.

NCMC:

This metric shows the client coverage and effectiveness of the mesh routers in covering the mesh clients within the WMN.

We consider, the following fitness function:

$$\text{Fitness} = \alpha \times \text{SGC}(\mathbf{x}, \mathbf{y}, \mathbf{r}) + \beta \times \text{NCMC}(\mathbf{x}, \mathbf{y}, \mathbf{r}). \quad (4)$$

The coefficients α and β represent the weight assigned to SGC and NCMC metrics, respectively. In our previous research [19], we evaluated the effect of weight coefficients. So, in this study, we use the weight coefficients $\alpha = 0.7$ and $\beta = 0.3$.

Both WMN-CS and WMN-PSO systems can generate different types of mesh client distributions. In this paper, we focus on two types of distributions (Normal and Uniform distribution), as shown in Fig. 1.

The Normal distribution, shown in Fig. 1(a), is categorized as a hot-spot distribution, where mesh clients are set in the center region of the considered area. On the other hand, the Uniform distribution, shown in Fig. 1(b), is categorized as a non-hot-spot distribution, where mesh clients are spreadly distributed in the area with no particular region having a higher traffic or resource demand concentration.

4.2 Design of WMN-CS System

In WMN-CS [4], nests represent solutions, and eggs represent the selected solutions. The process begins with the random generation of an initial solution.

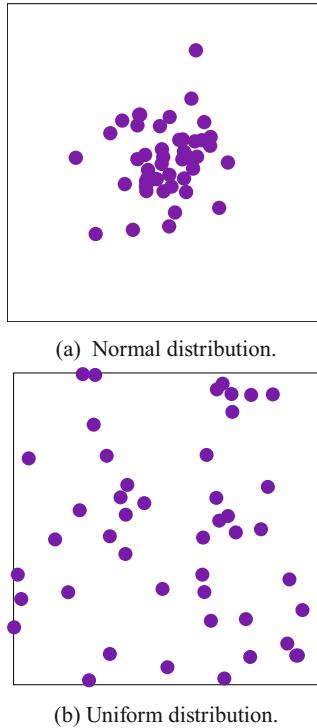


Fig. 1. Normal and Uniform distributions of mesh clients.

Then, cuckoos seek the best nest in which to lay their eggs. They follow Lévy flight to move and try to find the best nest for egg-laying. When a cuckoo finds a suitable nest, the cuckoo replaces the host bird eggs with its own. The WMN-CS evaluates these solutions using the fitness function that calculates a fitness value.

WMN-CS uses three basic hyperparameters: the number of nests, the host bird recognition rate, and the scale parameter. These three parameters were fine-tuned in our previous work [21, 22].

4.3 Design of WMN-PSO System

In WMN-PSO, a particle-pattern is a solution and a particle is a mesh router as shown in Fig. 2.

The WMN-PSO has a mechanism for moving particles. Each mesh router is defined by x and y coordinate positions and velocity. There are numerous movement methods in PSO field [9, 24, 26, 27]. We use the Linearly Decreasing Vmax Method (LDVM) in this paper. The PSO parameters in LDVM are set to an unstable region with $\omega = 0.9$ and $C_1 = C_2 = 2.0$. The V_{max} value, representing the maximum velocity of the particles, is linearly reduced [24].

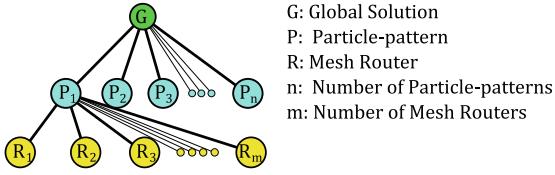


Fig. 2. Relationship among global solution, particle-patterns and mesh routers.

5 Simulation Results

We compare the performance of WMN-CS and WMN-PSO systems for a small-scale WMN (considered area of 32 by 32 units with 16 mesh routers and 48 mesh clients). For the simulations, we consider Normal and Uniform distributions of mesh clients. The communication distance of the mesh router is from 2.0 to 3.0. While the total iteration number is 2000 and the number of iteration per phase 10. We tuned the fitness function weight-coefficients by many simulations and used the values 0.7, 0.3 for α and β , respectively. While in Table 1 and Table 2 are shown the parameters for WMN-CS and WMN-PSO, respectively.

Table 1. Parameter settings for WMN-CS system.

Parameters	Values
Number of Nests	60
Host Bird Recognition Rate (p_a)	0.925
Scale Parameter (σ)	0.09

Table 2. Parameter settings for WMN-PSO system.

Parameters	Values
Number of Particle-patterns	60
Replacement Method	LDVM

Simulation results for SGC and NCMC in case of Normal distribution of mesh clients are shown in Fig. 3 and Fig. 4, respectively. The SGC and NCMC have to maximal values for both intelligent systems. However, WMN-CS system performs better than WMN-PSO system in terms of convergence speed.

We show the simulation results for SGC and NCMC in case of Uniform distribution of mesh clients in Fig. 5 and Fig. 6, respectively. In both systems, all mesh routers are connected, so the SGC is maximal. However for NCMC, the WMN-CS system has better performance than WMN-PSO system. Also, the convergence speed of WMN-CS system is faster than WMN-PSO system.

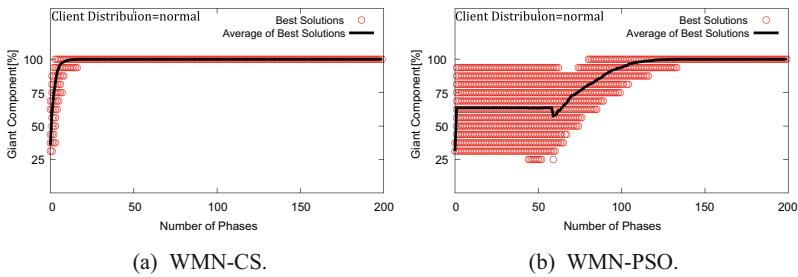


Fig. 3. Simulation results of SGC for Normal distribution of mesh clients.

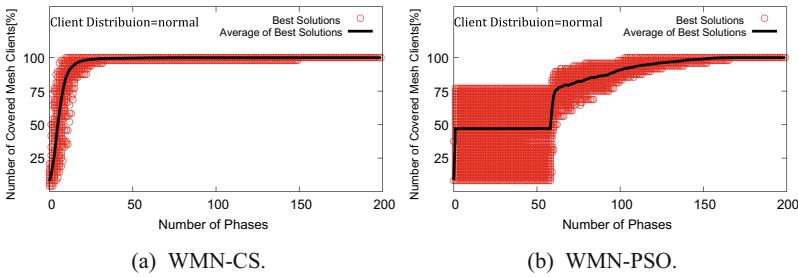


Fig. 4. Simulation results of NCMC for Normal distribution of mesh clients.

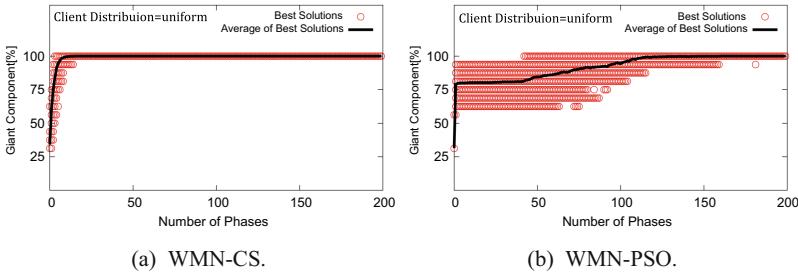


Fig. 5. Simulation results of SGC for Uniform distribution of mesh clients.

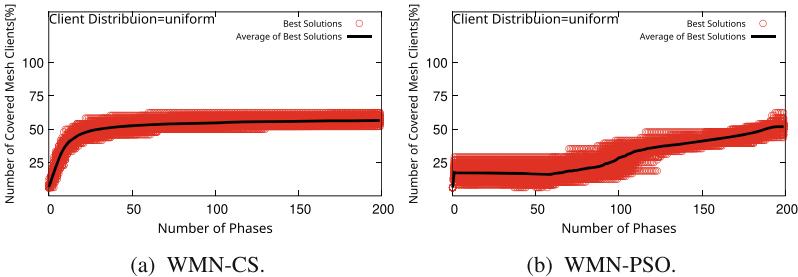


Fig. 6. Simulation results of NCMC for Uniform distribution of mesh clients.

6 Conclusions

In this paper, we compared the performance of WMN-CS and WMN-PSO simulation systems considering Normal and Uniform distributions of mesh clients for a small-scale WMN. The simulation results show that for Normal distribution the SGC and NCMC have to maximal values for both intelligent systems. However, WMN-CS system performs better than WMN-PSO system in terms of convergence speed. While in case of Uniform distribution, for both systems, all mesh routers are connected, but WMN-CS system has better performance than WMN-PSO system for NCMC. Also, the convergence speed of WMN-CS system is faster than WMN-PSO system.

In our future research, we plan to compare the performance of both simulation systems by considering various distributions of mesh clients.

References

1. Ahmed, A.M., Hashim, A.H.A.: Metaheuristic approaches for gateway placement optimization in wireless mesh networks: a survey. *Int. J. Comput. Sci. Netw. Secur. (IJCSNS)* **14**(12), 1 (2014)
2. Akyildiz, I.F., Wang, X., Wang, W.: Wireless mesh networks: a survey. *Comput. Netw.* **47**(4), 445–487 (2005)
3. Amaldi, E., Capone, A., Cesana, M., Filippini, I., Malucelli, F.: Optimization models and methods for planning wireless mesh networks. *Comput. Netw.* **52**(11), 2159–2171 (2008)
4. Asakura, K., Sakamoto, S.: A cuckoo search based simulation system for node placement problem in wireless mesh networks. In: Barolli, L. (ed.) CISIS 2023. Lecture Notes on Data Engineering and Communications Technologies, vol. 176, pp. 179–187. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-35734-3_18
5. Barolli, A., Bylykbashi, K., Qafzezi, E., Sakamoto, S., Barolli, L., Takizawa, M.: A comparison study of chi-square and uniform distributions of mesh clients for different router replacement methods using WMN-PSODGA hybrid intelligent simulation system. *J. High Speed Netw.* **27**(4), 319–334 (2021)
6. Barolli, A., Bylykbashi, K., Qafzezi, E., Sakamoto, S., Barolli, L.: Implementation of roulette wheel and random selection methods in a hybrid intelligent system: a comparison study for two islands and subway distributions considering different router replacement methods. *Appl. Soft Comput.* **131**(109), 805 (2022)
7. Basirati, M., Akbari Jokar, M.R., Hassannayebi, E.: Bi-objective optimization approaches to many-to-many hub location routing with distance balancing and hard time window. *Neural Comput. Appl.* **32**, 13267–13288 (2020)
8. Chang, X., Sakamoto, S., Oda, T., Ikeda, M., Barolli, L., Xhafa, F.: Node placement in WMNs for different movement methods: a hill climbing system considering exponential and weibull distributions. In: The 9th IEEE International Conference on Broadband and Wireless Computing, Communication and Applications, pp. 440–445. IEEE (2014)
9. Clerc, M., Kennedy, J.: The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans. Evol. Comput.* **6**(1), 58–73 (2002)
10. Coelho, P.H.G., do Amaral, J.F., Guimaraes, K., Bentes, M.C.: Layout of routers in mesh networks with evolutionary techniques. In: The 21st International Conference on Enterprise Information System (ICEIS-2019), pp. 438–445 (2019)

11. Elmazi, D., Oda, T., Sakamoto, S., Spaho, E., Barolli, L., Xhafa, F.: Friedman test for analysing WMNs: a comparison study for genetic algorithms and simulated annealing. In: 2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 171–178. IEEE (2015)
12. Gharehchopogh, F.S., Shayanfar, H., Gholizadeh, H.: A comprehensive survey on symbiotic organisms search algorithms. *Artif. Intell. Rev.* **53**, 2265–2312 (2020)
13. Ikeda, M., Oda, T., Sakamoto, S., Honda, T., Barolli, L.: Analysis of WMN-SA and WMN-GA simulation results: a comparison performance for wireless mesh networks. In: The 17th IEEE International Conference on Network-Based Information Systems, pp. 45–52. IEEE (2014)
14. Lee, S.C., Tan, S.W., Wong, E., Lee, K.L., Lim, C.: Survivability evaluation of optimum network node placement in a hybrid fiber-wireless access network. In: IEEE Photonic Society 24th Annual Meeting, pp. 298–299. IEEE (2011)
15. Lin, C.C.: Dynamic router node placement in wireless mesh networks: a PSO approach with constriction coefficient and its convergence analysis. *Inf. Sci.* **232**, 294–308 (2013)
16. Oda, T., Elmazi, D., Barolli, A., Sakamoto, S., Barolli, L., Xhafa, F.: A genetic algorithm-based system for wireless mesh networks: analysis of system data considering different routing protocols and architectures. *Soft. Comput.* **20**, 2627–2640 (2016)
17. Qiu, L., Bahl, P., Rao, A., Zhou, L.: Troubleshooting wireless mesh networks. *ACM SIGCOMM Comput. Commun. Rev.* **36**(5), 17–28 (2006)
18. Sakamoto, S.: A hybrid intelligent system for wireless mesh networks: assessment of implemented system for two instances and three router replacement methods using v max parameter. *Int. J. Web Grid Serv.* **19**(3), 389–400 (2023)
19. Sakamoto, S., Oda, T., Ikeda, M., Barolli, L., Xhafa, F., Woungang, I.: Investigation of fitness function weight-coefficients for optimization in WMN-PSO simulation system. In: The 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2016), pp. 224–229. IEEE (2016)
20. Sakamoto, S., Obukata, R., Oda, T., Barolli, L., Ikeda, M.: Implementation of an intelligent hybrid simulation system for node placement problem in WMNs considering particle swarm optimization and simulated annealing. In: The 31st IEEE International Conference on Advanced Information Networking and Applications (AINA-2017), pp 697–703. IEEE (2017)
21. Sakamoto, S., Asakura, K., Barolli, L., Takizawa, M.: An intelligent system based on cuckoo search for node placement problem in WMNs: tuning of scale and host bird recognition rate hyperparameters. In: Barolli, L. (ed.) BWCCA 2023. LNDECT, vol. 186, pp. 168–177. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-46784-4_15
22. Sakamoto, S., Barolli, L., Takizawa, M.: Performance evaluation of a cuckoo search based system for node placement problem in wireless mesh networks: evaluation results for computation time and different numbers of nests. In: Barolli, L. (ed.) EIDWT 2024. LNDECT, vol. 193, pp. 384–393. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-53555-0_36
23. Sanni, M.L., Hashim, A.H.A., Anwar, F., Naji, A.W., Ahmed, G.S.: Gateway placement optimisation problem for mobile multicast design in wireless mesh networks. In: 2012 International Conference on Computer and Communication Engineering (ICCCE), pp. 446–451. IEEE (2012)
24. Schutte, J.F., Groenwold, A.A.: A study of global optimization using particle swarms. *J. Global Optim.* **31**(1), 93–108 (2005)

25. Seetha, S., Anand John Francis, S., Grace Mary Kanaga, E.: Optimal placement techniques of mesh router nodes in wireless mesh networks. In: Haldorai, A., Ramu, A., Mohanram, S., Chen, M.-Y. (eds.) 2nd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing. EICC, pp. 217–226. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-47560-4_17
26. Shi, Y.: Particle swarm optimization. IEEE Connect. **2**(1), 8–13 (2004)
27. Shi, Y., Eberhart, R.C.: Parameter selection in particle swarm optimization. In: Evolutionary Programming VII, pp 591–600 (1998)
28. Taleb, S.M., Meraih, Y., Gabis, A.B., Mirjalili, S., Ramdane-Cherif, A.: Nodes placement in wireless mesh networks using optimization approaches: a survey. Neural Comput. Appl. **34**(7), 5283–5319 (2022)
29. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms. Luniver Press (2010)



A Fuzzy-Based System for Assessment of Tie Strength in Online Social Networks

Shunya Higashi¹(✉), Phudit Ampririt¹, Ermioni Qafzezi², Makoto Ikeda², Keita Matsuo², and Leonard Barolli²

¹ Graduate School of Engineering, Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
{mgm23108, bd21201}@bene.fit.ac.jp

² Department of Information and Communication Engineering, Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
qafzezi@bene.fit.ac.jp, makoto.ikd@acm.org, {kt-matsuo, barolli}@fit.ac.jp

Abstract. In the era of digital communication, Online Social Networks (OSNs) have become a fundamental aspect of daily life, influencing information dissemination, community formation, and social influence. Assessing the strength of ties between users in OSNs is critical for applications such as targeted marketing, recommendation systems, and social influence analysis. Traditional methods for evaluating tie strength often rely on metrics that fail to capture the nuanced nature of human relationships. To address this, we propose a Fuzzy-based System for Assessment of Tie Strength (FSATS) in OSNs. Our system integrates multiple dimensions of tie strength, including Interaction Time (IT), Level of Intimacy (LoI), and Emotional Intensity (EI), leveraging fuzzy logic to handle the inherent uncertainty and complexity of social interactions. Simulation results demonstrate the effectiveness of FSATS, showing that increase of EI, IT, and LoI positively correlate with higher Tie Strength (TS).

1 Introduction

In the era of digital communication, Online Social Networks (OSNs) have become a fundamental aspect of daily life, facilitating interactions across diverse populations and geographies. These platforms enable users to build and maintain relationships, share information, and engage in various forms of social exchange. The strength of ties between users in OSNs is a critical factor that influences information dissemination, community formation, and social influence [1,2].

Tie strength, a concept originating from sociology, refers to the closeness and interaction frequency between individuals in OSNs. Strong ties often represent close relationships, such as family and close friends, while weak ties are typically more casual acquaintances. Understanding and accurately assessing tie strength is essential for various applications, including targeted marketing, recommendation systems, and the detection of influential users [3].

Traditional methods for assessing tie strength often rely on straightforward metrics such as interaction frequency, message volume, or mutual friend count [4]. However,

these methods may fail to capture the nuanced and multifaceted nature of relationships in OSNs. To address these limitations, this research proposes a novel Fuzzy Logic (FL) based system for the assessment of tie strength in OSNs. The FL, with its ability to handle uncertainty and approximate reasoning, provides a robust framework for modeling the complex and imprecise nature of human relationships.

The primary objective of this study is to design and implement a FL based system that integrates multiple dimensions of Tie Strength (TS), including Interaction Time (IT), Level of Intimacy (LoI), and Emotional Intensity (EI). By leveraging FL, the proposed system aims to provide a more accurate and comprehensive assessment of tie strength compared to traditional approaches. Our proposed system is evaluated by computer simulations. The simulation results demonstrate a positive correlation between input parameters (IT, LoI, EI) and TS, which is the output parameter. Consequently, an increase in these parameters results in an increase in TS.

This paper is structured as follows. Section 2 reviews the existing literature on tie strength in OSNs and gives a short overview of the application of FL in this domain. Section 3 provides an overview of FL principles relevant to the proposed system. Section 4 details the design and implementation of the FL-based system for assessing tie strength. Section 5 presents the simulation results and evaluates the performance of the proposed system. Finally, Sect. 6 concludes the paper with a summary of findings and suggestions for future research directions.

2 Literature Review

2.1 Tie Strength in OSNs

Tie strength is a fundamental concept in social network analysis, originally introduced by Granovetter in his seminal work “The Strength of Weak Ties”. Granovetter posited that weak ties, despite being less intimate and less frequent than strong ties, play a crucial role in information diffusion and bridging disparate social groups [5]. Subsequent research has expanded on this idea, examining how tie strength impacts various aspects of OSNs, including trust, community formation, and social capital [6–8].

In the context of OSNs, the assessment of tie strength has become increasingly important. OSNs provide a rich source of interaction data that can be leveraged to quantify relationships between users. In [9], the authors utilized interaction frequency, message volume, and shared content to measure tie strength in OSNs. However, these traditional metrics often overlook the qualitative aspects of relationships, such as emotional intensity and intimacy, which are crucial for a comprehensive understanding of tie strength.

2.2 Application of FL in OSN Analysis

The FL has proven particularly useful in modeling the nuanced and subjective aspects of social relationships. For example, FL has been employed to assess trust in virtual communities, considering factors like reputation and past behavior. Similarly, it has been used to measure the strength of ties based on the overlap of social circles and shared interests [10].

In the domain of OSNs, the FL has been used to model user behavior and interactions. In [11], the authors proposed a FL system to predict user engagement on social media platforms, demonstrating improved accuracy over traditional linear models. Furthermore, a recent study explores the application of FL in OSNs, emphasizing its utility in capturing the complex nature of human interactions in digital environments [12].

The application of FL for tie strength assessment is a promising approach. By accommodating the imprecise nature of human relationships, FL provides a robust framework for capturing both quantitative and qualitative dimensions of tie strength. This study proposes a FL-based system that incorporates multiple dimensions of tie strength, including interaction time, level of intimacy, and emotional intensity.

In summary, the literature underscores the importance of accurately assessing tie strength in OSNs and highlights the potential of FL to enhance this assessment. The proposed FL-based system aims to address the limitations of traditional methods, providing a more comprehensive and nuanced evaluation of tie strength in OSNs.

3 FL Overview

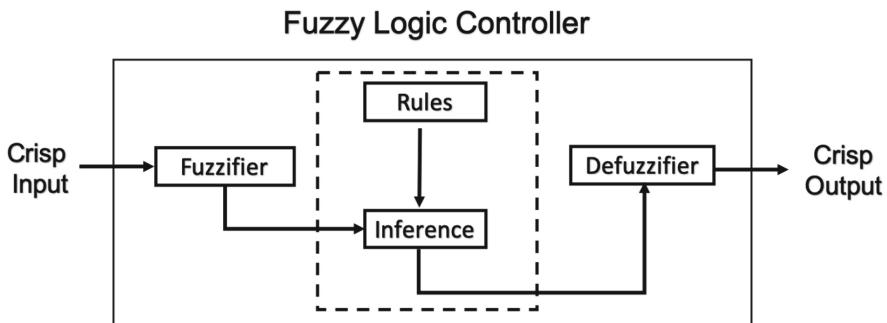
The FL is a mathematical framework that facilitates reasoning in situations characterized by uncertainty and imprecision. Unlike classical binary logic, which confines variables to true or false values, FL variables can take on a continuum of values between 0 and 1. This feature makes FL particularly well-suited for modeling complex systems where human-like reasoning is required [13].

FL includes fuzzy sets, which are classes with unsharp boundaries, allowing for partial membership. For instance, in the context of social interactions, a relationship can be slightly strong or slightly weak, reflecting the inherent ambiguity in human relationships.

A Fuzzy Logic Controller (FLC) is the main part of the proposed system. The FLC processes inputs and generates outputs using fuzzy inference rules. It comprises four main components, as shown in Fig. 1.

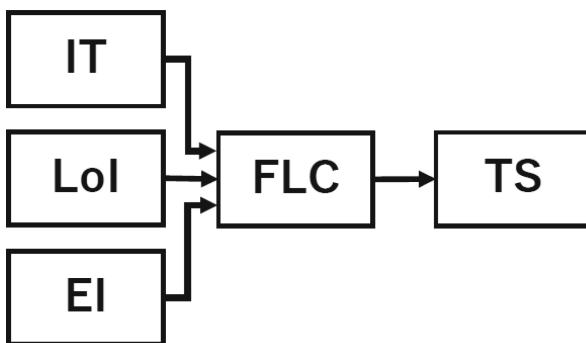
- **Fuzzification:** This step converts crisp input values into fuzzy values using pre-defined membership functions. Membership functions define how each input value maps to a degree of membership within a fuzzy set.
- **Rule Base:** The rule base is a collection of fuzzy IF-THEN rules that embody the knowledge and expertise required for decision-making. For example, a rule might state, “IF Interaction Time is Long, Level of Intimacy is Very High and Emotional Intensity is High, THEN Tie Strength is Very Strong”.
- **Inference Engine:** This component applies the fuzzy rules to the fuzzified inputs to produce fuzzy outputs. The inference engine integrates multiple rules to determine the relationship between inputs and outputs.
- **Defuzzification:** The final step converts the fuzzy output values back into crisp values, providing a concrete result that can be interpreted and utilized in decision-making processes.

The FLC structure enhances adaptability and effectiveness in handling the complexity and uncertainty inherent in social interactions. By utilizing fuzzy control rules and input variables, FLC can effectively assess tie strength in OSNs.

**Fig. 1.** FLC structure.

4 Proposed Fuzzy-Based System

This section introduces the Fuzzy-based System for Assessment of Tie Strength (FSATS) in OSNs. Utilizing FL, the FSATS determines tie strength through three input parameters: Interaction Time (IT), Level of Intimacy (LoI), and Emotional Intensity (EI). The output parameter is Tie Strength (TS). The system's structure is illustrated in Fig. 2.

**Fig. 2.** Proposed system structure.

- **Input Parameters**

- **Interaction Time (IT)**: It represents the duration of interactions between users in the network.
- **Level of Intimacy (LoI)**: It captures the closeness and depth of the relationship between users.
- **Emotional Intensity (EI)**: It measures the strength of emotional exchanges between users.

- **Output Parameter**:

- **Tie Strength (TS)**: It is the resulting measure of the strength of the relationship between users based on the input parameters.

4.1 Membership Functions

Membership functions translate precise input values into fuzzy sets, reflecting linguistic terms and their associated degrees. This conversion allows the system to manage the ambiguity and uncertainty present in social interactions. Figure 3 depicts the membership functions for Interaction Time, Level of Intimacy, Emotional Intensity, and Tie Strength.

Table 1 details the term sets associated with each input and output parameter. The term sets define the possible values that each parameter can take, facilitating the fuzzification process.

Table 1. Parameters and their term sets.

Parameters	Term Sets
Interaction Time (IT)	Short (Sh), Medium (Me), Long (Lo)
Level of Intimacy (LI)	Very low (Vl), Low (Lw), Medium (Md), High (Hg), Very high (Vh)
Emotional Intensity (EI)	Low (L), Medium (M), High (H)
Tie Strength (TS)	Very Weak, Weak, Slightly Weak, Moderate, Slightly Strong, Strong, Very Strong

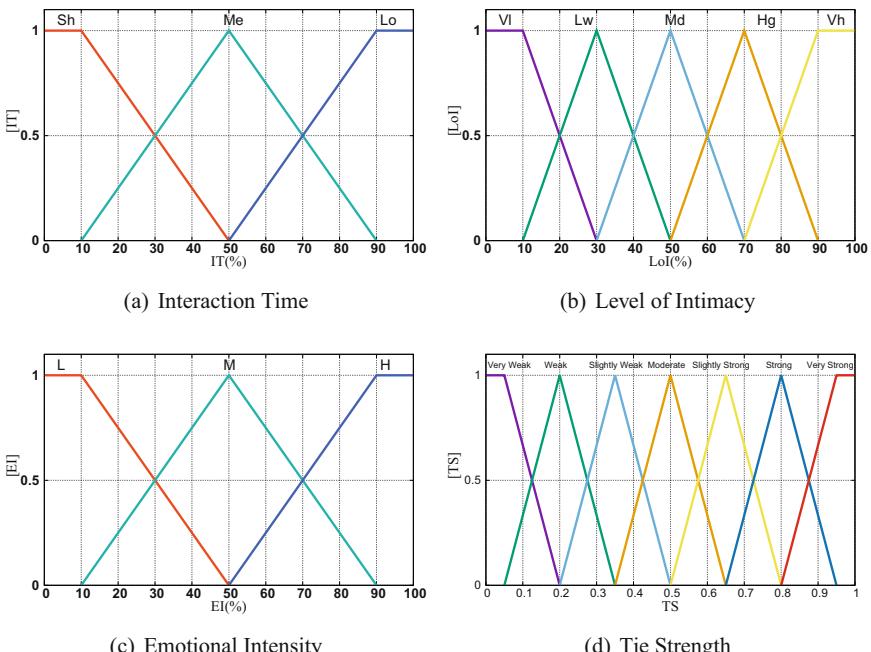


Fig. 3. Membership functions.

4.2 Fuzzy Rule Base

The Fuzzy Rule Base (FRB) forms the core of the FSATS's decision-making process. It comprises a set of “IF-THEN” rules that define the relationships between input parameters and the output parameter. Table 2 provides an excerpt of the FRB, which includes 45 rules, assuming three linguistic values for each input parameter. The FRB covers all possible combinations of input values, ensuring a comprehensive assessment of tie strength.

Table 2. FRB.

Rule No.	IF Condition	THEN Action
1	IF IT is Short AND LoI is Very low AND EI is Low	THEN TS is Very Weak
2	IF IT is Short AND LoI is Very low AND EI is Low	THEN TS is Very Weak
...
45	IF IT is Long AND LoI is Very high AND EI is High	THEN TS is Very Strong

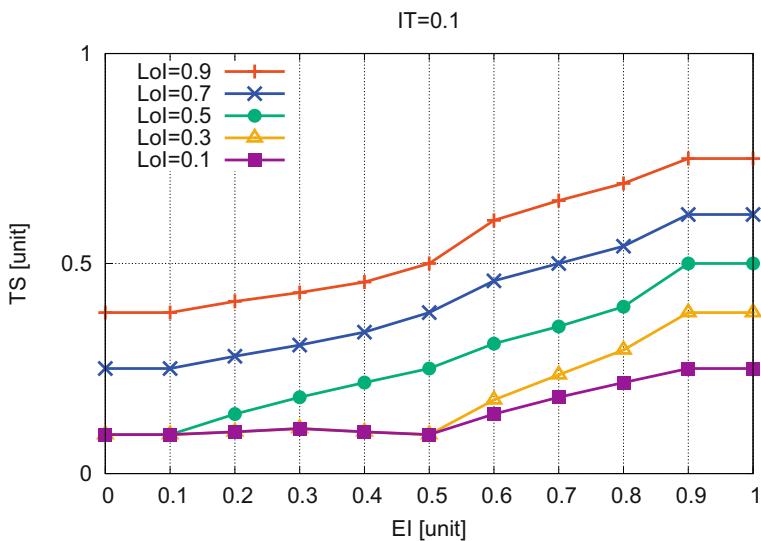
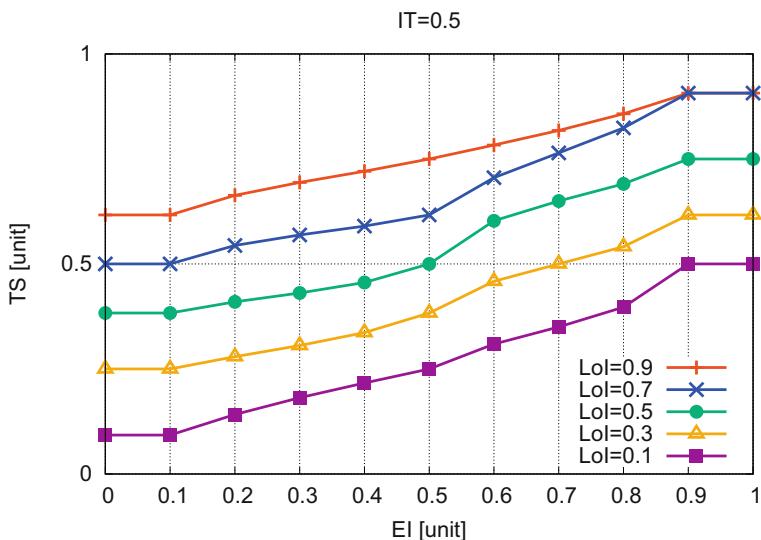
5 Simulation Results

This section presents the simulation results for FSATS. The simulations were conducted to evaluate how the IT, LoI, and EI affect the TS. The results are displayed in Fig. 4, Fig. 5 and Fig. 6 for different values of interaction time.

Figure 4 shows the simulation results for $IT = 0.1$. As the EI increases, the TS also increases for all LoI values. The relationship between EI and TS becomes more stronger for higher values of LoI, indicating that higher levels of intimacy amplify the effect of emotional intensity on tie strength. For instance, when the LoI is 0.9, TS increases by 37%, which is increased more than 26% comparing with the case when LoI is 0.1.

Figure 5 illustrates the simulation outcomes for $IT = 0.5$. Similar to the previous scenario, an increase in EI leads to an increase in TS. However, the impact of EI on TS is more substantial compared with the case when IT is 0.1. The results show that for a medium level of interaction time, the TS values are increased for higher levels of LoI. For example, in case when the LoI values are 0.7 and 0.9, TS is about 90% when EI is more than 0.9.

The results for $IT = 0.9$ are depicted in Fig. 6. In this scenario, the TS values are generally higher across all levels of EI and LoI compared to the previous scenarios. The increased interaction time significantly increases the tie strength, showing a strong correlation between prolonged interaction and stronger ties.

**Fig. 4.** Simulation results for IT = 0.1.**Fig. 5.** Simulation results for IT = 0.5.

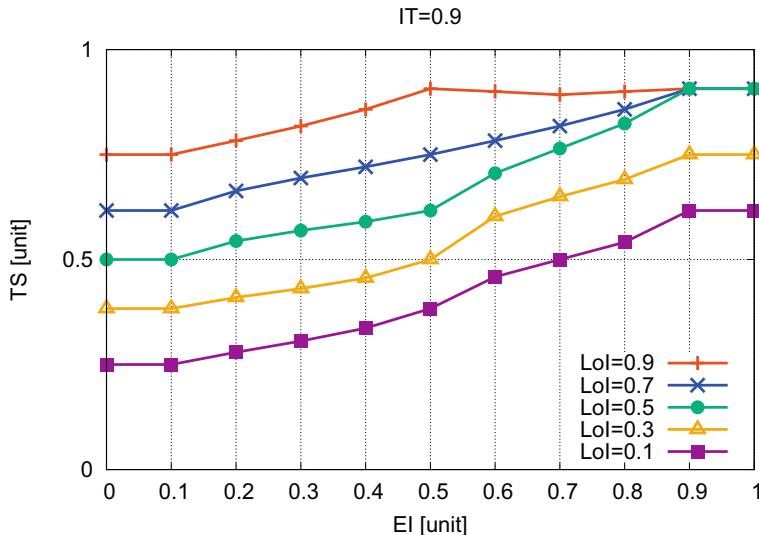


Fig. 6. Simulation results for IT = 0.9.

6 Conclusions and Future Work

In this study, we proposed the FSATS for OSNs. The objective was to develop a system that integrates multiple dimensions of tie strength including IT, LoI, and EI in order to provide a more accurate and comprehensive assessment compared to traditional methods.

Our simulation results demonstrated the effectiveness of the proposed FSATS in capturing the dynamics of tie strength in OSNs. Key findings from the simulations include:

- **Positive Correlation between EI and TS:** Across all scenarios, an increase in Emotional Intensity led to a higher Tie Strength, with more increased values at higher Levels of Intimacy.
- **Impact of IT on TS:** Higher Interaction Time significantly enhanced the Tie Strength, suggesting that prolonged interactions contribute to stronger relationships.
- **Amplification by LoI:** The Level of Intimacy amplified the effect of both Emotional Intensity and Interaction Time on Tie Strength, indicating that deeper relationships are more sensitive to these factors.

In the future, we will consider incorporating additional parameters to further enhance the system's accuracy and reliability. Extensive simulations will be conducted to evaluate the proposed system under various conditions and scenarios.

References

1. Arnaboldi, V., Conti, M., Passarella, A., Dunbar, R.I.M.: Online social networks and information diffusion: the role of ego networks. *Online Soc. Netw. Media* **1**, 44–55 (2017). <https://doi.org/10.1016/j.osnem.2017.04.001>
2. Ullah, F., Lee, S.: Social content recommendation based on spatial-temporal aware diffusion modeling in social networks. *Symmetry* **8**(9), 89 (2016). <https://doi.org/10.3390/sym8090089>
3. Rajkumar, K., Saint-Jacques, G., Bojinov, I., Brynjolfsson, E., Aral, S.: A causal test of the strength of weak ties. *Science* **377**, 1304–1310 (2022). <https://doi.org/10.1126/science.abl4476>
4. Jones, J.J., Settle, J.E., Bond, R.M., Fariss, C.J., Marlow, C.A., Fowler, J.: Inferring tie strength from online directed behavior. *PLoS ONE* **8**(1) (2013). <https://doi.org/10.1371/journal.pone.0052168>
5. Granovetter, M.S.: The strength of weak ties. *Am. J. Sociol.* **78**(6), 1360–1380 (1973). <https://doi.org/10.1086/225469>
6. Gilbert, E., Karahalios, K.: Predicting tie strength with social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 211–220 (2009). <https://doi.org/10.1145/1518701.1518736>
7. Kossinets, G., Watts, D.: Empirical analysis of an evolving social network. *Science* **311**, 88–90 (2006). <https://doi.org/10.1126/SCIENCE.1116869>
8. Ellison, N., Steinfield, C., Lampe, C.: The benefits of Facebook “friends:” social capital and college students’ use of online social network sites. *J. Comput.-Mediat. Commun.* **12**(4), 1143–1168 (2007). <https://doi.org/10.1111/J.1083-6101.2007.00367.X>
9. Xiang, R., Neville, J., Rogati, M.: Modeling relationship strength in online social networks. In: Proceedings of the 19th International Conference on World Wide Web, pp. 981–990 (2010). <https://doi.org/10.1145/1772690.1772790>
10. Adamic, L.A., Adar, E.: Friends and neighbors on the web. *Soc. Netw.* **25**(3), 211–230 (2003). [https://doi.org/10.1016/S0378-8733\(03\)00009-1](https://doi.org/10.1016/S0378-8733(03)00009-1)
11. Liu, Y., Kliman-Silver, C., Mislove, A.: The tweets they are a-changin: evolution of twitter users and behavior. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, pp. 305–314 (2014). <https://doi.org/10.1609/icwsm.v8i1.14508>
12. Atif, Y., Al-Falahi, K., Wangchuk, T., Lindström, B.: A fuzzy logic approach to influence maximization in social networks. *J. Ambient. Intell. Humaniz. Comput.* **11**, 2435–2451 (2019). <https://doi.org/10.1007/S12652-019-01286-2>
13. Dadaz-Corts, M.-A., Jimnez, E.V.C., Glvez, J., Camarena, O.: A new metaheuristic optimization methodology based on fuzzy logic. *Appl. Soft Comput.* **61**, 549–569 (2017). <https://doi.org/10.1016/j.asoc.2017.08.038>



An Efficient Algorithm to Prevent Procrastination in Spatial Crowdsourcing

Naren Debnath¹(✉), Sajal Mukhopadhyay¹, and Fatos Xhafa²

¹ Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur, 713209 West Bengal, India
nd.20cs1104@phd.nitdgp.ac.in, smukhopadhyay.cse@nitdgp.ac.in

² Department of Computer Science, Universitat Politècnica de Catalunya, 08034 Barcelona, Catalonia, Spain
fatos@cs.upc.edu

Abstract. Spatial crowdsourcing entails many research works that address the issue of meeting deadlines (for task submission) by task executors (TEs) where they can submit the tasks at any point of time within the given deadline. In accordance with the deadline, TEs may hold up their task submission until the penultimate day. Such a behavioural phenomenon is known as procrastination, when exercised by TEs, may create a problematic situation for the task provider (TP) who is also bounded by some deadline. Though procrastination has not at all been addressed in spatial crowdsourcing scenarios, one work addresses this issue by proposing a procrastination-aware scheduling algorithm in a bipartite graph environment. But the balancing effect of task distribution in different schedules (slots) has not been taken care of. In this paper, we have proposed a mechanism to prevent procrastination in spatial crowdsourcing scenario considering the balancing effect. Extensive simulation is done, and in that our main finding is that the proposed mechanism performs significantly better than the existing method in terms of balancing effect.

1 Introduction

When a large set of tasks is brought to an open forum and an independent crowd is asked to execute them is called Crowdsourcing [10, 16]. Some examples of crowdsourcing (cs) applications are agriculture [17], IoT [14], etc. A kind of crowdsourcing that specifically exploits sensors and mobile devices to acquire data is called mobile crowdsourcing [9, 22, 25]. A few applications of MCS are smart tourism and smart cities [13], IoT [14], crime watch [18].

In MCS, tasks at some locations might not attain adequate agents or TEs [8], which leads to the emergence of location-based crowdsourcing or spatial crowdsourcing (SC) where the task is allocated to the agents with spatiotemporal constraints. In SC, tasks are assigned to a TE with its location information and the deadline, and she has to be physically present at that location to execute the task within the given deadline. Examples of SC are food delivery service [15], on-demand vehicle service [8], collecting geo-spatial data [23], journalism [8] etc.

The literature of SC [8, 15, 23] concentrate on allocating tasks floated by TP to the TEs. Here TP and TEs both place their demands to a common platform. These literature of SC did not consider the fact that the submission of tasks by the TEs is getting delayed *i.e.*, the submission of tasks happening at the penultimate day or at the final day of the deadline. Such a behavioural phenomena is known as procrastination [11, 12, 20], when exercised by TEs, may create a problematic situation for the task provider (TP) who is also bound by some deadline and as the TEs are in haste, the submission quality may be poor as well. So, prevention of procrastination is necessary in SC environment and it is noticed that there is dearth of literature in SC to address this issue. One research work (though not in SC environment) by *Wang et al.* in a bipartite graph setting has proposed a procrastination-aware scheduling in [21]. But they have not considered the balancing of tasks to avoid procrastination in their paper.

Our proposed mechanism begins with the fact that the platform has already allocated tasks to the TEs. However, the TEs may procrastinate within the given deadline. In this mechanism the platform (a cognitive agent) perceives that the TEs may procrastinate and efficiently distributes tasks into different schedules (slots with equal time frames) in a balanced manner to avoid procrastination. We have defined *schedules* in Sect. 3. The idea of distributing the tasks in a balanced manner will help us prevent procrastination in a systematic manner.

The rest of our paper is organised as follows. Section 2 presents the most relevant literature review. Section 3 provides the system model and objective function. Section 4 elaborates the proposed mechanism and provides the time complexity analysis. In Sect. 5, we have furnished the simulation. We have ended our paper with some concluding remarks and future directives in Sect. 6.

2 Literature Review

Procrastination could be comprehended from the story of the famous economist G. Akerlof mentioned in [1, 11, 20]. In brief, his story narrates that he procrastinated while returning his friend's luggage for months due to the long postal process. This story brings out the fact that procrastination is a common trait of human behaviour based on which systematic mechanisms could be designed. Our paper is motivated in this direction by the above-mentioned fact.

There are a salient number of research works have been carried out on *task scheduling* in crowdsourcing environments with load balancing and deadline [2–4, 7, 19, 24, 26]. In [3, 19], and [4], spatial task scheduling mechanisms were proposed where tasks were scheduled, maintaining the load balance among the TEs. [2] have addressed task scheduling in IoT/Mobile Crowdsourcing environments where they have ensured the balanced scheduling of tasks, which is deadline and budget-sensitive. [7, 24], and [26] have proposed task assignment scheduling mechanisms where the task is constrained by the deadline. The above review addresses load balancing and time constraints in spatial crowdsourcing, but procrastination has not at all been addressed.

A task-scheduling mechanism that explicitly addresses the issue of procrastination but not specifically in crowdsourcing is found in [21]. They have proposed a Procrastination-aware-Scheduling algorithm (*OFFPSP*) in which all the tasks or jobs are allocated into a set of schedules in some order where each schedule has a threshold. This algorithm produces an unbalanced set of schedules, which can cause irregular completion and submission of tasks. This irregular submission of tasks might be problematic for the task requester (or TP) as she will not be able to get ample time to verify all the submitted tasks. We have addressed the issue of balancing of tasks so that all the tasks are completed and submitted uniformly to the TP. In addition to that, we have pruned the choices of TEs to prevent procrastination.

3 System Models

As stated earlier, our model begins with the fact that the tasks are already allocated to the TEs and how we can prevent the procrastination of TEs within the given deadline \mathbb{R} . For example, suppose a TE is already assigned a set of 10 tasks and it was given a deadline $\mathbb{R} = 21$ days. It may be the case that the TE may submit the task on the last day (\mathbb{R}^{th} day) or as late as possible. This procrastination issue of TE is not addressed in earlier literature on spatial crowdsourcing (SC). In our model, we have proposed a mechanism to prevent the procrastination of an arbitrary TE i and thereby, it is applicable to all the TEs in the same way. In our model, TEs are heterogeneous in nature, *i.e.*, they can perform different types of jobs and the capacity of performing of jobs is also different. The key constituents here are a set of schedules (Λ), a set of Tasks (\mathfrak{I}), and a set of locations (\mathbb{L}), discussed below.

1. Notations and their Usage

Schedules and Deadline: A set of schedules (or slots) $\Lambda = \{\alpha_1, \alpha_2, \dots, \alpha_l\}$ is created based on the overall deadline \mathbb{R} . Each of these schedules or slots contains a subset of tasks (defined next). As the schedules (slots) are created based on \mathbb{R} , each schedule has a deadline also. The deadline for each schedule is created by dividing \mathbb{R} into l sub-deadlines of equal length, represented as $\mathbb{R} = \{\theta_1, \theta_2, \dots, \theta_l\}$, where $\theta_1 = \theta_2 = \dots = \theta_{|\Lambda|}$ and $l = |\Lambda|$. $|\Lambda|$ is defined as $|\Lambda| = \frac{\mathbb{R}}{\theta_i}$ where $\theta_i \in \mathbb{R}, \theta_i \neq 0$, and $\theta_i \leq \mathbb{R}$. The duration of θ_i is application-dependent, and accordingly, our proposed mechanism is pliable enough for any such application.

Set of Tasks: The set of tasks is defined as $\mathfrak{I} = \{\iota_1, \iota_2, \dots, \iota_n\}$ where $n = |\mathfrak{I}|$ and the costs of these tasks is represented as $\kappa = \{\varkappa_1, \varkappa_2, \dots, \varkappa_n\}$ where $\varkappa_i \in \kappa$ corresponds to the task $\iota_i \in \mathfrak{I}$. The cost could be the power consumption of the mobile device(s), time to complete the task, internet charge, etc. Each of these tasks will be allocated to a schedule α_i . The total cost of any schedule α_i will be denoted as α_i^c . Note that any two schedules α_i and $\alpha_j, i \neq j$, will not be allocated with the same task, *i.e.*, $\alpha_i \cap \alpha_j = \emptyset, \forall_{(\alpha_i, \alpha_j) \in \Lambda}$.

Location: A set of locations defined as $\mathbb{L} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ constitutes the SC area where in each location a set of tasks is cluttered. That means the location $\lambda_e \in \mathbb{L}$ is populated with a set of tasks \mathfrak{S} . Each task $\iota_i \in \mathfrak{S}$ is associated with a cost \varkappa_i and a location λ_e . So, the set of tasks \mathfrak{S} could be redefined as $\mathfrak{S} = \{\langle \iota_1, \varkappa_1, \lambda_e \rangle, \langle \iota_2, \varkappa_2, \lambda_e \rangle, \dots, \langle \iota_n, \varkappa_n, \lambda_e \rangle\}$. Our proposed mechanism is based on an arbitrary location λ_i and could be applied to all such m locations.

2. Objective

Given a set of tasks \mathfrak{S} and a set of schedules (slots) Λ , at the location $\lambda_e \in \mathbb{L}$ our objective is to distribute \mathfrak{S} tasks into different schedules uniformly so that all the schedules become balanced in terms of total cost and number of tasks. In other words, our objective is to minimize the standard deviation of each schedule in terms of total cost and number of tasks. Such uniformity ensures the distribution of higher cost tasks into all the schedules which prohibits the accumulation of such higher cost tasks into one schedule. This will prevent delay in the execution of those higher cost tasks along with the other tasks of that schedule.

As our objective relies on uniform (or balanced in terms of the objective function defined below) distribution of tasks, the deadline for each schedule must also be uniform so that each schedule gets equal amount of time to complete the tasks assigned to it. To achieve that, we have divided \mathbb{R} into $|\Lambda|$ equal sub-deadlines for each schedule and each sub-deadline is represented by θ_i . We have summarised our objective in Eq. 1 where, σ_c and σ_d represent the *standard deviations* (given in Eq. 2 and Eq. 3) in terms of cost and number of tasks of each schedule, respectively.

$$\text{Minimize } \sigma_c\{\Lambda\} \text{ and } \sigma_d\{|\alpha_1|, |\alpha_2|, \dots, |\alpha_l|\}, \text{ Subject to } \theta_i = \frac{\mathbb{R}}{|\Lambda|} \quad (1)$$

$$\text{where, } \sigma_c = \sqrt{\frac{\sum_{i=1}^{|\Lambda|} (\alpha_i^c - \frac{\sum_{k=1}^{|\Lambda|} \alpha_k^c}{|\Lambda|})^2}{|\Lambda|}} \quad (2)$$

$$\sigma_d = \sqrt{\frac{\sum_{i=1}^{|\Lambda|} (|\alpha_i| - \frac{\sum_{k=1}^{|\Lambda|} |\alpha_k|}{|\Lambda|})^2}{|\Lambda|}} \quad (3)$$

4 Proposed Mechanism

This section discusses the proposed mechanism Uniform Cost Scheduling Mechanism for Prevention of Procrastination(UCSMPP) in two ways. First, a schematic diagram of the proposed mechanism is provided. Second, the algorithms are discussed. Finally, an example with an arbitrary dataset is furnished to illustrate our proposed mechanism.

4.1 Schematic Diagram of the Proposed Mechanism

Figure 1 shows the flow of our proposed mechanism UCSMPP. In Block 1, the agent submits n tasks to the platform where they are arranged in ascending order according to their costs (Block 2). In Block 3, the sorted tasks are allocated into

l schedules in the following manner - $\iota_1, \iota_n \rightarrow \alpha_1$, $\iota_2, \iota_{n-1} \rightarrow \alpha_2$, and so on upto $\iota_l, \iota_{n-l+1} \rightarrow \alpha_l$. This process is repeated for all the tasks. Block 4 generates the schedules with allocated tasks. Finally, Block 5 publishes schedules with heaviest schedule first manner.

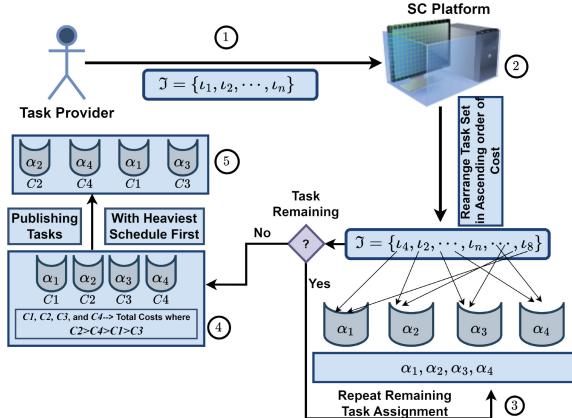


Fig. 1. Flow Diagram of the Proposed Mechanism

4.2 Uniform Cost Scheduling Mechanism for Prevention of Procrastination

4.2.1 Uniform Task Distribution as per Costs (UTDC)

This algorithm distributes the set of tasks into l number of schedules after reordering the set of tasks in ascending order (Algorithm 1). The input to this algorithm is \mathfrak{I} and \mathbb{R} , and the output is Λ . Each $\lambda_i \in \Lambda$ is allocated with some tasks where each task has a cost. So, the total cost of each schedule is the sum of the costs of all tasks allocated to it. At the end of the Algorithm 1, the schedules are published to the TEs in reverse order of their total cost.

The proposed algorithm divides \mathbb{R} into several schedules with the same deadline (defined in Sect. 3) (choice reduction) and distributes higher cost tasks along with other tasks into those schedules uniformly (balancing effect). By choice reduction, we mean that \mathbb{R} is divided into smaller and equal time frames (schedules), and allocated tasks in each schedule are bound to be completed within the time frame of that schedule. Thus the choices of TEs for submitting tasks are reduced from the larger time frame \mathbb{R} to smaller time frames θ so that they can periodically submit the tasks. This is one of the key ideas of our proposed algorithm.

So, we can see that the balancing effect and choice reduction in our proposed algorithm prevent procrastination. Also, the reverse order of publishing schedules benefits TP in the sense that when TEs submit the tasks, TP will have enough time to verify the heaviest schedule first and so on.

Algorithm 1. Uniform Task Distribution as per Costs (UTDC)**Require:** \mathfrak{S}, \mathbb{R} **Ensure:** Λ with allocated tasks

```

1:  $S\_list \leftarrow call\ TSRAO(\mathfrak{S}, 1, |\mathfrak{S}|)$                                  $\triangleright$  sorted list as per cost
2:  $l \leftarrow \frac{|\mathfrak{S}|}{\theta}, i \leftarrow 0, j \leftarrow |\mathfrak{S}| - 1$            $\triangleright \theta : \text{deadline of one schedule}$ 
3: Initialise  $\alpha^k \leftarrow \{\}$  for  $k = 1$  to  $l$                           $\triangleright l : \text{number of schedules}$ 
4: while  $i \leq \lfloor \frac{|\mathfrak{S}|}{2} \rfloor$  do
5:    $k \leftarrow 0$ 
6:   while  $k < l$  do
7:     if  $i == j$  then                                               $\triangleright : if |\Lambda| == |\mathfrak{S}|$ 
8:        $\alpha^k \cup S\_list_i, break$ 
9:     else if  $i > j$  then                                               $\triangleright : if |\Lambda| < |\mathfrak{S}|$ 
10:     $break$ 
11:   else
12:      $\alpha^k \cup S\_list_i, \alpha^k \cup S\_list_j$ 
13:      $k += 1, i += 1, j -= 1$ 
14:   end if
15: end while
16: end while
17:  $\Lambda$  is published in reverse order of the cost of each  $\alpha_k \in \Lambda$ 

```

Algorithm 2. Subroutine $TSRAO(A, 1, |A|)$

```

1: if  $|A| > 1$  then
2:    $mid \leftarrow \frac{|A|}{2}$ 
3:    $TSRAO(A, 1, mid)$ 
4:    $TSRAO(A, mid + 1, |A|)$ 
5:    $merge(A, 1, mid, |A|)$                                                $\triangleright$  Merge sort is used here
6: end if
7: return  $A$ 

```

4.2.2 Task Set Rearrangement in Ascending Order (TSRAO)

Algorithm 1 calls the subroutine $TSRAO()$ (Algorithm 2) to rearrange the tasks in \mathfrak{S} in ascending order of their costs by implementing merge sort algorithm.

Example 1. Let $\kappa = \{89, 57, 6, 80, 85, 53, 20, 98, 10, 32, 34, 11, 23, 50, 60, 14, 18\}$ and $nosch = 4$. Rearranging \mathfrak{S} in ascending order of the cost, we get $S_{list} = \{6, 10, 11, 14, 18, 20, 23, 32, 34, 50, 53, 57, 60, 80, 85, 89, 98\}$. Now, from the new list S_{list} , 1st and 17th tasks are assigned to α_1 , 2nd and 16th tasks are assigned to α_2 , and so on up to α_4 . It is repeated until all the tasks are assigned to the respective schedules. The final task allocation is, $\alpha_1 = \{98, 6, 60, 18, 34\}$, $\alpha_2 = \{89, 10, 57, 20\}$, $\alpha_3 = \{85, 11, 53, 23\}$, $\alpha_4 = \{80, 14, 50, 32\}$ where the total cost for $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are 216, 176, 172, 176 respectively. After task allocation, each schedule is published to the TEs according to the decreasing order of their costs (see Sect. 4.1). Each schedule will get $\frac{|\mathfrak{S}|}{4} = 1$ week. In 1st week, tasks in α_1 , in 2nd week tasks in α_2 , in 3rd week tasks in α_4 and in 4th week tasks in α_3 will be executed.

4.3 Time Complexity Analysis

Lemma 1. *The time complexity by UCSMPP is $O(|\mathfrak{S}| \log |\mathfrak{S}|)$.*

Proof. Algorithm 1 will iterate atmost $\lceil \frac{|\mathfrak{S}|}{2} \rceil$ times including the inside *while* loop. As each pair of elements is assigned exactly once, so over $\lceil \frac{|\mathfrak{S}|}{2} \rceil$ number of iterations, atmost $|\mathfrak{S}|$ elements will be assigned. As each assignment takes $O(1)$ time, the algorithm takes $O(|\mathfrak{S}|)$ time. This analysis is motivated by one of the methods of amortised analysis, termed as aggregated analysis as mentioned in [6]. In our algorithm, the preprocessing (*merge sort*) takes $O(|\mathfrak{S}| \log |\mathfrak{S}|)$ and the main segment takes $O(|\mathfrak{S}|)$ time. \square

5 Experiments and Results

5.1 Setup

Overall Deadline: \mathbb{R} for all the tasks is considered 4 weeks, i.e., $\mathbb{R} = 4$ weeks.

Schedules and Their Deadlines: Here, i.e., $\Lambda = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$, $|\Lambda| = 4$, $\theta_i = 1$ week. Therefore, deadline of each α_i is $\theta_i = \frac{\mathbb{R}}{|\Lambda|} = 1$ week.

System Specification: The simulation platform has 5th generation microprocessor, 8 GB of RAM, Ubuntu as the operating system and *Google Colab* platform (with Python).

5.2 Dataset

Synthetic Dataset: The synthetic dataset, shown in Table 1, has been generated randomly with 200 jobs using normal distribution (given in Eq. 4),

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \text{ with } (\mu) = 75, (\sigma) = 80 \text{ and } x = 200. \quad (4)$$

Table 1. Sample of Synthetic Dataset

Sl #	Job (\mathfrak{S})	Job Cost (\varkappa)
1	ι_1	10
2	ι_2	40
3	ι_3	25
4	ι_4	100
5	ι_5	89

Table 2. Sample of BDD Dataset

Sl#	Bus.ID	Tot.Dur
1	1	1145
2	2	1110
3	3	850
4	4	1090
5	5	840

Bus Driver Scheduling Dataset 1: In BDD [5], *Bus_Line_ID* and *Duration* columns are extracted. Then for each *Bus_Line_ID* (each bus), total running cost is calculated for the entire day by summing up all the values in *Duration* column. For instance, for *Bus_Line_ID*= 1, the total running time is 1145 minutes. In such a way, a total of 359 bus records have been collected, some are shown in Table 2.

5.3 Simulation Study of UCSMPP

For our simulation, we have used the following parameters.

1. Total Cost of Each Schedule: With this simulation, we have shown that the total cost of each schedule is almost balanced. Each segment of the chart in Fig. 2 shows that each schedule is almost balanced in terms of the total cost of jobs.

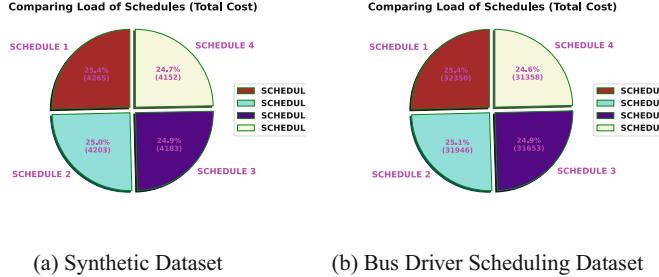


Fig. 2. Total Cost of Each Schedule with Synthetic and BDD Datasets

2. Number of Tasks: This simulation shows that the number of tasks in each schedule is also uniform (Fig. 3). We have determined the number of tasks in each schedule in the following way. Let N denote the number of pairs of tasks where $N = \lceil \frac{|\mathfrak{S}|}{2} \rceil$ and let $M = |\mathcal{A}|$. 1^{st} pair contains 1^{st} and $|\mathfrak{S}|^{th}$ tasks, 2^{nd} pair contains 2^{nd} and $(|\mathfrak{S}| - 1)^{th}$ tasks and so on upto N^{th} pair which contains $\frac{|\mathfrak{S}|}{2}^{th}$ task, if $2 \nmid |\mathfrak{S}|$, else it contains $\frac{|\mathfrak{S}|}{2}^{th}$ and $(\frac{|\mathfrak{S}|}{2} + 1)^{th}$ tasks. Such pairing is only considered when \mathfrak{S} is sorted in ascending order.

$$\alpha_i = \begin{cases} \frac{N}{M} & \text{if } M \mid N, \text{ for } i = 1 \text{ to } M \\ \frac{N}{M} + 1 & \text{if } M \nmid N, \text{ for } i = 1 \text{ to } N \bmod M \\ \frac{N}{M} & \text{if } M \nmid N, \text{ for } i = (N \bmod M) + 1 \text{ to } M \end{cases} \quad (5)$$

With such setting, Algorithm 1 will allocate 1^{st} pair to 1^{st} schedule, 2^{nd} pair to 2^{nd} schedule and so on upto M^{th} schedule, repeatedly for all pairs. So, the number of pairs in each schedule is atleast $\lfloor \frac{N}{M} \rfloor$. The number of tasks (pair-wise) allocated to each schedule is given in Eq. (5) where \bmod finds the remainder.

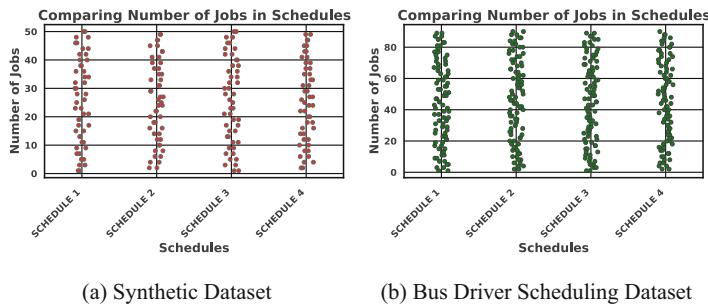


Fig. 3. Number of Tasks in Each Schedule with Synthetic and BDD Datasets

5.4 Comparative Study: UCSMPP vs OFFPSP

Our proposed mechanism, UCSMPP, is compared with an existing mechanism called OFFPSP on the basis of **Average Cost** and **Number of Tasks** in each schedule.

When compared with the OFFPSP, UCSMPP shows better balancing effect in terms of both average cost (Fig. 4) and number of tasks (Fig. 5) of each schedule. In each of the Fig. 4 and Fig. 5, we can see that in UCSMPP, the average cost and number of tasks of each schedule is almost similar. Whereas in OFFPSP, the average cost and number of tasks of each schedule varies significantly.

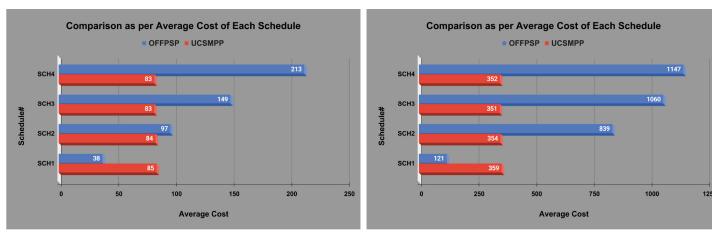


Fig. 4. Comparison Between UCSMPP and OFFPSP on Average Cost of Each Scheduling Solution in NS2

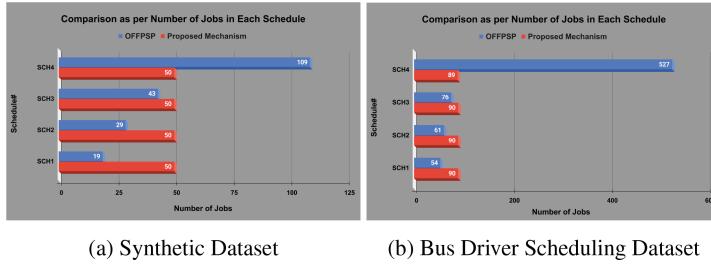


Fig. 5. Comparison Between UCSMPP and OFFPSP on Number of Jobs in Each Schedule using Synthetic and BDD Datasets

6 Conclusion and Future Works

We claim that our proposed mechanism produces balanced schedules in terms of: *i*) total cost, *ii*) average cost, and *iii*) number of tasks in each schedule. The deadline of the tasks in each schedule is obtained by equally dividing \mathbb{R} . The schedules are published to the TEs in *heaviest schedule first* order so that the TEs will be able to perform with less load as they proceed with the schedules and the task provider will get enough time for verification of submitted tasks of heavier schedule. Through our proposed mechanism we have achieved, *i*) **Choice Reduction:** the choices for submitting tasks are reduced for TEs by equally dividing \mathbb{R} into schedules, *ii*) **Balancing Effect:** all the schedules are balanced in terms of total cost and number of tasks, and *iii*) **Benefit of Task Provider:** task provider is benefited as all the schedules are published in reverse order of their total cost. Through rigorous simulation with synthetic and real datasets we have shown that our proposed mechanism performs better than the existing one in terms of balancing effect.

The possible future directive could be to use randomised mechanism to allocate tasks into schedules. Giving incentives to the TEs along with choice reduction could be another direction to prevent procrastination.

References

1. Akerlof, G.A.: Procrastination and obedience. *Am. Econ. Rev.* **81**(2), 1–19 (1991)
2. Al-muqarm, A.M.A., Hussien, N.A.: Dynamic cost-optimized resources management and task scheduling with deadline constraint for mobile crowd sensing environment. *Int. J. Intell. Eng. Syst.* **16**(3), 201–219 (2023)
3. Alabbadi, A.A., Abulkhair, M.F.: Task-scheduling based on multi-objective particle swarm optimization in spatial crowdsourcing. *J. King Abdulaziz Univ. Comput. Inf. Technol. Sci.* **8**, 45–57 (2019)
4. Alabbadi, A.A., Abulkhair, M.F.: Multi-objective task scheduling optimization in spatial crowdsourcing. *Algorithms* **14**(3), 77 (2021)
5. Constantino, A.A., de Mendonca, C.F.X., de Araujo, S.A., Landa-Silva, D., Calvi, R., dos Santos, A.F.: Solving a large real-world bus driver scheduling problem

- with a multi-assignment based heuristic algorithm. *J. Univ. Comput. Sci.* **23**(5), 479–504 (2017)
- 6. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: *Introduction to Algorithms*. PHI Learning (Originally MIT Press) (2010)
 - 7. Deng, D., Shahabi, C., Zhu, L.: Task matching and scheduling for multiple workers in spatial crowdsourcing. In: *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–10 (2015)
 - 8. Gummidi, S.R.B., Xie, X., Pedersen, T.B.: A survey of spatial crowdsourcing. *ACM Trans. Database Syst. (TODS)* **44**(2), 1–46 (2019)
 - 9. Hao, S., Duan, L.: To save mobile crowdsourcing from cheap-talk: a game theoretic learning approach. *IEEE Trans. Mob. Comput.* **23**(8), 8418–8430 (2024)
 - 10. Howe, J.: The rise of crowdsourcing. *Wired Mag.* **14**(6), 1–4 (2006)
 - 11. Kleinberg, J., Oren, S.: Time-inconsistent planning: a computational problem in behavioral economics. In: *Proceedings of the fifteenth ACM conference on Economics and computation*, pp. 547–564 (2014)
 - 12. Kleinberg, J., Oren, S., Raghavan, M.: Planning with multiple biases. In: *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 567–584 (2017)
 - 13. Kontogianni, A., Alepis, E., Virvou, M., Patsakis, C.: Mobile applications in smart tourism and smart cities based on crowdsourcing. In: *Smart Tourism-The Impact of Artificial Intelligence and Blockchain*. Intelligent Systems Reference Library, vol. 249, pp. 33–52. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-50883-7_3
 - 14. Lashkari, B., Rezazadeh, J., Farahbakhsh, R., Sandrasegaran, K.: Crowdsourcing and sensing for indoor localization in IoT: a review. *IEEE Sens. J.* **19**(7), 2408–2434 (2018)
 - 15. Liu, Y., et al.: FoodNet: toward an optimized food delivery network based on spatial crowdsourcing. *IEEE Trans. Mob. Comput.* **18**(6), 1288–1301 (2018)
 - 16. Mukhopadhyay, J., Singh, V.K., Mukhopadhyay, S., Pal, A.: A balanced dissemination of time constraint tasks in mobile crowdsourcing: a double auction perspective. In: Barolli, L., Takizawa, M., Yoshihisa, T., Amato, F., Ikeda, M. (eds.) *3PGCIC 2020. LNNS*, vol. 158, pp. 74–85. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-61105-7_8
 - 17. Rahman, M., Blackwell, B., Banerjee, N., Saraswat, D.: Smartphone-based hierarchical crowdsourcing for weed identification. *Comput. Electron. Agric.* **113**, 14–23 (2015)
 - 18. Shah, S., Bao, F., Lu, C.-T., Chen, I.-R.: CrowdSafe: crowd sourcing of crime incidents and safe routing on mobile devices. In: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 521–524 (2011)
 - 19. Sun, D., Gao, Y., Yu, D.: Efficient and load balancing strategy for task scheduling in spatial crowdsourcing. In: Song, S., Tong, Y. (eds.) *WAIM 2016. LNCS*, vol. 9998, pp. 161–173. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47121-1_14
 - 20. Tim, R.: CS269I: incentives in computer science lecture# 19: time-inconsistent planning. Stanford Course (2016)
 - 21. Wang, L., Tong, Y., Hu, C., Chen, L., Li, Y.: Procrastination-aware scheduling: a bipartite graph perspective. In: *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 1650–1653. IEEE (2019)

22. Wang, X., Tushar, W., Yuen, C., Zhang, X.: Promoting users' participation in mobile crowdsourcing: a distributed truthful incentive mechanism (DTIM) approach. *IEEE Trans. Veh. Technol.* **69**(5), 5570–5582 (2020)
23. Wang, Z., Li, Y., Zhao, K., Shi, W., Lin, L., Zhao, J.: Worker collaborative group estimation in spatial crowdsourcing. *Neurocomputing* **428**, 385–391 (2021)
24. Zhao, Y., Zheng, K., Li, Y., Han, S., Liu, J., Zhou, X.: Destination-aware task assignment in spatial crowdsourcing: a worker decomposition approach. *IEEE Trans. Knowl. Data Eng.* **32**(12), 2336–2350 (2019)
25. Zhao, Y., et al.: Mobile crowdsourcing quality control method based on four-party evolutionary game in edge cloud environment. *IEEE Trans. Comput. Soc. Syst.* **11**(3), 3652–3666 (2024)
26. Zhou, J., Fan, J., Wang, J.: Task scheduling for mobile edge computing enabled crowd sensing applications. *Int. J. Sens. Netw.* **35**(2), 88–98 (2021)



A Learning Web System for Website Development

Aino Nakamura and Masaki Kohana^(✉)

Faculty of Global Informatics, Chuo University, 1–18 Ichigaya-Tamachi, Shinjuku-ku,
Tokyo 162–0843, Japan
a21.pb4y@chuou-u.ac.jp, kohana@tamacc.chuo-u.ac.jp and

Abstract. Some schools teach web page development as part of their programming education. However, there are many problems, such as the lack of teachers and not all schools are able to implement the programming education. This study develops a web system that consists of the coding editor and the page viewer for purposes of learning elementary web page development. This system also provides the learning contents for HTML, CSS, and JavaScript in the form of conversational text. A user can learn the contents and immediately develop web pages based on what they have learned. Using this system, students can learn web page development without setting up a development environment and needing help from teachers.

1 Introduction

In the new Guidelines for the Course of Study for Senior High Schools to be implemented from the 2022 academic year, “Computers and Programming” is included in “Information I,” a common required subject within the Information Science course at senior high schools. The subjects of Information Science have been reorganized, and “Information I,” which all students are required to take, has been newly established, making contents such as programming and the basics of networks and databases (data utilization) mandatory. In addition to “Information I,” an elective “Information II” has been established. Based on the contents learned in “Information I”, this course aims to develop the ability to use information systems and various data appropriately and effectively, as well as the ability to create content[1]. As a result, all high school students will be introduced to programming from the 2022 academic year, and students who choose “Information II” will learn more advanced programming.

In addition, from the 2021 school year, the new Courses of Study have been fully implemented in junior high schools nationwide, and programming education has become mandatory[2]. In “information technology”, which is included in the “technology” classes, instruction is given to the design and production of digital works, including “knowing the characteristics and usage of the media and being able to design productions.

Programming in elementary schools has been made mandatory since the 2020 school year. However, the goal of elementary school programming is not to

acquire the language by actually writing code, but rather to educate students in “Computational Thinking”. The interest in programming in educational institutions has clearly increased in recent years, as evidenced by the “mandatory learning” [3]. The purpose of making programming mandatory is to cultivate the qualities and abilities needed to participate proactively in the information society, and to make ICT more accessible as a clue to the development of human resources in the face of rapid ICT technology development.

In this way, learning programming is becoming increasingly important, but several problems have arisen in current programming education. Typical examples are the lack of teachers and the difficulty in setting up a development environment. For example, while it is sufficient to have a text editor and a browser environment to use HTML, the pre-installed text editor is inconvenient. Even if Atom or Visual Studio Code is installed, it is not possible to start writing immediately. Beginning students will probably have to start by learning the intricacies of using developer software. In this study, we built a system that allows users to learn website creation without the need to install a development environment, which can be difficult for beginning students, by using conversational content. Simply by eliminating the need to install a development environment, one hurdle to implementing coding for beginning students can be removed. This system is being studied to help middle and high school students understand contents creation and information practice conducted by them, while also including an approach to the problem of teacher shortage.

2 Related Research

This section refers to existing programming learning services and previous studies on e-learning. Koganey (2005) focuses on the importance of output in programming learning and describes the construction of a practical training environment on a website[5]. He states that the restrictions of time and place for practical training are eliminated, allowing for free coding. The only requirement that the hardware used by the learner must meet is that he/she can use a Web browser, and the study of implementation is based on three issues: system maintenance, efficient allocation of computing resources, and appropriate management of sessions. Yaegashi et al. (2005) developed an interactive e-learning system to solve the problem of maintaining and motivating students to learn[4]. It has been confirmed that the ability to see other learners’ feedback on the course content and to feel their presence provides a high level of satisfaction with the course. In addition, Ide (2022) showed that more students found the classes using programming learning services to be “fun” and “easy” in terms of website creation than those using lecture-based classes[6]. It is inferred that having students learn the contents dealing with code through the service as e-learning, rather than through materials and verbal explanations as in other subjects, may have made them want to learn the contents and enjoy the process of learning[7].

3 Purpose of Development

The purpose of developing this system is twofold:

- To provide middle and high school students with coding knowledge, and to have them enjoy coding by eliminating barriers to coding.
- To contribute to information classes, where there is a shortage of teachers, in the form of educational materials that do not require an instructor.

Many students want to learn programming but are puzzled by the complexity of the language. For such beginners, the structure of web page creation, where they can easily see the code they have written visually, makes it easy to sustain their motivation. Those who feel barriers to coding can discover anew the joy of coding. Today's middle and high school students have more opportunities to get involved in programming and coding than ever before. To prevent them from falling behind quickly, we have developed a learning system that allows them to understand, through learning programming, what languages they can now learn in the system and where they are actually used. Additionally, we included basic explanations of things to know before creating a web page, going beyond just grammar and other fundamental aspects. In addition, the learning system was designed to include as few technical terms and esoteric code as possible.

4 Questionnaire for Middle and High School Students

In May 2023, three of the four schools affiliated with Chuo University (1,010 respondents) were surveyed regarding web page production and programming learning. The purpose was to determine how much knowledge middle and high school students actually have about learning to code, whether they are interested in learning to code, and in which areas there is demand for students who are willing to learn.

Table 1 shows how many students were interested in programming. Fifty-two percent of the students responded “not interested” and 48% responded “interested”, indicating that only about half of all students are interested in programming.

Table 1. Answer to “Are you interested in programming?”

Choices	Responses
Never heard	12
Have heard but not interested	516
Have heard and are interested	447
Do private programming	35

Table 2. Answer to “Are you interested in website development?”

Choices	Responses
Not interested	360
Interested but have never thought about creating	380
Interested and would like to create, but don't know how	225
Interested and have create websites	45

Table 3. Answer to “Have you ever heard of the “HTML” language used in website development?”

Choices	Responses
Never heard	652
Have heard	265
Have heard and used	93

Table 4. If you answered “Interested in website creation, but have never thought about creating” or “Interested and would like to make, but don't know how”, why? (Multiple answers are possible.)

Choices	Responses
Don't know how to learn	293
Seems difficult	438
No one around me is doing	192
Busy with club activities or studying	275
I'll get bored in the middle of the course	164
(Free answer) Don't have a clear purpose for making	7
Don't have the environment to do	2
Can't use a PC well	1

Table 2 set up a question item regarding interest in web page production. Here, 36% of the students responded that they are not interested, while the remaining 67% responded that they are interested. The difference between interest in programming and website development is due to the presence of students who are “not interested in programming but interested in website development. There were 218 students who were not interested in programming but were interested in website design (while 64 students were not interested in website design but were interested in programming). While we guessed that many students are interested in website development, as shown in Table 3, we also found that many students do not know the languages necessary for website development.

Next, Table 4 asks the reasons for not (or can't) create a website. The most common response was “no clear reason to create”, followed by “don't have the

environment to do. The most common response was “It looks too difficult”, indicating that many students feel barriers not only to website creation but also to coding. This was followed by “don’t know how to learn” and “busy with club activities and studying. Overall, it was found that many students are slightly more interested in website creation than in programming, and that the reasons for not getting involved include concerns about the complexity of coding, not being equipped with a computer runtime environment, and the idea that they do not see a purpose for it. Based on these results, We intend to develop this system with the following aims in mind.

- Give them with a sense of purpose
- Teach how to create a website as the first step in coding
- Provide easy-to-understand explanations for beginners without including difficult content

5 Implementation of this System

5.1 Conversation-Based Learning Content

This section describes the implementation of this system. Figure 1 shows an interaction diagram of this system and users.

In setting up the learning content, we prepared conversational content with reference to an existing programming learning application. Figure 2 and Fig. 3 show the editor and preview classes, respectively. In the “editor” class, the text for HTML, CSS, and JavaScript is shown in the top right corner to indicate which language is currently in use. When JavaScript is active, it appears in the top right corner.

5.2 Implementation of the Editor Part

5.2.1 Editor with CodeMirror This system allows the learned code to be executed within the system. This helps the learner skip building the execution environment. We implemented our editor using CodeMirror, a library that enables code editors in the browser. It provides syntax highlighting, line numbering, code folding, and other features like most code editors. Then, its UI is so simple and prevents user confusion.

The editor is separated into three parts, HTML, CSS, and JavaScript. When users edit codes, the codes are immediately displayed on the preview screen.

Figure 4 shows an overview of combining HTML, CSS, and JavaScript codes. The code entered into each editor is retrieved as a string and inserted into the preview’s ‘doc’ property. Then, it’s encoded as an HTML element. This document inserts the retrieved HTML into the ‘body’ tag, the CSS into the ‘style’ tag, and the JavaScript into the ‘script’ tag. This combined code is like creating a web page using only HTML. It is then used to display the HTML on the preview screen. Figure 5 shows the JavaScript code.

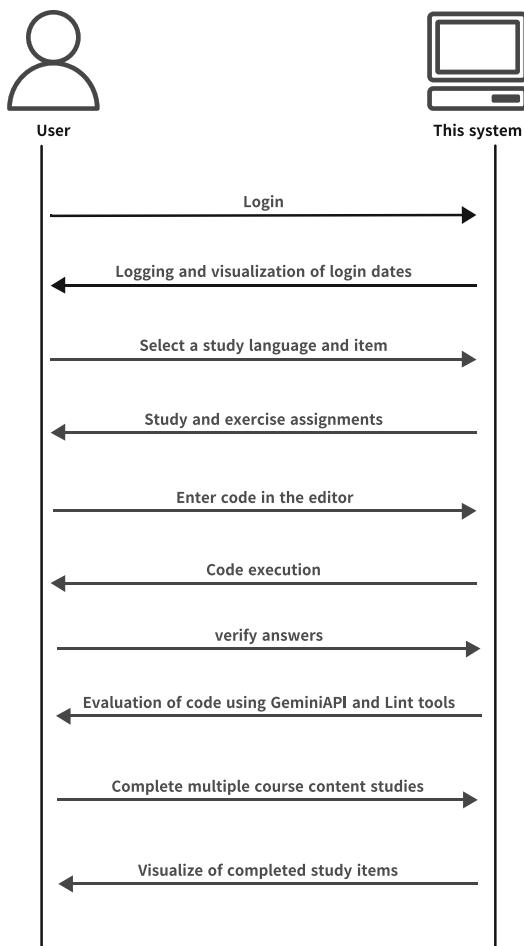


Fig. 1. Interaction diagram between the user and this system after registration

Users can view exercises and see the output of their code. Users can go back and forth between the exercise page and the lecture page with a single button. Even if users forget the content, he/she can check it immediately. Users-written code can be saved to their own PC as an HTML file by clicking the “Save” button. Furthermore, users can visually check how it will be displayed in an actual browser.

**Fig. 2.** Editor Class

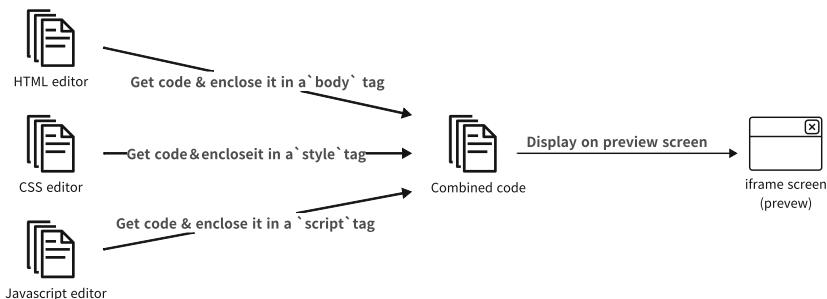
The screenshot shows a user interface for previewing web pages. On the left, there is a circular icon of a teacher figure with the word "teacher" below it. At the top, there is a "PREVIEW" button. Below it, the content of the preview is displayed:

Hello,World!

Let's challenge yourself to create web pages!

Learn the basics of HTML, CSS, and simple Javascript to complete your own web page.

[Click here to learn HTML](#)

Fig. 3. Preview Class**Fig. 4.** How Codes Appear on the Preview Screen

5.2.2 Use of Images

In the image selection field below the code entry field, users can select an image file or drag and drop an image file to register it in the system. This allows users to use any image in the exercise page's editor using the 'img' tag. Figure 6 shows an overview of the image file processing. The name, format, and URI of the selected image are registered in local storage. Users could write the URI directly into the editor to display the image. However, URI are very long strings. It's inconvenient

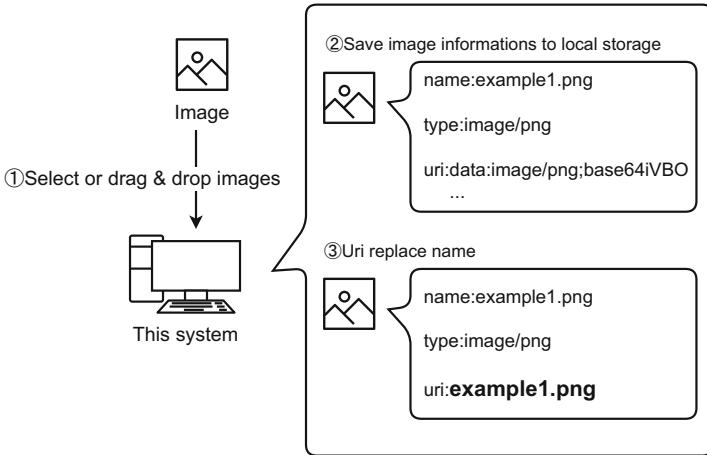
```
1 function getCode() {
2     const html = htmlView.state.doc.toString();
3     const css = cssView.state.doc.toString();
4     const javascript= javascriptView.state.doc.toString();
5
6     return {
7         html, css, javascript,
8         doc: '<!DOCTYPE html>
9             <html>
10                 <head>
11                     <meta charset="UTF-8">
12                     <style>${css}</style>
13                 </head>
14                 <body>${html}
15                     <script>${[javascript]}</scr$['ipt']> //escaping
16                 </body>
17             </html>,
18     };
19 }
```

Fig. 5. Combining Each Codes

for them to type URI in manually. To solve this problem, the image's name is encoded in base64 and replaced with the URI. Users simply enter the name of the image in the src attribute value and the image is displayed in the editor. Users are able to use images like a editor software. When images are deleted by pressing the X button in the upper right corner of the image selected in the image selection field, the information of that image is erased from this system. This prevents the image information from remaining in local storage. Also, it ensures that images not displayed in the image selection field are unavailable for use.

5.2.3 Code Evaluation Function Using GeminiAPI

Users can have their code evaluated by GeminiAPI by clicking the “Evaluate” button. Each exercise page has separate assignments to make users output what they have learned in each section. The console is shown in Fig. 7. Users are asked to scrutinize whether they have faithfully followed the assignments (e.g., whether they have used only one h1 tag instead of two or three, as instructed to use “one” h1 tag, etc.). If correct, the system displays, ‘You are correct! Please move on to the next study.’ If incorrect, it says, ‘Let’s try again! Check to make sure you are doing what the exercise instructs you to do, and that you are writing the tags correctly.’

**Fig. 6.** Image Insertion and Display

2.[Exercise page]Let's write a sentence!
『p-tag and h-tag』

Piyopiyo
I'll give you some chick trivia.

In the HTML file, type "Piyopiyo" using the h2 tag.

In the HTML file, type "I'll give you some chick trivia" using the p tag.

Writing code that is not directed or doesn't follow the assignment

Try again! Check to see if it matches what the exercise instructs you to do, and make sure you are writing the tags correctly.

HTML

```
1 <!DOCTYPE html>
2 <html><head></head>
3 <h1>Piyopiyo</h1>
4 <p>I give you some chick trivia.</p>
```

CSS

Fig. 7. Response of GeminiAPI

5.2.4 Syntax Checking of Code Using Lint Tools

GeminiAPI can determine if users are writing code according to the task. However, it's difficult for GeminiAPI to detect common beginner mistakes, such as incorrect use of the/(slash) in closing tags. Instead of this, lint tools is used for detailed syntax checks of the code. An overview is shown in Fig. 8. For HTML, CSS, and JavaScript, we decided to use the Lint tools HTMLHint, CSS Tree Validator, and JSLint, respectively. CodeMirror, which is used as the editor for the exercise pages, has a feature called Linting that draws a red wavy line anywhere in the code and pops up a message. By using this feature, errors and warnings in the code can be visually highlighted. HTMLHint returns error messages in the format “there is an error in the Xth column of the Xth line,” while CodeMirror’s Linter function highlights errors in the format of “display (the error) from

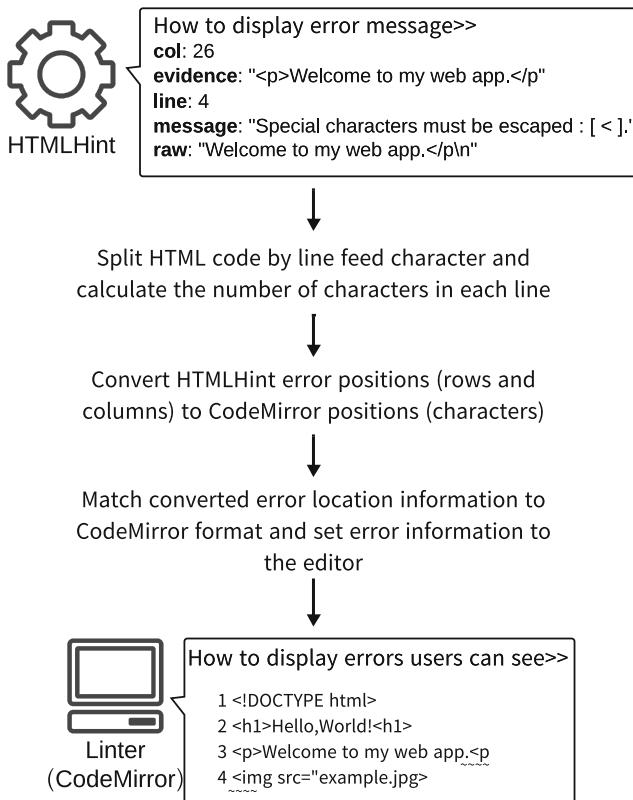


Fig. 8. Interaction Diagram of Lint Tool and CodeMirror

the Xth character to the Xth character. Since the HTMLHint test results are returned as error information including line and column numbers, the HTML code is divided into lines, and the number of characters in each line is calculated. The total number of characters up to the error line is then calculated, and the start and end positions of the error are calculated on a character count basis, so that a wavy line to be drawn where the error actually occurred, indicating that it is a mistake. By calculating the total number of characters up to the line where the error occurred and identifying the position by adding the column number within that line, the HTMLHint error position (line and column) is converted to a CodeMirror position (character count).

We have implemented two features to encourage continuous learning, which was one of our goals. When the calendar is displayed on the home screen and the current date is indicated on the calendar, a “today” class is assigned to that date. Furthermore, since the login date is recorded in a “loginDateArray”, the corresponding date is assigned a “login-date” class. This visually highlights the date the user logged in and the date of the day so that it can be easily

identified on the calendar. The three circles below the calendar are completed by the learning status of each content. Data on access to each lecture page will be acquired. Calculate the progress rate for each course based on the acquired data and update the style of each progress bar. The progress rate for each course is calculated by dividing the number of applicable courses completed by the total number of courses and converting it to a percentage. SQLite is used as the database for both login records and progress management.

6 Summarize

We are developing a website creation learning system for beginning students, allowing them to create a website from input to output within the system, without the need for an instructor or an execution environment. The editor can output what users have learned. Users can check how the code entered by them is displayed on the preview screen. The system uses a database to visualize learning status and dates of use. The record of learning is effective in keeping users' learning. This system features member registration, learning contents, progress visualization, and a code editor. However, we still need to improve in several ways. For example, making it possible to check GeminiAPI replies in the browser and providing content that gives a sense of purpose. In the future, We plan to implement a friend function to differentiate the system from existing programming learning applications and to enhance the sense of purpose.

References

1. Ministry of Education, Culture, Sports, Science and Technology: Courses of Study for Senior High Schools (2018). https://www.mext.go.jp/content/20220324-mxt-kouhou02-000021499_1.pdf
2. Ministry of Education, Culture, Sports, Science and Technology: Key Points for Revision of Courses of Study for Senior High Schools (2018). https://www.mext.go.jp/content/1421692_2.pdf
3. Ministry of Education, Culture, Sports, Science and Technology: Courses of Study for Junior High Schools: The Power to Live (2021). <https://www.mext.go.jp/a-menu/shotou/new-cs/youryou/chu/gika.htm>
4. Yaegashi, F., Kitamura, S., Hisamatsu, S., Sakai, T., Mochizuki, T., Yamauchi, Y.: iPlayer: development and evaluation of an interactive streaming player for e-learning (2005). t29_KJ00004286880.pdf
5. Koganeya, H.: Realization of a programming training environment on a web site (2005). https://www.jstage.jst.go.jp/article/konpyutariyoukyouiku/18/0/18_121/-pdf/-char/ja
6. Ide, H.: Practice of programming lessons using programming education service in high schools and verification of educational effectiveness (2022). https://www.jstage.jst.go.jp/article/jaeis/14/1/14_39_.pdf/-char/ja
7. Payton, J., et al.: The positive impact of social and emotional learning for kindergarten to eighth-grade students: findings from three scientific reviews. Technical Report (2008). <https://eric.ed.gov/?id=ED505370>



A Community Web System for LGBTQ+ Students with Identification

Miyu Sato and Masaki Kohana^(✉)

Faculty of Global Informatics, Chuo University, 1-18 Ichigaya-Tamachi, Shinjuku-ku,
Tokyo 162-0843, Japan

a21.7nca@g.chuo-u.ac.jp, kohana@tamacc.chuo-u.ac.jp

Abstract. In recent years, there has been a growing global trend toward recognizing diversity. Although this trend can be seen in Japan, it is not yet well understood. As one of these, the sexual minorities area has some issues. There are a number of communities and services for LGBTQ+, but these do not sufficiently cover young people. There are also many issues with the services available to minors, such as slander, identity theft, and use for the purpose of dating. This study develops a community site for young people to provide a place where they can communicate with persons concerned. Furthermore, an identification verification will be introduced to prevent the participation of users with inappropriate purposes, with the aim of creating a system that can be used with peace of mind by the concerned persons.

1 Introduction

In recent years, there has been a growing trend to recognize diversity around the world. The Sustainable Development Goals (SDGs), adopted at the UN Summit in 2015, set 17 goals to achieve a sustainable and better world by 2030. These goals include initiatives related to diversity, such as “GOOD HEALTH AND WELL-BEING” “GENDER EQUALITY” and “REDUCED INEQUALITIES”. This shows that diversity is attracting global attention.

Chuo University (Japan) has a “Diversity Center” to promote diversity in three areas: “Disability Area” concerning disabilities and diseases, “Gender and Sexuality Area” concerning gender and sexuality, and “Global Area” concerning multicultural coexistence [1]. This center has experts in various fields on staff who can provide individual advice and support. Moreover, the center also provides a place for educational activities.

However, many issues still remain and have not been resolved. As one of these, the sexual minorities area has some issues in Japan. Sexual minorities are mainly categorized into “Sexual Orientation” and “Gender Identity”. According to Ministry of Health, Labour and Welfare (2019), “Sexual Orientation” means which gender a person’s love or sexuality targets. Some people like a different gender from themselves, some like the same gender as themselves, some like that person regardless of the other person’s gender, and some have no romantic or

sexual feelings for anyone [2]. Also, according to Ministry of Health, Labour and Welfare (2019), “Gender Identity” refers to the perception of one’s own gender. There are many combinations of the two. In Thailand, which is known as a “leading LGBTQ+ country,” there are 18 different genders.

There are various ways to refer to sexual minorities, such as “LGBT” and “LGBTQIA”. In this paper, “LGBTQ+” is used to refer to sexual minorities. “Ally” indicates a person who is not a LGBTQ+ person, but who understands and supports them. There are some cases where even Ally is having trouble because they do not know how to interact with LGBTQ+ person.

Japan still lacks understanding of the sexual minorities. Educational institutions have begun to educate students about diversity, but have not solved specific problems such as changing clothes and toilets. Furthermore, there is “Outing”, which is the exposure of a person’s sexuality to a third party without his or her permission, as well as discrimination and bullying regarding LGBTQ+. Therefore, there are many issues that lead to “Come out”. There are some cases where even Ally is having trouble because they do not know how to interact with LGBTQ+ people.

This research develops a community website that will serve as a place for consultation and exchange targeting LGBTQ+ and Ally, especially young people who face such problems daily. On the other hand, many social media do not encourage the use of social media by those under the age of 13. This is to prevent children who are unable to make appropriate decisions from getting into trouble. For example, the U.S. has the Children’s Online Privacy Protection Act (COPPA), which requires parental consent for the collection of personal information from children under the age of 13 [3]. This law has had a significant impact on age restrictions on social media. Considering this situation, the system will be available for registration to “students” aged from 13 to 22 years old who are positioned as LGBTQ+ or Ally. Through this system, we would like to provide an opportunity for LGBTQ+ people to eventually be able to live their lives in a more personal way by sharing advice and concerns with each other.

2 LGBTQ+ Trends

This section focuses on trends regarding LGBTQ+ around the world. In the United States, Congress passed the Respect for Marriage Act in 2022 [4]. In the U.S., where case law is the rule, the Supreme Court previously decided whether same-sex marriages would be recognized. However, this bill has clearly legalized it through legislation. The Netherlands was the first country in the world to legalize same-sex marriage in 2001, and at the same time legalized adoption rights for same-sex couples.

Next, this section focuses on trends in Japan. In Japan, the LGBT Understanding Promotion Act went into effect in June 2023, and interest in diversity has increased. However, even today, same-sex marriage is not recognized. Instead, the partnership systems are spreading in the various municipalities. Under this system, new services, such as applying to move into metropolitan housing, will

be available [5]. However, unlike marriage, it is not a legal act, and thus has no legal effect. As a result, the person is not considered a “spouse”, and problems have arisen, such as not being able to inherit. In terms of education, the elementary school guidelines established by the Ministry of Education, Culture, Sports, Science and Technology (2017) do not include “sexual diversity” or “LGBTQ+” [6]. However, based on social trends, elementary school textbooks to be used from 2024 will include significantly more content on “sexual diversity. On the other hand, since the guidelines are revised every 10 years, it is expected that “sexual diversity” and “LGBTQ+” will not be included in the guidelines until around 2028.

In this way, Japan overall lags behind other developed countries, and many problems remain. A recognition survey towards LGBTQ+ people by Hidaka (2016) reports the results of an analysis of 15,064 cases of people living in Japan[7]. This report states that approximately 60% of all respondents experienced “bullying” in their school life (primary, junior high and high school). It also states that 63.8% of them were victims of verbal bullying and 18.3% of them were victims of being stripped of their clothes. Furthermore, only 13.6% of respondents answered “helpful” when asked if their teachers were helpful in resolving bullying when they were being bullied. This is believed to be due to inadequate support systems in schools[8]. It can be seen that even teachers in educational institutions do not always play an adequate role in supporting LGBTQ+ students. In this way, while the educational environment has improved, the situation of dealing with bullying against them remains challenging. There are still many years to improve the educational environment regarding sexual diversity, including the improvement of the school guidelines. This means that they must be protected in an inadequate educational environment for at least several years. For these backgrounds, the system now provides for LGBTQ+ and Ally only.

3 Related Research

In conducting this study, there are many existing social media sites that target LGBTQ+ people as references. However, most of them are for the purpose of dating and there are those where the target age is over 18 years old. Even those that are not, many of them are used for dating purposes or used for sexual crimes as a result. It is heard that, there was a social media platform that was “for students only” in the past. However, there were many users who were not students, and it became a hotbed of sexual crimes, so this platform has ended. Based on these considerations, this study will cover young people who have not been previously targeted, while at the same time setting an upper age limit for use and implementing age verification. By doing so, it will prevent the site from becoming a breeding ground for dating and sexual crimes, and ensure that people use the site safely and securely in a manner consistent with the purposes of this study.

4 System Overview

This system adopts anonymity and consists of Profile, Posting, Consultation, Discussion, and Friend functions.

The profile function allows users to check their own icons, self-introductions, past posts, etc. The posting function allows users to freely post what they want to say and to view the posts of other users they are following (“friends”). Users can freely comment on their own posts and those of their friends. The consultation function provides a place where users can consult with each other, including other users who are not their friends. The discussion function provides a place for all users to exchange opinions on LGBTQ+ topics raised by the administrator of this system, divided arbitrarily into agree and disagree groups. Finally, the friend function provides users with the ability to follow or block certain other users at their discretion. The system introduces identity verification using student IDs to eliminate impersonation, harassment, and sexual misconduct, and to restrict users to LGBTQ+ and Ally student.

5 Function Classification According to Authentication Status

This system implements voluntary identity verification to ensure safety and reliability. To maintain a secure site, available functions will be divided according to each user’s identity verification status. The correspondence is shown in Table 1. In addition, accounts that have verified their identity will be given badges to make them visible to other users.

- **Status 1:** The user has not verified the identity or not used an e-mail address provided by the educational institution.
- **Status 2:** The user has registered using an e-mail address provided by the educational institution.
- **Status 3:** Identity verification has been completed using the student ID card.

Table 1. Corresponding Relationship between Identification Methods and Functional Limitations

	Self-Declaration (Status: 1)	School Email Address (Status: 2)	Student ID + School Email Address (Status: 3)
Available Time	6:00–23:00	6:00–23:00	No Restriction
Posting	△ (View Only)	△ (View Only)	○
Consultation	×	△ (View Only)	○
Discussion	×	×	○

6 System Implementation

6.1 User Registration

User registration requires a unique ID, password, e-mail address, and date of birth. E-mail address and birth date are used for age restriction and to determine identification status. If there is a registered e-mail address provided by an educational institution (e.g. “ac.jp”, “ed.jp”), the identification status is set to “2” and the information is recorded in the database. For other e-mail addresses, it is determined to be in status “1”.

6.2 Identification

The system will verify the identity of the user to prevent the participation of users with impersonation or sexual crime purposes. The system uses a “student ID card” since the age range for registration is set as “student” between 13 and 22 years old. Authentication is carried out by comparing the birth date extracted from the student ID card with the birth date entered at the time of member registration, and by checking whether the name of the school is real or not. The authentication is basically performed automatically, and if it fails, the authentication is performed manually.

Figure 1 shows an overview of the identification process. A user sends an image that is the student ID card. The system receives the file, and converts it to a PNG file. Then, the system retrieves the birth date and school name. The retrieved birth date is checked against one entered at the time of user registration. After that, search the database for the existence of the retrieved school name. Finally, the users authentication status is changed to “3”.

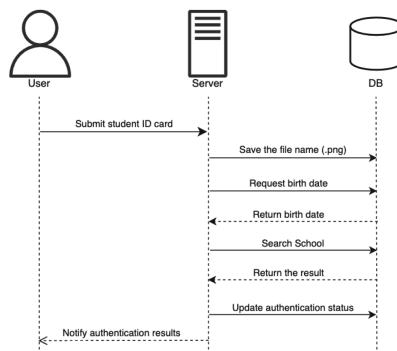


Fig. 1. Interaction diagram at Identification process

6.2.1 Capture and Submit Image of Student ID Card In order to verify a user's identity, an image of the student ID card must be submitted. From the Settings page, users can go to the identification page, which will launch the camera if using a smartphone, or display the Explorer if using a PC, and select a file to submit. This page will not be displayed in the case of authentication status 3. The “capture” attribute is used to take a picture, and the attribute value is set to “environment” to activate the outside camera.

6.2.2 File Format Conversion The uploaded file is converted to PNG file using the sharp module. At the same time, the file names are stored in the database. Then, the image file before conversion is deleted and the image path is defined for text extraction of the converted file.

6.2.3 Google Cloud Vision API The Google Cloud Vision API (“Vision API”) is an image recognition service provided by the Google Cloud Platform [9]. Vision API provides image labeling, face and landmark detection, optical character recognition (OCR), and tagging of inappropriate content, and so on. There is also the possibility of using a text search engine to find text that is not appropriate for the site. After that, the detected text is extracted, new-lined and concatenated into a single string.

6.2.4 Filtering To verify that the entered birth date is correct and to confirm that the user is a student, the system retrieves the birth date and school name from the student ID card. At first, the birth date is extracted. Find the year, month, and day of the Japanese or Western calendar that follows the string “birth date” in the string. If the matched birth date information is in Japanese calendar format, it is converted to a Western calendar format. Then, the extracted portions are stored in the respective variables “year”, “month”, and “day”. Next, the name of the school is extracted. In this case, “university,” “vocational school”, and so on are matched to the line where there is a string. The matched strings are returned, and the next process proceeds.

6.2.5 Check Against Database (Request Birth Date, Search School, Update Identification Status) The extracted birth date information is compared with the birth date information in the database. If a match is found, proceed to school name verification. In this case, the list of school names published by the Ministry of Education, Culture, Sports, Science and Technology was obtained, and the educational institutions with secondary education and above were registered in the database in advance [10]. The information in the database is checked to determine if there are any schools that match the extracted school names. If there is a school with a match, the user's authentication status is changed to “3”, the image file used is deleted, and the identification process is completed.

6.2.6 Processing When Authentication Fails In the following cases, an authentication failure will be registered in the database.

- Something other than a student ID was submitted.
- Text regarding birth date or school name or both was not extracted.
- The extracted birth date did not match the registered one.
- The name of the extracted school did not actually exist.

If the authentication fails, the uid, image file name, problem summary, time of occurrence, and response status are registered in the database. Then, the process moves to a manual authentication process by the administrator. During this time, the identification page disappears from the user's screen, and the user can continue to use other functions.

Identification		
Issue List		
UID	Issue Overview	Time
mippi	Date of birth was not extracted	2024/07/20/03:42:04

Fig. 2. Problem List Page

The screenshot shows a web page titled "identification". Below it is a section titled "Rejection / Notification". It displays a message: "User name : mippi" and "reason for rejection : The extracted birth date did not match the registered one." At the bottom of this section is a button labeled "Send Notification".

Fig. 3. Create notification page

6.2.7 Get a List of Issues Figure 2 shows a list of problems that occurred during the identification process. After logging in at the login page for the administrator, the administrator can check the list (in ascending order). To manually verify identity, click on the uid of the target user.

6.2.8 Manual Authentication Clicking on the uid will display the birth date information entered by the target user at the time of new member registration and the image of the submitted student ID card. There is also a field to search for the name of the school, which allows the administrator to enter the full name of the school on the student ID card to search for the school name in the database and confirm its existence. After confirming the date of birth and the name of the school, the authentication is completed by pressing the "Authenticate" button and changing the authentication status to "3". On the other hand, if the authentication is not accepted, press the "Reject" button.

6.2.9 Notification to the Use Figure 3 shows the page when creating the notification to the user. Selecting the “Reject” button takes this page. By selecting the reason for rejection from the options and clicking the “Send Notification” button, a template notification is sent to the user as an HTML file according to the reason for rejection. At the same time, the contents of the sent notice are recorded in the database.

Your identity verification was not approved.

To : mippi
 From : Administrative Staff
 Date : 2024/07/16 09:30

 Thank you for always using our services.
 Your identity verification was not approved for the following reason.
 The date of birth on your student ID card does not match the one you registered.
 Therefore, we request that you take the following actions
 Register again with the correct date of birth.

Fig. 4. Example of the notice

6.2.10 Display Notifications on the User Side Figure 4 shows an example of a notice delivered to the user. On the user page, a notice from the administrator will appear in the system’s notices section, and the user will be able to access the identity verification page again.

6.3 Profile

The profile page displays the user name, icon image, background image. These images can be edited. This page also displays their sexuality and self introduction. These can also be edited, and the sexuality can be hidden.

6.4 Posting

The posting page shows the posts from the user and the friends in descending order. To post a message, it can be entered by the user on the posting page, with a minimum of 1 character and a maximum of 200 characters. Then, the submission is completed by pressing the “Submit” button at the bottom of the page. When a post is completed, post ID (automatically assigned), uid, contents of the post, and posting time are recorded in the “timeline table” in the database, and the view page is updated. To comment on a post, the user can press the “Balloon mark” button that appears when the target post is displayed, which allows the user to enter text in the same way as when posting the post. After that, the comment can be posted by pressing the “Post” button at the bottom of the page. On the server side, uid, post content, posting time, and post ID of the comment destination are recorded in the “timeline table” in the database.

6.5 Consultation

When the consultation page is accessed, past consultations of all users are pulled from the “consultation table” in the database and displayed on the page in descending order. To consult, it can be entered by the user on the consultation page with a minimum of 1 character and a maximum of 400 characters. Then, the submission of the consultation is completed by clicking the “Submit” button at the bottom of the page. When the posting is completed, post ID, uid, contents of the posting, time of posting, and role (questioner) are recorded in the “consultation table” in the database, and the browse page is updated. In addition, the consultation function allows users to optionally select “Post Anonymously” for each post to make it easier for users to post their conversations. By selecting “Post Anonymously” when posting, the question will be made public to all users, but the user name, icon, and other profile information will be kept private. This selection is saved to local storage.

To answer the consultation, the “Answer” button allows the user to move to the answer page. The same way as when posting a consultation, between 1 and 400 characters can be entered. Then, the answer is completed by pressing the “Submit” button at the bottom of the page. On the server side, the uid, contents of the post, posting time, role (respondent), and post ID of the commenter are recorded in the “consultation table” in the database.

6.6 Friend

Users are free to follow or block other users through the posting or consultation functions (except for anonymous posters in the consultation function). When a user blocks another user, the blocked user’s post will no longer be visible to the user. Additionally, the blocked user will not be able to see the user’s posts and profile. To follow or block, first click on the icon of the contributor of either of the two functions and go to the profile page of the user. Figure 5 shows an image of the user’s profile page. On the profile page, click the “Follow” button in the center right to complete the follow. When pressed again, the follow status will be canceled. To block, click on the “—” bar in the upper right corner.

When these actions are performed, the server side records the uid and follow status flags of the actor and the recipient of the action in the “friendship table” in the database. The flags correspond as follows.

1. The actor is following the recipient of the action.
2. The actor is blocking the recipient of the action.

If a user is unfollowed, the corresponding record is deleted. Users’ follow and block status can be checked from the friend page of settings.



Fig. 5. Profile Page of other users

7 Challenges and Future Prospects

There are four issues in this system. The first issue is the content problem and copyright infringement. At this time, the system does not have a function to upload images and videos for situations other than identity verification. However, there is a possibility that this will be implemented in the future. In that case, there is a risk that age-restricted content may be uploaded, so it is necessary to consider some kind of mechanism to automatically block such contents. Similarly, there is a risk of copyright infringement if uploaded, so appropriate measures should be taken.

Next, responses to problems must also be considered. Although the implementation of identification can prevent to some extent the use of the site for impersonation and dating purposes, it is difficult to prevent it completely. In addition, there is no mechanism to automatically exclude people who are not LGBTQ+ and Ally, so there is the potential for outing, slandering, and trolling. Furthermore, there is a risk of leakage of personal information by users themselves or other users. Therefore, there is a need to put in place a mechanism to respond in the way of alerting and suspending accounts as necessary, using a system of reporting.

Next, the formulation of a privacy policy and terms of use must be considered. There is a need to establish both to prevent problems. When creating this document, there are plans to refer to existing social media. In addition, all risks that may occur at the site will be assumed in advance, and countermeasures will be considered accordingly. To make it easier to respond to problems and reduce the risk of the site itself being blamed in the event of trouble, there is a policy of creating a detailed and strict privacy policy and terms of use.

Finally, improving the accuracy of the identity verification function is a challenge. The implemented function allows for authentication as long as the name of the school and date of birth information is provided. This makes it possible to authenticate with forged student ID. Some technical measures are needed to prevent it.

8 Conclusion

This study develops a community website that provides a site for consultation and interaction for students who are sexual minorities, especially LGBTQ+ and Ally who understand and support them. In Japan, diversity is still not as well understood compared to other countries. Although there are several social media for LGBTQ+, they are all dating apps for people over the age of 18. In addition, there are many issues with services available to minors, such as slander, identity theft, and use for the purpose of dating. Therefore, the purpose of this system is to target LGBTQ+ and Ally students and provide them with a safe place to interact and share their concerns. To achieve it, our system restricts users to student. Our system conducts identity verification using e-mail address and student ID card to restrict users. However, some issues remain, such as trouble prevention and handing, the formulation of terms of use, and improving the accuracy of the identity verification function. In the future, there are plans to complete the parts that have not been implemented, as well as to strengthen measures to deal with the operational parts of the system, such as preventing problems and formulating terms of use. Through this research, we hope to ultimately provide an opportunity for LGBTQ+ people to be able to live more true to themselves.

References

1. Chuo University. Chuo University Declaration of Diversity (2017). https://www.chuo-u.ac.jp/uploads/2024/03/campuslife_diversity_declaration_02.pdf. Accessed 20 July 2024
2. Ministry of Health, Labour and Welfare. Shokuba ni okeru daibashiti suishin jigyo hokokusho in Japanese, p. 5 (2019). <https://www.mhlw.go.jp/content/000673032.pdf>. Accessed 20 July 2024
3. Federal Trade Commission. Children's Online Privacy Protection Rule (COPPA) (2013). <https://www.ecfr.gov/current/title-16/chapter-I/subchapter-C/part-312>. Accessed 20 July 2024
4. U.S. Congress. Respect for Marriage Act, H.R. 8404, 117th Cong (2022). <https://www.congress.gov/bill/117th-congress/house-bill/8404/text>
5. Tokyo Metropolitan Government Human Rights Division. Partnership Oath System (2023). https://tokyo-partnership.metro.tokyo.lg.jp/s/?language=en_US. Accessed 20 July 2024
6. Ministry of Health, Labour and Welfare. Shogakko gakushu shido yoryo in Japanese (2017). https://www.mext.go.jp/a_menu/shotou/new-cs/youryou/syo/tai.htm. Accessed 20 July 2024
7. Hidaka, Y.: LGBT touzisyano ishikityousa in Japanese (2016). <https://health-issue.jp/reach>. Accessed 8 Sept 2024
8. Matsumoto, Y., Kirifu, M., Fujiwara, M.: Support System for LGBT Persons in Junior High Schools (2022). <https://kyujyo.repo.nii.ac.jp/records/448>. Accessed 8 Sept 2024
9. Google. OCR documentation. <https://cloud.google.com/vision/docs/ocr?hl=ja>
10. Ministry of Health, Labour and Welfare. Gakko code in Japanese (2024). https://www.mext.go.jp/b_menu/toukei/mext_01087.html. Accessed 20 July 2024



Single Sign-on System with Local Personal Information Store

Yoshiki Hosoda and Masaki Kohana^(✉)

Faculty of Global Informatics, Chuo University, 1–18 Ichigayatamachi, Shinjuku-ku,
Tokyo 162–8478, Japan

a21.5g3y@g.chuo-u.ac.jp, kohana@tamacc.chuo-u.ac.jp

Abstract. Because of the increasing web services, users are required to manage a lot of account information. To resolve this problem, some single sign-on systems are proposed. However, these systems are cloud-based system. In this type, personal information is managed by a certain service. This situation has some risks. For example, if the service is closed, the user cannot use the system. And the personal information is managed on cloud server, which is a potential risk of information leak. This research proposes a single sign-on system that locates the personal information onto local machine and the information is managed by the user.

1 Introduction

Traditionally, a combination of an ID and password has been used for user authentication and remains widely in use today. With the widespread use of web services, users are now required to manage an increasing number of accounts and passwords. As a result, many users face challenges in managing multiple IDs and passwords daily. To simplify this, some users resort to reusing the same password across multiple sites or choosing passwords that are easily guessable. However, to protect valuable assets, it is necessary to create complex, unique passwords for each account—something that is practically impossible to memorize for multiple accounts.

This is where single sign-on (SSO), a mechanism that allows users to log into multiple systems with a single authentication procedure, comes in [1,3]. By using SSO, users can access various web services with a single ID and password, thereby enhancing convenience. For example, if a user logs into one application or service, the same credentials can be applied to access other related services without the need for additional authentication procedures. When implemented within an organization, SSO typically requires a central server for authentication and authorization. Although this setup can be costly, it improves both the security and user experience within the organization.

In contrast, when individuals use SSO, they often rely on web services provided by corporations (hereinafter referred to as “SSO providers”). A common example is the “Sign in with Google” option that appears on the login pages of many web services. This allows users to access multiple services under a single

account, reducing the burden of remembering and typing in numerous passwords. With fewer credentials to manage, users can create stronger, more secure passwords and minimize the risk of password fatigue. In this way, SSO not only improves user convenience but also enhances security.

While the concept of SSO is sound, concerns arise regarding its use by individuals. Specifically, when individuals use SSO, they entrust the SSO provider with managing their data. These providers handle sensitive information, such as authentication credentials, personal data, and details about the services being accessed. If this data were to be compromised, users' "assets" could be at risk.

From a security and privacy standpoint, such scenarios should be avoided. However, alternatives to corporate-managed SSO solutions are limited for individual users. There is a need for a mechanism that allows individuals to manage their sensitive information without relying on third-party SSO providers. Furthermore, managing personal data is critical from the perspective of privacy protection.

When using web services, users are often required to register personal details such as their address, name, and phone number. These pieces of information are stored by each service in association with the user's account, making it difficult to track where and what data has been provided. If personal information could be managed alongside the account itself, it would contribute to greater privacy assurance. However, current SSO mechanisms lack such functionality.

Therefore, this research proposes a new local-based SSO system. It will record credentials or personal information, which will be stored locally and visible to the user, in order to increase confidentiality and availability and reduce the user's disadvantage. Then, the proposed system will be implemented and a technical evaluation will be conducted. This will result in clarifying its practical issues.

2 Previous Research

2.1 Differences from Traditional Password Managers

This section describes how the system in this study differs from a product called a password manager. Today, many products have been created to manage IDs and passwords on behalf of users. They are called password managers, which automatically recognize when a user opens a screen to sign in to a web application and fills in a form with stored IDs and passwords. Account data is encrypted and stored on the device or on a cloud server. Whether or not the password manager recognizes that the sign-in screen has been opened and fills in the values in the appropriate form depends on whether the page has the proper structure or whether the password manager's guessing function works well enough. In addition, the ability to generate passwords during the sign-up process is also important. Today's password managers have the ability to automatically generate passwords when users sign up. Since the password is recorded by the password manager, the user is able to generate a secure password automatically. While this is a useful function, whether it works or not also depends on the website. This is due to the fact that some websites place restrictions on password strings. In

other words, even if a password manager generates a password, it may not register if it does not have the proper rules (e.g., more than 8 characters). These have a significant impact on usability.

According to an experiment by A Hutchinson et al. that tested password generation on 100 websites using four different password managers, more than a quarter of the websites encountered some kind of usability problem, although the causes of these problems varied [8]. Poor usability may cause users having high awareness of security to reuse their passwords again. The SSO in this study differs from such password managers in that it exchanges information more strictly. We aim to create a mechanism that is stricter than a password manager, but also more flexible.

2.2 OAuth2.0 and OpenID Connect

OAuth2.0 and its extension, OpenID Connect [2], are technical standards used by various SSO web services, including Google and Facebook. It is a mechanism that allows a website or application to access user's resources hosted by other web applications. However, it is sometimes pointed out that the mechanism is too simple to be completely secure and that insecure implementations are likely to be generated if developers do not have a deep understanding of web security[7]. In addition, it is the web service provider that provides this API that manages the user's information.

2.3 SSO Architecture Using HTTP Cookies

V. Samar analyzes the security issues of Session Cookies and SSO Cookies and proposes three approaches to SSO architecture using HTTP cookies: centralized, distributed, and centralized. While this solution allows selectively building *depending upon the customer requirements of deployability, performance and centralized management, an appropriate single-sign-on solution can be chosen for web applications*, it is the Cookie Server that manages the credentials [4].

3 Overview

3.1 Definition of Roles and Designations

For the purposes of this study, roles and designations are defined as follows.

Credentials	A set of information used to authenticate a person, such as user IDs or passwords.
Personal Information	A set of personal information, such as a phone number or e-mail address, that can be used to identify a person to some degree.
General User	Persons who possess credentials or personal information and use this system.

Partner	Third-party services that receive credentials or personal information.
Client	Web applications that can be operated by general users.
Agent	Intermediation web servers between partners and clients.
Datastore	XML files created in a general user's device that contain credentials or personal information.
Remote Datastore	Storage servers needed to move data stores across devices.

3.2 Sign-in Flow

To simplify the overall picture of the new SSO system, the system process flow when a general user signs in is described here as an example of a function. The flow chart is shown in the Fig. 1.

1. A general user visiting a partner's website begins sign-in. Figure 2 shows an example of a partner's website. A user enters the ID and the password and pushes the button.
2. The partner redirects the general user to the client and at the same time sends information about the partner (partner information) to the agent server. When this occurs, the agent temporarily stores the information.
3. The general user chooses the datastore as Fig. 3 and types in the password at the client to decrypt datastore as Fig. 4. The client fetches the partner information from the agent. If both are successful, The client can retrieve his/her credentials for sign-in and personal information from decrypted datastore.
4. The client sends those information to the partner and then redirects the general user to the partner's website.

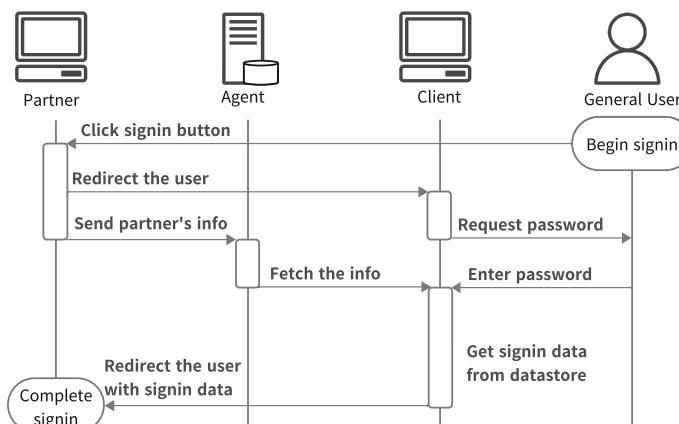


Fig. 1. The system process flow when a general user signs in

3.3 Expected Personal Information and Credentials

The expected credentials are an ID and a password. These are random alphanumeric characters of 10 to 20 characters, including upper and lower case letters and numbers. The personal information to be expected is Email (ID), Password, Phone number, Name and Nickname, Address and Post number. Each item was determined by collecting personal information required for membership registration from web services with the largest number of users in 2021.



Fig. 2. Starting sign-in on a partner website



Fig. 3. Selecting a data-store on client app



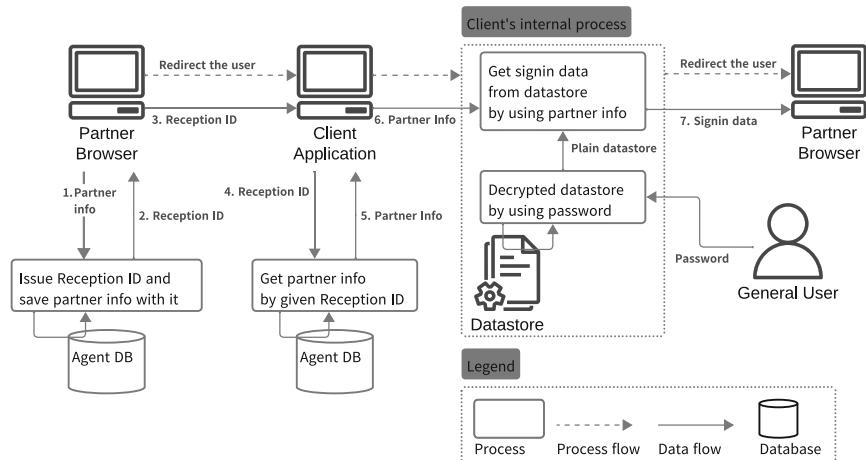
Fig. 4. Entering pass to decrypt and submit accounts

4 Implementation

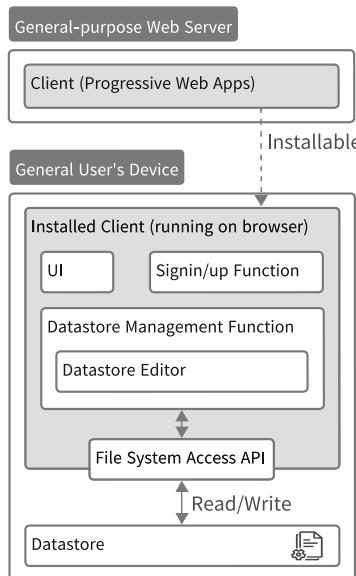
4.1 Overall Architecture

General users of SSO may either (1) have never used a partner service before and sign up for a partner account, or (2) have used a partner service before and sign in to a partner account. In this section, we will use pattern (2). We will describe the overall architecture of the system when a partner gets the sign-in data including credentials and personal information (i.e., the general user is signed in). Refer to the Fig. 5.

- When a general user begins sign-in at a partner's website, the partner service sends partner information to the agent.
- The agent issues a reception ID, stores it in the DB with the partner information, and sends the only reception ID to the partner.
- The partner redirects the general user to the client with the reception ID.
- The client attempts to obtain partner information by making a query to the agent for the received reception ID.
- The agent retrieves partner information from the DB based on the reception ID and sends it to the client.

**Fig. 5.** SSO control flow at sign-in

6. Upon receiving the partner information, the client collects the needed sign-in data from the datastore based on that information.
7. The client sends the collected sign-in data to the partner and redirects the general user to an endpoint assigned by the partner.

**Fig. 6.** Client Architecture

In the case of sign-up, the process of generating credentials and writing them to the datastore is added in (6), but the architecture is basically like this.

4.2 Client Architecture and Technical Details

This section describes the client architecture. The architecture of the client is shown in Fig. 6. The client is a static web application composed of HTML, CSS, and JavaScript. It is the only application that handles credentials or personal information. The application is designed as a PWA, which means that it is a near-native application [5, 6]. Therefore, the application can be installed on a device and run stand-alone by placing the files on a general-purpose web server. While an Internet connection is required for sign-in/sign-up, the application can be used offline when there is no need to go through the Internet, such as when just browsing a datastore. Vue.js, a front-end framework, was used for the implementation. There are two reasons for this. One is that Vue.js is suited for this application, which is mainly functional and has little UI, because it allows the UI and the program to be written in one place. Also, using a front-end framework such as Vue.js allows the project to export static, plain HTML, CSS, and JavaScript.

4.3 Agent Architecture and Technical Details

The agent was implemented with Google Cloud Firestore and Google App Script. The server has a single endpoint, which can exchange information in the DB via HTTP GET and POST. When a partner sends partner information to the agent, they send a POST method with the parameters redirect_uri, scope, type, requester_id, and key.

redirect_uri	This is the URL to which the client sends sign-in/sign-up data. General users are also redirected here.
scope	A list of labels for the requested personal information. The label is a string such as fullname or phonenumber.
type	A method in which the partner authenticates the user. The current implementation supports only the ID/Password method.
requester_id	An unique string to identify partners.
key	A string that partners use to prove their own validity. It is used when agent verifies partners.

If either Requester ID or Key is incorrect, the partner's request is rejected. When a client inquires partner information from an agent, it uses the GET method and includes the reception_id as a parameter.

reception_id	A string issued by the agent. It is used during sign-in/sign-up and serves as an identifier for the agent to retrieve partner information.
--------------	--

When partner information is received, the agent adds time information to the information and stores it in the DB. Using this time information, the validity period of the partner information is limited to 10 min, and once the partner information has been obtained, it cannot be obtained again. This prevents the same reception ID from being used repeatedly. This is effective against URLs that lead to clients via spam mail, etc.

4.4 Partner Architecture and Technical Details

Partner is a web server, implemented using the Express.js framework. It has a sign-in/sign-up screen and was created to experiment with information exchange with SSO.

4.5 Remote Datastore Architecture and Technical Details

The remote datastore was implemented with Google Cloud Firestore and Google App Script. The server has one endpoint, which can exchange information in the DB via HTTP GET and POST. When the client sends a datastore to the remote datastore, it uses the POST method and includes the parameters: devicefileid, ‘key_salt’, ‘key_iv’, and ‘content’.

```
<?xml version="1.0" encoding="UTF-8" ?>
<root>
    <application> <name>Account-Store</name> </application>
    <meta>
        <devicefileid>5db99067-cb2a-44e3-927d-ab5375c462a3</devicefileid>
    </meta>
    <data>/FA6SkL831eSwfnOP84mizpC2p3gRR+oz+mXPy2v6cbh178==</data>
</root>
```

Fig. 7. Example of datastore

devicefileid	A unique string that identifies the datastore. UUID assigned during datastore creation.
key_salt	Salt of the key used to encrypt the datastore.
key_iv	Initial vector of keys used to encrypt the datastore.
content	Datastore Contents.

When the client downloads a datastore from a remote datastore, it sends a GET request with the devicefileid as a parameter. When the datastore is received, the remote stores it in the DB with time information in addition to the datastore information. This time information is used to limit the lifetime of the datastore to 10 min, and once the partner information is retrieved, it cannot be retrieved again. General users can upload the datastore only when they want to move it, and download it immediately to prevent unintentional data leakage.

4.6 Datastore Architecture and Technical Details

Datastore is a XML local file for a general user to store his/her credentials and personal information. That's XML structure in the root node is divided into application block, meta block and data block. The contents of a sample datastore are shown in the Fig. 7.

The objectives of each block are as follows.

application	is used to indicate that the file is a data store. When reading the file, the client first checks this block.
meta	is used to record information about the datastore. Each datastore has a unique identifier (UUID) generated at the time of file creation and it is stored in this block.
data	is used to record credentials and personal information. This block has child XML nodes, but there are encrypted and encoded as Base 64.

```
<data>
  <personalinfo>
    <fullname>Alice</fullname> <email>alice@hogehoge.com</email>
  </personalinfo>
  <services>
    <service>
      <id>something-partner-id</id> <scope>email<scope>
      <credentials>
        <id>hogehoge</id> <password>123456</password>
      </credentials>
    </service>
  </services>
</data>
```

Fig. 8. Example of data blocks in a datastore (image)

The data block contains credentials, such as ID and password, and personal information, such as fullname and phonenumber. These information are also described in XML, as shown in the Fig. 8. Actually, these strings in the data node are encrypted with encryption key and encoded into Base 64. The key is generated with key derivation function at the time of writing data to datastore. General users are required to types in password every time the client writes something to datastore and the key also changes. The salt and initialization vector used for key generation are associated with the file's UUID, stored in the browser's Local Storage.

5 Challenges and Future Prospects

5.1 Account Restoration

This system transmits personal information each time a user signs in, and do not allow the partner to store that information. However, with this system, if a

general user loses their datastore and credentials, it may be difficult to recover them. This is because e-mail addresses are often used to reset account's password. The URL for password recovery is sent to the user via email, and the user resets the password by using that link.

5.2 Data Loss

In this system, all data is lost due to deleting the browser cache or losing the device. If the browser cache is deleted, a part of the encryption key is lost, which means that data decryption is no longer possible. In addition, credentials exist only in the local datastore, and if lost, all information would be lost. This would result in losing usability, and therefore some solution is necessary.

5.3 Password Generation Constraints

Password constraints vary from web service to web service. In order to issue a proper password, the client must know about the constraints.

6 Summary

In this research, SSO was developed to simplify sign-in and sign-up, and to investigate the usefulness of the system that allows users to store credentials or personal information locally and manage it themselves. This system was proposed to reduce the risks that are growing today. The risks are that the traditional method of memorizing passwords causes security concerns that people reuse passwords or privacy concerns that personal information are leaked. On the other hand, there still remain issues such as how to guarantee data in case of device loss.

References

1. AWS. What is SSO?. Amazon Web Services. <https://aws.amazon.com/jp/what-is/sso/>
2. Sakimura, N., Bradley, J., Jones, M., de Medeiros, B., Mortimore, C.: OpenID Connect Core 1.0 incorporating errata set 2 — CodeFlowSteps. The OpenID Foundation (2023). <https://openid.net/specs/openid-connect-core-1.0.html#CodeFlowSteps>
3. Radha, V., Reddy, D.H.: A survey on single sign-on techniques. Procedia Technol. 4, 134–139 (2012). <https://www.sciencedirect.com/science/article/pii/S2212017312002988>
4. Samar, V.: Single sign-on using cookies for Web applications. In: IEEE 8th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE 1999) (1999). <https://ieeexplore.ieee.org/document/805192>
5. Progressive Web Apps (PWA). Mdn web docs. https://developer.mozilla.org/ja/docs/Web/Progressive_web_apps. Accessed July 2024
6. LePage, P., Steiner T.: What are Progressive Web Apps — Get results from PWAs (2020). <https://web.dev/articles/what-are-pwas?hl=ja>

7. Hossain, N., Hossain, M.A., Hossain, M.Z., Sohag, M.H.I., Rahman, S.: OAuth-SSO: a framework to secure the OAuth-based SSO service for packaged web applications (2018). <https://ieeexplore.ieee.org/abstract/document/8456096>
8. Hutchinson, A., Tang, J., Aviv, A.J., Story, P.: Measuring the prevalence of password manager issues using in-situ experiments (2024). <https://www.ndss-symposium.org/wp-content/uploads/usec2024-94-paper.pdf>



A Data Platform for the Integration of Smart City Subsystems

Stefano Silvestri^(✉), Giuseppe Tricomi, Emanuele Damiano, Mario Sicuranza,
and Mario Ciampi

Institute for High Performance Computing and Networking of National Research
Council of Italy (ICAR-CNR), via Pietro Castellino 111, 80131 Naples, Italy
`{stefano.silvestri,giuseppe.tricomi,emanuele.damiano,
mario.sicuranza,mario.ciampi}@icar.cnr.it`

Abstract. Advanced smart city environments are usually based on a framework composed of several heterogeneous distributed subsystems, including Digital Twins that model infrastructures, assets, and social aspects of the city, as well as analytics and forecasting modules, and IoT sensor networks. Each of these systems must collect and process heterogeneous data formats, and share the results of their elaborations with the other modules. Moreover, they also require interacting with external data sources, such as databases, or other external systems. The integration of these systems into a framework requires dedicated ICT solutions able to support the collection, management, sharing and integration of data, which are also able to facilitate the interactions among them. In this paper, we propose a data lake-based system that has the purpose of collecting, storing, managing, integrating and sharing all the data used by the smart city modules, from both internal and external sources. Moreover, the data lake offers a standardized interface to all subsystems, facilitating their full integration. This system has been successfully tested in a real smart city system in Italy.

1 Introduction

Smart City systems are complex frameworks where several different systems interact to provide advanced and smart solutions for the improvement of the policy making, well being of the citizens, sustainability, management and administration of the services of the city, exploiting advanced information and communication technologies to realize these services [1]. This is possible thanks to the integration of state-of-the art technologies, giving rise to the Urban Intelligence (UI) paradigm, an ecosystem of technologies tailored to improve urban environment, well-being, quality of life [2], exploiting Digital Twins (DT) able to represent a digital counterpart of the city systems, infrastructures, and processes.

This technological infrastructure in the UI of a smart city must also facilitate and support the integration of the DTs, as well as the other subsystems and modules of this framework, which usually include IoT sensor networks, information systems, dedicated network infrastructures, cloud, edge or fog-computing

environments, analytics and data processing modules, and others [3]. In addition, specific data ingestion, storing and management systems must be included in this environment, to the end of effectively acquiring and integrating several and heterogeneous data generated from this complex framework of interacting assets and IT infrastructures, capturing in this way the most crucial aspects of a city, such as mobility, weather, pollution, social data, touristic flows, etc. [4,5], and sharing the obtained information within the smart city DTs, as well as with other internal and external systems.

Therefore, in this scenario, data storage and management is a crucial task, but some challenges must be faced, due to the huge size and heterogeneity of data, the different features of the data sources, the highly variable data rates, the need of notification of specific events related to the data, the security and privacy issues, and the retrieving and processing requirements demanded by each subsystem [6,7]. Furthermore, the data management infrastructure of the smart city must also allow to easily share the information between its subsystems, which often need to interact with each other, to provide specific analytics and model aspects of the city requested by the users [8,9]. Finally, it must guarantee scalability features, for real-time and historical data visualization, processing, and actuation in the city [10].

This paper presents the architecture of an IT platform, whose main purposes are storing, managing and sharing the data in a smart city environment. The architecture includes customized functionalities to integrate heterogeneous data and extract specific information. Moreover, it supports the effective interaction and integration of the subsystems and modules of the smart city into a single framework. The proposed solution is based on a data lake and a REST API, and also includes specific Extract, Transform, and Load (ETL) functionalities for data aggregation, format conversion and information extraction. Furthermore, the processing performance requirements are respected thanks to the adoption of Big Data Analytics (BDA) technologies. Finally, it provides an asynchronous notification service, able to notify data update events to the specific subsystems of the smart city that are interested in.

The proposed architecture can be implemented using open source technologies, providing an open platform that can be easily replicated to support the development of smart city infrastructure, acting as data collector and integrator. In our use case, it has been successfully implemented in real-world smart city frameworks, demonstrating its effectiveness and performances.

The paper is organized as follows. The next Sect. 2 shows the recent works presented in literature related to data platforms specifically suited for smart city applications. Section 3 describes the architecture of the proposed platform, while Sect. 4 shows the implementation of the platform, based on open source tools and frameworks. Section 5 describes a real-world use case, where our architecture has been successfully adopted in a smart city project. Finally, Sect. 6 draws out the conclusions and future works.

2 Related Works

The research related to the integration of data and subsystem of the smart cities recently proposed several solutions and approaches, to the end addressing the issues related to the management of heterogeneous data [11], and the integration of different systems and technologies [12, 13]. Likewise, the literature demonstrated that efficient data and technology management is a critical step to fully exploit the potential of smart cities [14].

Data lake-based solutions are commonly used as a data management and storing platform in smart cities, allowing for the effective use of large and heterogeneous datasets generated in such an environment [15]. A data lake can be seen as a vast and unified repository, which allows to store heterogeneous data in without any predefined schema, both structured, unstructured, semi-structured, or binary, coming from diverse sources [16]. In the domain of smart cities, the authors of [17] proposed a data lake approach where SQL and NoSQL database are integrated with Online Analytics Processing (OLAP) and Online Transaction Processing (OLTP) capabilities, allowing to effectively manage heterogeneous data. In [18], the authors studied the features and capabilities of NoSQL technologies applied to the implementation of a data lake for smart cities, considering in particular performance, scalability, accuracy, and complexity as main indicators. The results of their experiments revealed that MongoDB was the most stable NoSQL database, assessing its use in the implementation of a smart city data platform. A semantic data lake for smart cities has been proposed by [19], incorporating semantic metadata within a data lake to ease the annotation and enrichment with metadata, and exploiting a Multi-Dimensional Ontology [20] to check their conformance.

In [21], a Model-Based Systems Engineering (MBSE) approach to design a model of an integrated smart city system has been proposed, bringing all subsystems to operate together as one system and focusing on the information perspective of a city system. The study presented in [22] faced the issues related to the management of multiple and heterogeneous data sources and systems in a smart city, proposing a platform-based solution to share data and merge the subsystems into a single infrastructure. This platform is based on intelligent agents and leverages ontologies for the adaptation to the specific context, to respond to unforeseen situations.

The architecture described in this paper proposes a simple, scalable, and effective method to integrate both data and subsystems of a smart city, and can be easily implemented and replicated using exclusively open source technologies. Its main innovative aspects with respect to the recent literature is the integration of a dedicated Big Data engine to perform the most complex ETL processing in real-time, as well as the capability of performing both data and subsystems integration.

3 Data Platform Architecture

Figure 1 depicts the layers of the architecture of the proposed data platform, which includes two horizontal layers and one vertical layer, whose details are described below.

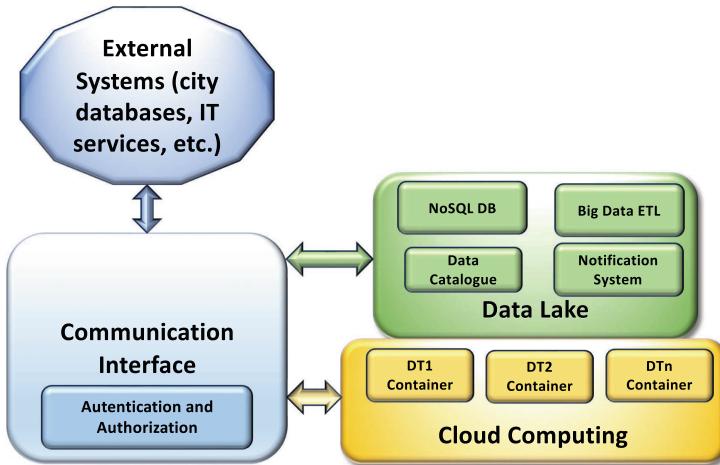


Fig. 1. Layers of the architecture of the data lake.

3.1 Cloud Computing Layer

The lowest horizontal layer is the *Cloud Computing*, which provides all the functionalities to run the various software subsystems of the smart city (DTs, analytics, forecasting, and other required services) also on a cloud-based environment, facilitating their deployment, portability, and integration by exploiting a container-based approach. Also, the software related to of the other layers is deployed here, in their own containers. In summary, this layer oversees the dynamic coordination and orchestration of the hardware and software resources of the modules, as well as provides the environment to run the container-based version of each subsystem of the smart city, including the data lake and its API.

3.2 Data Lake Layer

The upper horizontal layer is the *Data Lake*, which includes all the functionalities for data gathering and storing, also offering customized ETL capabilities for the integration, aggregation and transformation of heterogeneous data formats, extracting the required information and structuring them as needed by the other subsystem of the framework for their processing. It also acts as a data collector for all the information of the smart city platform, acquiring and sharing any related data. Moreover, it allows to easily query and retrieve the data stored.

This layer is based on a NoSQL database, which does not require a strict and predefined data structure, allowing to easily store and manage heterogeneous data, and to extend the data catalogue and data types whenever required. Moreover, it also integrates a Distributed File System (DFS), which provides the capability to store and index large binary data collections (images, and other kinds of raw data).

A data catalogue is also included, which has the purpose of preventing data swamp degeneration [23] and facilitating the data retrieving and its semantic interpretability, enriching the data with tailored metadata, such as measurement units, value ranges, data types, and others. Furthermore, the data catalogue can be also queried, promoting data sharing within the subsystems of the smart city, as well as with external systems.

The ETL functionalities included in this layer exploit BDA technologies, using the Big Data ETL approach previously described by [24], which allows the BDA engine to directly access the NoSQL database, to efficiently execute complex queries on large datasets in almost real-time [25], leveraging a distributed and scalable execution environment to respect any time or size requirement of the DTs and other data processing modules.

The data lake layer also incorporates an asynchronous notification subsystem, which has the purpose of notifying data updates in the Data Lake, selecting the recipients of these notifications, following a publish/subscribe approach, optimizing in this way both the bandwidth usage and the performances of the DB [26].

3.3 Communication Interface Layer

The vertical layer of the architecture is connected to both horizontal layers and contains the *Communication Interface*, which acts as a common and standardized interface among the data lake and all the subsystems of the smart city framework, supporting the integration of its modules. It also provides the data ingestion and gathering capabilities from external sources and systems, as well as it allows the connections from external systems to the data lake.

The main function implemented in this layer is represented by a REST API, which exposes the data retrieving and acquisition functionalities in a standardized and common form to each subsystem of the smart city to interact with the data lake without the need of knowing any detail of its implementation. Moreover, the functions exposed by the API are specifically tailored for each module that must send or get data, including, when required, also simple ETL functions, such as data format transformation, data aggregation or conversion, information extraction, and others, with the main purpose of providing data for different modules of the smart city in their required format. In this way, it also acts as an integration module, because all the subsystems of the smart city platform interact among them using the facilities provided by the API. To this end, the data lake through the API collects, integrates, transforms, and shares data from any internal source of the smart city environment, not only from IoT sensors, but also including any data produced as results by each of the modules of the framework, sharing them among the other modules, which could need results

from other subsystems to produce their outputs. Similarly, any data acquired from external systems is also stored in the data lake and shared through the API.

Finally, this layer includes the authentication and authorization functionalities, ensuring the security and privacy of the system and the data.

4 Implementation

As explained in previous Sect. 3, the data lake is based on a NoSQL database and a Distributed Files system. In our implementation, we used MongoDB [27], which supports automatic scalability features, also in distributed environments [28], providing the required performance levels for smart city applications [18]. Moreover, its scheme-free data facilitates the integration of heterogeneous data, without the need of a previous definition of their structure. MongoDB natively supports json-like data, as well as geojson data, allowing the storage and management of the information of the smart city [29]. Finally, it provides all the required indexing, retrieval, and querying functionalities, which, in our case, are exclusively accessed through the customized API in the communication interface layer.

The data collected in the data lake is enriched with metadata, applied, when required, by the API during the data gathering phase in the communication interface layer. A data catalogue is directly implemented in a dedicated database and collection in MongoDB, and contains, for each type of stored data, the fields and the metadata, with a brief description and other useful information, such as the value range, or the measurement units, facilitating in this way data sharing and exchanging.

To the end of also allowing the management and storing of binary files, such as images, 3D models of the city, and others, we adopted a Distributed File System (DFS), based on GridFS [30], which is natively integrated within MongoDB. Therefore, the indexing of the files, enriched, if required, with metadata, happens directly in the database, including the information related to these files in the data catalogue too. Finally, GridFS can leverage the same dynamic distributed scalability functionalities of MongoDB [31].

The ETL functions are also provided by this architecture. While simplest data aggregation or transformations are performed directly by the API in the communication layer, when data is acquired, the most complex ETL tasks are obtained through a BDA framework, based on Spark [32] and SparkSQL [33], speeding up the requests on large collection in the database [34]. In detail, SparkSQL can be integrated, with a dedicated connector, with MongoDB, improving the execution time of queries on large collections, exploiting the approach proposed in [24], thanks to the distributed Big Data processing features of Spark.

The asynchronous notifications of data updates leverages an asynchronous mechanism based on a publish/subscribe approach, using the Message Queue Telemetry Transport (MQTT) standard protocol [35]. MQTT allows users or software modules which require data updates notifications to subscribe to them.

When subscribed data is updated, the MQTT server automatically sends a notification to the subscribers, in an asynchronous way, allowing to receive the notifications of the data changed when not online too, optimizing both the database performances, and the network bandwidth. The adopted implementation is based on the Eclipse Paho MQTT Python library [36].

The communication interface layer is realized by a dedicated REST API, which has the purpose of standardizing any interaction between the data lake layer and the other modules of the smart city, and with external systems that could need the data collected in the smart city framework. In this way, it provides a common, simplified and standard method to send or collect data, which must be respectively stored or gathered from the data lake. Any user, or software module which has to interact with the data, does not need to deal with MongoDB, GridFS, or SparkSQL, but tailored API functions have been specifically implemented to provide any required functionality to each module. Moreover, as explained above, if data produced by a module or a sensor requires a simple ETL process, such as a format conversion or information extraction, the API directly executes this task, storing both raw and processed data in the data lake. The implementation of the REST API is based on the Flask Python library [37].

The communication interface also provides security, authentication, and authorization mechanisms. These are based on the use of both certificates and name and password, managed through Keycloak [38], an open-source identity and access management framework.

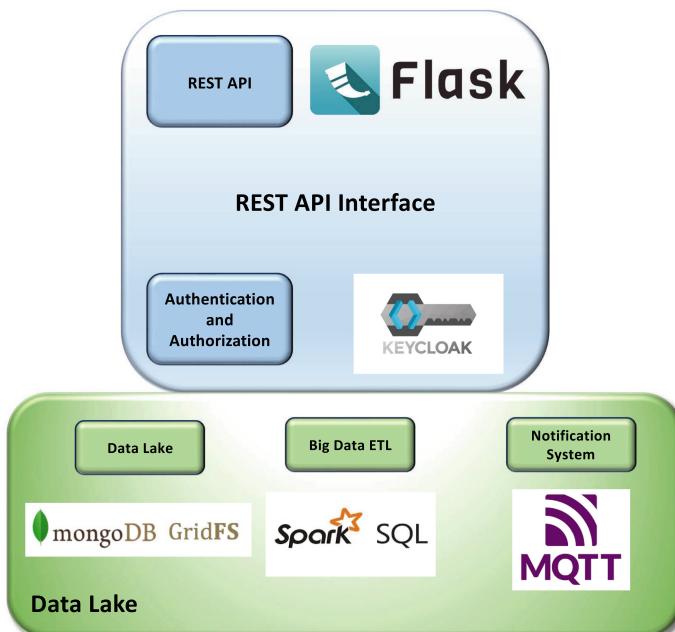


Fig. 2. Key technologies used to implement the proposed data lake and its REST API.

We underline that all the adopted technologies are open source, allowing an easy reproduction and a wide adoption of the same architecture. Figure 2 summarizes the tools and software used to implement our proposed architecture.

5 Use Case

The proposed architecture, implemented using open source software as described in previous Sect. 4, has been successfully adopted to realize the smart city framework of the city of Matera in the South of Italy, allowing it to be tested in a real-world environment. Matera is an important touristic city, which includes the UNESCO World Heritage property of the Sassi of Matera and the Park of the Rupestrian Churches, hosting millions of visitors each year. The recently implemented smart city platform includes several DTs which allow to simulate and forecast many aspects of the city, such as traffic, occupancy of the Points of Interests (POIs), smart paths (based on shadowing, slopes, presence commercial activities, etc.), energetic emissions of the buildings, 3D model of the city, graph representation of the city, solar heights, and other important aspects, creating many useful intelligent services for tourists, citizens and policy makers.

In this system, many modules must access huge and heterogeneous data collection, acquired from multiple sources, both internal (i.e., traffic and pollution IoT sensors, or databases with city information) and external (i.e., satellite data, mobile phone data, etc.). These modules must also interact with each other, sharing or exchanging the results of their processing, to calculate further results and information. The architecture allows the integration of the modules and the data, providing a data collector with a common interface, used by all the modules of the platform, and sharing the data, previously specifically tailored for each module thanks to ETL pipelines. For example, the smart path generator requires the shadowing data, calculated using the 3D model and the solar radiation data from Copernicus satellite [39], to provide the tourists with the optimal shadowed path on a hot summer day. Therefore, the involved modules must exchange the required information, previously stored in the data lake, using the API of the communication interface of the proposed architecture. The proposed data platform makes the modules of the smart city unaware of any implementation detail related to the data gathering, sharing, integration, transformation, or storing, providing them with all these functionalities. In this way, the development of the other modules of the smart city becomes easier and faster. Moreover, it also promotes their integration, thanks to the availability of a common data collector and a standardized interface.

As example, a result of the ETL process is depicted in Fig. 3. In this case, the raw data from environmental sensors (temperature, pollution, wind, rain, etc.) showed in Fig. 3a is processed using a simple ETL pipeline integrated in the API, making them more easily understandable for external users and other subsystems, transforming the fields name, separating information from different sensors, as well as structuring the data as expected by other modules of the platform, obtaining the structure shown in Fig. 3b.

```

_id: ObjectId('6578617be75f4f05ab96eb88')
Device_serial: 19820
Date: 2023-10-22T00:00:00.000+00:00
Temperature_1_(Medium)_C: 15.4
Humidity_2_(Medium)_RH: 83
Wind_Direction_4_(Medium)_GN: 181
Wind_Direction_4_(Standard_Deviation)_GN: 86
Wind_Direction_4_(Mean_Square_Deviation)_GN: 48
Wind_Speed_9_(Medium)_m/s: 0.5
NO_35_(Medium)_ppm: 0
NO2_37_(Medium)_ppm: 0
O3_38_(Medium)_ppm: 0.02
SO2_39_(Medium)_ppm: 0
Auxiliary_Measurement_41_(Medium)_ppm: 0
Atmospheric_pressure_2013_(Medium)_hPa: 965.2
Auxiliary_Measurement_91_(Medium)_ppm: 12.6
Auxiliary_Measurement_141_(Medium)_ppm: 20.7
Auxiliary_Measurement_191_(Medium)_ppm: 40.3
CO_34_(Medium)_ppm: 0

```

```

_id: ObjectId('68b0ef34580549b3ae396b94')
sensor_id: ObjectId('19820d2e3cf1148411119820')
timestamp: 2023-10-22T00:00:00.000+00:00
temperature: 15.4
humidity: 83
Wind_dir: 181
Wind_dir_STD: 86
Wind_dir_Mean_Square: 48
Wind_speed: 0.5
NO: 0
NO2: 0
O3: 0.02
SO2: 0
VOC: 0
atm_pressure: 965.2
pm1: 12.6
pm2_5: 20.7
pm10: 40.3
CO: 0

```

(a) Raw sensor data

(b) Sensor data after the ETL process

Fig. 3. Example of the ETL process applied to environmental sensors data.

In this use case, the implemented data platform is capable of acquiring and processing data rates higher than 300 MB/minute with no issue and in almost real-time, as soon as they are produced by the IoT sensors installed in the city and by the other data sources. The tests on the implemented platform also demonstrated that it can share data at the same time with all the modules of the smart city, thanks to the parallel, reliable, and scalable architectures provided by MongoDB and Spark.

In summary, the functional and performance tests carried out in the use case demonstrated that the proposed data platform architecture is capable of acquiring high rates and volumes of data from the IoT sensor network of the city and other sources without any issue, sharing them at the same time among the modules of the smart city. Furthermore, it is able to store and integrate the huge volume of data from both internal and external sources. The REST API facilitates the integration of the subsystems of the smart city, providing a common, standard and simplified interface to share and exchange data, also with external systems, ensuring the interaction of all the modules as a single framework in a distributed environment.

6 Conclusion

This paper presented the architecture of a data platform for smart cities, devoted not only to acquire, store, and manage the data, but also to facilitate the effective integration and interactions between the various subsystems of the smart city, providing a common interface for their interactions. Moreover, this architecture includes customized functionalities to integrate heterogeneous data, enriching with metadata and extracting the required information. The proposed solution leverages a data lake and a communication interface, respectively based on a NoSQL database and a REST API. It also includes specific ETL functionalities

for data aggregation, format conversion and information extraction, integrated with a BDA engine, to ensure the required performance levels in the case of queries on huge data collections. It also provides an asynchronous notification service, to notify specific subsystems of the smart city when data update events happen, and, finally, includes the authorization and authentication features for the security of the system.

The proposed architecture has been implemented using open source technologies and tools in a real-world smart city project in Matera in Italy, demonstrating its effectiveness. As future works, further developments are currently under development, testing different technologies to improve the functionalities of the platform. Moreover, a possible integration with a workflow engine is under investigation, to the end of providing a further orchestration and coordination tool for a deeper integration of the different modules of the smart city.

Acknowledgements. This work is supported by the European Union - NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) - Project: “SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics” - Prot. IR0000013 - Avviso n. 3264 del 28/12/2021.

The authors would like to thank Simona Sada and Giuseppe Trerotola for the technical and administrative support provided.

References

1. Rajab, H., Cinkelr, T.: IoT based smart cities. In: 2018 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–4 (2018). <https://doi.org/10.1109/ISNCC.2018.8530997>
2. Castelli, G., et al.: Urban intelligence: a modular, fully integrated, and evolving model for cities digital twinning. In: 2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT), pp. 033–037 (2019). <https://doi.org/10.1109/HONET.2019.8907962>
3. Habibzadeh, H., Kaptan, C., Soyata, T., Kantarci, B., Boukerche, A.: ACM Comput. Surv. **52**(2) (2019)
4. Deng, T., Zhang, K., Shen, Z.J.M.: J. Manage. Sci. Eng. **6**(2), 125 (2021)
5. White, G., Zink, A., Codec, L., Clarke, S.: Cities **110**, 103064 (2021)
6. Jeong, S., Kim, S., Kim, J.: City data hub: Implementation of standard-based smart city data platform for interoperability. Sensors **20**(23), 7000 (2020). <https://doi.org/10.3390/s20237000>, <https://www.mdpi.com/1424-8220/20/23/7000>
7. Cheng, B., Longo, S., Cirillo, F., Bauer, M., Kovacs, E.: Building a big data platform for smart cities: Experience and lessons from santander. In: 2015 IEEE International Congress on Big Data, pp. 592–599 (2015). <https://doi.org/10.1109/BigDataCongress.2015.91>
8. Silvestri, S., Tricomi, G., Bassolillo, S.R., De Benedictis, R., Ciampi, M.: An urban intelligence architecture for heterogeneous data and application integration, deployment and orchestration. Sensors **24**(7) (2024). <https://doi.org/10.3390/s24072376>, <https://www.mdpi.com/1424-8220/24/7/2376>
9. Tricomi, G., D'Agati, L., Longo, F., Merlino, G., Puliafito, A., Silvestri, S.: Paving the way for an urban intelligence OpenStack-based architecture. In: 2024 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 284–289 (2024). <https://doi.org/10.1109/SMARTCOMP61445.2024.00069>

10. Vítor, G., Rito, P., Sargent, S., Pinto, F.: A scalable approach for smart city data platform: support of real-time processing and data sharing. *Comput. Netw.* **213**, 109027 (2022). <https://doi.org/10.1016/j.comnet.2022.109027>, <https://www.sciencedirect.com/science/article/pii/S1389128622001839>
11. Bibri, S.E.: Data-driven smart sustainable cities of the future: urban computing and intelligence for strategic, short-term, and joined-up planning. *Comput. Urban Sci.* **1**(1), 1–29 (2021). <https://doi.org/10.1007/s43762-021-00008-9>
12. Puliafito, A., Tricomi, G., Zafeiropoulos, A., Papavassiliou, S.: Sensors **21**(10), 3349 (2021)
13. Goumopoulos, C.: Smart city middleware: a survey and a conceptual framework. *IEEE Access* **12**, 4015–4047 (2024). <https://doi.org/10.1109/ACCESS.2023.3349376>
14. Prabowo, O.M., Mulyana, E., Nugraha, I.G.B.B., Supangkat, S.H.: IEEE Access **11**, 120157 (2023)
15. Ramos, G.S., Fernandes, D., Coelho, J.A.P.d.M., Aquino, A.L.L.: Toward data lake technologies for intelligent societies and cities. In: da Silva Portela, C.F. (eds.) Sustainable, Innovative and Intelligent Societies and Cities. EAI/Springer Innovations in Communication and Computing, pp. 3–29. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-30514-6_1
16. Cherradi, M., EL Haddadi, A.: Data lakes: a survey paper. In: Ben Ahmed, M., Boudhir, A.A., Karas, İR., Jain, V., Mellouli, S. (eds.) SCA 2021. LNNS, vol. 393, pp. 823–835. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-94191-8_66
17. Miloslavskaya, N., Tolstoy, A.: Procedia Comput. Sci. **88**, 300 (2016)
18. Nurhadi, Kadir, R.B.A., Surin, E.S.B.M.: In: Kim, H., Kim, K.J., Park, S. (eds.) Information Science and Applications. Lecture Notes in Electrical Engineering, vol. 739, pp. 383–392. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-6385-4_35
19. Bianchini, D., De Antonellis, V., Garda, M.: A semantics-enabled approach for personalised Data Lake exploration. *Knowl. Inf. Syst.* **66**(2), 1469–1502 (2024). <https://doi.org/10.1007/s10115-023-02014-1>
20. Alicante, A., Benerecetti, M., Corazza, A., Silvestri, S.: A distributed architecture to integrate ontological knowledge into information extraction. *Int. J. Grid Util. Comput.* **7**(4), 245–256 (2016). <https://doi.org/10.1504/IJGUC.2016.081011>, <https://www.inderscienceonline.com/doi/abs/10.1504/IJGUC.2016.081011>
21. Muvuna, J., Boutaleb, T., Baker, K.J., Mickovski, S.B.: A methodology to model integrated smart city system from the information perspective *Smart Cities* **2**(4), 496–511 (2019). <https://doi.org/10.3390/smartcities2040030>, <https://www.mdpi.com/2624-6511/2/4/30>
22. Aguilar, J., Jerez, M., Mendonça, M., Sánchez, M.: Performance analysis of the ubiquitous and emergent properties of an autonomic reflective middleware for smart cities. *Computing* **102**(10), 2199–2228 (2020). <https://doi.org/10.1007/s00607-020-00799-5>
23. Hai, R., Geisler, S., Quix, C.: Constance: an intelligent data lake system. In : Proceedings of the 2016 International Conference on Management of Data, SIGMOD’16, pp. 2097–2100. Association for Computing Machinery, New York (2016). <https://doi.org/10.1145/2882903.2899389>
24. Silvestri, S., Esposito, A., Gargiulo, F., Sicuranza, M., Ciampi, M., De Pietro, G.: A big data architecture for the extraction and analysis of EHR data. In: 2019 IEEE World Congress on Services (SERVICES), vol. 2642, pp. 283–288. IEEE (2019)

25. Karras, A., Karras, C., Pervanas, A., Sioutas, S., Zaroliagis, C.: SQL query optimization in distributed NoSQL databases for cloud-based applications. In: Foschini, L., Kontogiannis, S. (eds.) ALGOCLOUD 2022. LNCS, vol. 13799, pp. 21–41. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-33437-5_2
26. Hunkeler, U., Truong, H.L., Stanford-Clark, A.: MQTT-S—A publish/subscribe protocol for wireless sensor networks. In: 2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08), pp. 791–798 (2008). <https://doi.org/10.1109/COMSWA.2008.4554519>
27. Mongodb. <https://www.mongodb.com/>. Accessed 20 Jul 2024
28. Huang, C.W., Hu, W.H., Shih, C.C., Lin, B.T., Cheng, C.W.: The improvement of auto-scaling mechanism for distributed database - a case study for MongoDB. In: 2013 15th Asia-Pacific Network Operations and Management Symposium (APNOMS), Hirshima, Japan, pp. 1–3. IEEE (2013)
29. Kazmi, A., Jan, Z., Zappa, A., Serrano, M.: In: Podnar Žarko, I., Broering, A., Souratos, S., Serrano, M. (eds.) Interoperability and Open-Source Solutions for the Internet of Things, pp. 20–35. Springer, Cham (2017)
30. GridFS. <https://mongodb.github.io/node-mongodb-native/3.4/tutorials/gridfs/>. Accessed 15 Jul 2024
31. Wang, S., Li, G., Yao, X., Zeng, Y., Pang, L., Zhang, L.: A distributed storage and access approach for massive remote sensing data in MongoDB. ISPRS Int. J. Geo-Inf. **8**(12), 533 (2019). <https://doi.org/10.3390/ijgi8120533>. <https://www.mdpi.com/2220-9964/8/12/533>
32. Zaharia, M., et al.: Commun. ACM **59**(11), 56–65 (2016)
33. Armbrust, M., et al.: Spark SQL: relational data processing in spark. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, SIGMOD'15, pp. 1383–1394. Association for Computing Machinery, New York (2015). <https://doi.org/10.1145/2723372.2742797>
34. Karras, A., Karras, C.N., Pervanas, A., Sioutas, S., Zaroliagis, C.D.: SQL query optimization in distributed NoSQL databases for cloud-based applications. In: Foschini, L., Kontogiannis, S. (eds.) ALGOCLOUD 2022. LNCS, vol. 13799, pp. 21–41. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-33437-5_2
35. MQTT: The standard for IoT messaging. <https://mqtt.org>. Accessed 30 Jan 2024
36. Eclipse Paho. <https://eclipse.dev/paho/>. Accessed 31 Jan 2024
37. Flask. <https://flask.palletsprojects.com/en/3.0.x/>. Accessed 31 Jan 2024
38. Keycloak. <https://www.keycloak.org>. Accessed 30 Jan 2024
39. Cams solar radiation time-series. Copernicus atmosphere monitoring service (cams) atmosphere data store (ads). <https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-solar-radiation-timeseries?tab=overview> Accessed 20 Jul 2024



Minimization of Transfer Time for User Files Through Read Control for Backup with Deadline Time

Futa Takahashi[✉] and Takayuki Kushida

Graduate School of Computer Science, Tokyo University of Technology, Tokyo, Japan
g212301977@edu.teu.ac.jp, kushida@acm.org

Abstract. Employees in video production companies transfer files to a file server. The backup software in the file server transfers the business data to a backup server. Employee file transfer overlaps with a backup period as a project deadline approaches. The file server is required to minimize the file transfer time for employees when complete backups that have deadline time. The proposed method maximizes file writes by controlling file reads to complete backups by the deadline time. It controls file reading because the calculated and measured backup speeds are different. It also increases the backup speed when the backup speed is decreased according to the number of users transferring files at the same time. The evaluation compares the proposed method with employee file transfers during backup. Three users transfer 100 [GB] files to the file server. The backup deadline is 60 min. File transfer time was reduced from 48 min 46 s to 44 min 3 s. The experimental results show that the proposed method reduced the user's file transfer time by approximately 9.7%.

1 Introduction

Companies are required to archive data to avoid data loss [5]. Data loss occurs due to hardware or software failures or natural disasters [8]. Companies perform backups in the middle of the night and finish within a period. The backup period varies from company to company. An example of the backup start time is the business end time. An example of the backup deadline time is the business start time.

Toei Animation backs up the business data. Toei Animation develops video, television, and character products. Image and video files are shared as business data for the production of content for these development projects.

Three papers report fairness in multi-user environments to prevent only particular users from waiting too long for data processing requests [1, 4, 6].

Computing systems can process data at more exponential rates in the future. On the other hand, the performance of the storage infrastructure is much slower than that of the computing system. A supercomputer Mira is an example of a system with different processing speeds [2, 3, 9]. The network bandwidth transfer

rate is 1536 [GB/s] for Mira's computing nodes. On the other hand, the execution speed of the storage system is 250 [GB/s]. This situation is due to I/O congestion. This situation causes I/O congestion and reduces application performance. System performance is reduced when the system is accessible during backup [7]. The disk bandwidth used by the backup process competes with other service processes.

User file transfer time increases when employees transfer video files while the file server is backed up in a video production company. Backup is not completed when the backup is interrupted for file transfer. The file server is required to minimize the file transfer time for users during backup and complete time-limited backup. Figure 1 shows the summary of issues.

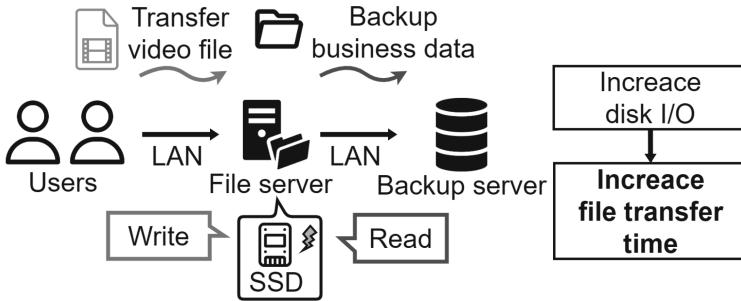


Fig. 1. Increase user file transfer time due to backup

The user transfers the video file to the file server via LAN. The file server writes the video files to the SSD. The file server also backs up business data to a backup server via LAN. The file server reads the business data to the SSD. User file transfer time increase due to conflict between writing and reading to the SSD.

Users can reduce file transfer time by suspending backups. On the other hand, backups are prolonged when the backup is interrupted each time a user transfers a file. A prolonged backup violates the Recovery Point Objective (RPO). The RPO defines how long data loss is tolerated. Data integrity is not maintained when the backup is not completed until the next backup.

2 Related Study

The scheduling method to transfer a large data over optical grid network is studied [1]. Four dynamic scheduling algorithms show the efficiency and fairness of user file transfer time. On the other hand, the study does not show that users can reduce both file transfer time and backup time.

A reliable multi-cast based approaches were studied [2]. The problem is that the central file system becomes a bottleneck when downloading data at the data

center. This issue also occurs when multiple Virtual Machines (VM) download from a file server. The proposed method of this study is called the MCD system. The MCD system solves the bottleneck of the central file system. The evaluation shows that the proposed method is 9.9 to 14.29 times faster than the existing system. On the other hand, the MCD systems only focus on downloading data, so upload time is not evaluated. There is also no mention of the possibility of the backup time exceeding the deadline. The use case in this paper is a video production company. The file server is used for backups and uploads. Therefore, the MCD systems cannot solve this paper's problem.

3 Proposed Method

The purpose of the proposed method is to increase the amount that a file server can write to disk in order to reduce file transfer time for users. The proposed method leaves bandwidth for users by completing backups on time to increase the amount of disk writing on the file server. Figure 2 shows the Solution summary and the backup bandwidth control by the proposed method.

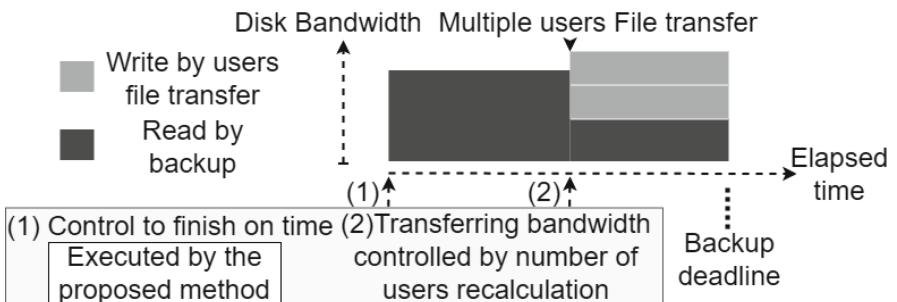


Fig. 2. The proposed method controls the backup to finish on time

The vertical axis is disk bandwidth. The horizontal axis is elapsed time. The proposed method controls the bandwidth in (1) Control to finish on time so that the backup ends on time. (2) Transferring bandwidth controlled by a number of users recalculation recalculates the bandwidth depending on the number of users during the transfer and controls the backup to finish on time more accurately. The proposed method uses the following calculation formula for backup control. Let S [MB/s] be the upper limit of the amount of disk read per second. Let C [MB] be the capacity of business data to be backed up. Let $L[s]$ be the time until the backup deadline. Let the maximum the backup speed be B_{max} [MB/s], and the actual backup speed during file transfer from the user be B_{real} [MB/s]. Let C [MB/s] be the difference between B_{max} [MB/s] and B_{real} [MB/s]. Equations (1) and (2) show the limit amount of disk reading.

$$S = \frac{C}{L} + R \quad (1)$$

$$R = B_{max} - B_{real} \quad (2)$$

The purpose of Eq. (1) is to calculate the backup speed S at which backups are completed by the deadline. S is the backup speed limit. The disk bandwidth available to users is increased by limiting backup rates. User file transfer time is reduced by increasing a number of files that can be written per unit of time.

The backup speed to complete the backup by the deadline is determined by the capacity of the business data to be backed up and the time until the backup deadline (hereinafter referred to as the Garrulus method). On the other hand, the actual backup speed is only sometimes at the upper limit. The file is transferred once through the buffer. The proposed method prevents the backup deadline from being exceeded by adding a correction R that increases the upper limit of the backup speed.

The R in the formula (1) used in the Garrulus method is calculated based on preliminary experiments to increase the backup speed. Preliminary experiments measure backup speeds that decrease with file transfers from users. R ensures that the backup end time is within the deadline. On the other hand, the backup speed is slowed down when the number of users transferring files at the same time increases. Contention between disk reads and writes reduces the amount of files that can be read per unit of time.

The proposed method calculates the backup speed when users transfer files at the same time. R is a value that measures how much the backup reading speed decreases depending on the number of users.

The difference between the proposed and the Garrulus method is that the backup speed is calculated when multiple users transfer files simultaneously. The maximum backup speed B_{max} [MB/s] is the uncorrected disk read amount calculated from the amount of business data to be backed up and the time until the backup deadline. The actual backup speed B_{real} [MB/s] is the backup speed calculated in the experiment when the backup is limited by the maximum backup speed B_{max} [MB/s]. Based on a case study of an actual video production company, The experiment transfers up to three files simultaneously from an actual video production company example. The file size to be transferred in the experiment was set at 100 GB based on an example from a video production company. A dummy file containing a list of random text is used for the transfer. Dummy files are a substitute for user files and business data that are backed up. The reason for using a dummy file is to prepare a video file of exactly 100 GB. The backup time deadline in the experiment was set to one hour. This deadline is the default backup interval of existing backup software. The use case is a video production company. The correction is calculated outside of business hours to avoid conflicts with user file transfers. The correction calculation takes approximately three hours at a time, so the correction is calculated twice outside of business hours. The maximum backup time is used to prevent backup deadlines from being exceeded. The uncorrected backup times are shown in the following Table 1.

Table 1. Backup time and speed without correction

Experimental method	About Backup time [min:sec]	About Transfer speed [MB/s]
1 User+Backup	62:00	28.8
2 Users+Backup	63:42	28.1
3 Users+Backup	62:28	28.8

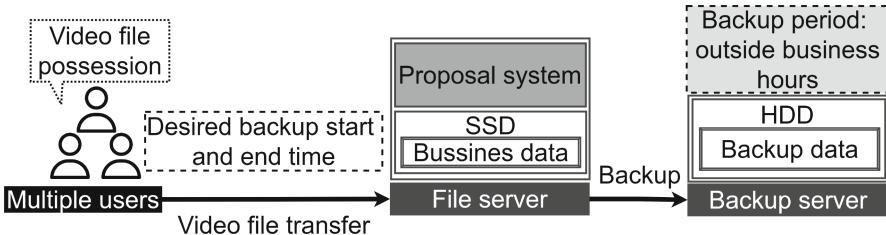
**Fig. 3.** How to use the proposal method in a video production company

Figure 3 shows a use-case summary. The proposed method is used in video production companies. The correction in the proposed method is calculated outside of business hours. Employees in a video production company transfer files to a file server during the backup period due to late-night overtime work. Employees can reduce the increase in file transfer time during backup by inputting the backup start and end times into the proposed system on the file server.

4 Experiments

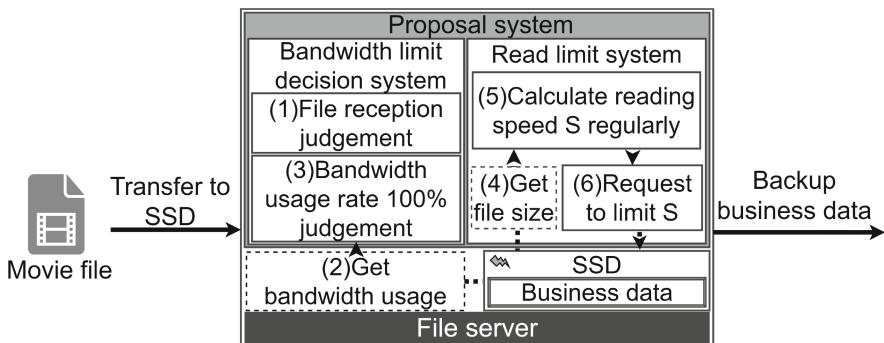
**Fig. 4.** Flow of the proposed method to control backup speed

Figure 4 shows the implementation of the system. Software implementing the proposed method is deployed on a file server. (1) File reception judgement and

(2) Get bandwidth usage are used as a bandwidth limit decision system. (1) File reception judgment determines whether video files are received during the backup period. The disk bandwidth utilization is obtained in (2). (3) Bandwidth usage rate 100% judgement in the Bandwidth limit decision system uses the disk bandwidth in (2) to determine whether the bandwidth usage rate is 100%. The proposed system starts controlling the read speed so the backup finishes on time when both (1) and (3) are satisfied. The reading control system includes (4) Get file size, (5)Calculate reading speed S regularly, and (6) Request to limit S. (4) Get file size obtains the Business data size. (5) Calculate reading speed S regularly calculates reading speed S. (6)Request to limit S request a limit on the backup read speed S after calculating the read speed.

Table 2. VM Composition

Role	vCPU [Core]	RAM [GB]	HDD [GB]	Average Transfer Speed [MB/s]
Users	2	8	120	112
Fileserver	2	8	400	112
Backupserver	2	8	120	112

Table 2 shows VM composition. VMs are used in the experiment. The roles of each VM are user A, user B, file server, and backup server, respectively. The average transfer rate is the average network speed measured by Iperf. Iperf is a tool for measuring network performance. The capacity of the file server is larger than the other VMs because it stores the files of users A and B and backs up their files.

Figure 5 shows the detail of the experiment. The experiment confirmed the increase in file transfer speed with the proposed method and whether the backup

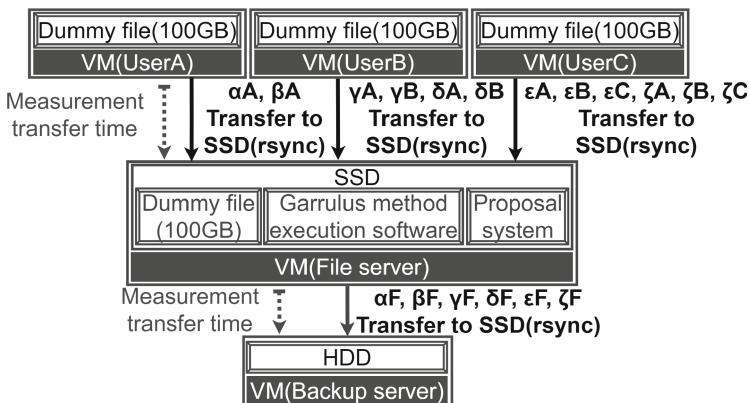


Fig. 5. Experiment to compare file transfer times of three methods for each number of users

time exceeds the deadline. Figure 5 shows that α to ζ are experimental methods. The experiment measures the file transfer time for the user. The experiment also measures the time that the file server takes the start of the backup to the end of the backup. Each experiment runs five times. This paper uses the median transfer time as the measurement time because the user's transfer time varies from experiment to experiment. Each server transfers 100 GB of files from an actual video production company. The experiment uses a dummy file containing a list of random texts, similar to the proposed basic experiment. The user transfers a 100 GB dummy file to the file server's SSD. The file server transfers the 100 GB dummy file to the backup server. User A transfers a dummy file to the file server in experiment α . The file server also transfers the dummy file to the backup server. β is an experiment in which the file server applies the proposed method in addition to experiment α . α_A to ζ_A are the file transfer time for user A. γ_B to ζ_B are the file transfer time for user B. α_F to ζ_F are the backup time. ϵ_C and ζ_C are the file transfer time for user C. Experiment γ differs from α in that user B also transfers a dummy file to the file server, and the file server applies the Garrulus method. Experiment δ differs from γ in that the file server applies the proposed method instead of the Garrulus method. Experiment ϵ differs from γ in that user C also transfers a dummy file to the file server. Experiment ζ differs from δ in that user C also transfers a dummy file to the file server.

5 Experimental Results

Figure 6 compares user file transfer times in experiment α and β . The vertical axis shows the user's file transfer time. The horizontal axis shows the experimental method. α_A takes about 20 min and 54 s. β_A was reduced by about 9.2% compared to α_A . The difference between the Garrulus and the proposed methods is that the correction for backup is changed according to the number of users transferring files. Therefore, the Garrulus and the proposed method gives the same result in the case of file transfer by one user.

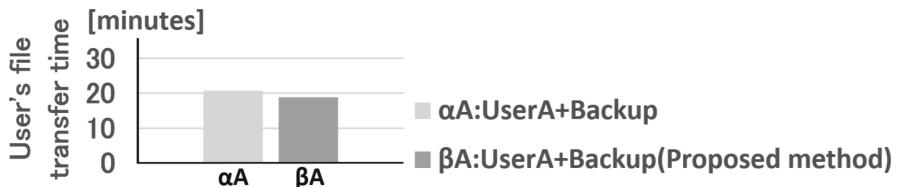


Fig. 6. Comparison of user file transfer times in experiment α and β

Figure 7 compares user file transfer times in experiment γ and δ . γ_A takes about 33 min and 48 s. δ_A takes about 31 min and 25 s. δ_A is reduced by about

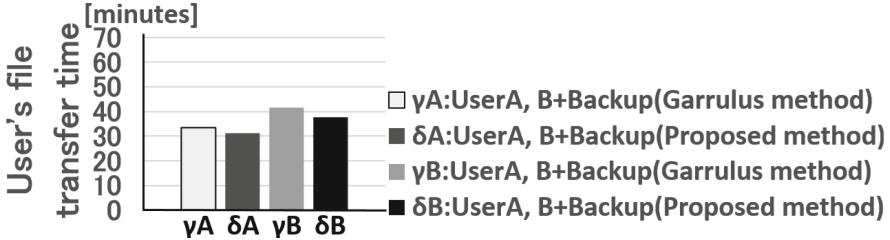


Fig. 7. Comparison of user file transfer times in experiment γ and δ

7.1% compared to γA . γB takes about 41 min and 48 s. δB takes about 38 min and 3 s. Therefore, δB is about 9.0% shorter than γB .

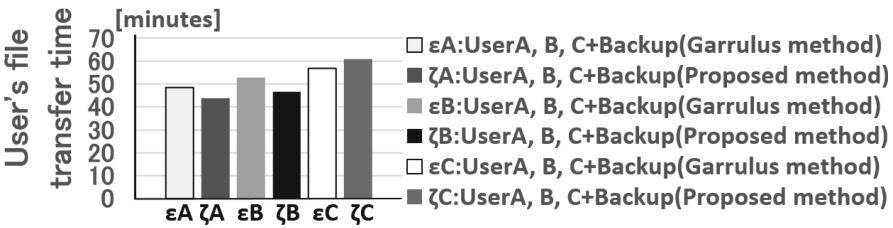


Fig. 8. Comparison of user file transfer times in experiment ϵ and ζ

Figure 8 compares user file transfer times in experiment ϵ and ζ . ϵC takes about 57 min and 13 s. ζC takes about 61 min and 25 s. ζC increased by about 6.8% compared to ϵC .

The user's file transfer time does not necessarily increase because the file transfer time of each user differs from experiment to experiment. Figure 8 shows that user C's file transfer time increased as user A's file transfer time decreased.

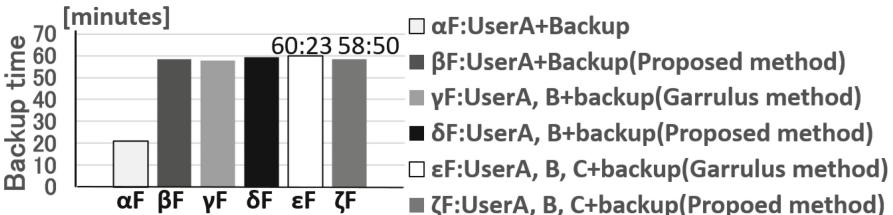


Fig. 9. Comparison of backup time

Figure 9 compares backup times. αF takes about 21 min and 7 s. βF takes about 59 min and 1 s. γF takes about 58 min and 8 s. δF takes about 59 min and 42 s. ϵF takes about 60 min and 23 s. ζF takes about 58 min and 50 s.

Table 3. Comparison of backup time for experiment ϵ and ζ

	Experiment ϵ [min:sec]	Experiment ζ [min:sec]
Backup time	60:23	58:50
Difference from the deadline	+00:23	-1:10

Table 3 compares backup times for experiment ϵ and ζ . Experiment ϵ shows that the Garrulus method exceeded the backup deadline of 60 min. The Garrulus method exceeded the deadline by 23 s. The backup time of the experiment ζ with the proposed method completed 1.9% faster than the backup deadline of 60 min. The proposed method completed the backup 1 min and 10 s earlier than the deadline. Therefore, the proposed method is superior to the Garrulus method because it strictly meets the backup deadline. On the other hand, the backup time was not on time, leaving room for improvement. The purpose of the proposed method is to complete the backup on time.

6 Discussion

The proposed method does not support bandwidth allocation for each user. The use case of this proposal is a video production company. User dissatisfaction with long file transfer times can be resolved by knowing which project the transferred files are from and prioritizing the transfer of files from the highest priority project.

The backup completion time increases as the number of users increases. The backup completion time during file transfer by one user is approximately 59 min and 1 s. The backup completion time during file transfer by two user is approximately 59 min and 42 s. Therefore, the backup time exceeds the deadline when four or more users transfer files simultaneously. Disk bandwidth can be secured by determining which project files are prioritized and interrupting file transfers that are not prioritized.

7 Conclusion

File transfers and backups increase file transfer time for users due to competing disk read and write operations. The proposed method maximizes disk writes by setting the backup transfer rate for disk reads to complete the backup until the deadline. The experimental results showed that the user's file transfer time was reduced from 48 min and 46 s to 44 min and 3 s by applying the proposal, the reduction of approximately 9.7%. The backup time did not exceed the deadline.

Acknowledgements. This work was supported by JSPS KAKENHI Grant Numbers JP23K11073, JP23K11087.

References

1. Hu, M., Guo, W., Hu, W.: Dynamic scheduling algorithms for large file transfer on multi-user optical grid network based on efficiency and fairness. In: 2009 Fifth International Conference on Networking and Services, pp. 493–498 (2009)
2. Kaiser, J., Meister, D., Gottfried, V., Brinkmann, A.: MCD: overcoming the data download bottleneck in data centers. In: 2013 IEEE Eighth International Conference on Networking, Architecture and Storage, pp. 88–97 (2013)
3. Malensek, M., Pallickara, S.L., Pallickara, S.: Alleviation of disk I/O contention in virtualized settings for data-intensive computing. In: 2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC), pp. 1–10 (2015)
4. Qin, C., Guo, W., Sun, W., Jin, Y., Hu, W.: Scheduling strategies for multiple optical grid applications based on scheduling span and fairness. In: Hu, W., Liu, S.-K., Ichi Sato, K., Wosinska, L. (eds.) Network Architectures, Management, and Applications VI , Vol. 7137, International Society for Optics and Photonics, SPIE, p. 713715 (2008)
5. Rahumed, A., Chen, H.C., Tang, Y., Lee, P.P., Lui, J.C.: A secure cloud backup system with assured deletion and version control. In: 2011 40th International Conference on Parallel Processing Workshops, IEEE, pp. 160–167 (2011)
6. Su, G.-M., Han, Z., Wu, M., Liu, K.R.: A scalable multiuser framework for video over OFDM networks: fairness and efficiency. IEEE Trans. Circ. Syst. Video Technol. **16**(10), 1217–1231 (2006)
7. Xia, R., Yin, X., Alonso Lopez, J., Machida, F., Trivedi, K.S.: Performance and availability modeling of ITSystems with data backup and restore. IEEE Trans. Dependable Secure Comput. **11**(4), 375–389 (2013)
8. Yao, J., Lu, P., Zhu, Z.: Minimizing disaster backup window for geo-distributed multi-datacenter cloud systems. In: 2014 IEEE International Conference on Communications (ICC), pp. 3631–3635 (2014)
9. Zhou, Z., Yang, X., Zhao, D., Rich, P., Tang, W., Wang, J., Lan, Z.: I/O-aware batch scheduling for petascale computing systems. In: 2015 IEEE International Conference on Cluster Computing, pp. 254–263 (2015)



Distributing Energy Consumption in Multi-interface Networks: Dimension of Cycle Space

Alessandro Aloisio^{1(✉)} and Diletta Cacciagrano²

¹ Department of International Humanities and Social Sciences, University of International Studies of Rome, UNINT, Via Cristoforo Colombo 200, 00147 Rome, Italy

alessandro.aloisio@unint.eu

² Division of Computer Science, University of Camerino, Via Madonna delle Carceri 9, 62032 Camerino, Italy

diletta.cacciagrano@unicam.it

Abstract. Some modern networks are set up using highly heterogeneous wireless devices. To make them work properly, selecting a subset of the available interfaces is required. This practical problem can be described by one of the well-known Multi-Interface network models. Among them is the Coverage model, where the main goal is to activate the cheapest subset of interfaces to establish all the desired links. Here, “cheapest” refers to energy consumption. This work focuses on the well-known Coverage in Multi-Interface network model. The network is represented by an undirected graph $G = (V, E)$, where each node corresponds to a device and each edge denotes a desired connection. Additionally, each node is equipped with a set of interfaces, and the objective is to find a subset of them such that every node has at least one common interface, minimizing the total energy cost. Since this problem has been proven to be NP -hard, we decided to analyze the case with respect to the dimension of the cycle space of G . Specifically, we provided a deterministic algorithm that returns a solution for the decision version of the problem, running in FPT-time relative to the sum of the number of available interfaces and the dimension of the cycle space.

1 Introduction

Nowadays, the world is teeming with diverse devices that can communicate with each other through one or more networks. Typically, each network relies on a single protocol, like Wi-Fi. However, there are situations where using multiple protocols is more practical to achieve the desired communication. For instance, consider a network set up when standard communications fail due to an issue,

This work is partially supported by the project ‘Soluzioni innovative per il problema della copertura nelle multi-interfacce e relative varianti’ - UNINT, and by the Italian National Group for Scientific Computation (GNCS-INDAM).

such as a blackout. In such scenarios, a variety of wireless devices can be used to create a network using the available interfaces each device offers. Once the network's structure is determined, we must select the appropriate protocols for each link.

This involves choosing a subset of interfaces that ensure reliable communication within the network. Additionally, since these devices are battery-powered, it is crucial to opt for a solution that minimizes energy consumption. This problem belongs to the class known as *Coverage* in Multi-Interface Networks, which has been extensively investigated over the last decade [1, 15, 17, 20, 21]. The main reason is that while the problem is easy to understand, finding a solution remains challenging even without considering energy usage.

A network of devices can be formalized as a graph $G = (V, E)$, in which the set of devices is denoted by V , and the set of potential connections is denoted by E . These connections are contingent upon the distance between the devices and their shared interfaces. The set of accessible interfaces for each device $v \in V$ is indicated by $\delta(v)$. The entire set of interfaces that are accessible throughout the network is denoted by $\bigcup_{v \in V} \delta(v)$, and is also known as $\{1, \dots, k\}$. If two devices' endpoints on the same edge share at least one active interface, a link is formed between them. Upon activation at a node v , an interface α uses $c(\alpha)$ of energy to remain active, hence enabling maximum communication bandwidth with any nearby devices that share the interface α .

The Coverage model, which takes into account the maximum number of interfaces that a device can activate in addition to the global cost, is the main emphasis of this study [17, 20, 21]. This extra restriction seeks to regulate the energy consumption of particular devices, providing a more nuanced view on cost optimization when limits are present [13].

A field where this model could find practical application is in military tactical networks [34].

1.1 Related Work

Recent years have seen a significant amount of study on multi-interface networks, with an emphasis on the advantages of using several interfaces per device in different circumstances. This study reexamines the core problems with conventional network optimization, particularly with regard to network connectivity and routing (e.g., [23, 25, 26, 28, 29]). Since [24], combinatorial issues in multi-interface wireless networks have been studied. Papers like [1, 13, 15, 27, 32] have examined the Coverage problem and its variations.

Problems with *Connectivity* have been dealt with in [22, 33]. The challenge is figuring out how to guarantee network connectivity at the lowest possible cost. Stated differently, its goal is to determine, at each node, a subset of the available interfaces that need to be activated in order to minimize the total cost of all the activated interfaces across the network and ensure a path between every pair of nodes in G .

See [2, 4–6, 13, 17, 19–22, 28, 33] for more information on the Coverage model. There, in some of them, the authors extended the model by adding two profit

functions—one for the devices and another for the links—that encourage the activation of interfaces inside the network. This new way of looking at things leads to stronger bonds while consuming less energy overall. In military tactical networks, ensuring network longevity is crucial, and the proposed models help improve energy distribution among nodes [34].

If the Coverage problem is approached in a decentralized manner, where devices act as agents with their own utility functions—such as the profits discussed in [16–18]—a different dynamic arises. This scenario can be modeled as a Coordination game or its more general form, known as Polymatrix games [3, 8, 9, 11], or as a more specific variant known as Hedonic games [10].

1.2 Our Results

This work is related to the extensively studied Coverage in Multi-Interface problem. More specifically, we explore a model that aims to balance energy consumption by limiting the number of interfaces that can be activated on each device. As in several previous papers [13, 18, 20], we will continue using the notation where p represents the maximum number of interfaces allowed on a device.

This problem has been shown to be difficult to solve even for $p = 2$. For this reason, we apply Fixed Parameter Tractability (FPT) Theory to better understand the model's complexity. Following the approach of previous works, we introduce a new parameter composed of the dimension of the cycle space plus the number of available interfaces.

This theory, of which we recall some basic definitions, seeks to identify one or more parameters to incorporate into the time complexity analysis, rather than relying solely on the size of the instance. This approach leads to more complexity classes and a more precise understanding of the problem's complexity.

Our results present an algorithm that demonstrates $CMI(2)$ is polynomial-time FPT-tractable when the parameter is the sum of the dimension of the cycle space and the number of available interfaces.

1.3 Outline

Section 2 presents some preliminary notations and definitions, as well as the formal model of our problem. Then, Sect. 3 describes the main result of the work, which is the inclusion of $CMI(2)$ in the polynomial fixed-parameter tractability class. Finally, we discuss the conclusions and suggest some ideas for future investigations.

2 Preliminaries

We will use standard notation for an undirected graph $G = (V, E)$, where the vertices are in V and the undirected edges are in E . Each node represents a device, and each edge represents a connection to be established. We also assume no multiple edges and no loops. As is usually done in graph theory, we will use

n to denote $|V|$ and m to denote $|E|$. To formally define the problem, we will use the following definitions.

Definition 1 (Assignment function). The subset of interfaces available on each device is determined by an assignment function $\delta: V \rightarrow 2^{\{1, \dots, k\}}$, where $\delta(u) \cap \delta(v) \neq \emptyset$ must hold for each edge $\{u, v\}$ in E .

Definition 2 (Activation function). The subset of interfaces active on each device is determined by an activation function $\delta_A: V \rightarrow 2^{\{1, \dots, k\}}$, where $\delta_A(u)$ is a subset of $\delta(u)$ for every vertex in V .

Definition 3 (Feasible Activation function). A feasible activation function is defined as $\delta_A: V \rightarrow 2^{\{1, \dots, k\}}$, where $\delta_A(u)$ is a subset of $\delta(u)$ for every vertex in V , and $\delta_A(u) \cap \delta_A(v) \neq \emptyset$ for every edge $\{u, v\}$ in E .

Definition 4 (Cost function). The cost associated with each type of interface is determined by a cost function $c: \{1, \dots, k\} \rightarrow \mathbb{R}_{>0}$.

We are now ready to give the formulation of $CMI(p)$.

<hr/> <i>CMI(p): Coverage in Multi-Interface Networks</i> <hr/>	
Input:	A undirected graph $G = (V, E)$, an assignment function $\delta: V \rightarrow 2^{\{1, \dots, k\}}$, an interface cost function $c: \{1, \dots, k\} \rightarrow \mathbb{R}_{>0}$, and an integer $p \geq 1$.
Coverage:	A feasible activation function $\delta_A: V \rightarrow 2^{\{1, \dots, k\}}$ such that $ \delta_A(v) \leq p$.
Task:	Find a coverage that minimize the total cost $c(\delta_A) = \sum_{v \in V} \sum_{\alpha \in \delta_A(v)} c(\alpha)$.

In Fig. 1 is shown a graph $G = (V, E)$ depicting a sample network of five devices, where each interface type is represented by a number from 1 to 4. We can consider two variations of the problem discussed: one where the cost function c can take any value in $\mathbb{R}_{>0}$, and another where $c(\alpha) = 1$ for all $\alpha \in \{1, \dots, k\}$ (known as the *unit cost case*). In both scenarios, we assume $k \geq 2$, since the case where $k = 1$ is trivial, with all nodes needing to activate their only available interface.

As shown in [13], $CMI(p)$ is a particular case of the more general $CMI \infty$ problem (see [26, 32]), in which each node can activate up to p interfaces. Interestingly, the simple case where $p = 2$ is generally more difficult than $CMI \infty$. Nevertheless, some classes of graphs are easier to handle. For example, in trees and complete graphs, $CMI \infty$ has been proven to be APX-hard and not approximable within $O(\log k)$, respectively, while $CMI 2$ can be solved in polynomial time.

CMI(p): Coverage in Multi-Interface Networks (decision version)

Input: A undirected graph $G = (V, E)$, an assignment function $\delta: V \rightarrow 2^{\{1, \dots, k\}}$, an interface cost function $c: \{1, \dots, k\} \rightarrow \mathbb{R}_{>0}$, and two integers $p, l \geq 1$.

Coverage: A feasible activation function $\delta_A: V \rightarrow 2^{\{1, \dots, k\}}$ such that $|\delta_A(v)| \leq p$.

Question: Is there a coverage with a total cost $c(\delta_A) = \sum_{v \in V} \sum_{\alpha \in \delta_A(v)} c(\alpha) \leq l$?

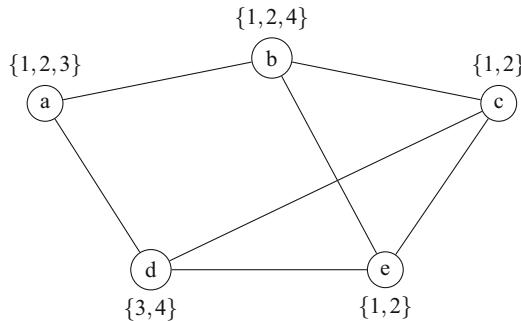


Fig. 1. A graph $G = (V, E)$ depicting a sample network of five devices, where each interface type is represented by a number from 1 to 4.

2.1 Parameterized Problems

We will briefly recall some theory that we will use in Sect. 3. As noted in [30], a decision problem can be represented as $Q \subseteq \Sigma^*$, where Σ^* is the set of all possible strings that can be formed using a non-empty alphabet Σ .

Definition 5 ([30]). Given a finite alphabet Σ :

- We call $\kappa: \Sigma^* \rightarrow \mathbb{N}$, which is polynomial-time computable, a parameterization of Σ^* .
- We call (Q, κ) a parameterized problem (over Σ), where Q is a subset of Σ^* , Σ^* represents all the strings that can be formed with Σ , and κ is a parameterization.

An instance of a parameterized problem (Q, κ) is represented by $x \in \Sigma^*$, with the corresponding parameter denoted as $\kappa(x)$.

2.2 Fixed-Parameter Tractability

We will also recall the formal definition of an FPT algorithm, which we will use in Sect. 3.

Definition 6 ([30]). Let Σ be a finite alphabet and $\kappa : \Sigma^* \rightarrow \mathbb{N}$ a parameterization:

- An algorithm \mathbb{A} with input alphabet Σ is an FPT-algorithm with respect to κ if there is a computable function $f : \mathbb{N} \rightarrow \mathbb{N}$ and a polynomial $p \in \mathbb{N}_0[X]$ such that for every $x \in \Sigma^*$, the running time of \mathbb{A} on input x is at most

$$f(\kappa(x)) \cdot p(|x|)$$

- A parameterized problem (Q, κ) is fixed-parameter tractable if there is an FPT-algorithm with respect to κ that decides Q . FPT denotes the class of all fixed-parameter tractable problems.

Informally, fixed-parameter tractability theory offers a more sophisticated and comprehensive approach to determining the complexity of problems. Unlike classical complexity theory, which is one-dimensional, this approach is two-dimensional, as it seeks to identify a parameter that influences the time complexity function. Specifically, the polynomial part of the complexity should depend only on the parameter and not on the size of the instance.

For the case where f is also a polynomial, \mathbb{A} is denoted as *polynomial FPT-algorithm* with respect to κ [31].

3 FPT Algorithm

In this section, we describe a deterministic algorithm that solves the decision version of the problem in FPT-time relative to the number of available interfaces plus the dimension of the cycle space. In Fig. 2 are shown in black color the edges of the graph of Fig. 1, which induce a spanning tree. While in Fig. 3 are shown cycle bases for the graph of Fig. 1. The FPT algorithm can be used to solve the minimization version of the problem.

Theorem 1. *There is a deterministic algorithm that, given an instance of $dCMI(2)$ with $G = (V, E)$ as the underlying graph, computes an optimal solution in $O(t \cdot k^4 \cdot \Delta \cdot n)$ time, where t is the dimension of the cycle space of $G = (V, E)$.*

Proof. Let t be the dimension of the cycle space of G . We can find a subset F of t edges whose removal leaves a forest T in polynomial time. Any coverage of G needs to use one or two interfaces to allow connection throughout any edge e in F . Let \mathcal{A} be the family of t subsets $A(e)$ of at most two interfaces contained in $\delta(u) \cap \delta(v)$, where u and v are the endpoints of e .

For a given family \mathcal{A} , we can solve $dCMI(2)$ on the forest T using an algorithm based on the one provided in [14, 15] for trees.

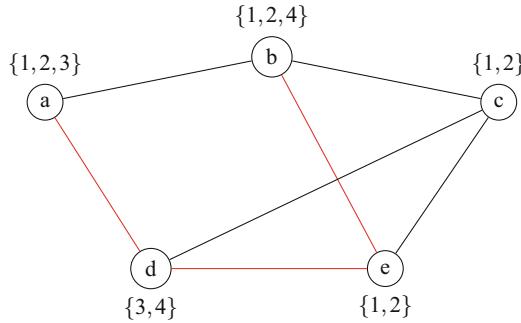


Fig. 2. The edges of the graph in Fig. 1 that induce a spanning tree are shown in black.

We can assume T is a tree; otherwise, we can solve each subtree of the forest individually. We arbitrarily choose an internal node r of T as the root and orient all tree edges away from the root. We will continue to refer to this directed tree as T .

Let v be a node of T . We use $T(v)$ to denote the subtree consisting of v and all its descendants. Let $f(v, B(v))$ represent the minimum cost of $dCMI(2)$ on $T(v)$, where $B(v)$ is a set of at most 2 interfaces belonging to $\delta(v)$ and containing every $A(e)$ for each edge e that includes v . The second requirement is to ensure communication through each e of F , adhering to the constraints specified by $A(e)$.

Case: leaf.

For every leaf v of T , we can compute $f(v, B(v))$ for a given $B(v) \subseteq \delta(v)$ such that:

- $|B(v)| = 1$ if there are no edges in F incident to v ;
- $|B(v)| \leq 2$ otherwise.

In the first case, we only need to activate one interface at v because the degree of v is one. In the second case, where the degree of v is at least two, we can use up to two interfaces. We can now compute $f(v, B(v))$ for a leaf v and a specific $B(v)$ as follows:

$$\begin{aligned}
 f(v, B(v)) &= c(B(v)) && \text{if there are no edges in } F \text{ incident to } v; \\
 f(v, B(v)) &= c(B(v)) && \text{if there is at least one edge in } F \text{ incident to } v \text{ and} \\
 &&& A(e) \subseteq B(v) \text{ for every } e \text{ in } F \text{ with } v \in e; \\
 f(v, B(v)) &= +\infty && \text{otherwise.}
 \end{aligned}$$

Case: internal node.

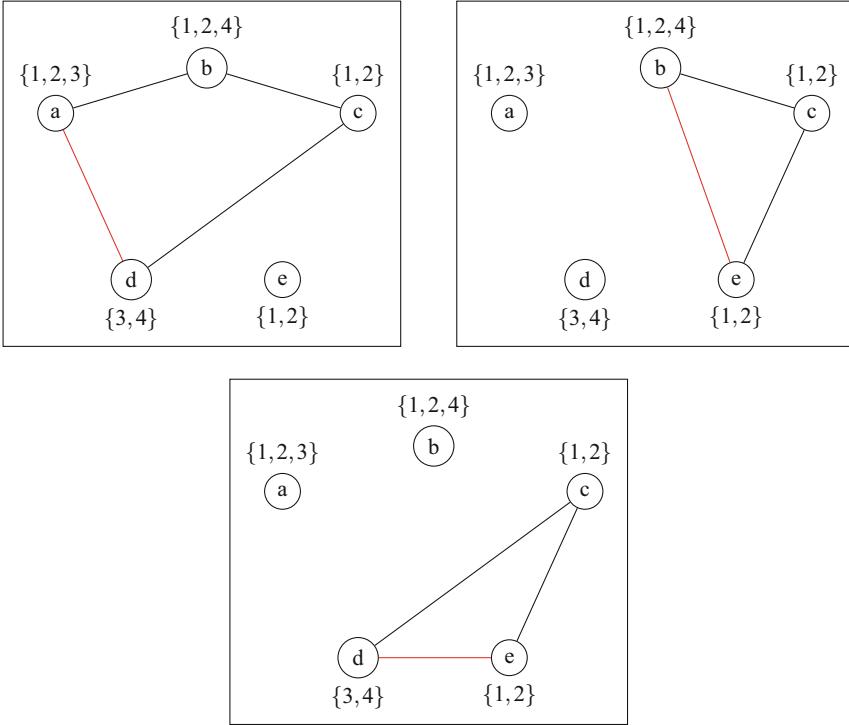


Fig. 3. Cycle bases for the graph of Fig. 1.

For every internal node v of T , and any subset $B(v) \subseteq \delta(v)$ of at most 2 interfaces, we solve the following minimization problem.

$$\begin{aligned} \min \quad & \sum_{u \in S(v)} f(u, B(u)) + c(B(v)) \\ \text{s.t.} \quad & B(u) \cap B(v) \neq \emptyset, \quad \forall u \in S(v) \\ & A(e) \subseteq B(v), \quad \forall e \in F: v \in e \end{aligned} \tag{1}$$

The objective function is the sum of the minimum costs over all children u of v , plus the cost associated with $B(v)$. The first constraint ensures that v can communicate with each of its children, denoted by the set $S(v)$. The second constraint ensures that v can communicate across every edge e incident to it using the interfaces contained in $A(e)$.

At the end of the algorithm, we have a positive answer for $dCMI(2)$ if the minimum value $f(r, B(r))$ for every admissible subset $B(r) \subseteq \delta(r)$ with a cardinality of at most 2 is less than or equal to l ; otherwise, the answer is negative. After computing $f(r, B(r))$, a coverage can be found by tracing back to the leaves using standard dynamic programming techniques.

Computational Complexity.

We now analyze the time complexity of the algorithm. For a leaf v , we consider at most $k + \binom{k}{2}$ subsets $B(v)$, which includes all subsets of cardinality one or two. This results in a time complexity of $O(k^2 \cdot t)$ for processing a leaf since the time complexity for each set is $O(t)$.

For an internal node v , we also consider $k + \binom{k}{2}$ subsets $B(v)$ and solve a problem as described in Problem (1), which requires $O(\deg(v))$ time. This leads to a time complexity of $O(k^2 \cdot \Delta \cdot n)$ for a specific family of subsets \mathcal{A} , where Δ is the maximum degree of G .

Since there are at most $O(k^{2t})$ families \mathcal{A} , we can conclude that the overall time complexity of the algorithm is $O(k^{2t+2} \cdot \Delta \cdot n)$.

Corollary 1. *The problem $dCMI(2)$ is in FPT with respect to $t + k$, where t is the dimension of the cycle space of the underlying graph.*

We conclude the section with the following corollary.

Corollary 2. *There is an deterministic algorithm that, given an instance of $CMI(2)$ with G as the underlying graph, computes an optimal solution in $O(k^{2t+2} \cdot \Delta \cdot n)$ time, where t is the dimension of the cycle space of G .*

4 Concluding Remarks

This study delves into one of the classic problems in the class of Multi-Interface networks: Coverage. We chose a model with a limit on the number of interfaces that can be activated on each device. This constraint aims to balance energy consumption across the network, thereby reducing battery drain.

This model has been proven to be hard to solve even when $p = 2$. Therefore, we continued the analysis of the problem through the lens of Fixed Parameter Tractability (FPT) theory. FPT provides tools that can offer a more refined complexity analysis of the problem than classical complexity theory.

Among all possible parameters, we combined the number of available interfaces with the dimension of the cycle space. The cycle space is also an important measure for describing the structure of a network. Since this model can be applied in various practical situations where setting up a heterogeneous wireless network is necessary, it is crucial to understand how to find optimal solutions when the network presents certain structural characteristics.

Future lines of research include finding more parameters and structural properties of graphs to identify both positive and negative results. For example, we believe it is worthwhile to consider local treewidth and clique-width. Another branch of investigation could involve the use of nonlinear analysis (e.g., quadratic) or the introduction of multiple objective functions. Finally, as mentioned above, we believe it is beneficial to explore the similarities between $CMI(p)$ viewed as a game and classic games like polymatrix and its variations.

References

1. Aloisio, A.: Coverage subject to a budget on multi-interface networks with bounded carving-width. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) WAINA 2020. AISC, vol. 1150, pp. 937–946. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44038-1_85
2. Aloisio, A.: Coverage subject to a budget on multi-interface networks with bounded carving-width. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) WAINA 2020. AISC, vol. 1151, pp. 937–946. Springer, Cham (2020)
3. Aloisio, A.: Distance hypergraph polymatrix coordination games. In: Proceedings of 22nd Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pp. 2679–2681 (2023)
4. Aloisio, A.: Algorithmic aspects of distributing energy consumption in multi-interface networks. In: Conference on Advanced Information Networking and Applications (AINA), vol. 204, pp. 114–123. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-57942-4_13
5. D'Angelo, G., Di Stefano, G., Navarra, A.: Min-Max coverage in multi-interface networks. In: Černá, I., Gyimóthy, T., Hromkovič, J., Jefferey, K., Králović, R., Vukolić, M., Wolf, S. (eds.) SOFSEM 2011. LNCS, vol. 6543, pp. 190–201. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-18381-2_16
6. Aloisio, A.: On balancing energy consumption in multi-interface networks. In: Proceedings of the 26th Italian Conference on Theoretical Computer Science, Torino, Italy, 11–13 September 2024. vol. 3811 pp. 158–165. Springer, Heidelberg (2024)
7. Aloisio, A.: Min-Max Coverage in Multi-interface Networks: pathwidth. In: Proceeding of the 19th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2024), San Benedetto, Italy, 13–15 November 2024. Springer (2024, in press)
8. Aloisio, A., Flammini, M., Kodric, B., Vinci, C.: Distance polymatrix coordination games. In: Proceedings of 30th International Joint Conference on Artificial Intelligence (IJCAI), pp. 3–9 (2021)
9. Aloisio, A., Flammini, M., Kodric, B., Vinci, C.: Distance polymatrix coordination games (short paper). In: SPIRIT co-located with 22nd International Conference on AIIXIA 2023, Rome, Italy, 7–9 November 2023, vol. 3585 (2023)
10. Aloisio, A., Flammini, M., Vinci, C.: The impact of selfishness in hypergraph hedonic games. In: Proceedings of 34th Conference on Artificial Intelligence (AAAI), pp. 1766–1773 (2020)
11. Aloisio, A., Flammini, M., Vinci, C.: Generalized distance polymatrix games. In: Fernau, H., Gaspers, S., Klasing, R. (eds.) SOFSEM 2024, vol. 14519, pp. 25–39. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-52113-3_2
12. Aloisio, A., Piselli, F.: Min-Max coverage in multi-interface networks: series-parallel graphs. In: Barolli, L. (ed.) Advances on Broad-Band Wireless Computing, Communication and Applications, pp. 212–222. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-76452-3_21
13. Aloisio, A., Navarra, A.: Balancing energy consumption for the establishment of multi-interface networks. In: Proceedings of 41st International Conference on Current Trends in Theory and Practice of Computer Science, (SOFSEM), vol. 8939, pp. 102–114 (2015)
14. Aloisio, A., Navarra, A.: Balancing energy consumption for the establishment of multi-interface networks. In: Italiano, G.F., Margaria-Steffen, T., Pokorný, J., Quisquater, J.-J., Wattenhofer, R. (eds.) SOFSEM 2015. LNCS, vol. 8939, pp.

- 102–114. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46078-8_9
- 15. Aloisio, A., Navarra, A.: Budgeted constrained coverage on bounded carving-width and series-parallel multi-interface networks. *Internet Things* **11**, 100259 (2020)
 - 16. Aloisio, A., Navarra, A.: Budgeted constrained coverage on series-parallel multi-interface networks. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) AINA 2020. AISC, vol. 1151, pp. 458–469. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44041-1_41
 - 17. Aloisio, A., Navarra, A.: Constrained connectivity in bounded X-width multi-interface networks. *Algorithms* **13**(2), 31 (2020)
 - 18. Aloisio, A., Navarra, A.: On coverage in multi-interface networks with bounded pathwidth. In: Barolli, L. (ed.) AINA 2024, vol. 204, pp. 96–105. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-57942-4_11
 - 19. Aloisio, A., Navarra, A.: Parameterized complexity of coverage in multi-interface iot networks: pathwidth. *Internet Things* **28**, 101353 (2024)
 - 20. Aloisio, A., Navarra, A., Mostarda, L.: Distributing energy consumption in multi-interface series-parallel networks. In: Barolli, L., Takizawa, M., Xhafa, F., Enokido, T. (eds.) WAINA 2019. AISC, vol. 927, pp. 734–744. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15035-8_71
 - 21. Aloisio, A., Navarra, A., Mostarda, L.: Energy consumption balancing in multi-interface networks. *J. Ambient. Intell. Humaniz. Comput.* **11**(8), 3209–3219 (2020)
 - 22. Athanassopoulos, S., Caragiannis, I., Kaklamanis, C., Papaioannou, E.: Energy-efficient communication in multi-interface wireless networks. *Theory Comput. Syst.* **52**, 285–296 (2013)
 - 23. Bahl, P., Adya, A., Padhye, J., Walman, A.: Reconsidering wireless systems with multiple radios. *SIGCOMM Comput. Commun. Rev.* **34**(5), 39–46 (2004)
 - 24. Caporuscio, M., Charlet, D., Issarny, V., Navarra, A.: Energetic performance of service-oriented multi-radio networks: issues and perspectives. In: Proceedings of 6th International Workshop on Software and Performance (WOSP), pp. 42–45. ACM (2007)
 - 25. Cavalcanti, D., Gossain, H., Agrawal, D.: Connectivity in multi-radio, multi-channel heterogeneous ad hoc networks. In: Proceedings of 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1322–1326. IEEE (2005)
 - 26. D'Angelo, G., Di Stefano, G., Navarra, A.: Multi-interface wireless networks: complexity and algorithms. In: Ibrahem, S.R., El Emery, M.M. (eds.) Wireless Sensor Networks: From Theory to Applications, pp. 119–155. CRC Press, Taylor & Francis Group (2013)
 - 27. D'Angelo, G., Stefano, G.D., Navarra, A.: Minimize the maximum duty in multi-interface networks. *Algorithmica* **63**(1–2), 274–295 (2012)
 - 28. Draves, R., Padhye, J., Zill, B.: Routing in multi-radio, multi-hop wireless mesh networks. In: Proceedings of 10th International Conference on Mobile computing and networking (MobiCom), pp. 114–128. ACM (2004)
 - 29. Faragó, A., Basagni, S.: The effect of multi-radio nodes on network connectivity—a graph theoretic analysis. In: Proceedings of 19th International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC), pp. 1–5. IEEE (2008)
 - 30. Flum, J., Grohe, M.: Parameterized Complexity Theory. Springer, Heidelberg (2006). <https://doi.org/10.1007/3-540-29953-X>
 - 31. Gurski, F., Rehs, C., Rethmann, J.: Knapsack problems: a parameterized point of view. *Theoret. Comput. Sci.* **775**, 93–108 (2019)

32. Klasing, R., Kosowski, A., Navarra, A.: Cost minimization in wireless networks with a bounded and unbounded number of interfaces. *Networks* **53**(3), 266–275 (2009)
33. Kosowski, A., Navarra, A., Pinotti, M.: Exploiting multi-interface networks: connectivity and cheapest paths. *Wirel. Netw.* **16**(4), 1063–1073 (2010)
34. Perucci, A., Autili, M., Tivoli, M., Aloisio, A., Inverardi, P.: Distributed composition of highly-collaborative services and sensors in tactical domains. In: Ciancarini, P., Mazzara, M., Messina, A., Sillitti, A., Succi, G. (eds.) SEDA 2018. AISC, vol. 925, pp. 232–244. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-14687-0_21



Min-Max Coverage in Multi-interface Networks: Pathwidth

Alessandro Aloisio^(✉)

Department of International Humanities and Social Sciences, University of International Studies of Rome, UNINT, Via Cristoforo Colombo 200, 00147 Rome, Italy
alessandro.aloisio@unint.eu

Abstract. This paper delves into one of the problems within the class known as Multi-Interface Networks, specifically studying Coverage in Multi-Interface Networks. This class of problems has been extensively investigated by scholars due to its many applications in real-world scenarios. In this paper, we take into consideration the pathwidth of the networks. The primary reason for this focus is that the problem is challenging to solve on general networks. Additionally, we chose to analyze it using Fixed Parameter Tractability (FPT) theory. We show that the problem is in FPT with respect to the number of available interfaces plus the pathwidth.

1 Introduction

We decided to study the Coverage in Multi-Interface Networks problem, which belongs to the vast class known as Multi-Interface Networks. This class was introduced less than twenty years ago and quickly received significant attention due to its theoretical challenges and wide range of practical applications. Among the possible applications is the field of the Internet of Things (IoT), as one of the goals of Multi-Interface Networks is to build networks of heterogeneous wireless devices that exploit different communication protocols and enhance overall performance.

When dealing with wireless devices, a crucial aspect is energy consumption, as the majority of these devices are battery-powered. This consideration becomes even more relevant when the devices are small sensors.

Another field of application arises when standard networks fail due to a calamity, creating the need to quickly establish a wireless network using the available devices. These devices are often different from one another and have various types of communication interfaces.

Among the vast class of Multi-Interface Networks, we are tackling the problem known as Min-Max Coverage in Multi-Interface Networks [31]. Its main goal is to establish the desired communication links while minimizing the maximum energy consumption across all devices. This problem is commonly represented using an undirected

This work is partially supported by the project ‘Soluzioni innovative per il problema della copertura nelle multi-interfacce e relative varianti’ - UNINT, and by the Italian National Group for Scientific Computation (GNCS-INDAM).

graph $G = (V, E)$, where each vertex states for a device, and each edge states a connection to be established. Additionally, each vertex is associated with a set of available interfaces, typically represented as binary variables (activated or deactivated). A communication link is successful when the devices at both ends activate the same type of interface. Each interface is associated with a cost that defines its energy consumption. The energy consumption on a specific device is the sum of the costs of the active interfaces.

The problem we address in this paper has been shown to be *NP*-hard, even in the case of bounded unit costs. This raises the question of the underlying source of the problem's difficulty. To explore this, we applied the Fixed Parameter Tractability (FPT) theory [37], which offers a more detailed understanding of a problem's complexity than traditional computational complexity theory. Unlike classical approaches, which define tractability as the ability to solve problems in polynomial time, FPT broadens this definition by allowing algorithms that may behave non-polynomially, but only in a way that is dependent on a specific parameter.

Furthermore, due to the problem's inherent difficulty, we chose to investigate the class of networks by considering pathwidth, a well-established measure in graph and network theory.

1.1 Related Work

In recent years, significant research has been dedicated to Multi-Interface Networks, emphasizing the advantages of utilizing multiple interfaces per device across different scenarios. This research revisits core problems in traditional network optimization, with a particular emphasis on network connectivity and routing (e.g., [25, 30, 32, 35, 36]). Since the study in [28], combinatorial problems in Multi-Interface wireless networks have become a focal point of investigation.

Connectivity problems have been addressed in [23, 42], where the main difficulty is ensuring network connectivity at minimal cost. This involves determining, at each node, which subset of available interfaces should be activated to minimize the total cost across the network, while still ensuring a communication path between every pair of nodes in G .

Several studies, such as [1, 7, 14, 15, 22, 34, 40], and [2, 4, 6, 17, 19–21, 40] examined the *Coverage* problem and its various forms. Additionally, works like [16–18, 20] expand the model by introducing two profit functions—one for devices and another for links—that incentivize the activation of interfaces within the network. This approach strengthens connectivity while optimizing overall energy consumption. A problem that appears quite similar but has significant differences is the weighted edge coloring problem [5, 13, 27].

Both Connectivity and Coverage have been studied in terms of minimizing the maximum cost incurred by any individual node [34]. The *Cheapest path* problem, an extension of the classic Shortest Path problem in traditional networks, is explored in [42]. Moreover, classical problems like Maximum Matching [41] and Flow [24, 33] have been investigated within this framework.

When the Coverage problem is approached in a decentralized manner, where devices act as agents with their own utility functions—such as the profits discussed

in [16, 18]-a different dynamic arises. This scenario can be modeled as a Coordination game or its more general form, Polymatrix games [3, 8, 9, 11, 12], or as the more specific variant, Hedonic games [10].

In military tactical networks, network longevity is paramount, and the proposed models contribute to better energy balance among nodes [43].

1.2 Our Results

We chose to explore the well-studied problem of Min-Max Coverage in Multi-Interface Networks [31]. Our focus is specifically on the scenario where the network is represented as an undirected graph with a path decomposition and no restrictions on the number of available interfaces. We refer to this model as *MMCov*, consistent with the terminology in [31]. Since this problem was proven to be *NP*-hard in [31], we opted to study it using recent advances in Fixed-Parameter Tractability (FPT) theory [37] to gain deeper insight into the underlying complexity.

In recent years, parameterized complexity theory, which evaluates running time by considering specific parameters of the input instance rather than just the input size, has gained popularity due to its ability to provide a more precise analysis of computational complexity.

We introduce a deterministic algorithm that resolves the decision version of the problem in FPT-time, with respect to the dimension of the pathwidth plus the total number of available interfaces. This algorithm is also capable of addressing the optimization version of the problem.

1.3 Outline

In Sect. 2, we describe some preliminary notation and the formal definition of *MMCov* problem. Then, the main results are given in Sect. 3, where we provide a detailed description of an optimal algorithm that runs in FPT time. The last section (Sect. 4) gives final consideration and describe some new research directions and open problems.

2 Preliminaries

As commonly used in graph theory, we use $G = (V, E)$ to denote a graph, which we suppose to be undirected, with no loops, and without multiple edges. We also assume that there are no isolated vertices. Moreover, the number of vertices $|V|$ is also denoted by n while the number of edges $|E|$ is also denoted by m .

The following definitions are used to specify the set of interfaces available on each device and those that are activated in a particular solution.

Definition 1 (Availability function). The set of interfaces available on each device/vertex is specified by the availability function $\delta: V \rightarrow 2^{[k]}$.

Definition 2 (Activation function). The set of active interfaces on each device/vertex is defined by an activation function $\delta_A: V \rightarrow 2^{[k]}$, which depends on an availability function δ , such that $\delta_A(u) \subseteq \delta(u)$.

Definition 3 (Feasible Activation function). An activation function $\delta_A : V \rightarrow 2^{[k]}$, defined over an availability function δ is called feasible if $\delta_A(u) \cap \delta_A(v) \neq \emptyset$ for every edge $\{u, v\}$ in E .

The energy consumption of an active interface is defined by the following function, which assigns the same cost to each interface of the same type.

Definition 4 (Cost function). The cost of each (type) of interface is defined by $c : [k] \rightarrow \mathbb{R}_{>0}$.

We are ready to give the formal definition of our problem (*MMCov*).

MMCov: Min-Max-Cost Coverage in Multi-Interface Networks

Input: An undirected graph $G = (V, E)$, an availability function $\delta : V \rightarrow 2^{[k]}$ such that $\delta(u) \cap \delta(v)$ for each edge $\{u, v\} \in E$, and an interface cost function $c : [k] \rightarrow \mathbb{R}_{>0}$.

Coverage: A feasible activation function $\delta_A : V \rightarrow 2^{[k]}$ with respect to δ .

Goal: Find a coverage that minimizes the maximum cost of the active interfaces across all vertices, where the cost is given by $c(\delta_A) = \max_{v \in V} \sum_{\beta \in \delta_A(v)} c(\beta)$.

In Fig. 1 is shown a sample of a heterogeneous network. It's essential to examine two distinct scenarios for the problem at hand: the first where the cost function can assume any positive real number, and the second where $c(\beta) = 1$ for all $\beta \in [k]$ (referred to as the *unit cost scenario*). In both scenarios, we assume $k \geq 2$, as the case when $k = 1$ is straightforward, having a single solution where each node activates its sole interface. The complexity of the problem in these scenarios is summarized by the following two theorems.

Theorem 1 ([31]). *MMCov is NP-hard even when restricted to the bounded unit cost case, for any fixed $\Delta \geq 5$ and $k \geq 16$.*

Theorem 2 ([31]). *In the unit cost case with $k \leq 3$, MMCov is optimally solvable in $O(n)$ time.*

We will now present the formal definition of the decision version of our problem, referred to as *dMMCov*, which is the decision variant of *MMCov*.

dMMCov: Min-Max-Cost Coverage (decision version)

Input: An undirected graph $G = (V, E)$, an availability function $\delta : V \rightarrow 2^{[k]}$ such that $\delta(u) \cap \delta(v)$ for each edge $\{u, v\} \in E$, an interface cost function $c : [k] \rightarrow \mathbb{R}_{>0}$, and an integer $l \geq 1$.

Coverage: A feasible activation function $\delta_A : V \rightarrow 2^{[k]}$ with respect to δ .

Question: Is there a coverage with cost $c(\delta_A) = \max_{v \in V} \sum_{\beta \in \delta_A(v)} c(\beta)$ less than or equal to l ?

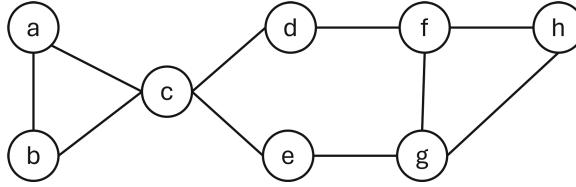


Fig. 1. Sample of a heterogeneous network.

2.1 Parameterized Problems

We will use Σ to represent an alphabet, and Σ^* to indicate the set of all possible strings that can be formed from a non-empty alphabet Σ . Additionally, the subset $Q \subseteq \Sigma^*$ will be used to describe a decision problem, as detailed in [37].

Definition 5 ([37]). Given a finite alphabet Σ :

A function $\kappa: \Sigma^* \rightarrow \mathbb{N}$ that can be computed in polynomial time is called a parameterization of Σ^* ;

A pair (Q, κ) , where Q is a subset of Σ^* (strings over Σ), is called a parameterized problem.

Furthermore, for a given parameterized problem (Q, κ) , we refer to each $x \in \Sigma^*$ as an *instance* of the problem, with $\kappa(x)$ representing its corresponding *parameters*.

2.2 Fixed-Parameter Tractability

The following definition will be used to solve our problem.

Definition 6 ([37]). Let Σ be a finite alphabet and $\kappa: \Sigma^* \rightarrow \mathbb{N}$ a parameterization:

An algorithm \mathbb{A} with input alphabet Σ is an FPT-algorithm with respect to κ if there is a computable function $f: \mathbb{N} \rightarrow \mathbb{N}$ and a polynomial $p \in \mathbb{N}_0[x]$ such that for every $x \in \Sigma^*$, the running time of \mathbb{A} on input x is at most

$$f(\kappa(x)) \cdot p(|x|)$$

A parameterized problem (Q, κ) is fixed-parameter tractable if there is an FPT-algorithm with respect to κ that decides Q . FPT denotes the class of all fixed-parameter tractable problems.

Fixed-parameter tractability (FPT) theory can be seen as an extension of traditional complexity theory. It focuses on identifying specific parameters within a typically hard decision problem that, when kept small, make the problem tractable. In this framework, the input size contributes to a polynomial function $p(|x|)$, while the parameters are associated with a function $f(\kappa(x))$. If f is also a polynomial, then the algorithm \mathbb{A} is classified as a *polynomial FPT-algorithm* with respect to the parameter κ [39].

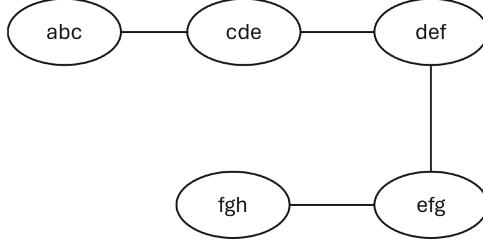


Fig. 2. A path decomposition of the sample network given in Fig. 1.

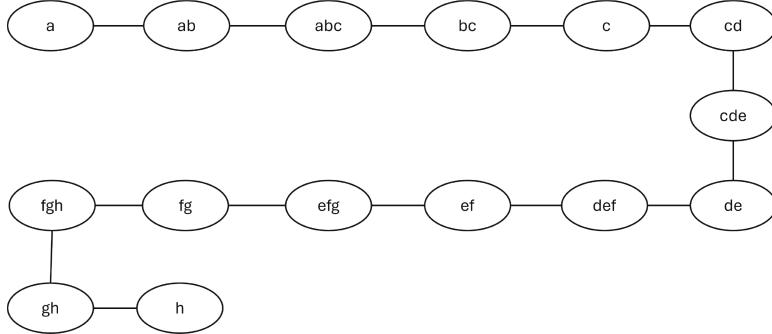


Fig. 3. A nice path decomposition of the sample network given in Fig. 1.

2.3 Graphs with Bounded Pathwidth

In this section, we revise a formal definition of the *pathwidth* of a graph.

Definition 7 ([38]). A path decomposition of a graph $G = (V, E)$ is a set $\mathcal{P} = (Y_1, \dots, Y_r)$ of subsets of V , that is $Y_i \subseteq V$ for each $i \in [r]$, called *bags*, such that: (i) for every $u \in V$ there exists $i \in [r]$ with $u \in Y_i$; (ii) for every $\{u, v\} \in E$, there exists $i \in [r]$ with $\{u, v\} \in Y_i$; (iii) for every three bags Y_i , Y_j , and Y_k , with $i \leq j \leq k$, it holds that $Y_i \cap Y_k \subseteq Y_j$.

The *width* of a path decomposition \mathcal{P} is defined as the maximum number of vertices in any single bag of \mathcal{P} minus one, mathematically expressed as $\max_{i \in [r]} |Y_i| - 1$.

The *pathwidth* of a graph G is the minimum width among all possible path decompositions of G . For clarity, throughout this paper, we will refer to elements of \mathcal{P} as nodes and elements of V as vertices.

A key feature of path decomposition, as noted in [38], is the *pathwidth separator* property, which is utilized in our algorithm. This property states that for any three nodes Y_i , Y_j , and Y_k , where Y_j is situated between Y_i and Y_k , every path in G that connects a vertex in $Y_i \setminus Y_j$ to a vertex in $Y_k \setminus Y_j$ must pass through a vertex in Y_j . Essentially, Y_j acts as a separator, keeping the vertices in $Y_i \setminus Y_j$ distinct from those in $Y_k \setminus Y_j$. Our algorithm employs a specific type of path decomposition known as *nice*, which is advantageous for designing dynamic programming algorithms. Additionally, there is a linear time

algorithm that can convert any path decomposition into a nice path decomposition while maintaining the same width, as described in [38].

Definition 8 ([38]). A path decomposition of a graph $G = (V, E)$ is *nice* if $|Y_1| = |Y_r| = 1$, and for every $i \in \{1, 2, \dots, r-1\}$ there is a vertex $v \in V$, such that either $Y_{i+1} = Y_i \cup \{v\}$ (*introduce node*), or $Y_{i+1} = Y_i \setminus \{v\}$ (*forget node*).

According to the previous definition, a *nice* path decomposition contains $2|V| + 1$ nodes. This result arises from property (iii) in Definition 7, which asserts that each vertex $v \in V$ appears in a sequence of consecutive bags.

3 Solving MMCov

We present an FPT algorithm for *dMMCov* taking in consideration the pathwidth of the underlying graph. In Fig. 2 is shown a path decomposition of the sample network given in Fig. 1. While in Fig. 3 is shown a nice path decomposition of the sample network given in Fig. 1. Our algorithm leverages the decomposition defined in Definition 8, with particular emphasis on the pathwidth separator.

For a given *MMCov* instance, it is possible to obtain a path decomposition of $G(V, E)$ with width h in linear time [26, 29]. Following this, the algorithm constructs a nice path decomposition $\mathcal{P} = (Y_1, \dots, Y_r)$ with the same width h , which can also be achieved in linear time [38]. Let $G(Y_i)$ denote the subgraph induced by the vertices in $\bigcup_{j=1}^i Y_j$. We will define a constrained version of our problem that we will use in our algorithm. Specifically, this new version of the problem include a constraint on the interfaces that are active on the vertices contained in bag Y_i .

$\mu(Y_i, \mathcal{B})$: Min-Max-Cost Coverage in Multi-Interface Networks (const.)

Input: A subgraph $G(Y_i)$ induced by the vertices in $\bigcup_{j=1}^i Y_j$ related to an instance of *MMCov* having a path decomposition $\mathcal{P} = (Y_1, \dots, Y_r)$ related to the underlying graph. A collection \mathcal{B} of $|Y_i|$ subsets $B(u)$ of the available interfaces $\delta(u)$, with $u \in Y_i$.

Coverage: An activation function $\delta_A: V(u) \rightarrow 2^{[k]}$ with respect to δ such that for each edge $\{u, v\} \in E$, $\delta_A(u) \cap \delta_A(v) \neq \emptyset$, and the interfaces active on the vertices in Y_i are the ones in \mathcal{B} .

Goal: Minimize the maximum cost of the active interfaces among all the vertices, $c(\gamma_A) = \max_{v \in V(u)} \sum_{\beta \in \delta_A(v)} c(\beta)$.

The main idea of the algorithm is to compute the values $\mu(Y_i, \mathcal{B})$ for every set \mathcal{B} at each node Y_i of the path decomposition \mathcal{P} . The method clearly begins at Y_1 and concludes at Y_r . If the constrained version of *dMMCov* has no feasible solution, we set $\mu(X_i, \mathcal{B}) = +\infty$.

At the first step, node Y_1 , we only need to check the following condition, as $G(Y_1)$ consists of a single vertex u with no edges.

$$\mu(Y_i, \mathcal{B}) = c(B(u)) \quad \text{if } |B(u)| \subseteq \delta(u).$$

In other words, at node Y_1 , we check every possible subset of $\delta(u)$. Furthermore, we do this to create all the possible partial solutions needed to form an optimal solution for the entire network.

Introduce node

In any *introduce node* $Y_{i+1} = Y_i \cup \{v\}$, the value $\mu(Y_{i+1}, \mathcal{B})$, for a specific collection \mathcal{B} consisting of sets of active interfaces, is computed by solving the following constrained minimization problem, which uses the values $\mu(Y_i, \mathcal{C})$ already computed in the previous node Y_i .

$$\begin{aligned} \min \quad & \max\{\mu(Y_i, \mathcal{C}), c(B(v))\} \\ \text{s.t.} \quad & C(u) = B(u) \quad \forall u \in Y_i \\ & B(v) \cap C(u) \neq \emptyset \quad \forall u \in Y_i \cap N(v) \end{aligned} \tag{1}$$

with $N(v)$ being the set of vertices neighbouring v . The new collection \mathcal{B} must be the same as the previous collection \mathcal{C} for all vertices except the newly added vertex v (first constraint). The second constraint ensures communication between vertex v and its neighbors in $G(Y_{i+1})$. The objective function is the maximum of the previously computed optimal value $\mu(Y_i, \mathcal{C})$ and the cost of the interfaces $B(v)$ activated for the new vertex v . It is important to note that once \mathcal{B} is fixed, there are no variables left in the problem, reducing it to simply checking the constraints.

Forget node

We use the following minimization problem to compute the value $\mu(Y_{i+1}, \mathcal{B})$ at each *forget node* $Y_{i+1} = Y_i \setminus \{v\}$. We solve this for every possible collection of interface subsets $B(u)$, where $u \in Y_i$.

$$\begin{aligned} \min \quad & \mu(Y_i, \mathcal{B} \cup C(v)) \\ \text{s.t.} \quad & B(u) \cap C(v) \neq \emptyset \quad \forall u \in Y_i \cap N(v) \\ & C(v) \subseteq \delta(v) \end{aligned} \tag{2}$$

The first constraint is needed to guarantee that the forget node v can communicate with all its neighbors belonging to Y_i . The object function takes the minimum value among the one already computed for Y_i .

Once the algorithm concludes, *dMMCoV* has a yes answer if the minimum $\mu(Y_r, \mathcal{B})$ for the unique vertex $u \in Y_r$, taken over all possible collections of interface subsets $B(u) \subseteq \delta(u)$ is at most l .

Computational complexity

To conclude this section, we analyze the time complexity of the described procedure. For each introduce node, the number of problems in (1) that we need to solve is bounded above by $(\sum_{t \in \Delta} \binom{k}{t})^{(h+1)}$. This bound arises from the possible collections \mathcal{B} , which are limited to at most $(\sum_{t \in \Delta} \binom{k}{t})^{(h+1)}$. This is because there are at most $h+1$ subsets $B(u)$ related to a bag, and up to $\sum_{t \in \Delta} \binom{k}{t}$ ways to select t interfaces from $[k]$, with no more than Δ interfaces required per vertex.

At any *forget node*, the complexity is also at most $(\sum_{t \in \Delta} \binom{k}{t})^{(h+1)}$, given by all the subsets of active interfaces $B(u)$, with $u \in Y_{i+1}$, and all the possible subsets of active interface $C(v)$ for the forget vertex v . In conclusion, since there are at most n introduce nodes and n forget nodes, and solving Problems (1) and (2) takes $O(h)$ time, the time complexity of the dynamic programming algorithm is $O((\sum_{t \in \Delta} \binom{k}{t})^{(h+1)} \cdot h \cdot n)$. We can then state the following:

Theorem 3. *Given an instance of dMMCov with a graph G , and a pathwidth decomposition \mathcal{P} of width h for G , dMMCov is solvable in the minimum between $O(2^{k(h+1)} \cdot h \cdot n)$ and $O(k^{\Delta(h+1)} \cdot h \cdot n)$ time.*

Corollary 1. *Given an instance of dMMCov with a graph G , a pathwidth decomposition \mathcal{P} of width h for G , there exists an FPT-algorithm with respect to the parameter $k + \Delta + h$ that solves it.*

Corollary 2. *Given an instance of dMMCov with a graph G , a pathwidth decomposition \mathcal{P} of width h for G , there exists an FPT-algorithm with respect to the parameter $k + h$ that solves it.*

Corollary 3. *Given an instance of MMCov with a graph G , and a pathwidth decomposition \mathcal{P} of width h for G , MMCov is solvable in the minimum between $O(2^{k(h+1)} \cdot h \cdot n)$ and $O(k^{\Delta(h+1)} \cdot h \cdot n)$ time.*

4 Concluding Remarks

In this study, we explore a variant of the well-known Min-Max Coverage problem within the context of multi-interface networks. The objective is to identify the most cost-effective strategy for establishing all connections specified by an input graph by selectively activating subsets of interfaces at network nodes. Although this problem is widely acknowledged as *NP-hard*, we present an optimal fixed-parameter tractable (FPT) algorithm taking into account the pathwidth of the graph, demonstrating that the decision version of *MMCov* can be effectively managed with respect to the number of available interfaces, the network's maximum degree, and its pathwidth. Additionally, our results show that *MMCov* is in FPT when parameterized by the sum of the maximum number of interfaces and the pathwidth.

Future research directions could include examining *MMCov* in relation to other parameters such as local treewidth and cliquewidth, aiming to derive both positive and negative findings. Another promising avenue of investigation involves analyzing the multi-interface coverage problem from a game-theoretic perspective. In this scenario, the problem is decentralized, with devices acting as agents [3, 8, 10] that seek to maximize their utility (e.g., connections) while minimizing overall energy consumption. This approach could employ one of the models discussed in [9, 11, 12] and investigate how social welfare degrades in relation to different standard metrics, such as the price of anarchy and stability.

References

1. Aloisio, A.: Coverage subject to a budget on multi-interface networks with bounded carving-width. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) WAINA 2020. AISC, vol. 1150, pp. 937–946. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44038-1_85
2. Aloisio, A.: Coverage subject to a budget on multi-interface networks with bounded carving-width. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) WAINA 2020. AISC, vol. 1151, pp. 937–946. Springer, Cham (2020)
3. Aloisio, A.: Distance hypergraph polymatrix coordination games. In: Proceedings of 22nd Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pp. 2679–2681 (2023)
4. Aloisio, A.: Algorithmic aspects of distributing energy consumption in multi-interface networks. In: Conference on Advanced Information Networking and Applications (AINA), vol. 204, pp. 114–123. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-57942-4_13
5. Aloisio, A.: Fixed-parameter tractability for branchwidth of the maximum-weight edge-colored subgraph problem. In: Barolli, L. (ed.) AINA 2024, vol. 204, pp. 86–95. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-57942-4_10
6. Aloisio, A.: On balancing energy consumption in multi-interface networks. In: Proceedings of the 26th Italian Conference on Theoretical Computer Science, Torino, Italy, 11–13 September 2024. Springer, Heidelberg (2024)
7. Aloisio, A., Cacciagrano, D.: Distributing energy consumption in multi-interface networks: dimension of the cycle space. In: Proceedings of the 19th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2024), San Benedetto, Italy, 13–15 November 2024. Springer, Heidelberg (2024)
8. Aloisio, A., Flammini, M., Kodric, B., Vinci, C.: Distance polymatrix coordination games. In: Proceedings of 30th International Joint Conference on Artificial Intelligence (IJCAI), pp. 3–9 (2021)
9. Aloisio, A., Flammini, M., Kodric, B., Vinci, C.: Distance polymatrix coordination games (short paper). In: SPIRIT co-located with 22nd International Conference on AIxIA 2023, Rome, Italy, 7–9 November 2023, vol. 3585 (2023)
10. Aloisio, A., Flammini, M., Vinci, C.: The impact of selfishness in hypergraph hedonic games. In: Proceedings of 34th Conference on Artificial Intelligence (AAAI), pp. 1766–1773 (2020)
11. Aloisio, A., Flammini, M., Vinci, C.: Generalized distance polymatrix games. In: Fernau, H., Gaspers, S., Klasing, R. (eds.) SOFSEM 2024, vol. 14519, pp. 25–39. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-52113-3_2
12. Aloisio, A., Flammini, M., Vinci, C.: Generalized distance polymatrix games. In: Proceedings of the 26th Italian Conference on Theoretical Computer Science, Torino, Italy, 11–13 September 2024. Springer, Heidelberg (2024)
13. Aloisio, A., Mkrtchyan, V.: Algorithmic aspects of the maximum 2-edge-colorable subgraph problem. In: Barolli, L., Woungang, I., Enokido, T. (eds.) AINA 2021. LNNS, vol. 227, pp. 232–241. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75078-7_24
14. Aloisio, A., Navarra, A.: Balancing energy consumption for the establishment of multi-interface networks. In: Italiano, G.F., Margaria-Steffen, T., Pokorný, J., Quisquater, J.-J., Wattenhofer, R. (eds.) SOFSEM 2015. LNCS, vol. 8939, pp. 102–114. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46078-8_9
15. Aloisio, A., Navarra, A.: Budgeted constrained coverage on bounded carving-width and series-parallel multi-interface networks. Internet Things **11**, 100259 (2020)

16. Aloisio, A., Navarra, A.: Budgeted constrained coverage on series-parallel multi-interface networks. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) AINA 2020. AISC, vol. 1151, pp. 458–469. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44041-1_41
17. Aloisio, A., Navarra, A.: Constrained connectivity in bounded X-width multi-interface networks. *Algorithms* **13**(2), 31 (2020)
18. Aloisio, A., Navarra, A.: On coverage in multi-interface networks with bounded pathwidth. In: Barolli, L. (ed.) AINA 2024, vol. 204, pp. 96–105. Springer, Heidelberg (2024). https://doi.org/10.1007/978-3-031-57942-4_11
19. Aloisio, A., Navarra, A.: Parameterized complexity of coverage in multi-interface iot networks: pathwidth. *Internet Things* **28**, 101353 (2024)
20. Aloisio, A., Navarra, A., Mostarda, L.: Distributing energy consumption in multi-interface series-parallel networks. In: Barolli, L., Takizawa, M., Xhafa, F., Enokido, T. (eds.) WAINA 2019. AISC, vol. 927, pp. 734–744. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15035-8_71
21. Aloisio, A., Navarra, A., Mostarda, L.: Energy consumption balancing in multi-interface networks. *J. Ambient. Intell. Humaniz. Comput.* **11**(8), 3209–3219 (2020)
22. D’Angelo, G., Di Stefano, G., Navarra, A.: Min-Max coverage in multi-interface networks. In: Černá, I., et al. (eds.) SOFSEM 2011. LNCS, vol. 6543, pp. 190–201. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-18381-2_16
23. Athanassopoulos, S., Caragiannis, I., Kaklamanis, C., Papaioannou, E.: Energy-efficient communication in multi-interface wireless networks. *Theory Comput. Syst.* **52**, 285–296 (2013)
24. Audrito, G., Bertossi, A., Navarra, A., Pinotti, C.: Maximizing the overall end-user satisfaction of data broadcast in wireless mesh networks. *J. Disc. Algor.* **45**, 14–25 (2017)
25. Bahl, P., Adya, A., Padhye, J., Walman, A.: Reconsidering wireless systems with multiple radios. *SIGCOMM Comput. Commun. Rev.* **34**(5), 39–46 (2004)
26. Bodlaender, H.L.: A linear-time algorithm for finding tree-decompositions of small treewidth. *SIAM J. Comput.* **25**(6), 1305–1317 (1996)
27. Cao, Y., Chen, G., Jing, G., Stiebitz, M., Toft, B.: Graph edge coloring: a survey. *Graphs Comb.* **35**(1), 33–66 (2019)
28. Caporuscio, M., Charlet, D., Issarny, V., Navarra, A.: Energetic performance of service-oriented multi-radio networks: issues and perspectives. In: Proceedings of 6th International Workshop on Software and Performance (WOSP), pp. 42–45. ACM (2007)
29. Cattell, K., Dinneen, M.J., Fellows, M.R.: A simple linear-time algorithm for finding path-decompositions of small width. *Inf. Process. Lett.* **57**(4), 197–203 (1996)
30. Cavalcanti, D., Gossain, H., Agrawal, D.: Connectivity in multi-radio, multi-channel heterogeneous ad hoc networks. In: Proceedings of 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1322–1326. IEEE (2005)
31. D’Angelo, G., Di Stefano, G., Navarra, A.: Minimizing the maximum duty for connectivity in multi-interface networks. In: Wu, W., Daescu, O. (eds.) COCOA 2010. LNCS, vol. 6509, pp. 254–267. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17461-2_21
32. D’Angelo, G., Di Stefano, G., Navarra, A.: Multi-interface wireless networks: complexity and algorithms. In: Ibrahem, S.R., El Emery, M.M. (eds.) Wireless Sensor Networks: From Theory to Applications, pp. 119–155. CRC Press, Taylor & Francis Group (2013)
33. D’Angelo, G., Di Stefano, G., Navarra, A.: Flow problems in multi-interface networks. *IEEE Trans. Comput.* **63**, 361–374 (2014)
34. D’Angelo, G., Stefano, G.D., Navarra, A.: Minimize the maximum duty in multi-interface networks. *Algorithmica* **63**(1–2), 274–295 (2012)

35. Draves, R., Padhye, J., Zill, B.: Routing in multi-radio, multi-hop wireless mesh networks. In: Proceedings of 10th International Conference on Mobile Computing and Networking (MobiCom), pp. 114–128. ACM (2004)
36. Faragó, A., Basagni, S.: The effect of multi-radio nodes on network connectivity—a graph theoretic analysis. In: Proceedings of 19th International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC), pp. 1–5. IEEE (2008)
37. Flum, J., Grohe, M.: Parameterized Complexity Theory. Springer, Heidelberg (2006). <https://doi.org/10.1007/3-540-29953-X>
38. Fomin, F.V., Kaski, P.: Exact exponential algorithms. *Commun. ACM* **56**(3), 80–88 (2013)
39. Gurski, F., Rehs, C., Rethmann, J.: Knapsack problems: a parameterized point of view. *Theor. Comput. Sci.* **775**, 93–108 (2019)
40. Klasing, R., Kosowski, A., Navarra, A.: Cost minimization in wireless networks with a bounded and unbounded number of interfaces. *Networks* **53**(3), 266–275 (2009)
41. Kosowski, A., Navarra, A., Pajak, D., Pinotti, C.: Maximum matching in multi-interface networks. *Theor. Comput. Sci.* **507**, 52–60 (2013)
42. Kosowski, A., Navarra, A., Pinotti, M.: Exploiting multi-interface networks: connectivity and cheapest paths. *Wirel. Netw.* **16**(4), 1063–1073 (2010)
43. Perucci, A., Autili, M., Tivoli, M., Aloisio, A., Inverardi, P.: Distributed composition of highly-collaborative services and sensors in tactical domains. In: Ciancarini, P., Mazzara, M., Messina, A., Sillitti, A., Succi, G. (eds.) SEDA 2018. AISC, vol. 925, pp. 232–244. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-14687-0_21



A Scalable State Channel for IoT Using Interactive Consistency Protocols

Gianmarco Mazzante¹, Leonardo Mostarda², Alfredo Navarra²,
and Davide Sestili¹(✉)

¹ Department of Computer Science, University of Camerino, Camerino, Italy
{gianmarco.mazzante,davide.sestili}@unicam.it

² Department of Mathematics and Computer Science, University of Perugia, Perugia,
Italy
{leonardo.mostarda,alfredo.navarra}@unipg.it

Abstract. Scalability remains a fundamental challenge when using blockchains for building Internet of Things (IoT) applications. The throughput of blockchains is not suitable to process the high volume of data produced by IoT devices and public blockchain fees can be prohibitive when IoT devices generate lots of data. This paper proposes a novel state channel that is suitable for building IoT applications. This facilitates the processing of numerous transactions outside the main blockchain. Similarly to any other layer-2 scaling solution, the state channel is designed to decrease direct interactions with the main blockchain by processing and computing tasks off-chain. This reduces fees while increasing the throughput. Unlike existing state channels, the proposed IoT state channel does not require monitoring or synchronising with the main blockchain. In addition, very few data are exchanged between the processes composing the state channel and the main blockchain. Our IoT state channel makes use of a solution to the well-known interactive consistency problem and requires the majority of off-chain processes to act correctly. This ensures that correct processes reach the same sequence of states throughout the entire state channel lifetime. An evaluation of two state channel implementations using two distinct interactive consistency protocols is provided.

1 Introduction

The popularity of Bitcoin and its growing adoption has identified the blockchain as a promising technology for enabling trust and security in a decentralised system. While Bitcoin [15] enables payments without a centralised trusted party, modern blockchains such as Ethereum [4] also allow the definition of Turing-equivalent smart contracts (SCs) [3]. These automatically execute, control or legally document important events and actions according to the contract terms. The adoption of SCs has received an increasing interest for building secure IoT applications. SCs can enhance auditability, security, reliability, and anonymity on the Internet of Things (IoT) applications where billions of devices are connected to the Internet in order to facilitate daily life and provide personalised

services. Integrating IoT and blockchain technologies avoids the use of a centralised authority [1, 2, 12], which may lead to data ownership problems, lack of transparency and security issues. Blockchain technology has the potential to enable machine to machine communication, allowing the implementation of truly decentralised and secure IoT applications.

Blockchains, however, do not scale well. Their transaction throughput is often quite low [14], and running SCs on public blockchains can require the payment of a high amount of fees. These problems are exacerbated on IoT applications, where the amount of data produced can be very high. Several solutions for increasing blockchain scalability have been proposed. The most common ones are referred to as layer-2 solutions. These are designed to decrease direct interactions with the main blockchain by processing and computing tasks off-chain. This reduces fees while increasing the throughput. For instance, well known second layer solutions are presented in [7, 10, 18]. These solutions allow secure execution of smart contracts off-chain. However, they often require that data to the smart contracts are committed on-chain. This means that these solutions may not be suitable for IoT applications where elaboration of continuous streams of data is required. Very promising layer-2 solutions are *state channels* [16]: state channels refer to the process in which users transact with one another directly outside of the blockchain and greatly minimize their use of ‘on-chain’ operations. These generally require only an initial and final state to be committed to the blockchain, which can be interleaved by several state updates off-chain. Existing state channels have some limitations when used for building IoT applications. Quite often, state channels can be run amongst a maximum of two parties. When more than two parties are allowed, they must take turns in performing state updates [6]. A solution that tries to overcome these limitations has been described in [5], however, a single process failure requires all correct processes to commit a lot of data to the blockchain. Generally speaking, state channels can require continuous monitoring or synchronisation with the main blockchain. In addition, blockchain interaction can be required in case of failures or incorrect process behaviour.

This paper proposes a novel multi-party state channel for the scalable execution of IoT smart contracts. The proposed IoT state channel does not require monitoring or synchronising with the main blockchain. The presence of incorrect or faulty processes do not necessitate any blockchain interaction when the majority of processes behave correctly. Our state channel requires the use of a generic solution to the *Interactive consistency problem* [17] in order to guarantee that the input data of any correct party influences the state transitions performed by the state channel. We also provide an evaluation of two different implementations of our state channel. The two implementations differ on the interactive consistency protocol used, highlighting the importance of selecting a suitable interactive consistency protocol according to the specific needs of the IoT application.

The paper is structured as follows:

- Section 2 introduces the background needed to understand our state channel solution;
- Section 3 describes the proposed state channel protocol and system model;
- Section 4 provides a performance evaluation of two distinct implementations of the proposed state channel protocol;
- Section 5 provides conclusions to the work.

2 Background

This section defines concepts that are going to be used in the rest of the paper, i.e., the *Byzantine broadcast*, *Byzantine consensus* and *interactive consistency* definitions. The *Byzantine broadcast problem* can be described in the following way:

Definition 1. Let a system be composed of n processes where a distinguished process p_i (in the following also referred to as general) holds an initial value v_i and at most k processes are incorrect. A protocol Π executed by the n processes is a *Byzantine broadcast protocol*, tolerating k incorrect processes if it satisfies the following properties:

- *Agreement*: every correct process outputs the same value v ;
- *Validity*: if p_i is correct, then the value output by correct processes will be v_i ;
- *Termination*: every correct process will eventually output a value v .

The problem of Byzantine broadcast is often described along with the similar problem of Byzantine consensus. The Byzantine consensus has not a general holding an initial value used by other processes to reach a consensus, but every process holds its initial value. This problem can be defined as follows:

Definition 2. Let a system be composed of n processes where each process p_i holds an initial value v_i and amongst these n processes at most k of them can be incorrect. A protocol Π executed by the processes is a *Byzantine consensus protocol*, tolerating k incorrect processes if it satisfies the following properties:

- *Agreement*: every correct process p_i outputs the same value v ;
- *Validity*: if every correct process holds the same initial value v then they will output v ;
- *Termination*: every correct process will eventually output a value v .

Another problem is the *interactive consistency problem* where correct processes need to reach an agreement on the same vector of values V of size n (the number of processes in the system). This must contain all initial values held by correct processes:

Definition 3. Let a system be composed of n processes where each process p_i holds an initial value v_i and amongst these n processes at most k of them can be incorrect. A protocol Π executed by the processes is an *Interactive consistency (IC) protocol* tolerating k incorrect processes if it satisfies the following properties:

- *Agreement*: Every correct process will output the same vector V ;
- *Validity*: if the initial value of a correct process p_i is v_i then the element of the vector V at index i output by every correct process will be equal to v_i (i.e., $V[i] = v_i$);
- *Termination*: every correct process will eventually output a vector of values V .

Each of the aforementioned problems has a *binary* variant and a *multivalue* variant. The *binary* variant of each of these problems requires that the initial value of any process is chosen between two options, i.e., 0 or 1. The multivalue variant, however, does not pose this limitation. A solution to a binary problem can be used to solve the multivalue problem via *sequential* or *parallel* composition. For instance, a multivalue Byzantine broadcast protocol for broadcasting a message of size ℓ may be constructed by running ℓ sequential or parallel instances of binary Byzantine broadcast protocol. However, the construction of such multivalue protocols from binary protocols is not always straightforward because Byzantine protocols may not be secure under parallel or sequential composition unless they make use of unique identifiers [13]. Multivalue Byzantine protocols constructed via sequential or parallel composition with their binary counterparts are generally not efficient. For this reason, *extension protocols* have been devised. Extension protocols [8, 9] allow the construction of Byzantine broadcast or consensus protocols for long messages of size ℓ by using either a binary Byzantine broadcast or consensus protocol or a *short* message Byzantine broadcast or consensus protocol (generally a short message is as long as a hashed value).

In synchronous settings when cryptography can be used, it is possible to design Byzantine consensus protocols tolerating a maximum of $k < n/2$ incorrect processes, while the problem of Byzantine broadcast can be solved for any number of incorrect processes [11]. However, the two problems are mutually reducible to one another as long as $k < n/2$. In fact, a Byzantine broadcast protocol tolerating $k < n/2$ incorrect processes can be implemented with a Byzantine consensus protocol tolerating the same number of incorrect processes by making the general multicast its initial value to the other processes which will then use the received value as their initial value in the Byzantine consensus protocol. A Byzantine consensus protocol tolerating $k < n/2$ incorrect processes, instead, can be constructed with the use of a Byzantine broadcast protocol tolerating the same number of incorrect processes by running n parallel Byzantine broadcast protocol instances, each of which having a different process acting as general that will broadcast its initial value and, when these Byzantine broadcasts terminate, each process outputs the most broadcast value among the ones it received. Interactive consistency protocols can also be implemented with the use of Byzantine broadcast protocols by running n parallel Byzantine broadcast protocol instances each of which having a different general broadcasting its initial value. The output values of the Byzantine broadcast instances can then be ordered in an array as required by the interactive consistency problem. In Sect. 3, we describe our state channel assuming that a multivalue interactive consistency protocol tolerating $k < n/2$ incorrect processes exists. Afterwards, in Sect. 4, we

compare two different implementations of the interactive consistency protocol used in the solution. The first solution is implemented by making each process broadcast its initial value bit per bit using the authenticated version of the Lamport binary Byzantine broadcast protocol [11]. The second solution to the interactive consistency problem is implemented by using a Byzantine consensus extension protocol described in Ganesh et al. [9]. This is an extension protocol that requires the use of a short message Byzantine broadcast protocol. This is provided by using the parallel composition of the binary Lamport protocol. The Ganesh consensus protocol is then transformed into a Byzantine broadcast protocol using the transformation previously described. The Byzantine broadcast protocol will then be used to implement an interactive consistency protocol.

3 State Channel Environment and Description

In our system there is a set of n processes $P = \{p_1, p_2, \dots, p_n\}$ where $k < n/2$ may be incorrect. The processes are assumed to be fully connected, that is, each process can directly communicate with the other ones. The following network model is assumed:

- Synchrony: there is a known δ such that for every pair of processes p_a and p_b , if both p_a and p_b are correct and p_a sends a message m to p_b at time T , then p_b receives the message m at time $T' \leq T + \delta$.
- Integrity: if p_a and p_b are correct and p_b receives a message m for which $sender(m) = p_a$, then m was really sent by p_a and it has not been altered.

There is a sequence of sets W_1, \dots, W_t called *windows*, each window contains a set of inputs (*e.g.*, IoT data) $W_i = \{m_{i,1}, \dots, m_{i,j}\}$. Each input is said to be an element of the *set of possible inputs* M . We use S to denote the set of all possible states of the state channel; s_1, \dots, s_n are elements in S . There is a function $f : W \times S \longrightarrow S$ called *transition function* that maps a state and a window to a new state in S . Effectively, f implements the IoT application. There is a state $s_0 \in S$ called the *initial state*. Each process knows both f and s_0 . There is a blockchain holding a smart contract that acts as an adjudicator and outputs the final state reached by the state channel. Processes in the system can communicate with the on-chain smart contract by querying it or by sending transactions to it.

The state channel protocol (Listing 1) proceeds in rounds, starting from round 1 to round t . At each round $1 \leq i \leq t$ each process $p_q \in P$ has a set of inputs $W_{q,i} \subseteq W_i$ and a state s_{i-1} . The process p_q must enrich its local window $W_{q,i}$ with all inputs that belongs to the local windows of other correct processes. The process p_q uses the enriched window and its state s_{i-1} to compute a new state s_i by applying the transition function f . It is worth noticing that the enriched window ensures that all correct processes have the same input and reach the same state. At the end of the round t , a blockchain has to be convinced that the last state s_t is indeed the last state reached by the state channel. To do this, at any round i the processes perform the *interactive consistency*

phase (Listing 1, lines 5–8). In the interactive consistency phase processes are first required to use a multivalue interactive consistency protocol tolerating a maximum of $\lfloor(n - 1)/2\rfloor$ incorrect processes (Listing 1, line 6). The input value that any process $p_q \in P$ will use for the execution of the interactive consistency protocol is its local window $W_{q,i}$. At the end of this phase correct processes will have reached an agreement on the same vector of values V_i that will contain as elements at least every correct process' local i -th window. Once processes have reached an agreement on the same vector of windows V_i , they can construct a set of inputs called *union window* W_i^U as the union of every element of V_i (1):

$$W_i^U = \bigcup_{j=1}^n V_i[j] \quad (1)$$

This union window is assured to contain every input of every correct node. This will ensure that the input of any correct process is not lost and will be taken into consideration for moving the state of the state channel. Next, every process computes the following state of the channel s_i (2).

$$s_i = f(s_{i-1}, W_i^U) \quad (2)$$

When $i = t$ the state channel has to be closed. In order to do that, processes require to generate a proof that a given state s_t is indeed the state reached by the state channel. This is done by making each process p_a multicast to every other node a *proof-piece message* msp_a (3)

$$msp_a = \langle H(s_t) \rangle : p_a \quad (3)$$

where H is a hash function and $: p_a$ is the signature of process p_a applied to the message. This proof-piece allows any process to construct a *proof-state ps* (4)

$$ps = \langle s_t, ppSet \rangle \quad (4)$$

where *ppSet* is a set of proof-pieces containing proof-pieces from a majority of processes. The proof state will then be able to convince the blockchain that s_t is the t -th state reached by the state channel. In order to reduce the amount of data to be sent to the smart contract in the form of transactions, the *proof-state* p_i will be sent to the smart contract by a single process. This is done by making each process wait a different time (Listing 1, lines 12–13) at the end of which the smart contract is queried to know whether a valid proof-piece has already been sent. If the check does not pass, then the process will send its proof-piece. The smart contract (Listing 2) will then check whether the received proof-state is valid (Listing 2, lines 3–5) and if the checks pass it outputs the state contained in the proof-state message (Listing 2, line 6).

```

1  $\delta$ : period of time in seconds
2  $s_0$ : initial state
3 State channel protocol:
4    $i \leftarrow 0$ 
5   Interactive consistency phase:
6     Step 1: Each process  $p_q$  performs the multivalue interactive consistency protocol with
      the other processes in  $P$  using  $W_{q,i}$  as its input, each correct process will then
      output the same vector  $V_i$ 
7     Step 2: each process  $p_q$  calculates  $W_i^U = \bigcup_{j=1}^n V_i[j]$  and  $s_i = f(s_{i-1}, W_i^U)$ 
8     Step 3: if  $i < t$  then  $i \leftarrow i + 1$  and move back to Interactive consistency phase, else
      move to proof-phase
9   Proof phase:
10    Step 1: each process  $p_q$  sends to every other process a proof-piece  $msp_q = \langle H(s_t) : p_q$ 
11    Step 2: upon receiving a number of proof-pieces equal to  $|P|/2$ , each process  $p_q$ 
      assembles a proof-state  $ps_q = \langle s_t, ppSet \rangle$  with  $ppSet$  being a set of proof-pieces
      signed by different processes each of which has the same hash and  $|ppSet| \geq n/2$ 
12    Step 3: Create a mapping from processes to integers from 0 to  $|P| - 1$   $map : P \rightarrow$ 
       $\{0, \dots, |P| - 1\}$ 
13    Step 4: Each process  $p_q$  waits  $map(p_q) \cdot \delta$  seconds, then checks whether a valid proof-
      piece has already been sent to the smart contract. If a valid proof-piece has not
      already been sent then  $p_q$  sends  $ps_q$  to it.

```

Listing 1. State channel protocol pseudocode.

```

1 Smart contract:
2 Upon receiving a proof-state  $ps = \langle s_t, ppSet \rangle$  perform the following checks:
3   Check 1: check that  $|ppSet| > n/2$ 
4   Check 2: check that every  $msp \in ppSet$  is correctly signed by a process in  $P$ 
5   Check 3: check for every  $msp \in ppSet$  with  $msp = \langle hashedState : signature \rangle$  that
       $H(s_t)$  is equal to  $hashedState$ 
6   If every check passes then output  $s_t$  and terminate, else, wait until a valid proof-state is
      received.

```

Listing 2. Smart contract pseudocode.

4 Performance Evaluation

We evaluate the efficiency of our state channel construction using two different concrete interactive consistency protocols. The two protocols will influence only the communication amongst processes, and not the communication from processes to the blockchain. Since the number of bits that a process will send to the other processes in the proof-phase are very few, our evaluation will focus on the amount of bits that a process sends towards other processes during the interactive consistency phase. The number of bits that a process will send in the proof phase is shown in Fig. 1(a) and can be calculated as shown in Eq. 5:

$$(n - 1)(hashLength + sigLength) \quad (5)$$

where $hashLength$ is the size of the hash produced by the chosen hash function and $sigLength$ is the size of the signature. The number of bits received by the blockchain is shown in Fig. 1(b) and can also be easily calculated as shown in Eq. 6:

$$size(s_t) + \lceil (n + 1)/2 \rceil (hashLength + sigLength) \quad (6)$$

where $size(s_t)$ is the number of bits of the final state.

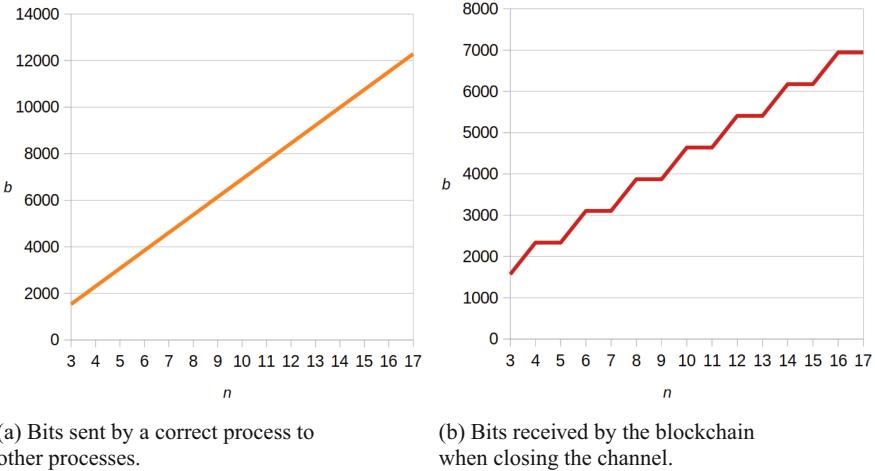


Fig. 1. Bits communicated during the proof phase, calculated using the $\text{size}(S_t) = 32$, the $\text{hashLength} = 256$ and the $\text{sigLength} = 512$.

Before moving on we will give a short description of the multivalue interactive consistency protocols used. Both of them require the execution of n parallel instances of multivalue Byzantine broadcast protocols. However, they differ in the Byzantine broadcast protocol used. In one IC-protocol, the multivalue Byzantine broadcast protocol is implemented via parallel composition of Lamport's binary Byzantine broadcast protocol [11]: we call the concrete IC-protocol produced in this way the *extended Lamport protocol*. In the other IC-protocol, the multivalue Byzantine broadcast is performed by the extension protocol tolerating $k \leq n/2$ incorrect processes described in Ganesh et al. [9]. The binary Byzantine broadcast protocol provided to the extension protocol is the binary Byzantine broadcast protocol of Lamport. The IC-protocol produced in this way will simply be called the *Ganesh protocol* from now on. The hash function used in the state channel protocol is assumed to be the *keccak256* hash function, meaning that hashes are long 256 bits. Signatures instead, are assumed to be long 512 bits.

4.1 Comparison of Interactive Consistency Protocols

This section shows the number of bits that each correct process needs to send to each other during the interactive consistency phase with the extended Lamport protocol and with the Ganesh protocol. In the rest of the section, we will refer to the number of bits sent by a single correct process with b . This analysis shows the growth of b when both the length of the message ℓ and the number of processes $n = |P|$ vary. For both protocols we tested the best and worst case scenarios. We define as the *best-case scenario* the situation in which all the processes are correct. The *worst-case scenario* is defined as the situation in which there are

$\lceil n/2 \rceil - 1$ incorrect processes that will make the other participants send the most possible bits to come to an agreement. The results we obtained are plotted in the 3-D chart of Fig. 2. These charts make it possible to compare how different protocols behave with varying number of processes n and varying lengths of the message ℓ in different scenarios. For a final comparison and a more precise evaluation, two line charts are provided in Fig. 3: Fig 3(a) shows how the number of bits sent by a correct process increases as n increases with a fixed $\ell = 2^{15}$ while Fig. 3(b) shows how the number of bits sent by a correct process grows when ℓ grows and $n = 17$.

- **Extended Lamport protocol, comparison between the best and worst case scenarios (Fig. 2(a)):** In this protocol, both values n and ℓ influence the number of bits b . If one value is high but the other value is low, b gains a negligible growth. Instead, if both values are high, b shows a significant growth. The difference in performance between the best-case scenario and the worst-case scenario becomes significant especially when n is high and ℓ is high.
- **Ganesh protocol, comparison between the best and worst case scenarios (Fig. 2(b)):** In this protocol, when n varies, b grows significantly, even with a small ℓ ; while when ℓ varies, b grows imperceptibly. This is evident in Fig. 3(b). This is due to the fact that using this protocol, hashes are sent instead of the whole message. In the best-case, the curve shows a significant improvement with respect with the worst case one when n increases.
- **Best-case scenarios comparison (Fig. 2(c)):** It is clear that the Ganesh protocol performs better than the extended Lamport protocol when the length of the message is high. More precisely, we found that above a message length of $\approx 2^{12}$ Ganesh performs better. In fact, Lamport uses raw messages while Ganesh uses hashed messages, hence performing better on longer messages.
- **Worst-case scenarios comparison (Fig. 2(d)):** Similarly to the best-case scenarios, the Ganesh protocol performances are better than the extended Lamport protocol when the message length is more than $\approx 2^{12}$. Figure 3(b) shows the intersection point where the two approaches produce the same output b .

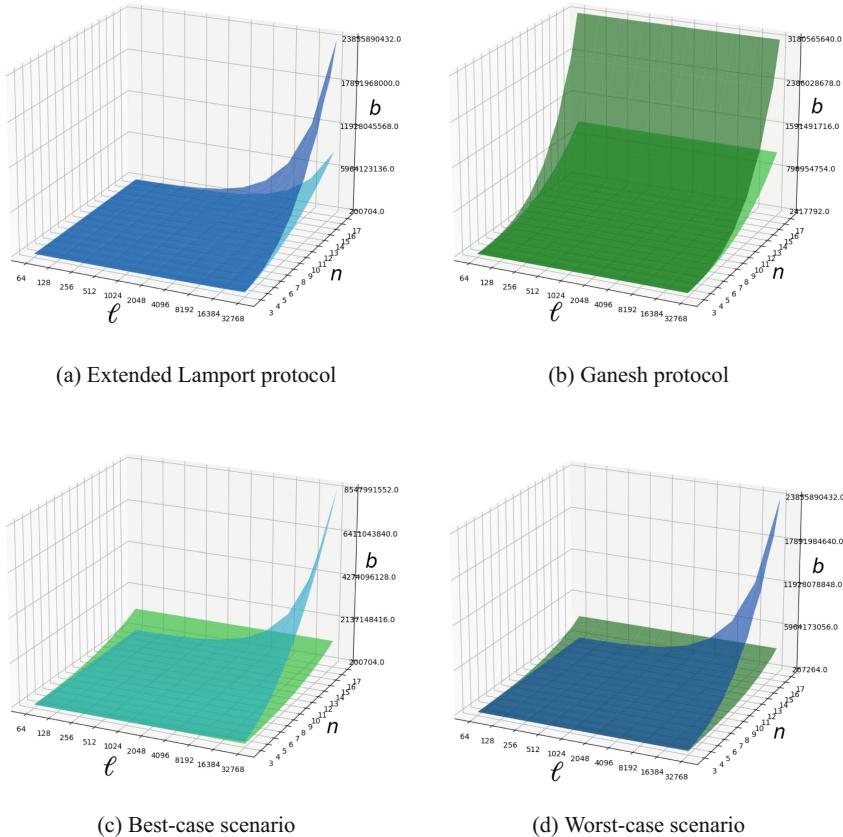


Fig. 2. Bits to communicate using the extended Lamport protocol (blue) and the Ganesh protocol (green) for solving the interactive consistency problem when varying both the message length ℓ and the number of processes n .

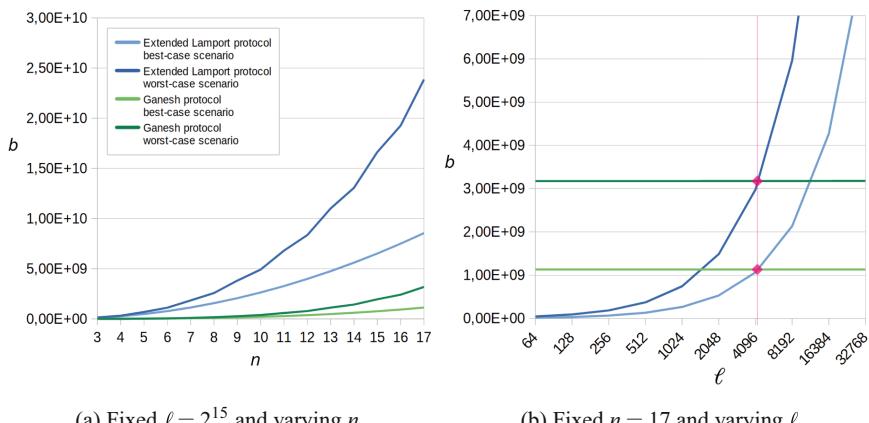


Fig. 3. Bits to communicate using the extended Lamport protocol (blue) and the Ganesh protocol (green) for solving the interactive consistency problem.

5 Conclusions

This paper proposes a novel multi-party state channel particularly suited for the IoT. The proposed state channel allows users to execute complex smart contracts off-chain without any fees and with minimal interaction with the blockchain of reference. This allows the proposed state channel to be freed both from the burden of storing the input data to the smart contract and to the burden of processing such data. The solution, however, requires a majority of participants to behave well, limiting thus the applicability of such a solution to those settings where a majority of participants can be expected to behave correctly. The proposed state channel requires that processes solve an interactive consistency problem in order to progress to a new state. This guarantees that the input of every correct process will affect the state transition. However, this part of the protocol is also the most expensive one in terms of communication complexity and may hinder the performance of the overall protocol. For this reason, we analysed the behaviour of two concrete interactive consistency protocols with varying number of participants and varying input length. The analysis shows that the communication complexity of certain solutions is more affected by the number of participants while others are more affected by the size of the input messages. This makes clear that the choice of the interactive consistency protocol to be used should be done on a case-by-case basis according to the peculiarities of the specific application developed.

References

1. Bigiotti, A., Mostarda, L., Navarra, A.: Blockchain and IoT integration for air pollution control. In: Barolli, L. (eds.) *Advances on P2P, Parallel, Grid, Cloud and Internet Computing . 3PGCIC 2023. Lecture Notes on Data Engineering and Communications Technologies*, vol. 189, pp. 27–38. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-46970-1_3
2. Bistarelli, S., Marcozzi, M., Mazzante, G., Mostarda, L., Navarra, A., Sestili, D.: Blockchain and IoT integration for pollutant emission control. In: Barolli, L., Hussain, F., Enokido, T. (eds.) *AINA 2022. LNNS*, vol. 451, pp. 255–264. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-99619-2_25
3. Bistarelli, S., Mazzante, G., Micheletti, M., Mostarda, L., Sestili, D., Tiezzi, F.: Ethereum smart contracts: analysis and statistics of their source code and opcodes. *Internet Things* **11**, 100198 (2020)
4. Buterin, V., et al.: A next-generation smart contract and decentralized application platform. *White Paper* **3**(37), 2–1 (2014)
5. Cacciagran, D., Corradini, F., Mazzante, G., Mostarda, L., Sestili, D.: Off-chain execution of IoT smart contracts. In: Barolli, L., Woungang, I., Enokido, T. (eds.) *AINA 2021. LNNS*, vol. 226, pp. 608–619. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75075-6_50
6. Close, T., Stewart, A.: ForceMove: an n-party state channel protocol. *Magmo, White Paper* (2018)
7. Das, S., Ribeiro, V.J., Anand, A.: Yoda: enabling computationally intensive contracts on blockchains with byzantine and selfish nodes. *arXiv preprint arXiv:1811.03265* (2018)

8. Fitzi, M., Hirt, M.: Optimally efficient multi-valued byzantine agreement. In: Proceedings of the Twenty-fifth Annual ACM Symposium on Principles of Distributed Computing, pp. 163–168 (2006)
9. Ganesh, C., Patra, A.: Optimal extension protocols for byzantine broadcast and agreement. *Distrib. Comput.* **34**, 59–77 (2021)
10. Kalodner, H., Goldfeder, S., Chen, X., Weinberg, S.M., Felten, E.W.: Arbitrum: scalable, private smart contracts. In: 27th USENIX Security Symposium (USENIX Security 18), pp. 1353–1370 (2018)
11. Lamport, L., Shostak, R., Pease, M.: The byzantine generals problem. *ACM Trans. Program. Lang. Syst.* **4**(3), 382–401 (1982)
12. Lao, L., Li, Z., Hou, S., Xiao, B., Guo, S., Yang, Y.: A survey of IoT applications in blockchain systems: architecture, consensus, and traffic modeling. *ACM Comput. Surv. (CSUR)* **53**(1), 1–32 (2020)
13. Lindell, Y., Lysyanskaya, A., Rabin, T.: On the composition of authenticated byzantine agreement. *J. ACM (JACM)* **53**(6), 881–917 (2006)
14. Mostarda, L., Pinna, A., Sestili, D., Tonelli, R.: Performance analysis of a BESU permissioned blockchain. In: Barolli, L. (ed.) AINA 2023. LNNS, vol. 655, pp. 279–291. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-28694-0_26
15. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system (2008)
16. Negka, L.D., Spathoulas, G.P.: Blockchain state channels: a state of the art. *IEEE Access* **9**, 160277–160298 (2021). <https://doi.org/10.1109/ACCESS.2021.3131419>
17. Pease, M., Shostak, R., Lamport, L.: Reaching agreement in the presence of faults. *J. ACM (JACM)* **27**(2), 228–234 (1980)
18. Wüst, K., Matetic, S., Egli, S., Kostiainen, K., Capkun, S.: ACE: asynchronous and concurrent execution of complex smart contracts. In: Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, pp. 587–600 (2020)



Digital Twins for Improving Buildings Performances: A Literature Review Methodology Use Case

Ionica-Larisa Puiu¹ and Teodor-Florin Fortis^{1,2(✉)}

¹ West University of Timișoara, 300223 Timișoara, Romania
ionica.puiu@e-uvt.ro

² Research Institute e-Austria Timișoara, 300223 Timișoara, Romania
florin.fortis@e-uvt.ro

Abstract. In the pursuit of sustainability and efficiency in the built environment, digital twin technology has emerged as a transformative tool. Our approach employs bibliometric techniques to systematically evaluate and synthesize recent research, demonstrating the benefits of digital twins in optimizing space management, improving energy distribution for heating and lighting and supporting comprehensive lifecycle sustainability assessments. We provide a detailed discussion on the methodology and bibliometric techniques used, showcasing their application in creating a comprehensive and impactful bibliographic portfolio.

Keywords: Digital Twin · Building Information Modelling · BIM · Sustainability · Building maintenance · Bibliometric techniques

1 Introduction

The built environment significantly impacts global energy consumption and greenhouse gas emissions, thus contributing to climate change and environmental degradation. The International Energy Agency (IEA) reports that buildings represent a substantial portion of global energy use and carbon emissions. Effective maintenance and management of building systems are crucial for optimizing energy consumption, reducing operational costs and minimizing environmental impact. The digital twin technology has emerged as a revolutionary approach to enhancing building performance and sustainability. This model provides valuable insights into energy usage, system performance and maintenance needs, facilitating proactive decision-making and efficient resource management.

Bibliometric techniques are essential to effectively highlight the benefits of digital twins in building performance and sustainability. These methods systematically evaluate extensive and growing bodies of literature, identify key trends, influential studies and knowledge gaps. By offering quantitative insights into the evolution of scientific research, bibliometric techniques guide researchers through systematic literature reviews and highlight significant contributions in specific

fields. These techniques are central for developing a strong foundation, ensuring new research builds on a comprehensive understanding of existing work.

The paper is structured as follows: Sect. 2 reviews relevant bibliometric techniques and tools. Section 3 provides a detailed discussion on the methodology and bibliometric techniques. The results of applying these techniques to a specific case study involving digital twin technology are presented in Sect. 4, demonstrating their practical applications and benefits in advancing scientific investigations, while Sect. 5 concludes with key insights and recommendations for future research. By demonstrating the practical applications and benefits of digital twins, this study underscores their essential role in advancing sustainable building management and achieving long-term environmental goals.

2 Background Information

Bibliometric techniques are essential tools for systematically analyzing and evaluating scientific literature. These methods use statistical and mathematical indicators to measure the impact and evolution of research within a specific field. By examining publication patterns, citation counts and the relationships between authors and topics, bibliometric analysis provides a comprehensive overview of scientific progress and trends.

2.1 Tools for Supporting Bibliometric Investigations

First introduced by Aria and Cuccurullo, in 2017, *bibliometrix*¹ was offered as an *R* package, which “provides a set of tools for quantitative research in bibliometrics and scientometrics” [4]. It offers a suite of functions for quantitative research in scientometrics and bibliometrics, such as data collection, analysis, and visualization, as recommended by [27]. It allows researchers to perform various bibliometric analyses, including co-citation, bibliographic coupling, collaboration, or keyword analysis.

ASReview is an active learning tool that improves systematic review efficiency by combining machine learning with traditional methods to prioritize studies for screening. The tool was used in the development of a systematic review regarding predictive maintenance using Digital Twins [9], significantly reducing workload by focusing on the most informative studies. The chosen setup uses TF-IDF for feature extraction, Naive Bayes as a classifier, Maximum Query Strategy for selecting uncertain cases and Dynamic Resampling to balance the training dataset, enhancing the model’s ability to distinguish relevant studies.

AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews) is a critical appraisal tool used to evaluate the methodological quality of systematic reviews, particularly those involving randomized or non-randomized studies of healthcare interventions. AMSTAR 2 comprises 16 items that assess various aspects of the review process, including the protocol, search strategy, selection criteria,

¹ <https://github.com/massimoaria/bibliometrix>.

data extraction, and risk of bias. This tool provides a structured framework to determine the reliability and validity of systematic reviews.

Data visualization is increasingly being used to enhance bibliometric research. *VOSviewer*², which was first introduced in 2006, is visualisation tool for constructing and exploiting bibliometric networks, enabling researchers to create maps based on co-authorship, co-citation and bibliographic coupling [11]. *VOSviewer* was instrumental in a bibliometric analysis of global research trends in digital twin technologies, highlighting key researchers and networks [8].

*Gephi*³, an open-source network analysis and visualization solution, is effective for exploring and understanding complex networks. Introduced in [5], Gephi supports the visualization of bibliometric data, allowing researchers to map relationships and interactions within scientific literature, dynamic filtering, statistical analysis and rich visualization features, which enable the examination of co-authorship, citation and keyword co-occurrence networks.

2.2 Bibliometric Research Methodologies

A plethora of methodologies exist for supporting the development of bibliometric research, including ProKnow-C (the Knowledge Development Process-Constructivist), Methodi Ordinatio, or PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). These approaches offer the methodological steps needed for conducting bibliometric research and systematic reviews, such as literature systematic analysis, identification of research trends and highlight influential contributions.

PRISMA comprises a comprehensive set of guidelines for conducting and reporting systematic reviews and meta-analyses with transparency and reproducibility. This interest in standardized methodologies is well-documented, as [7] mentions. The methodology includes a 27-item checklist and a three-phase flow diagram guiding researchers through identifying, screening plus assessing eligibility and including studies. It was employed to support reviewing the growth and challenges of the Industrial Internet of Things (IIoT) or highlighting the potential of Digital Twins for predictive maintenance [9,19], with a focus on eligibility criteria, search strategy, data extraction, and synthesis of results.

The Proknow-C method “is divided into two main phases, the first deals with the selection of the raw articles bank and the second the filtering process of the articles [26]. A comprehensive bibliometric approach, presented by [26], analyzes literature on performance indicators for energy management in Industry 4.0 using this method, with data collected both from Scopus and Web of Science, filtered for high-impact studies, and investigated using citation analysis, co-authorship networks and keyword co-occurrence. A systematic review identifies research gaps and trends, with visualization tools like VOSviewer and Gephi providing graphical representations of the data.

² <https://www.vosviewer.com/>.

³ <https://gephi.org/>.

The ProKnow-C methodology was applied for bibliometric analysis, involving systematic searches, descriptive bibliometric analysis, and detailed reviews to identify research gaps and trends in [21]. The Methodi Ordinatio approach ranks scientific outputs by citation count, publication year, and journal impact factor. Visualization tools enhance understanding of bibliometric data.

On the other hand, Methodi Ordinatio can be used for selecting and ranking scientific papers, [20]. The first five phases of Methodi Ordinatio adapt ProKnow-C's Selection of the Gross Bibliographic Portfolio. The last four phases replace ProKnow-C's citation criteria to prioritize publications. The process involves defining the research scope, conducting exploratory searches, finalizing keywords and collecting data from databases. Papers are filtered for quality, ranked using the InOrdinatio index based on impact factors, citation counts and publication years and systematically analyzed to ensure relevance.

3 Defining a Literature Review Methodology

Our literature review methodology, aligned with the PRISMA guidelines and the bibliometric methods suggested by [27], describes the main steps in conducting bibliometric investigations. Exemplified by a study on the integration of Digital Twins (DT) and Building Information Modeling (BIM), the methodology was applied to explore how digital twins enhance building maintenance efficiency. The methods, and the entire workflow, are detailed in Fig. 1, outlining the process from defining the research scope to running the bibliometric analysis and reporting the findings.

3.1 The Protocol Phase

3.1.1 Aim and Scope: The initial step involves clearly defining the study's aim and scope by formulating precise research questions and setting primary objectives. This crucial step determines specific areas of interest and measurable outcomes. Through detailed discussions, the study's boundaries are agreed upon, ensuring focus and relevance, avoiding unnecessary diversions, and targeting significant gaps in the field.

3.1.2 Keywords: Identifying appropriate keywords is crucial for a thorough literature search. This involves selecting terms that accurately represent the core concepts of the research topic. Proper keyword selection enhances the search strategy, ensuring it is comprehensive and targeted, thereby maximizing the retrieval of relevant studies.

3.1.3 Relevant Databases: Selecting the appropriate databases ensures comprehensive literature coverage. Databases were evaluated for relevance, accessibility and coverage. Scopus, known for its extensive coverage and high-quality indexing, is often chosen. This selection ensures the inclusion of all significant publications related to the research questions, providing a robust foundation for bibliometric analysis.

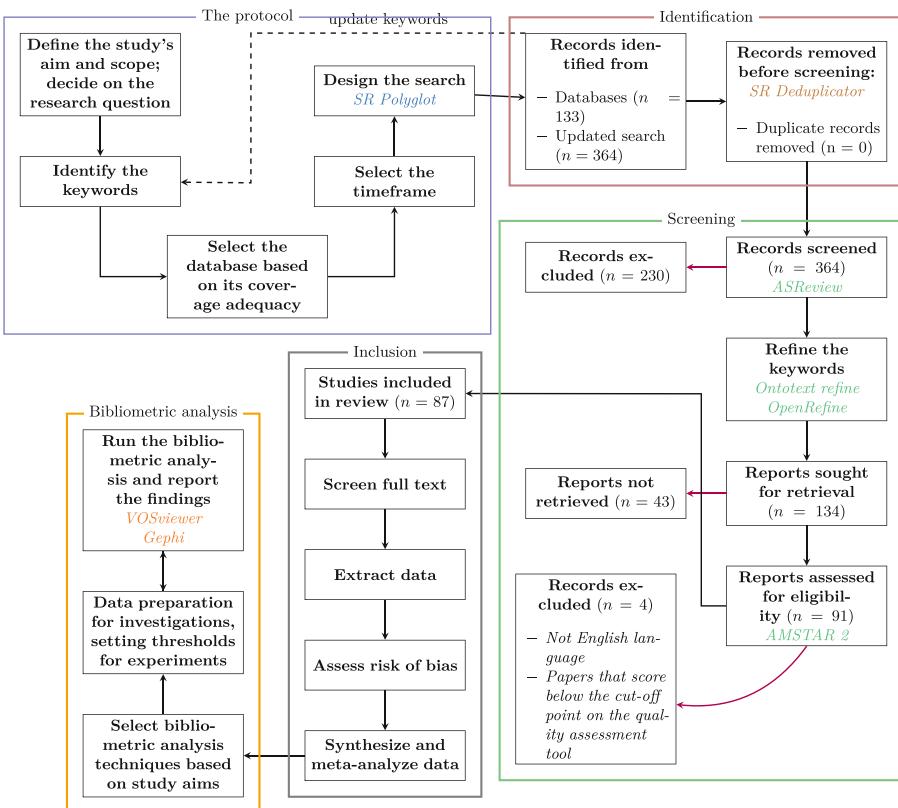


Fig. 1. The methodology workflow

3.1.4 Timeframe: Determining the publication date range ensures the inclusion of the most recent and relevant studies. A suitable timeframe captures the latest research developments and trends, maintaining the study's relevance and identifying current advancements and emerging issues within the research topic.

3.1.5 Search Strategy: Designing an effective search strategy involves using boolean operators and tools like 'SR Polyglot' to create and refine search strings. The strategy is tested and optimized to ensure a comprehensive search across multiple databases, capturing a wide array of relevant studies and minimizing the risk of missing crucial information.

3.2 The Identification Phase

3.2.1 Records Identified: This phase ensures a thorough and up-to-date collection of relevant literature, capturing a comprehensive dataset for analysis. An initial search using predefined keywords ('digital twin' and 'bim'), generated $n = 133$ entries, from the selected Databases. Subsequently, updating the keyword list and refining the search strategy resulted in $n = 364$ entries.

3.2.2 Records Removal: This step is crucial for maintaining dataset integrity and preventing skewed bibliometric analysis results. ‘SR Deduplicator’ is used to remove duplicate records, efficiently eliminating redundant entries and ensuring each study is included only once (no entries were removed).

3.3 The Screening Phase

3.3.1 Records Screened: Screening the records by assessing their relevance based on titles and abstracts. ‘ASReview’, an automated tool using machine learning, expedites this process alongside manual methods. The approach enhances efficiency and accuracy, resulting in the exclusion of 230 records ($n = 364$).

3.3.2 Refine the Keywords: ‘Ontotext Refine’ refines and optimizes keywords, ensuring dataset precision and consistency through data cleaning and transformation; ‘OpenRefine’ further enhances data quality. Together, these tools ensure keywords are accurate and relevant, improving the reliability and robustness of the bibliometric analysis.

3.3.3 Reports Considered: After the initial screening, relevant records are sought for full-text retrieval ($n = 134$). Attempts are made to access these reports for detailed evaluation, ensuring all potentially relevant studies are considered for the final analysis.

3.3.4 Reports not Retrieved: Several reports were not accessible due to subscription barriers, unavailability or other constraints ($n = 43$). This limitation is documented and alternative sources or similar studies are considered to mitigate the impact on the analysis.

3.3.5 Eligibility Assessment: Retrieved reports are assessed for eligibility based on methodological quality and relevance using tools like ‘AMSTAR 2’. Only high-quality, relevant studies are included, while reports not meeting criteria, such as those not in English, are excluded ($n = 91$ reports were evaluated).

3.4 The Inclusion Phase

3.4.1 Included Studies: A total of $n = 87$ studies met the inclusion criteria and were included in the review, forming the basis of the bibliometric and systematic analysis and providing a robust dataset for in-depth examination.

3.4.2 Full Text Screening: The full texts of the included studies are thoroughly screened to extract pertinent data, systematically collecting detailed information on study characteristics, outcomes, methodologies and key findings, ensuring all relevant information is captured for analysis.

3.4.3 Data Extraction: Data extraction gathers and organizes detailed information from the included studies, ensuring all necessary data points are accurately captured and systematically arranged for bibliometric and meta-analysis.

3.4.4 Assessing Risk of Bias: Evaluating bias in the studies ensures valid and reliable conclusions by examining potential biases in design, execution and reporting. Identifying and mitigating these biases is crucial for maintaining analysis integrity.

3.4.5 Synthesize and Meta-analyze Data: Data from multiple studies are synthesized and meta-analyzed to derive overall effect estimates and identify patterns. This involves using statistical techniques to combine data, providing a comprehensive overview of the research landscape, quantifying effects and identifying trends, thus enhancing the study's findings.

3.5 Bibliometric Analysis Phase

3.5.1 Bibliometric Analysis Techniques: Suitable bibliometric techniques, such as co-occurrence keyword analysis and bibliographic coupling, are selected to align with the research objectives. These methods help understand relationships and trends within the dataset, ensuring the analysis is relevant and insightful based on the study's aims and dataset nature.

3.5.2 Data Preparation: Data is prepared for bibliometric analysis by setting inclusion thresholds, like the minimum keyword occurrences and citation counts per document. This ensures the focus is on the most relevant and impactful studies, filtering out less significant ones for a more targeted and meaningful analysis.

3.5.3 Bibliometric Analysis and Reporting: ‘VOSviewer’ is employed for bibliometric analysis, offering advanced visualization capabilities for mapping and interpreting the research landscape. ‘Gephi’ is used to further enhance the visualization process. The tools help identify key themes, influential studies and research clusters. The findings are comprehensively reported, providing a visual and analytical overview of the study area and facilitating a deeper understanding of relationships and trends within the research domain.

By meticulously defining the research scope, employing robust search strategies and utilizing advanced bibliometric techniques, this detailed methodology ensures a systematic, transparent and reproducible approach to conducting a literature review and meta-analysis. This approach aims to produce high-quality, evidence-based insights that contribute meaningfully to the field of study, facilitating a deeper understanding of how digital twin technology can enhance building performance and sustainability.

4 Bibliometric Analysis and Results

Once the first four workflow stages, as depicted in Fig. 1, were fully implemented, additional tools for bibliometric investigations were considered. Keyword co-occurrence and bibliographic coupling analysis are frequently used methods in

bibliometric studies. Keyword co-occurrence helps identify relationships and thematic structures within a research field by analyzing how often specific terms appear together while Bibliographic coupling measures the similarity between documents based on shared references, revealing research clusters and influential studies.

4.1 Bibliometric Investigations

Selecting keywords co-occurrence and bibliographic coupling analysis for studying *building maintenance* using *digital twins* is driven by the need to fully understand and map the rapidly evolving research landscape in this field.

The keywords co-occurrence analysis enables the identification and visualization of the thematic structure of the literature, highlighting how key concepts such as ‘*digital twins*’, ‘*building maintenance*’, ‘*IoT*’ and ‘*sustainability*’ interrelate, as depicted in Fig. 2. This insight is crucial for recognizing emerging trends and core topics that drive advancements in building maintenance technologies. By examining the frequency and clustering of these keywords, the intellectual landscape can be uncovered.

Bibliographic coupling complements this by focusing on the connections and influences between studies through shared references. This method helps identify clusters of related research and trace the development of scientific knowledge. Additional tools enhance this process by providing powerful visualization and interaction capabilities: *bibliometrix* supports the generation of co-occurrence networks and maps citation networks, highlighting influential papers and emerging research fronts; *VOSviewer* offers clear visualization of how documents relate through shared references, identifying key contributions and research clusters; *Gephi* enables detailed and interactive exploration of citation networks, offering insights into the structure and dynamics of research communities.

Together, these tools and analyses provide a comprehensive understanding of the research landscape in building maintenance using digital twins. They help identify key themes, influential studies and emerging trends, facilitating informed and strategic research planning. The combination of keyword co-occurrence and bibliographic coupling analyses using these advanced tools ensures effective mapping, visualization and interpretation of complex bodies of scientific knowledge, thereby guiding future investigations and optimizing the integration and application of digital twins in building maintenance.

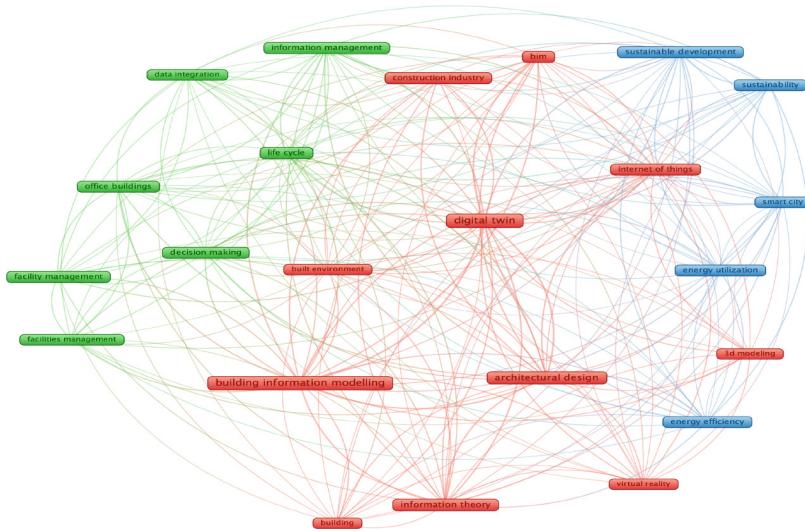


Fig. 2. The keyword co-occurrence network.

4.2 Insights of the Case Study

Digital twins have emerged as a transformative technology in various industries, including building management, contributing significantly to sustainability and reducing environmental and economic impacts.

By applying the methodology workflow (see Fig. 1), we were able to identify the most relevant $n = 87$ research entries, after the ‘Screening phase’. The key thematic areas that may be identified, based on the main corpus of papers, is discussed in current section. Exemplary papers are included in each paragraph.

4.2.1 Energy Efficiency and Resource Management ($n = 25$). A significant number of studies highlight the role of digital twins in enhancing energy efficiency and managing resources effectively. A digital twin approach in an energy hub resulted in substantial energy and cost savings, as explained in [10]. A retrofitting on a townhouse in Washington, DC was performed in [15]. It was shown that DT have the potential to reduce the intensity of energy use through various measures, like smart windows or renewable energy integration.

4.2.2 Sustainable Building Operations ($n = 15$). Digital twins can play a role in optimizing building operations, facilitating real-time monitoring and optimization of building operations, which leads to better energy management and reduced environmental impact [1]. Moreover, the use of DTs in managing HVAC systems and lighting in buildings has shown to significantly reduce energy consumption while maintaining occupant comfort [14].

4.2.3 Environmental Impact Reduction ($n = 12$). The integration of Building Information Modeling (BIM) and DTs allows for detailed analysis and optimization of energy consumption, thereby reducing greenhouse gas emissions [2, 15].

Also, digital twins aid in managing water resources more efficiently, contributing to overall environmental sustainability [10, 12].

4.2.4 Economic Benefits (n = 10). The ability to perform predictive maintenance and avoid unexpected failures reduces maintenance costs and extends the lifespan of building systems [6, 18]. The financial savings from optimized resource usage and improved operational efficiency are well-documented across several studies [13, 16].

4.2.5 Data-Driven Decision Making (n = 8). Digital twins can provide a robust platform for data collection and analysis, enabling informed decision-making. The integration of artificial intelligence and the IoT with DTs enhances their capability to analyze vast amounts of data and provide actionable insights [3, 25], thus supporting sustainable building management and policy formulation.

4.2.6 Enhanced Design and Construction Processes (n = 9). There are real benefits of digital twins in the design and construction phases, by facilitating better planning and reducing material wastage. They enable virtual testing of different design scenarios, leading to more efficient and sustainable building practices [22, 23]. The use of digital twins in the design phase ensures that sustainability considerations are integrated from the outset.

4.2.7 Smart Cities and Infrastructure (n = 8). The application of digital twins extends beyond individual buildings to include entire urban infrastructures, with specific integrations of DTs with smart city frameworks, transport optimization, utilities, and public services, and more sustainable urban environments [17, 24].

This comprehensive assessment illustrates that digital twins play a pivotal role in promoting sustainability and reducing both environmental and economic impacts in the building sector. The technology's ability to provide real-time data, optimize resources and support informed decision-making underscores its value in creating sustainable urban environments.

5 Conclusions

The methodological workflow used in our current paper, and described step-by-step, included analyses for keywords co-occurrence and bibliographic coupling, providing valuable thematic insights into key concepts like “digital twins”, “building maintenance”, “IoT”, and “predictive maintenance”, at the same time highlighting their interconnections and emerging trends.

The bibliographic coupling analysis identified influential studies and research clusters, offering a clear view of developmental pathways and pivotal contributions in digital twin technology for building maintenance. *Ontotext Refine* and *OpenRefine* ensure high data quality, the cleaning and curation steps thus enhancing the accuracy and reliability of the bibliometric analysis.

Visualisation solutions like *VOSviewer* and *Gephi* enable detailed and interactive visualizations of the research landscape, facilitating a deeper understanding of complex scientific relationships. This comprehensive strategy supports

informed and strategic research planning, guiding future studies and optimizing the application of digital twins in building maintenance. The systematic and reproducible approach used ensures high-quality, evidence-based insights that significantly advance the field.

Acknowledgements. The Romania Competitiveness Operational Programme partially supported this paper, under project number SMIS 120725 - SCAMP-ML (Advanced computational statistics for planning and tracking production environments). The research conducted in this paper was partially supported by the UVT 1000 Develop Fund of the West University of Timișoara.

References

1. Agostinelli, S.: Digital twin model for zero-energy districts: the case study of Anzio port, Italy. *WIT Trans. Ecol. Environ.* **260**, 357–363 (2022). <https://doi.org/10.2495/SC220291>
2. Agostinelli, S., Cumo, F., Nezhad, M.M., Orsini, G., Piras, G.: Renewable energy system controlled by open-source tools and digital twin model: Zero energy port area in Italy. *Energies* **15**(5), 1817 (2022). <https://doi.org/10.3390/en15051817>
3. Antonino, M., Nicola, M., Claudio, D.M., Luciano, B., Fulvio, R.C.: Office building occupancy monitoring through image recognition sensors. *Int. J. Saf. Secur. Eng.* **9**(3), 371–380 (2019). <https://doi.org/10.2495/SAFE-V9-N4-371-380>
4. Aria, M., Cuccurullo, C.: bibliometrix: an R-tool for comprehensive science mapping analysis. *J. Informat.* **11**(4), 959–975 (2017). <https://doi.org/10.1016/j.joi.2017.08.007>
5. Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 3, no. 1, pp. 361–362 (2009). <https://doi.org/10.1609/icwsm.v3i1.13937>
6. Boje, C., et al.: A framework using BIM and digital twins in facilitating LCSA for buildings. *J. Build. Eng.* **76**, 107232 (2023). <https://doi.org/10.1016/j.jobe.2023.107232>
7. Booth, A., James, M.S., Clowes, M., Sutton, A., et al.: Systematic Approaches to a Successful Literature Review, 3rd edn. SAGE Publications Ltd., London (2021)
8. Bortolini, R., Rodrigues, R., Alavi, H., Vecchia, L.F.D., Forcada, N.: Digital twins' applications for building energy efficiency: a review. *Energies* **15**(19), 7002 (2022). <https://doi.org/10.3390/en15197002>
9. van Dinter, R., Tekinerdogan, B., Catal, C.: Predictive maintenance using digital twins: a systematic literature review. *Inf. Softw. Technol.* **151**, 107008 (2022). <https://doi.org/10.1016/j.infsof.2022.107008>
10. Dulaimi, A., Hamida, R., Naser, M., Mawed, M.: Digital twin solution implemented on energy hub to foster sustainable smart energy city, case study of sustainable smart energy hub. *ISPRS Ann. Photogrammetry Remote Sens. Spatial Inf. Sci.* **X-4/W3-2022**, 41–48 (2022). <https://doi.org/10.5194/isprs-annals-X-4-W3-2022-41-2022>
11. van Eck, N.J., Waltman, L.: VOS: a new method for visualizing similarities between objects. In: Decker, R., Lenz, H.-J. (eds.) *Advances in Data Analysis*. SCDAKO, pp. 299–306. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-70981-7_34

12. Evangelou, T., Gkeli, M., Potsiou, C.: Building digital twins for smart cities: a case study in Greece. *ISPRS Ann. Photogrammetry Remote Sens. Spatial Inf. Sci.* **X-4/W2-2022**, 61–68 (2022). <https://doi.org/10.5194/isprs-annals-X-4-W2-2022-61-2022>
13. Franciosi, C., Miranda, S., Veneroso, C.R., Riemma, S.: Improving industrial sustainability by the use of digital twin models in maintenance and production activities. *IFAC-PapersOnLine* **55**(19), 37–42 (2022). <https://doi.org/10.1016/j.ifacol.2022.09.215>
14. Hosamo, H., Hosamo, M.H., Nielsen, H.K., Svennevig, P.R., Svidt, K.: Digital twin of HVAC system (HVACDT) for multiobjective optimization of energy consumption and thermal comfort based on BIM framework with ANN-MOGA. *Adv. Build. Energy Res.* **17**(2), 125–171 (2022). <https://doi.org/10.1080/17512549.2022.2136240>
15. Kaewunruen, S., Sresakoolchai, J., Kerinnonta, L.: Potential reconstruction design of an existing townhouse in Washington DC for approaching net zero energy building goal. *Sustainability* **11**(23), 6631 (2019). <https://doi.org/10.3390/su11236631>
16. Kaewunruen, S., Xu, N.: Digital twin for sustainability evaluation of railway station buildings. *Front. Built Environ.* **4**, 430624 (2018). <https://doi.org/10.3389/fbuil.2018.00077>
17. Lv, Z., Chen, D., Lv, H.: Smart city construction and management by digital twins and BIM big data in COVID-19 scenario. *ACM Trans. Multimed. Comput. Commun. Appl.* **18**(2s), 1–21 (2022). <https://doi.org/10.1145/3529395>
18. Massafra, A., Predari, G., Gulli, R.: Towards digital twin driven cultural heritage management: A HBIM-based workflow for energy improvement of modern buildings. *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.* **XLVI-5/W1-2022**, 149–157 (2022). <https://doi.org/10.5194/isprs-archives-XLVI-5-W1-2022-149-2022>
19. Nuaimi, M., Fourati, L.C., Hamed, B.B.: Intelligent approaches toward intrusion detection systems for Industrial Internet of Things: A systematic comprehensive review. *Journal of Network and Computer Applications* **215**, 103637 (Jun 2023) <https://doi.org/10.1016/j.jnca.2023.103637>
20. Pagani, R.N., Kovaleski, J.L., Resende, L.M.: Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. *Scientometrics* **105**(3), 2109–2135 (2015). <https://doi.org/10.1007/s11192-015-1744-x>
21. Pessin, V.Z., Yamane, L.H., Siman, R.R.: Smart bibliometrics: an integrated method of science mapping and bibliometric analysis. *Scientometrics* **127**(6), 3695–3718 (2022). <https://doi.org/10.1007/s11192-022-04406-6>
22. Ruperto, F., Strappini, S.: Complex works project management enhanced by digital technologies. In: *Building Information Modelling (BIM) in Design, Construction and Operations IV*. BIM21, vol. 1, p. 235–248. WIT Press (2021). <https://doi.org/10.2495/bim210201>
23. Sacks, R., Brilakis, I., Pikas, E., Xie, H.S., Girolami, M.: Construction with digital twin information systems. *Data-Centric Eng.* **1** (2020). <https://doi.org/10.1017/dce.2020.16>
24. Shaharuddin, S., Abdul Maulud, K.N., Syed Abdul Rahman, S.A.F., Che Ani, A.I.: Digital twin for indoor disaster in smart city: a systematic review. *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.* **XLVI-4/W3-2021**, 315–322 (2022). <https://doi.org/10.5194/isprs-archives-XLVI-4-W3-2021-315-2022>

25. Torrecilla-García, J.A., Pardo-Ferreira, M.C., Rubio-Romero, J.C.: Overall introduction to the framework of BIM-based digital twinning in decision-making in safety management in building construction industry. *Dirección y Organización* (74), 31–38 (2021). <https://doi.org/10.37610/dyo.v0i74.600>
26. Vieira, E.L., da Costa, S.E.G., de Lima, E.P., Ferreira, C.C.: Application of the Proknow-C methodology in the search of literature on performance indicators for energy management in manufacturing and Industry 4.0. *Procedia Manuf.* **39**, 1259–1269 (2019). <https://doi.org/10.1016/j.promfg.2020.01.343>
27. Zupic, I., Čater, T.: Bibliometric methods in management and organization. *Organ. Res. Methods* **18**(3), 429e28093472 (2014). <https://doi.org/10.1177/1094428114562629>



Blockchain and Digital Twin Integration for Remote Control of Cyber-Physical Systems

Alessandro Bigiotti¹(✉), Purav Shah², and Ramona Trestian²

¹ Division of Computer Science, University of Camerino, Via Madonna delle Carceri, 62032 Camerino, Italy
alessandro.bigiotti@unicam.it

² Department of Design Engineering and Mathematics, Middlesex University London, The Burroughs, London NW4 4BT, UK
{p.shah,r.trestian}@mdx.ac.uk

Abstract. In the era of Industry 4.0, Cyber-Physical Systems (CPS) have become integral to various domains, with the Digital Twin (DT) concept playing a crucial role in mirroring and optimising these systems' behaviours. The Digital Twin emulates the physical device behaviour based on the Internet of Things (IoT) sensor readings within CPSs. However, managing the CPS configurations presents significant challenges in terms of security, integrity, and traceability. This paper proposes an innovative integration of blockchain technology with Digital Twin systems to address these challenges. Blockchain's decentralised and immutable ledger allows for the secure storage and management of IoT sensor properties and CPS configurations. Furthermore, a threshold signature scheme is employed to enable the Digital Twin to autonomously modify CPS behaviour, with changes being verified through smart contracts. This approach ensures that configuration changes are applied only if authorised by the Digital Twin components, improving the security and resilience of the CPS. The proposed solution improves the security and integrity of sensor management while enabling more dynamic and adaptive responses to real-time data.

Keywords: Blockchain · Digital Twin · IoT device · Industry 4.0 · Smart contract

1 Introduction

In the rapidly evolving landscape of Industry 4.0, Cyber-Physical Systems (CPS) have become a cornerstone for smart manufacturing, autonomous vehicles, and

This research was funded by Ministero dell'Università e della Ricerca (MUR), issue D.M. 351/2022 “Borse di Dottorato” - Dottorato di Ricerca di Interesse Nazionale in “Blockchain & Distributed Ledger Technology”, under the National Recovery and Resilience Plan (NRRP).

advanced robotics [19]. One of the catalysts that has enabled this technological evolution lies in the Internet of Things (IoT), characterised by special sensors capable of conducting physical measurements in real-time [18]. Due to advancements in IoT sensor technology, the concept of the Digital Twin (DT) has emerged as a precise virtual representation that mirrors the physical system's behaviours, states, and interactions. The DT enables real-time monitoring, simulation, and predictive analysis, facilitating optimised decision-making and system efficiency [20]. The benefits of integrating IoT devices and DTs into manufacturing systems have been widely discussed in the literature. The main applications concern predictive maintenance, benchmarking, real-time tracking, energy efficiency, cost reduction, and operational intelligence [8]. A significant challenge we face today is enabling seamless interaction between the DT and the physical system. Facilitating the DT's ability to remotely and automatically control physical devices can optimise manufacturing processes and, importantly, prevent damage and failures caused by malfunctions or anomalies. However, this integration presents several challenges, particularly concerning the security of communications and the management of large volumes of data generated by IoT devices.

CPSs rely extensively on IoT sensors to collect and process data, making the integrity, security, and traceability of sensor data paramount. Traditional centralised approaches to managing sensor configurations are often susceptible to cyberattacks, tampering, and data loss. In this context, blockchain technology emerges as a promising solution. With its decentralised, immutable, and transparent ledger, blockchain provides a robust framework for securely storing and managing the configurations of IoT sensors within a CPS [15]. By integrating blockchain with a Digital Twin, the properties of these sensors, along with the configurations of the physical system, can be securely stored. Furthermore, blockchain can serve as a communication medium, facilitating the dynamic adjustment of the physical system based on real-time data and conditions.

This paper investigates the integration of blockchain within Digital Twin systems, with a focus on utilising blockchain to dynamically change the behaviour of the physical system. The proposed approach adopts an IoT-to-DT strategy, where IoT devices communicate directly with the digital counterpart. The DT operates as decentralised software that accurately mirrors the behaviour of the physical system. In the proposed framework, a (t of n) Threshold Signature Scheme (TSS) is employed to enable secure and collaborative decision-making by the DT. The TSS allows for the distributed signing of transactions, ensuring that changes to the physical system are authorised only when a predefined number of parties consent. These signatures are subsequently verified by a smart contract deployed on the blockchain, which ensures that any modifications meet the required criteria before being implemented on the physical CPS. The proposed approach not only enhances security but also enables the DT to autonomously adjust CPS behaviours in response to evolving operational environments, thereby optimising the performance and resilience of the entire system.

1.1 Outline

The paper is structured as follows: Sect. 2 shows the background, the related works, the motivations, and the contributions of the work. Section 3 proposes a solution for integrating blockchain with DTs to enhance the security of communications between DTs and the physical system. Section 4 shows the experimental results, presents the limitations associated with the current capabilities of modern blockchains and offers a critical discussion. Finally, Sect. 5 provides the conclusions.

2 Background and Related Work

2.1 Blockchain Technology for Industry 4.0

Blockchain technology, a decentralised and tamper-proof ledger, has the potential to significantly revolutionise Industry 4.0, the fourth industrial revolution characterised by the integration of new and emerging technologies into different social contexts [9]. Blockchain's ability to provide transparent, immutable records can enhance supply chain transparency and traceability, ensuring that all transactions and processes are verifiable and secure [17]. This transparency is crucial in smart factories, where automated systems can operate more efficiently with reliable data. Smart contracts favoured the application of the blockchain within Industry 4.0. Indeed, being self-executing contracts with terms directly coded, they enable automated, decentralised operations by enforcing agreements without intermediaries, reducing costs, and increasing efficiency [24]. For instance, smart contracts can automatically trigger maintenance, reorder supplies, or release payments when predefined conditions are met, ensuring that operations run smoothly and without human intervention. Collaboration among multiple stakeholders in Industry 4.0 is also enhanced through blockchain, as it provides a shared, trusted ledger where all parties can track progress and enforce agreements transparently. This fosters trust and reduces the need for costly intermediaries.

2.2 Digital Twin

A Digital Twin is a virtual replica of a physical object, system, or process that mirrors its real-world counterpart in real time using data from IoT sensor devices. This technology allows continuous monitoring, analysis, and real-time optimisation, offering significant potential across various industries [16]. In Industry 4.0, DTs play a central role by enhancing monitoring and maintenance, enabling predictive maintenance that reduces downtime and extends the lifespan of the equipment [23]. They also optimise operations by allowing manufacturers to simulate ‘what-if’ scenarios and test configurations virtually before applying them to physical processes, minimising disruptions and maximising efficiency. In product design and development, DTs accelerate innovation by enabling engineers to simulate and test designs virtually, reducing the need for physical prototypes.

and identifying potential issues early. This not only cuts costs but also improves product quality. Furthermore, DTs support data-driven decision-making, providing a comprehensive view of an asset's performance and helping businesses make informed strategies, especially in complex systems.

2.3 Threshold Signature

Threshold signatures are a cryptographic scheme that allows a group of n participants to collaboratively generate a digital signature without requiring all members to be involved. Only a predefined minimum number of participants (the threshold t) is needed to produce a valid signature. This method enhances security by distributing the private key among multiple participants, ensuring that no single party can generate the signature alone, thus reducing the risk of key compromise. In recent years, threshold signatures have attracted great interest in the scientific community. Some recent advancements made two well-known threshold signature schemes really versatile and efficient. The authors in [4] propose a threshold signature scheme based on the Boneh-Lynn-Shacham (BLS), improving the algorithm in charge of distributing the private keys and allowing the dynamic extension and reduction of the signing participants. The authors in [11] present a comparison of various threshold signature schemes based on the Elliptic Curve Digital Signature Algorithm (ECDSA) [10], whereas the authors in [21] propose a scheme for integrating an ECDSA-based threshold signature within blockchain technology. In distributed software, threshold signatures provide several key advantages. They enhance fault tolerance by allowing the system to continue functioning even if some participants are offline, as long as the required threshold is met. Thus, increasing the system resilience to failures. Moreover, threshold signatures support decentralised control, enabling collective decision-making without the need for unanimous agreement. This not only accelerates processes but also helps to avoid bottlenecks. This feature is particularly advantageous in distributed consensus protocols, where a subset of nodes can validate transactions or blocks. Additionally, threshold signatures improve privacy and confidentiality by ensuring that no single participant can reconstruct the entire private key, which is particularly valuable in collaborative environments where trust may be a concern.

2.4 Related Work

The integration of the blockchain with DTs has been widely explored in the literature. The characteristics of blockchain can bring numerous benefits, such as increasing the decentralisation of the system, increasing the transparency and traceability of the operations conducted, and increasing the integrity and reliability of the data produced by such systems.

Huang et al. [6] address the challenges of data management in DT systems within Product Life-cycle Management (PLM). The authors propose integrating blockchain technology to enhance data security, integrity, and sharing efficiency. By using blockchain, the approach ensures secure data storage, access,

and authenticity while preventing data loss through continuous updates in DTs. A case study on a turbine DT demonstrates the effectiveness of their method, showing how blockchain can manage life cycle data securely and efficiently.

Putz et al. [14] introduce EtherTwin, a blockchain-based decentralised application (DApp) for managing DT data in Industry 4.0. The solution ensures secure data sharing among multiple untrusted parties by leveraging blockchain's decentralised structure, fine-grained access control, and off-chain encrypted data storage. Validated through a prototype and industry use case, EtherTwin addresses the need for confidentiality, integrity, and availability in DT data management, offering a robust approach to secure life cycle data sharing.

Liu et al. [13] propose using blockchain to enhance data exchange and collaboration in digital twin manufacturing systems (DTMS). The study introduces a peer-to-peer data exchange mechanism via a manufacturing edge pool, reducing reliance on cloud systems and improving efficiency and flexibility. A case study on engine manufacturing validates the approach, highlighting blockchain's potential to solve cloud manufacturing challenges and improve DTMS collaboration.

Vairavasundaram et al. [1] propose a blockchain-based Proof of Authority (PoA) trust mechanism to secure data and enhance privacy in DT systems for the Industrial Internet of Things (IIoT). The authors address the challenges of trust, security, and data management in decentralised IIoT networks by integrating blockchain with digital twin technology. The proposed PoA mechanism reduces energy consumption and improves the efficiency and security of data transactions between IIoT nodes.

Hasan et al. [5] present a blockchain-based solution to securely manage the creation process of digital twins (DTs). The proposed approach leverages blockchain technology to ensure the traceability, accessibility, and immutability of DT transactions and data. The system incorporates smart contracts to govern interactions and employs decentralised storage via the Inter-Planetary File System (IPFS) to store and share DT information securely. The work addresses the limitations of traditional, centralised DT creation methods by offering a decentralised, secure, and transparent framework.

The literature analysis indicates that blockchain has primarily been used to track activities and manage data exchanged within CPS. This application undoubtedly enhances transparency in operations and facilitates data sharing across different organisations. An important feature that modern DTs should have is the capability to interact with the physical system, enabling remote control mechanisms to automate interventions and minimise the need for human involvement. However, it appears that this functionality has not yet been thoroughly explored in the existing literature.

2.5 Contribution

A key feature of modern Digital Twins is the ability to remotely and automatically control the physical system. However, the literature analysis reveals a lack of research addressing this functionality. Therefore, the contribution of this work is summarised as follows:

- C.1: Integrate blockchain within a cyber-physical factory to maintain a census of physical devices and their configurations;
- C.2: Propose a decentralised Digital Twin that monitors the state of the physical system and can automatically modify its behaviour;
- C.3: Enhance security in communications between the Digital Twin and the physical system by using blockchain and an ECDSA-based threshold signature scheme.

3 Cyber-Physical Factory: A Use Case

A cyber-physical factory is an advanced manufacturing environment in which physical processes and digital technologies are deeply integrated and interconnected. In such a factory, physical manufacturing processes are carefully monitored, managed, and optimised through digital means, enabling smarter, more efficient, and more flexible manufacturing. In such systems, physical components may include conveyor belts, cameras to detect the presence of defects, stations for heat or pressure treatment, and any other devices necessary to build a product. The physical components are equipped with IoT sensors that are intended to take measurements of the physical world and enable real-time monitoring of the entire system. Monitoring is carried out through a DT that, by recording data produced by sensors, is able to reflect changes in the physical world in real-time. The DT must be able to automatically modify the behaviour of the physical system. This allows to improve the production process and prevent any failures resulting from malfunctions.

3.1 Enabling Remote Control With Digital Twins

This section presents the proposed solution, which leverages blockchain as a communication medium between the DT and the physical system. Communications between the DT and the CPS are triggered by the detection of anomalies or potential failures, necessitating modifications to the physical system's behaviour to prevent further damage. The proposed solution follows an IoT-to-DT approach, designed to maximise sensor autonomy and eliminate centralised components responsible for managing communications. Given that IoT devices are generally hardware-constrained and susceptible to compromise, ensuring a minimum level of security in communications between the physical system and the DT is critical. To achieve this, IoT sensors employ a lightweight cryptography scheme [3], enabling them to securely sign messages sent to the DT.

In our approach, the DT must be able to modify the behaviour of the physical system to prevent failures or malfunctions. To this end, the blockchain plays a key role in the infrastructure, maintaining the census of the devices present in the system and maintaining the configurations of the physical system. These configurations might include parameters such as conveyor belt speed, temperature and timing for specific treatments, applied pressure on certain objects, and other operational settings. Blockchain ensures that these configurations are

tamper-proof while allowing them to be shared transparently and securely with the DT.

Figure 1 illustrates the proposed structure. Each sensor $S.i$ is responsible for monitoring the state of the component to which it is connected, with sensor readings tailored to the specific task being monitored. Once the data is collected, the sensor prepares a message and signs it using a lightweight cryptography scheme. The signed messages are then sent to the DT. In this framework, it is assumed that the DT shares the messages received from the sensors, allowing all parties to access the content to verify the correctness of the signatures and the associated data.

During operation, special messages (denoted by \star in Fig. 1) may be generated, indicating an anomaly that could lead to system or product damage. For example, these messages might report that the conveyor belt is moving at an incorrect speed or that the temperature for a heat treatment is too high. When such messages are detected, the DT must intervene to either restore proper operation or halt the assembly line to prevent damage. However, allowing a single DT component to modify the behaviour of the physical system poses significant risks. If the component is compromised or acts maliciously, it could intentionally alter the system's behaviour, potentially causing damage to individual products or the entire system. To enhance the security and resilience of the DT, an ECDSA-based threshold signature scheme has been introduced.

For example, suppose that in Fig. 1, the sensor $S.3$ detects a malfunction during a product construction phases. The message marked with \star is read by the DT component $D.3$. The message cannot be forged, as it is signed by the specific sensor, ensuring its authentication and integrity. Based on the message's content, the DT component must determine how to intervene on the physical system and prepare a message (*data*) containing the necessary adjustments.

At this point the digital twin works as follows:

1. The component $D.3$ sends a message (*data*) containing the changes to be made to the physical system, to the other components of the DT. The message is signed with its partial key $sk_{D.3}$, $(\sigma_{D.3}(data)) = (r, s)_{D.3}$, where r and s are the output of the ECDSA algorithm;
2. Each component $D.i$ of the DT verifies the integrity of the message related to the anomaly $\sigma_{S.3}(\text{msg}, H)$. If the message is invalid, the component $D.i$ rejects the request. If the message is valid, the component $D.i$ verifies the modification requested by the component $D.3$, first checking the received signature $(r, s)_{D.3}$ and then the content of the (*data*);
3. If the signature is invalid or if the requested changes are inconsistent with $D.3$'s request, the DT components reject the request. If the changes are validated, each component $D.i$ of the DT proceeds to sign $D.3$'s request using its own partial key $sk_{D.i}$, producing $\sigma_{D.i}(data) = (r, s)_{D.i}$, and then sends the partial signature back to $D.3$;
4. Component $D.3$ waits to receive at least t valid partial signatures $(r, s)_{D.i}$ from the other components $D.i$. Once the partial signatures are collected,

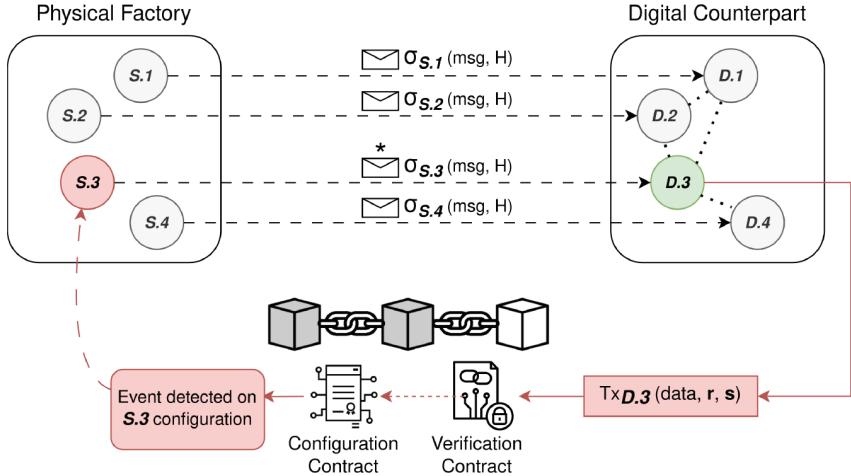


Fig. 1. The figure shows the integration of a blockchain for communication between the Digital Twin and its physical counterpart. $S.i$ sensors send signed messages to their Digital Twin $D.i$. The Digital Twin verifies the messages and can send transactions to the blockchain that must be signed using a threshold signature scheme.

component $D.3$ produces the threshold signature $\sigma(\text{data}) = (r, s)$ and constructs the transaction $(TxD.3(\text{data}, r, s))$ to be sent to the blockchain.

5. The transaction $TxD.3$ is directed to the smart contract responsible for verifying the threshold signature. If the signature is valid, the verification smart contract initiates an internal transaction to the configurations smart contract, which will then emit an event indicating the necessary changes to be made to the physical system.

The blockchain ensures that changes to the physical system are secure and tamper-proof due to its inherent mechanisms. Before any event containing instructions can be generated, the corresponding transactions must first be validated by the blockchain. This validation process includes the verification of a threshold signature, which ensures that no single component can unilaterally alter the behaviour of the physical system. Consequently, this approach not only enhances the security of communications between the DT and the CPS but also increases the resilience of the entire system.

4 Experimental Result

The examined use case does not present scalability issues, as the interaction with the blockchain occurs only in case of anomalies or malfunctions, which should be rare cases. The threshold signature verification has been implemented on smart contracts and does not constitute an issue. Thus, in order to understand the feasibility of the proposed solution, we measured the latency in communications

between the DT and the physical system. In particular, latency means the time elapsed between sending a transaction on the blockchain and its validation. The experiments were conducted on a Linux machine Pop-Os 6.9.3, RAM 32 GB Intel Core i7-10870H CPU @ 2.20 GHz, GPU 8 GB NVIDIA GeForce RTX 3070. The blockchain has been implemented in Hyperledger Besu (v.23.1.0-RC1) [7] using Docker (v.27.2.0) and docker compose (v. 1.29.2). The blockchain is composed of 5 validator nodes, implements a Proof of Authority consensus algorithm (IBFT 2.0), has a block generation time of 1 s and is gas free (maximum amount of gas per block). The smart contracts were written in Solidity (v. 0.8.12) [22] and are deployed using Truffle (v. 5.7.5). The DT was simulated using multi-threading processes and implements a (7 of 10) ECDSA-based threshold signature scheme. Latency was measured on 5000 transactions submitted to the blockchain at random time intervals. Figure 2 shows the latency distribution over 5000 transactions. From the experiments conducted we obtain that the introduction of the blockchain is a viable solution, obtaining an average latency of 0.627 s.

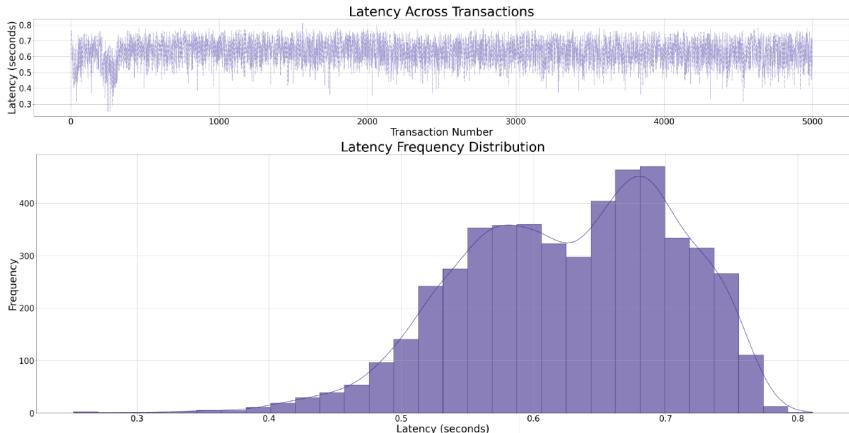


Fig. 2. The figure shows the latency in the transactions used for testing. Latency was calculated as the time between sending a transaction and the moment the transaction is recorded in a block.

4.1 Limitation, Challenge and Future Work

Currently, the most popular blockchains, such as Ethereum, do not natively support threshold signature schemes for generating transactions. This presents a challenge for the proposed approach, which can be addressed by implementing verification of such signatures within smart contracts (as illustrated by the Verification Contract of Fig. 1). Executing transactions on blockchain requires consensus from all participants, introducing several limitations. For instance, on EVM-based blockchains, each transaction is translated into a list of OPCODEs.

Each OPCODE has a specific cost known as *gas*. While there is no limit on the gas consumed per transaction, there is a limit on the total gas consumed within a block. Different blockchains have varying properties and configurations, with a wide range of options available today. Although this approach is feasible, verifying a signature on a smart contract requires an additional transaction, thereby increasing the workload of blockchain validator nodes and negatively impacting the overall blockchain performance. In the future, blockchains may natively integrate threshold signature schemes for generating transactions, which would significantly expedite verification of threshold signatures.

Another challenge is mitigating malicious behaviours from the DT. While decentralising the DT enhances fault tolerance, it also necessitates vigilance against potential malfunctions or malicious actions by individual components. The introduction of the threshold signature should ensure correct operation as long as at least t components of the DT remain intact and functional. To address malicious behaviours, one potential countermeasure is the implementation of a reputation system for the components of the DT [12]. However, identifying and analysing malicious behaviour is complex and may present hidden pitfalls, warranting further investigation.

The proposed approach has broader applications in various sectors of Industry 4.0. For example, the authors in [2] propose the use of blockchain for the registration of pollutants emitted by incinerators used for electricity production. By applying the proposed approach, it would be possible to enhance the control of pollutant emissions, automatically shutting down incinerators if emissions exceed the legally mandated thresholds. This could help avoid potential fines from regulatory bodies and, while promoting a more environmentally sustainable business model.

5 Conclusion

This work presents the integration of blockchain technology with the Digital Twin of a Cyber-Physical System in the manufacturing sector. The blockchain is responsible for maintaining the census of the IIoT devices and the configurations of the physical system. By utilising blockchain, the DT can modify the behaviour of the physical system by sending transactions to specific smart contracts. The execution of these transactions triggers events that contain instructions to adjust the physical system's behaviour. To enhance the security of these operations, the DT employs an ECDSA-based threshold signature scheme. The use of threshold signatures prevents single component from independently altering the behaviour of the physical system, thereby avoiding single points of failure and improving the overall security of the entire system. The proposed approach can be integrated in other Industry 4.0 sectors facing similar challenges, where there is a need for automatic intervention in physical system. Building on this framework, further research can explore additional methodologies or the integration of alternative signature schemes, such as Boneh-Lynn-Shacham (BLS).

References

1. Sasikumar, A., et al.: Blockchain-based trust mechanism for digital twin empowered industrial internet of things. Future Gener. Comput. Syst. **141**, 16–27 (2023). <https://doi.org/10.1016/j.future.2022.11.002>, <https://www.sciencedirect.com/science/article/pii/S0167739X22003636>
2. Bigiotti, A., Mostarda, L., Navarra, A.: Blockchain and IoT integration for air pollution control. In: Barolli, L. (eds.) Advanced Information Networking and Applications. AINA 2024. Lecture Notes on Data Engineering and Communications Technologies, vol. 200, pp. 27–38. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-57853-3_9
3. 54 El-hajj, M., Mousawi, H., Fadlallah, A.: Analysis of lightweight cryptographic algorithms on IoT hardware platform. Future Internet **15**(2), (2023). <https://doi.org/10.3390/fi15020054>, <https://www.mdpi.com/1999-5903/15/2/54>
4. Garg, S., Jain, A., Mukherjee, P., Sinha, R., Wang, M., Zhang, Y.: hints: threshold signatures with silent setup. Cryptology ePrint Archive, Paper 2023/567 (2023). <https://eprint.iacr.org/2023/567>
5. Hasan, H.R., et al.: A blockchain-based approach for the creation of digital twins. IEEE Access **8**, 34113–34126 (2020). <https://doi.org/10.1109/ACCESS.2020.2974810>
6. Huang, S., Wang, G., Yan, Y., Fang, X.: Blockchain-based data management for digital twin of product. J. Manuf. Syst. **54**, 361–371 (2020). <https://doi.org/10.1016/j.jmsy.2020.01.009>, <https://www.sciencedirect.com/science/article/pii/S0278612520300091>
7. Hyperledger: Hyperledger besu (2022). <https://www.hyperledger.org/use/besu>
8. Javaid, M., Abid Haleem, Pratap Singh, R., Rab, S., Suman, R.: Upgrading the manufacturing sector via applications of industrial internet of things (IIoT). Sens. Int. **2**, 100129 (2021). <https://doi.org/10.1016/j.sintl.2021.100129>, <https://www.sciencedirect.com/science/article/pii/S2666351121000504>
9. Javaid, M., Haleem, A., Pratap Singh, R., Khan, S., Suman, R.: Blockchain technology applications for industry 4.0: a literature-based review. Blockchain: Res. Appl. **2**(4), 100027 (2021). <https://doi.org/10.1016/j.bcra.2021.100027>
10. Johnson, D., Menezes, A., Vanstone, S.: The elliptic curve digital signature algorithm (ECDSA). Int. J. Inf. Secur. **1**(1), 36–63 (2001). <https://doi.org/10.1007/s102070100002>
11. Kachouh, B., Sliman, L., Samhat, A.E., Barkaoui, K.: Demystifying threshold elliptic curve digital signature algorithm for multiparty applications. In: Proceedings of the 2023 Australasian Computer Science Week, ACSW '23, pp. 112–121. Association for Computing Machinery, New York, NY, USA (2023). <https://doi.org/10.1145/3579375.3579389>
12. Lei, K., Zhang, Q., Xu, L., Qi, Z.: Reputation-based byzantine fault-tolerance for consortium blockchain. In: 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), pp. 604–611 (2018). <https://doi.org/10.1109/PADSW.2018.8644933>
13. Liu, S., Lu, Y., Li, J., Shen, X., Sun, X., Bao, J.: A blockchain-based interactive approach between digital twin-based manufacturing systems. Comput. Ind. Eng. **175**, 108827 (2023). <https://doi.org/10.1016/j.cie.2022.108827>, <https://www.sciencedirect.com/science/article/pii/S0360835222008154>
14. Putz, B., Dietz, M., Empl, P., Pernul, G.: EtherTwin: blockchain-based secure digital twin information management. Inf. Process. Manage. **58**(1), 102425

- (2021). <https://doi.org/10.1016/j.ipm.2020.102425>. <https://www.sciencedirect.com/science/article/pii/S0306457320309195>
- 15. Rejeb, A., et al.: Unleashing the power of internet of things and blockchain: a comprehensive analysis and future directions. *Internet Things Cyber-Phys. Syst.* **4**, 1–18 (2024). <https://doi.org/10.1016/j.iotcps.2023.06.003>, <https://www.sciencedirect.com/science/article/pii/S2667345223000366>
 - 16. Semeraro, C., Lezoche, M., Panetto, H., Dassisti, M.: Digital twin paradigm: a systematic literature review. *Comput. Ind.* **130**, 103469 (2021). <https://doi.org/10.1016/j.compind.2021.103469>, <https://www.sciencedirect.com/science/article/pii/S0166361521000762>
 - 17. Shakhrbulatov, D., Medina, J., Dong, Z., Rojas-Cessa, R.: How blockchain enhances supply chain management: a survey. *IEEE Open J. Comput. Soc.* **1**, 230–249 (2020). <https://doi.org/10.1109/OJCS.2020.3025313>
 - 18. Soori, M., Arezoo, B., Dastres, R.: Internet of things for smart factories in industry 4.0, a review. *Internet Things Cyber-Phys. Syst.* **3**, 192–204 (2023). <https://doi.org/10.1016/j.iotcps.2023.04.006>, <https://www.sciencedirect.com/science/article/pii/S2667345223000275>
 - 19. Tyagi, A.K., Sreenath, N.: Cyber physical systems: Analyses, challenges and possible solutions. *Internet Things Cyber-Phys. Syst.* **1**, 22–33 (2021). <https://doi.org/10.1016/j.iotcps.2021.12.002>, <https://www.sciencedirect.com/science/article/pii/S2667345221000055>
 - 20. Yaqoob, I., Salah, K., Uddin, M., Jayaraman, R., Omar, M., Imran, M.: Blockchain for digital twins: recent advances and future research challenges. *IEEE Netw.* **34**(5), 290–298 (2020). <https://doi.org/10.1109/MNET.001.1900661>
 - 21. Yu, H., Wang, H.: Elliptic curve threshold signature scheme for blockchain. *J. Inf. Secur. Appl.* **70**, 103345 (2022). <https://doi.org/10.1016/j.jisa.2022.103345>
 - 22. Zheng, G., Gao, L., et al.: Solidity (2016). <https://docs.soliditylang.org/>
 - 23. Zhong, D., Xia, Z., Zhu, Y., Duan, J.: Overview of predictive maintenance based on digital twin technology. *Heliyon* **9**(4), e14534 (2023). <https://doi.org/10.1016/j.heliyon.2023.e14534>, <https://www.sciencedirect.com/science/article/pii/S2405844023017413>
 - 24. Zou, W., et al.: Smart contract development: challenges and opportunities. *IEEE Trans. Softw. Eng.* **47**(10), 2084–2106 (2021). <https://doi.org/10.1109/TSE.2019.2942301>



Optimising Sea Rescue Missions by UAVs

Sajjad Ghobadi and Francesco Piselli^(✉)

Department of Mathematics and Computer Science, University of Perugia,
Perugia, Italy
sajjad.ghobadibabi@unipg.it, francesco.piselli@unifi.it

Abstract. This paper investigates the First Boat Rescue (FBR) problem whose objective is to rescue a set of boats at sea in the shortest time possible. This problem is motivated by sea emergencies where boats require urgent medical assistance. Since lifeboat missions are costly and time-consuming, the use of Unmanned Aerial Vehicles (UAVs) is highly recommended. UAVs can efficiently assist all boats by carrying necessary medical tools. However, their limited battery capacity necessitates frequent visits to gas stations, referred to as *buoys*, to recharge their battery. In this study, we evaluate the performance of existing algorithms for the FBR problem by considering various distributions of buoys. Using the several algorithms proposed for FBR, we demonstrate that changing the position of buoys can improve the algorithms' efficiency and overall performance. The obtained results highlight the importance of buoy positioning in improving rescue operations, optimising costs, and ensuring the boat rescue missions in the shortest time possible.

1 Introduction

Travelling by boat, for recreation, sport or simply to commute from one place to another, is nowadays more and more widespread all around the world. Obviously, as in all kinds of activities, the possibility of having problems should always be considered, especially in cases where aid is not at hand, e.g., when you are in open sea. It is also important to consider that most of the boats used in those situations have small or medium size, therefore they are more vulnerable to problems caused by bad weather. Moreover, most of the time, it is unlikely that those boats have on board medical experts or appropriate medical equipment to be able to intervene in the event of health problems of the crew members. These considerations are surely true worldwide and for example, in Italy, in 2020 about 71.192 boats were registered [2], with 99% of those with length of 24 m or less. Regarding accidents, the Italian statistics [2] reported 3632 calls for assistance which required the use of rescue lifeboats carrying medical personnel. Main attention should be placed on the number of boats that actually

The work has been supported in part by the Italian National Group for Scientific Computation GNCS-INdAM and by the “BREADCRUMBS” project funded by the PRIN 2022 PNRR under grant no. P2022K7ERB.

required to be examined on board by medical experts. In fact, out of those 3632 calls, 62,22% were false alarms, for which a few instructions could have been enough to provide help. Those situations can cause a great waste of resources and experts time, which could be used in a more efficient way or for some more urgent problems. To solve these kind of wastes, the perfect solution could come from the application of Unmanned Aerial Vehicles (UAVs), commonly known as drones. These vehicles have been used for a wide range of applications like, for example, rescue operations [5], goods delivery [3, 21], path planning [4], data collection from ground Internet of Things (IoT) sensors [12, 22], weather and field monitoring [6–8]. Therefore, using drones for sea rescue operations seems to be rather feasible.

In this paper, we investigate on the First Boat Rescue (FBR) problem, firstly introduced in [9], that aims on utilizing drones for rescue operations in open sea. A fleet of homogeneous drones is disposed along a coast and each drone is equipped with audio/video devices and medical sensors that can measure various data on the boats requesting help. A medical expert is remotely connected to the drones and can observe and evaluate the measurements taken on the boat crew. The drones' medical equipment can only provide a superficial picture of the actual status of the crew but being able to measure, for example, the heart rate level, temperature and oxygen level, and blood pressure, could still make a difference on understanding the type of treatment needed. Using such an approach, could avoid using rescue lifeboat for each call for help and could make possible to use those only when the medical help is really needed on board. Drones could be a great alternative to be able to verify the entity of the request and to intervene only with the essential resources. Since drones operate using a battery, they could finish their energy available during their flight. To solve this problem, energy sea buoys [23] are provided and are suitably disposed in the rescuing area to provide a wireless recharging station for the drones.

In FBR, each boat asking for help is required to be visited exactly once, i.e., by only one drone, and each drone can perform a multi-hop routing via the buoy grid system. Each drone starts its mission from a depot on the coast and always returns to a depot at the end of its tour. See Fig. 1 to have a general view of the operating environment.

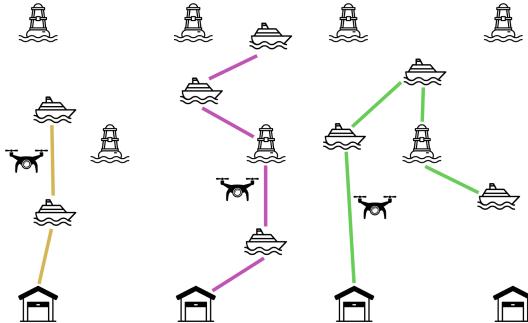


Fig. 1. An example representing a possible disposition of depots, buoys and boats requesting for help. Each drone starts its mission flying from a depot to the boats in its tour, possibly stopping on buoys to recharge. The same buoy can be visited multiple times by any drone, as long as it can provide energy.

1.1 Our Results

In this work, we study the FBR problem by addressing one of the open problems from [11]. In that work, the authors introduce and formally define the FBR problem and then they propose different solutions in two different scenarios: (i) drones with partial recharge at any buoy and, (ii) drones under Full Recharging Policy (FBRF).

For (i), they propose an Integer Linear Programming (ILP) to provide a solution to rescue the boats, followed by the algorithm for the Gas Station problem proposed in [15], to ensure that each drone returns to a depot.

For (ii), the authors propose an ILP formulation used together with the algorithm for the Gas Station problem from [15], in order to obtain an optimal solution to FBRF. The Uniform Cost Tour Gas Station (UGS) problem [15], is also used as a subroutine to provide a solution to FBRF with a provable approximation guarantee.

One of the main open problems in [11] that we want to investigate, is about the density and disposition of the buoys on the rescue area. In this work, we propose two different topologies for the buoy disposition: a triangular grid and, with respect to the one proposed in [11], a shifted rectangular grid. The number of buoys and their disposition is one of the main variables of the FBR problem, therefore it surely deserves main attention with the aim of reducing the resources used and speeding up the rescue operations.

1.2 Related Work

The FBR problem derives some of its most important features from the similarities with the Electric Vehicle Routing Problem (eVRP) [13, 17], which in turn extends the classical Vehicle Routing Problem (VRP) considering the more limited range of electric vehicles. Most of the models proposed for eVRP are

usually based on the assumption that the battery charging is linear and that the vehicles can partially or fully recharge their battery at each charging station. A work with some similarities to ours is [16] where the authors introduce the Electric Vehicle Routing Problem with Drones (EVRPD). This approach combines electric ground vehicles (EVs) and UAVs (drones) to deliver packages to customers: the drones are positioned on the EVs from which they start their flight to deliver the packages. Nonlinear battery charging is considered in [19], where the authors present a hybrid meta-heuristic which is tested on a set of realistic instances. Customers are served by an unlimited and homogeneous fleet of electric vehicles, able to recharge on stations that have no limitations on the number of vehicles that can recharge simultaneously. Each customer is visited exactly once and the vehicles start and end their route at a depot. The objective is to minimise the total route time. A different approach on the problem could reside in using different depots from where the vehicles can start their paths. This is done for example in [24] where the authors study a Multi-Depot Green Vehicle Routing problem. The objective is to minimise the total carbon emissions. To solve this problem the authors propose a two-stage Ant Colony System. In [18] the authors present a general eVRP with a formulation focused on the costs associated with time and electricity consumption. The proposed approach returns a solution with minimum cost in terms of travel time and energy. Our problem, and also eVRP, are closely related to the classical VRP [20] and to the Multi Depot Vehicle Routing (MVR) Problem [14], which is the generalization of the Travelling Salesman Problem. Those problems are known to be NP-hard and they aim to study how goods can be delivered from several depots, to customers, using a set of vehicles travelling on a road network. Another problem related to ours is the Steiner Travelling Salesman Problem introduced in [10]: each salesman is assigned to visit a set of nodes possibly passing through other nodes not required to be visited. The objective is to find minimum-weight closed walks. The most closely related problem to our work is the Uniform Cost Tour Gas Station (UGS) Problem studied in [15]: given a set of cities and a set of gas stations, the goal is to find the shortest tour that visits all the cities and, if it becomes necessary to purchase gas to avoid running out of it, gas stations. The authors also study the Gas Station Problem whose objective is to find the shortest possible path between two specific points. For UGS, they propose an approximation algorithm with a factor of $\frac{3(1+\alpha)}{2(1-\alpha)}$, for any $0 \leq \alpha < 1$, while for the Gas Station Problem they present an optimal algorithm.

1.3 Outline

In Sects. 2 and 3, we briefly introduce, respectively, the FBR and FBRF problems and offer a description of the algorithms proposed in [11]. In Sect. 4, we exploit the same instances from [11] and apply the algorithms with a new triangular grid and a shifted rectangular grid for the buoy disposition. We also offer a comparison between our new results and the ones from [11]. Finally, in Sect. 5 we provide concluding remarks and offer interesting directions for future works.

Remark 1. Detailed descriptions of the studied problems and the algorithms used can be found in [11] where those elements have been thoroughly explained. The following sections only provide a general overview of these topics.

2 Problem Formalisation

In this section, we provide the formalisation of the studied problem and a brief description of the algorithms proposed. In this work, differently from [11] where a rectangular grid **recta** was proposed, we consider two different fixed infrastructures of buoys: one is given by a triangular grid **triang** (see Fig. 1) and the other one is a shifted rectangular grid **s-recta**. Both grids are represented by a graph $G = (V, E)$ of n rows and m columns. The bottom row of the grids represents the base stations $D = \{d_1, \dots, d_m\}$, also called *depots*, located on the coast, from where the drones start and end their journeys. All the other vertices of the grid represent the recharging stations for the drones, i.e., the *buoys*. To denote the set of buoys we use $R = \{r_1, \dots, r_\kappa\}$, where $\kappa = (n - 1)m - 1$ for **triang**, and $\kappa = (n - 1)m - 2$ for **s-recta**.

The other elements in the system are the boats requesting for help and the fleet of homogeneous drones used for the rescue missions. The set of boats is denoted by $B = \{b_1, \dots, b_\ell\}$, the set of drones is denoted by $K = \{k_1, \dots, k_m\}$. Inside G , the operational area of the drones, boats can appear anywhere. The set of edges E is defined as follows. For any two vertices $i, j \in V$, with $i \neq j$, consider the edge (i, j) in E and let $t(i, j)$ be the time needed for a drone to fly from vertex i to vertex j . Moreover, for each boat $b \in B$, let s_b be the time needed for a drone to assist a boat b . We also use Q and P to denote, respectively, the drones battery capacity and the energy available at each buoy $r \in R$ before the rescue missions start.

Let us consider a drone $k \in K$ and its mission starting from a depot d_j , visiting one or more boats requesting for help, stopping at one or more buoys if needed to recharge its battery, and ending at a depot $d_{j'}$. Let this path/tour be $C_k = d_j \rightarrow b_{j+1} \rightarrow \dots \rightarrow b_l \rightarrow r_{l+1} \rightarrow b_{l+2} \rightarrow \dots \rightarrow b_p \rightarrow \dots \rightarrow d_{j'}$. Let b_p be the last boat in C_k served by the drone k . Before a mission starts, we assume that in each depot a drone with full battery is awaiting to be deployed in the rescue area. We also assume that multiple recharges are allowed at any buoy.

The total time spent by drone k in order to serve all boats in its flying tour C_k is

$$\tau(C_k) = \sum_{i=j}^p (t(v_i, w_{i+1}) + s_{v_i}),$$

where $t(v_p, w_{p+1}) = t(b_p, b_p) = 0$, $s_{v_{p+1}} = s_{b_p}$, and $v, w \in V = B \cup R \cup D$. Observe that $\tau(C_k)$ gives the time needed for a drone k in order to visit the last boat b_p in C_k , when the drone starts from depot d_j .

We now formally define First Boat Rescue problem (FBR) with triangular grid and shifted rectangular grid.

First Boat Rescue (FBR)

Input: A triangular grid `triang`, or a shifted rectangular grid `s-recta`, $G = (V; E; t : E \rightarrow \mathbb{R}^{\geq 0}; s : B \cup R \rightarrow \mathbb{R}^{\geq 0})$ and two integers Q and P .

Solution: For each drone, a trajectory with starting and finishing endpoints at a depot (not necessarily the same one) that is traversed by the drone without exhausting its battery.

Goal: Minimise the total time required to visit the last boat in the longest path/tour, among all obtained path/tour, and ensuring that all boats are served.

Note that, after all the boats are served, each depot will contain exactly one drone. Our investigation in this work aims to understand how the tours/paths executed by the drones change by using two topologies different from the one proposed in [11]. We propose a triangular grid and a shifted rectangular grid which have, respectively, one less buoy and two less buoys than the rectangular grid used in [11].

2.1 Optimal Solution to FBR

The proposed optimal solution to solve FBR is given by an Integer Linear Programming (ILP) model used together with the algorithm proposed to solve the Gas Station problem in [15]. We present here only the main elements and the objective function of this formulation called `ilp-partial`. A more in-depth view of all the constraints and their description is present in [11].

The elements of `ilp-partial` and its objective function are:

- R'_r : a set of dummy vertices to allow multiple visits to a buoy and another set of vertices to allow multiple charging at each buoy;
- g : a dummy destination vertex, necessary to ease the formalisation, where each drone ends its path;
- $A = B \cup R' \cup D$: a set of boats, dummy vertices and depots;
- x_{ij}^k : a binary variable to represent that drone k flies from $i \in A$ to node $j \in A$;
- Q_i^k : the battery load of drone k when it arrives at vertex $i \in A$;
- w_r^k : the amount of energy recharged by drone k at buoy r ;
- q_r : the amount of energy available at buoy $r \in R$;
- The objective function aims at minimising the length of the longest path among all obtained paths, i.e., minimising the time at which the last boat is served:

$$\min \max_{k \in K} \sum_{i \in A} \sum_{j \in A, i \neq j} x_{ij}^k (t(i, j) + s_j^k)$$

2.2 Heuristic HMVR for FBR

In this section, we describe a heuristic proposed for the FBR problem. First of all, an instance of the Multi-Depot Vehicle Routing (MVR) problem is solved. In practice, a preliminary solution for FBR by solving MVR with drones equipped with infinite battery is provided. Since MVR doesn't take into account the drones battery limitations, the solution obtained by solving MVR can be infeasible for FBR. To implement a realistic battery capacity of the drones, the code using genetic techniques provided in [1] is then used: the MVR solution previously obtained is processed in order to check whether the drones trajectories are actually feasible with limited battery or if they need to stop at some buoy to recharge. Backtracking methods are also applied when the drone trajectory cannot be traced, i.e., when some buoys around the current drone trajectory have already been assigned to another drone. This heuristic also optimises the time required for recharging activities on the buoys: the drones recharge on a buoy only the amount necessary to reach the next buoy/depot in its path.

3 Algorithms for FBRF

In this section, we study the FBR problem with full recharging policy, called FBRF, i.e., when a drone reaches a buoy it recharges until reaching its maximum capacity. The main reason to study this problem is that it allowed the authors of [11] to propose a provable approximation algorithm. Specifically, an algorithm with a guaranteed approximation ratio for FBRF can be provided, which represents an upper bound for any solution of FBR.

3.1 Optimal Approach to FBRF

The optimal solution for FBRF is an ILP algorithm, called `ilp-full`, which is quite similar to `ilp-partial`. However, the solutions obtained for FBRF never dominate those obtained for FBR. In fact, always recharging more than needed to reach the next buoy/depot leads the drones to waste time at the buoys.

3.2 Approximation Algorithm for FBRF

In this section, an approximation algorithm, called *Approximating using the UGS Problem* (`ApxGS`) is presented. The main aspect of this algorithm is that the results are bounded by a specific factor. In fact, `ApxGS` operates by exploiting the approximation algorithm for UGS introduced in [15] to obtain approximate shortest paths using a graph with weighted edges.

3.3 Heuristic heuGS for FBR

In this section, a heuristic for FBR called `heuGS` is presented. This algorithm exploits `ApxGS` by first executing it to obtain an approximate solution which

considers a full recharging policy. Then, each path obtained in the approximate solution is traversed to see if it is possible to remove unnecessary vertices (i.e., buoys) in the case of partial recharging policy. Finally, since the amount of energy used at each buoy is different from full to partial recharging policy, the recharged energy used by a drone on a buoy is updated with the amount necessary to reach the next buoy/depot on the tour. The rationale to propose also this heuristic is given by the fact that it can be used for experimental comparisons with heuristic HMVR even though it cannot guarantee an approximation ratio.

4 Experiments and Comparisons

In this section, we present the main result of this work which is an experimental study obtained by executing the algorithms to solve FBR and FBRF on the same instances of [11] by using the two new topologies infrastructures for the buoys, **triang** (Fig. 2) and **s-recta** (Fig. 3). We also present a comparison between the results obtained in the three topologies in Table 1.

4.1 Experimental Settings

The bottom row of each grid represents the starting and ending point of each drone, i.e., the depots on the coast. The drones start their flight with full battery capacity which can make them fly for 27 min at an average speed of 80 km/h. During their flight, the drones could need to recharge their battery on the buoys, represented on the grids by the two topmost lines or vertices. Each buoy is initially fully charged with twice the amount of the battery capacity of a drone, i.e., each buoy can offer energy to make a drone fly for 2×27 minutes. After a drone lands on a boat to offer medical help, it will not consume its battery used to fly but it will use an additional battery provided to each drone for this precise purpose. The medical examination on a boat lasts for 15 min on average. The proposed instances have, respectively, 5, 10, 15 and 20 boats asking for help in the operative area of $175 \text{ km} \times 40 \text{ km}$. The results are obtained by running each algorithm on 10 randomly generated instances of the same size and reporting the average value. The time required by a drone to accomplish its mission is given by the difference between the time at which it finishes serving the last boat on its flying path/tour, and the starting time. Even though we require the drones to return to their depots, the mission priority is to help the boats, hence for the results of the problems we do not consider the time needed by the drones to return to the depots.

4.2 Results

Table 1. Results obtained with the presented algorithms on instances of 5, 10, 15 and 20 boats using the three different buoys infrastructures **recta**, **triang** and **s-recta**.

Topology	Number of boats	Algorithms				
		ilp-full	ApxGS	heuGS	ilp-partial	HMVR
recta	5	55.8 ± 17.1	93.6 ± 26.1	85.8 ± 22.2	44.9 ± 6.5	109.9 ± 23.9
	10	72.3 ± 14.4	109.1 ± 17.6	102.3 ± 18.9	61.1 ± 9.1	136.4 ± 40.2
	15	90.2 ± 9.8	178.1 ± 71.9	151.7 ± 54.5	—	169.1 ± 57.9
	20	—	310.6 ± 141.3	251.4 ± 95.3	—	201.4 ± 44.2
triang	5	56.5 ± 17.9	90.2 ± 12.7	82.1 ± 11.5	46.1 ± 7.7	113.4 ± 28.7
	10	72.2 ± 14.0	119.9 ± 22.2	110.3 ± 18.1	59.8 ± 7.7	129.2 ± 40.3
	15	90.3 ± 11.0	173.0 ± 43.8	139.3 ± 25.8	—	173.4 ± 36.8
	20	—	262.3 ± 120.8	222.3 ± 92.3	—	201.4 ± 42.4
s-recta	5	56.5 ± 17.9	90.2 ± 12.7	81.6 ± 12.6	46.1 ± 7.7	118.3 ± 36.8
	10	72.2 ± 14	123.7 ± 29.1	116.2 ± 26.1	59.8 ± 7.7	135.3 ± 48.3
	15	90.3 ± 11.0	171.5 ± 44.1	147.2 ± 32.2	—	165.9 ± 46.8
	20	—	265.0 ± 139.9	226.7 ± 97.8	—	198.4 ± 39.8

In our results presented in Table 1 we can infer some interesting observations regarding the rescue time mission with respect of the number of buoys disposed and, hence, the amount of resources invested.

More in detail, in the instances with 5 and 10 boats, all the algorithms performed more or less with the same quality, with just few minutes differences between the three topologies. Figure 2 represents the trajectories obtained for an instance with 5 boats. For the instances of 15 boats, and in particular for algorithm **heuGS** in the **triang** grid, an improvement of 12 min with respect to the original **recta** grid is obtained. Even though this could not seem a great improvement, recall that we are dealing with possibly saving lives in emergency situations, hence even just a couple of minutes could make the difference. Meanwhile, algorithm **ilp-partial** was too much energy-demanding to be able to compute the results swiftly enough.

Finally, for the instances of 20 boats (see Fig. 3 for the trajectories obtained), again for execution time reasons **ilp-full** and **ilp-partial** could not be executed but great results can be observed for the other algorithms. Always with respect to results for the **recta** grid, algorithm **ApxGS** was 48 min faster in the **triang** grid and 45 min faster in the **s-recta** grid; algorithm **heuGS** was 29 min faster in **triang** and 25 min faster in **s-recta**; finally algorithm **HMVR** performed equally in all three grids with an improvement of 3 min for **triang**. Moreover, recall that all three topologies cover the same area but **triang** has one less buoy than **recta** and **s-recta** has two less buoys. This means that not only the rescue time improves or remains the same in the new two topologies, but also the

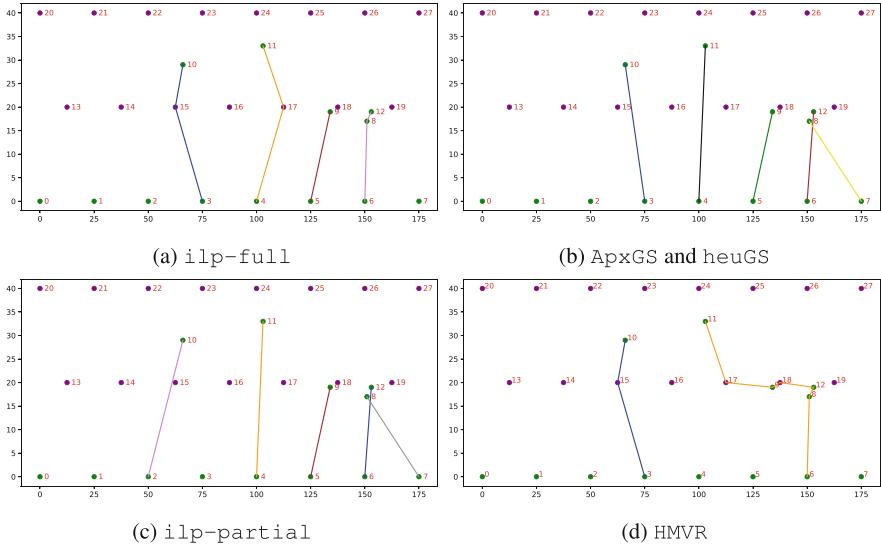


Fig. 2. Results obtained on a random instance with 5 boats using (a) **ilp-full**; (b) **ApxGS and heuGS**; (c) **ilp-partial**; (d) **HMVR**.

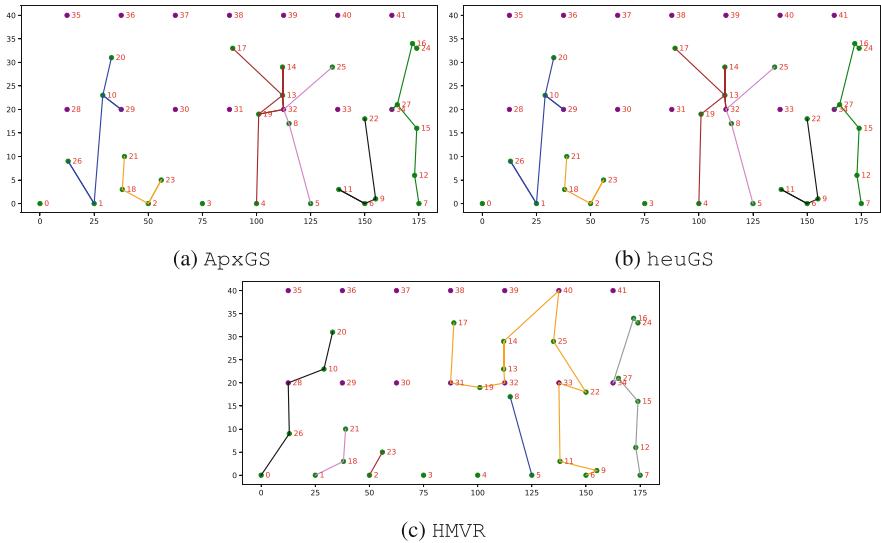


Fig. 3. Results obtained on a random instance with 20 boats using (a) **ApxGS**; (b) **heuGS**; (c) **HMVR**.

resources used are less, hence giving the opportunity to cover a greater area adding buoys or maybe using those “extra” resources to improve the existing buoys and drones technologies.

5 Conclusions and Future Works

In this paper we extended the work on the *First Boat Rescue* (FBR) problem, firstly approached in [9] and then thoroughly studied in [11]. This problem is a variant of the well-known *Electrical Vehicle Routing* Problem and offers a new interesting view on sea rescue missions by using a fleet of drones to offer first-aid medical support. Starting from the open problems of [11], we explored the problem implementing two new different grid topologies for the buoy infrastructure to observe whether an improvement on missions time or on the amount of resources used could be found. Thanks to the promising results of this paper, as future work, trying to find new different grid topologies to approach the problem still remains a good study direction, also considering in detail the data on the positions of previous rescue missions to be able to properly dispose the buoys. The FBR problem could also be studied with different settings, adding weather conditions such as wind or raging sea. Another approach could reside in an online version of the problem, where boats asking for help appear during the mission execution and the drones must compute new trajectories on-the-fly.

References

1. Hmvr implementation code. <https://github.com/brandhaug/multiple-depot-vehicle-routing-genetic-algorithm>
2. Il diporto nautico in italia - anno (2021). https://www.mit.gov.it/nfsmitsgov/files/media/pubblicazioni/2022-09/Diporto_Nautico_2021_WEB.con
3. Aurambout, J.P., Gkoumas, K., Ciuffo, B.: Last mile delivery by drones: an estimation of viable market potential and access to citizens across European cities. *Eur. Transp. Res.* **11**(1), 1–21 (2019)
4. Betti Sorbelli, F., Chatterjee, P., Corò, F., Ghobadi, S., Palazzetti, L., Pinotti, C.M.: A novel graph-based multi-layer framework for managing drone bvlos operations. *IEEE Trans. Netw. Serv. Manag.* (2024)
5. Calamoneri, T., Corò, F., Mancini, S.: A realistic model to support rescue operations after an earthquake via uavs. *IEEE Access* **10**, 6109–6125 (2022)
6. Chen, M., Liang, W., Das, S.K.: Data collection utility maximization in wireless sensor networks via efficient determination of UAV hovering locations. In: 19th IEEE International Conference on Pervasive Computing and Communications, PerCom 2021, pp. 1–10. IEEE (2021)
7. Chen, M., Liang, W., Li, J.: Energy-efficient data collection maximization for uav-assisted wireless sensor networks. In: IEEE Wireless Communications and Networking Conference, WCNC 2021, pp. 1–7. IEEE (2021)
8. Curry, J., Maslanik, J., Holland, G., Pinto, J.: Applications of aerosondes in the arctic. *Bull. Am. Meteor. Soc.* **85**(12), 1855–1861 (2004)
9. Teo, T.W., Choy, B.H.: in. In: Tan, O.S., Low, E.L., Tay, E.G., Yan, Y.K. (eds.) Singapore Math and Science Education Innovation. ETLPPSIP, vol. 1, pp. 43–59. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-1357-9_3
10. Gabrel, V., Mahjoub, A.R., Taktak, R., Uchoa, E.: The multiple steiner tsp with order constraints: complexity and optimization algorithms. *Soft. Comput.* **24**(23), 17957–17968 (2020)

11. Ghobadi, S., Mostarda, L., Navarra, A., Piselli, F.: Uavs missions for sea emergencies. Manuscript submitted for publication (2024)
12. Ghobadi, S., Pinotti, C.M.: Dispatching the minimum number of uavs in neighborhood iot networks. In: Georgiou, K., Kranakis, E. (eds.) *Algorithmics of Wireless Networks*, pp. 28–40. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-48882-5_3
13. Ghorbani, E., Alinaghian, M., Gharehpetian, G.B., Mohammadi, S., Perboli, G.: A survey on environmentally friendly vehicle routing problem and a proposal of its classification. *Sustainability* **12**(21), 9079 (2020)
14. Karakatic, S., Podgorelec, V.: A survey of genetic algorithms for solving multi depot vehicle routing problem. *Appl. Soft Comput.* **27**, 519–532 (2015)
15. Khuller, S., Malekian, A., Mestre, J.: To fill or not to fill: the gas station problem. *ACM Trans. Algor.* **7**(3), 36:1–36:16 (2011)
16. Kyriakakis, N.A., Stamadianos, T., Marinaki, M., Marinakis, Y.: The electric vehicle routing problem with drones: an energy minimization approach for aerial deliveries. *Cleaner Logist. Supply Chain* **4**, 100041 (2022)
17. Lalla-Ruiz, E., Erdelić, T., Carić, T.: A survey on the electric vehicle routing problem: variants and solution approaches. *J. Adv. Transport.* **2019**, 5075671 (2019). <https://doi.org/10.1155/2019/5075671>
18. Lin, J., Zhou, W., Wolfson, O.: Electric vehicle routing problem. *Transport. Res. Procedia* **12**, 508–521 (2016)
19. Montoya, A., Guéret, C., Mendoza, J.E., Villegas, J.G.: The electric vehicle routing problem with nonlinear charging function. *Transport. Res. Part B: Methodol.* **103**, 87–110 (2017)
20. Mor, A., Speranza, M.G.: Vehicle routing problems over time: a survey. *4OR* **18**(2), 129–149 (2020). <https://doi.org/10.1007/s10288-020-00433-2>
21. Sorbelli, F.B., Corò, F., Das, S.K., Palazzetti, L., Pinotti, C.M.: On the scheduling of conflictual deliveries in a last-mile delivery scenario with truck-carried drones. *Pervasive Mob. Comput.* **87**, 101700 (2022)
22. Sorbelli, F.B., Navarra, A., Palazzetti, L., Pinotti, C.M., Prencipe, G.: Wireless iot sensors data collection reward maximization by leveraging multiple energy-and storage-constrained uavs. *J. Comput. Syst. Sci.* **139**, 103475 (2024)
23. Viet, N.V., Wu, N., Wang, Q.: A review on energy harvesting from ocean waves by piezoelectric technology. *J. Model. Mech. Mater.* **1**(2) (2017)
24. Zhang, W., Gajpal, Y., Appadoo, S.S., Wei, Q.: Multi-depot green vehicle routing problem to minimize carbon emissions. *Sustainability* **12**(8), 3500 (2020)



Virtual Hazard Map for Disaster Prevention Education

Yumemi Fukushima and Tomoyuki Ishida^(✉)

Fukuoka Institute of Technology, Fukuoka 811-0295, Fukuoka, Japan
s20b1046@bene.fit.ac.jp, t-ishida@fit.ac.jp

Abstract. In this study, we developed a disaster prevention education support system termed “Virtual Hazard Map.” On the basis of three-dimensional city models, we realized a highly realistic virtual map that effectively combats virtual hazards. Using this map, a user can freely walk through the virtual hazardous space, where they encounter a character avatar using the keyboard and mouse. Furthermore, by implementing a quiz function, which is useful for disaster prevention and mitigation, we developed a virtual hazard map that incorporates a gamification element as well. In addition, we also implemented an user-initiated disaster prevention education element that utilizes this quiz function. To achieve this, we conducted an experiment on 30 subjects to evaluate the system response. As a result, several issues related to operability were clarified.

1 Introduction

In Japan, various natural disasters, such as typhoons, torrential rain, heavy snowfall, floods, landslides, earthquakes, tsunamis, and volcanic eruptions, occur frequently all across the country every year, causing severe damage to human lives. Although natural disasters cannot be avoided, it is important to prepare for such disasters in advance to respond to them if they occur. Thus, people must be fully aware of the vulnerabilities of their localities and prepare themselves for all types of natural disasters accordingly. One of the significant ways of preparing for such natural disasters is by adopting the “disaster prevention education” system, which is currently being implemented in most schools. According to a survey by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) on the implementation status of disaster prevention education in schools (2018) [1], 99.7% of schools provide adequate education and instructions on disaster safety to their students.

However, despite the ardent exercise of providing disaster safety guidance to students, several issues have been reported recently. The School Safety Comprehensive Support Project [2] launched by MEXT is facing the following four major challenges while promoting school safety:

- Significant plans and manuals in schools have not led to effective initiatives.
- School safety initiatives have led to several disputes among local communities, school founders, and school staff regarding the content of the manuals.

- Lack of nationwide promotion of the application of disaster prevention education, which may be useful in the future for combating any large-scale disaster like the 2011 Great East Japan Earthquake.
- Lack of collaboration with various local actors to influence children while promoting the safety measures.

2 Related Works

Sato et al. [3] developed and implemented teaching materials to infuse children with ability to create responsive actions during natural disaster. The teaching materials consisted of models of a three-dimensional map that the children could easily lay over their desks along with several information cards that narrated a brief history of recent natural disasters. This enabled the children to engage in experiential activities in which they choose to take actions and make decisions in the event of natural disasters by engaging in group activities.

Kikuchi et al. [4] developed an elaborate game with a verified learning effect to teach the mechanism of evacuation behavior which is essential for combating natural disasters. Through this game, players can touch objects placed on the map of a virtual facility and take the corresponding disaster prevention measures while evacuating and approaching their goals.

Shimbo et al. [5] developed an application called “Disaster Prevention Sugoroku” that can learn about evacuation routes available by taking into account traffic jams that may occur during a disaster. This application improves learning by motivating the players in groups to compete for more points. They are also inspired to take disaster prevention quizzes through which they can further learn to gain access to any evacuation center starting from a random point, each time by changing the routes. Thus, this application allows the users to experience simulated disaster in response to a disaster generated through VR.

Murakoshi et al. [6] examined the issue of assessing disaster risks on hazard and topographic maps, which is related to map literacy. The results revealed that university students without any deep understanding of the relationship between disasters and topography can easily assess potential disaster risks using these topographic maps.

Harada et al. [7] improved disaster prevention hazard maps using digital technologies such as debris flow prediction based on topographical data and metaverse and proposed effective operational methods. Harada [7] et al. revealed that the general public placed more importance on the “ease of viewing” and “ease of use” than on the “amount of information” provided on hazard maps.

3 Research Objective

In this study, we developed a disaster prevention education support tool named the “virtual hazard map for disaster prevention education.” By focusing on two key aspects: disaster prevention education based on knowledge and attitude, we aim to achieve practical disaster prevention education in preparation for large-scale disasters that are likely to occur in the future.

The proposed system has the following three significant characteristics:

- Reproduction of a realistic urban environment using a virtual hazard map based on three-dimensional city models.
- Implementation of behavioral learning theory under realistic conditions through walking in a virtual hazard map space with a character avatar.
- Gamification elements are realized by asking quiz questions related to disaster prevention and mitigation in the virtual hazard map space.

4 Virtual Hazard Map Configuration and Architecture

A schematic of the system configuration used in this study is presented in Fig. 1, and the system architecture is illustrated in Fig. 2. The system provides users (learners) with disaster prevention education through a gamification model through a walkthrough of a virtual hazard map based on the three-dimensional city model.

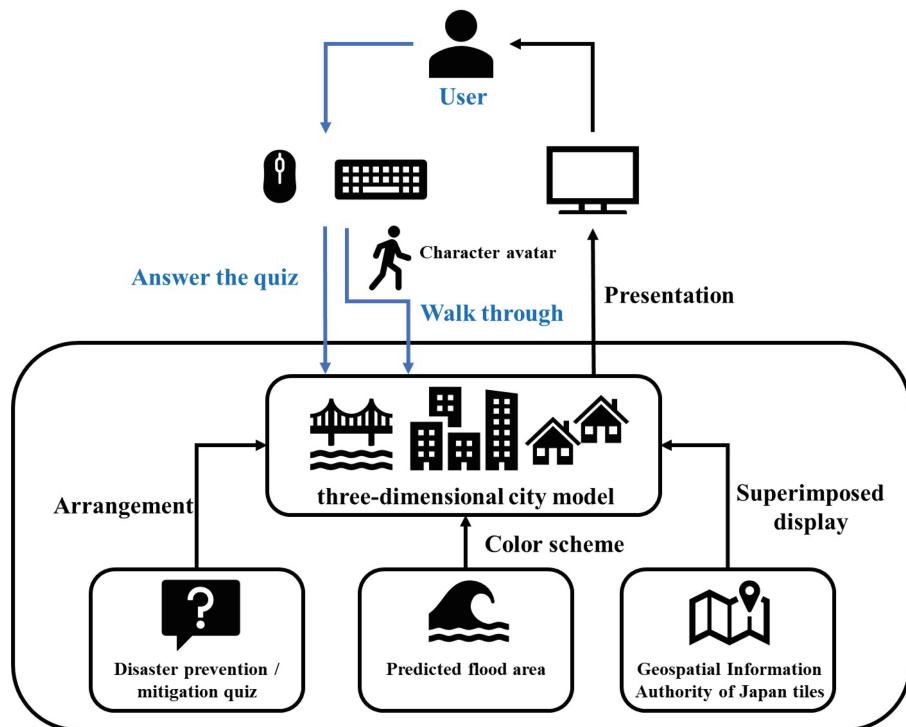


Fig. 1. Virtual hazard map configuration.

• Virtual map based on a three-dimensional city model

The hazard map is based on a three-dimensional city model. This is a virtual space where buildings are colored to correspond to the inundation rank of each potential flood area. The map comprises several disaster prevention/mitigation quiz items that are homogeneously laid throughout the map's virtual extent. The foundation of this

three-dimensional city model was created using PLATEAU [8], a three-dimensional city model development, utilization, and open data project led by the Ministry of Land, Infrastructure, Transport and Tourism of Japan. Moreover, to improve the authenticity of the hazards in the three-dimensional city model, we mapped the “Digital Basic Map of Japan (orthoimage) [9],” which represents the map data distributed by the Geospatial Information Authority of Japan (GSI) over the three-dimensional city model. Furthermore, the three-dimensional city model uses PLATEAU VIEW [10] as a reference and is extensively color-coded to indicate predicted flood areas. This allows users to experience any predicted flood area in three dimensions.

- **Walk through a function and character avatar**

We implemented a walkthrough function in the virtual hazard map space according to user operation of the character avatar. The user’s walkthrough is presented from the perspective of a third person. The user can walk forward, backward, left, and right by operating the cross key on the keyboard and can change the viewpoint by dragging the mouse.

- **Quiz function useful for disaster prevention and mitigation**

Users can answer quizzes at several locations within the three-dimensional city model while performing a walkthrough. The quiz features a question and option panels. After the user selects an answer option for a question, a correct/incorrect judgment follows.

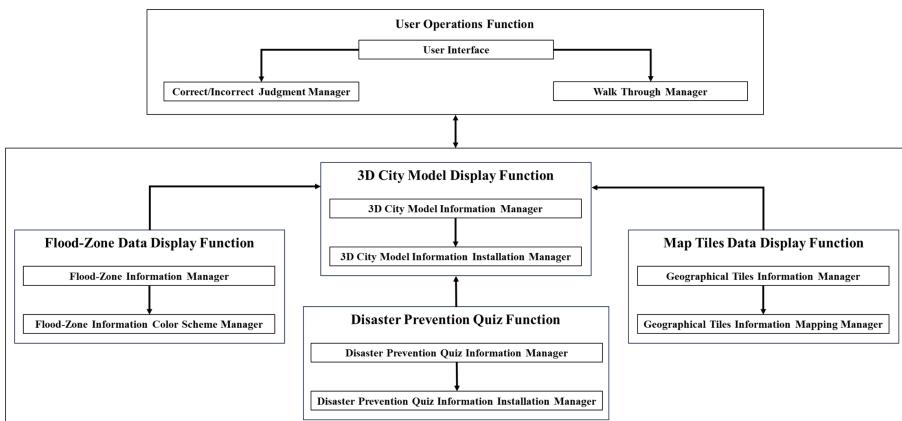


Fig. 2. Virtual hazard map architecture.

5 Virtual Hazard Map

To improve the reality of the virtual hazard map, in this study, we mapped the Geospatial Information Authority of Japan tiles onto a three-dimensional city model with specific color schemes for predicted flood areas (Fig. 3). Digital Japan Basic Map (orthoimage) [11] was used for the Geospatial Information Authority of Japan’s tiles. The orthoimage

provides image information that can be superimposed onto undistorted images using various geospatial information by converting aerial photographs and adding correct location information.

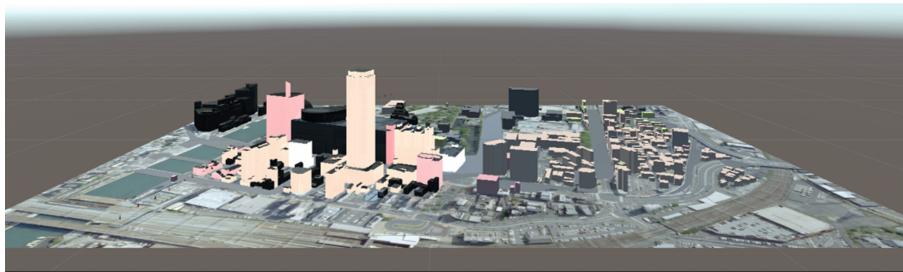


Fig. 3. Superimposed display of the geospatial information authority of Japan tiles in the unity scene.

Figure 4 shows the learning flow of users (learners) on the proposed system. When a user starts the system, the start screen appears on the display. As the user clicks the “START” button appearing on the screen, a virtual hazard map operation method screen is displayed, explaining how to operate the character avatar and answering quizzes within the map domain. After the user learns about the operation method and clicks the “Start” button, the virtual hazard map operation method screen changes to the virtual hazard map space. The user can then freely walk through the virtual hazard map space to check the predicted flood areas from any building. This allows users to adapt to adverse situations

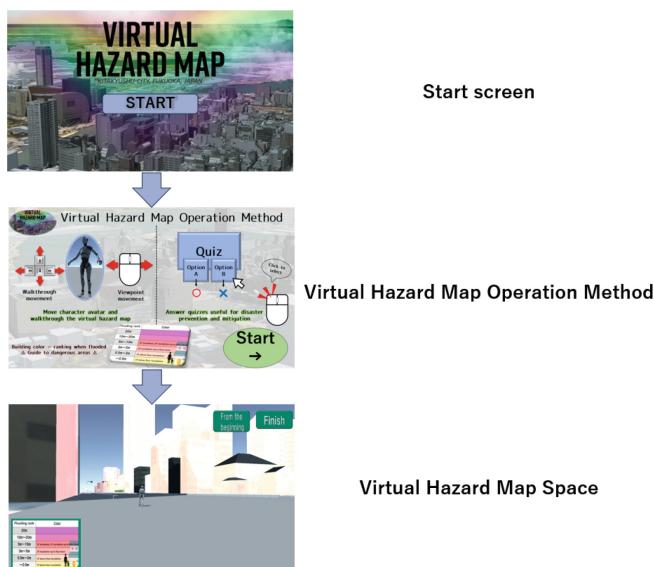


Fig. 4. User (learner) learning flow.

like natural disasters. Furthermore, users can gain knowledge about disaster prevention and mitigation by taking meaningful quizzes that are inserted into the three-dimensional city model space. The quiz has a two-choice question format, which the user answers by clicking on the panel that he or she thinks is correct between the two choices (Fig. 5). The table at the bottom left of the virtual hazard map space is a legend showing the flooding rank of the predicted flood area. To return to the starting point of the virtual hazard map space, the user must click the “From the beginning” button and then click the “Finish” button to leave the space.

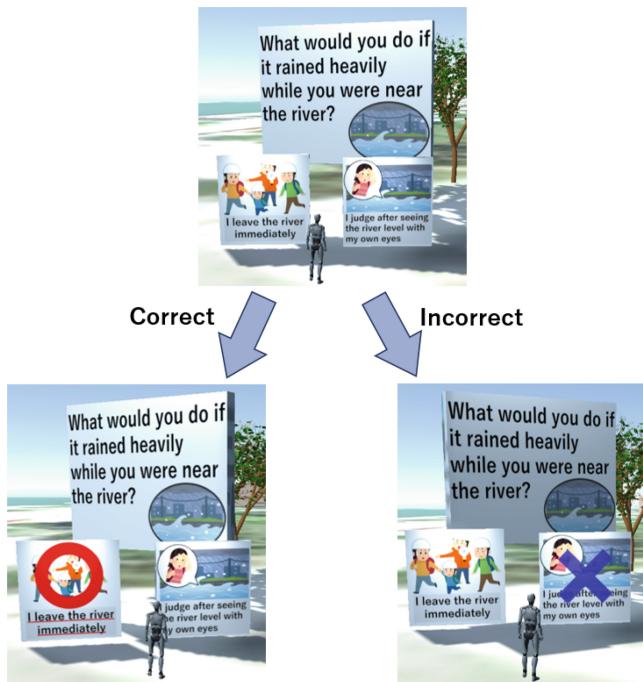


Fig. 5. Judging whether the quiz answers are correct or incorrect.

6 Virtual Hazard Map Evaluation

In this study, we assessed the operability, effectiveness, relevance, and applicability of a virtual hazard map by conducting an evaluation experiment on this system with 30 university students as our subjects.

Figure 6 shows the evaluation results of the operability of the virtual hazard map. It illustrates that more than 80% of the subjects answered in favor of “easy” or “somewhat easy,” whereas 16% answered in favor of “no opinion” or “somewhat difficult.” During the evaluation experiment, we noted that the subjects could not move their character avatars as much as they wanted. In this system, the user uses a mouse to shift the

perspective of the character avatar and to determine clicks for quizzes. Therefore, it is evident that some subjects may have faced operational difficulties. Figure 7 displays the evaluation results of the effectiveness of the virtual hazard map. More than 90% of the subjects answered in favor of “effective” or “somewhat effective.” Based on the evaluation results, we determined that the proposed system is effective for implementing disaster prevention education to prepare for future emergencies. Figure 8 shows the evaluation results of the relevance of the virtual hazard map. Of the subjects, 100% responded in favor of “relevant” or “somewhat relevant.” Based on these evaluation results, we can confirm that the proposed system is an effective tool for improving disaster prevention education. Figure 9 shows the applicability of the virtual hazard map. More than 90% of the subjects responded in favor of “possible” or “somewhat possible.” Based on this evaluation result, we also confirmed that the proposed system can be applied in areas other than disaster prevention and mitigation.

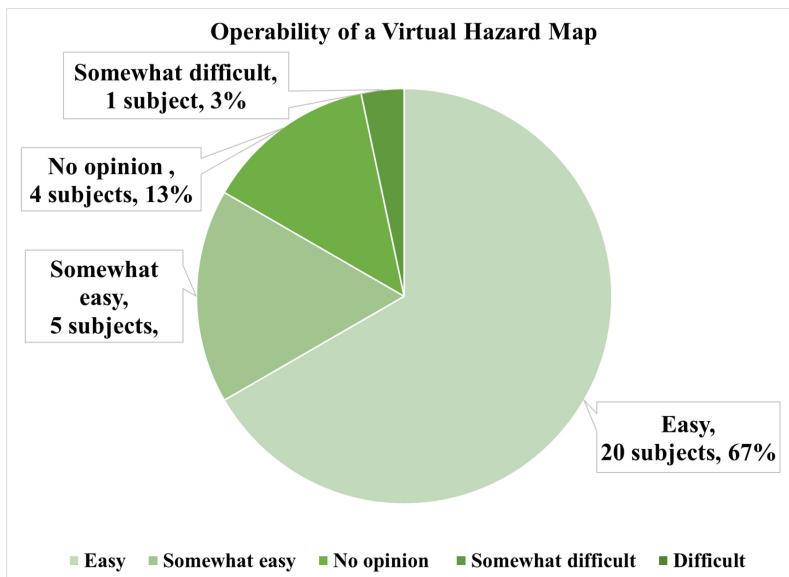


Fig. 6. Operability of the virtual hazard map ($n = 30$).

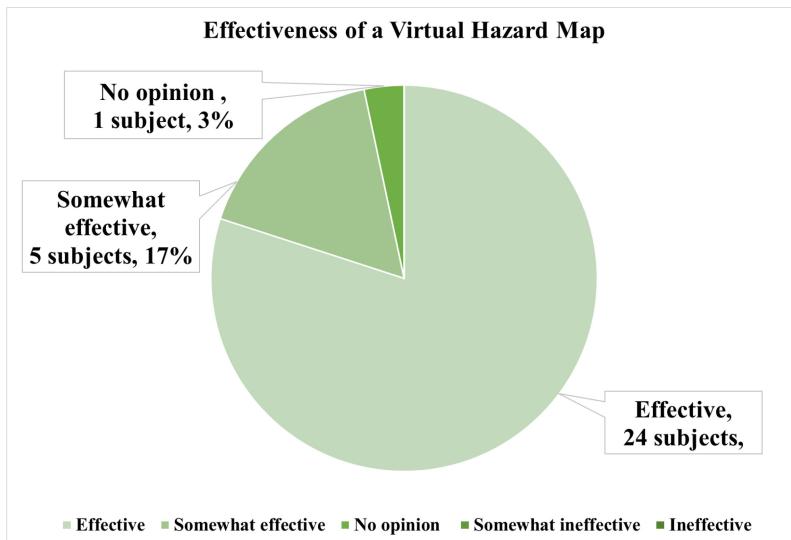


Fig. 7. Effectiveness of the virtual hazard map ($n = 30$).

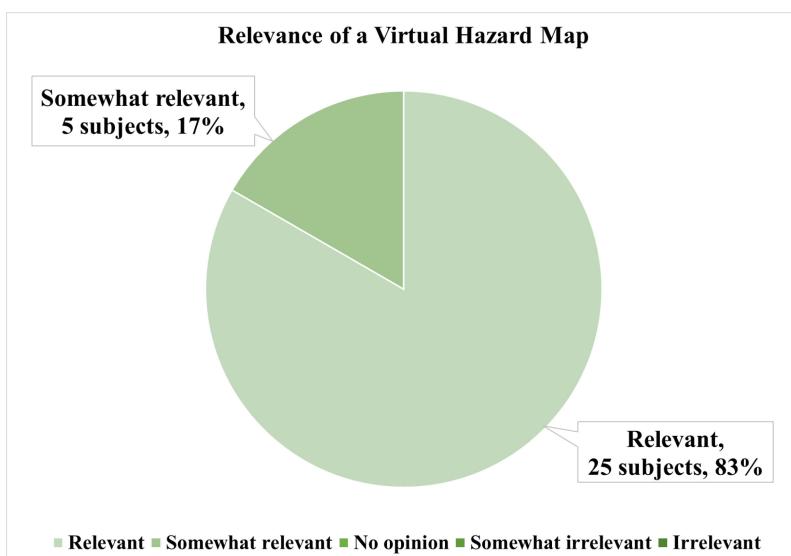


Fig. 8. Relevance of the virtual hazard map ($n = 30$).

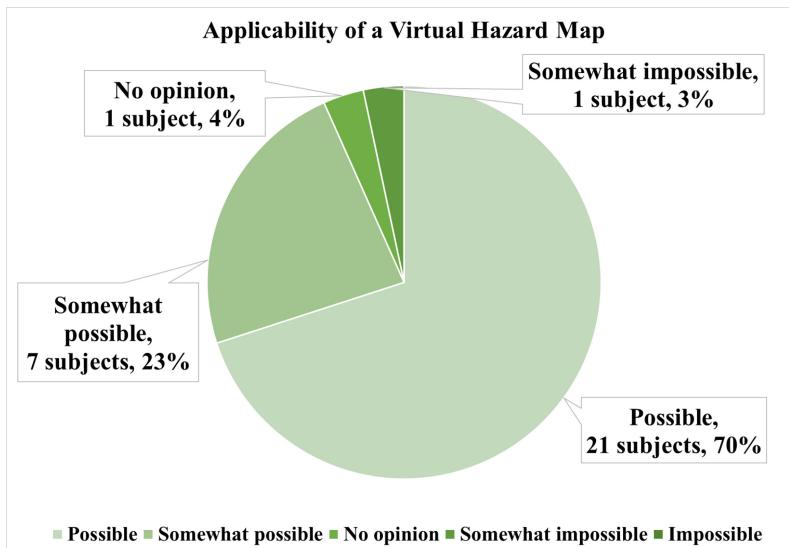


Fig. 9. Applicability of the virtual hazard map ($n = 30$).

7 Conclusion

This study describes the development and evaluation of the virtual hazard map. The virtual hazard map is constructed on the basis of a three-dimensional city model, through which a user can seamlessly walk through the map using a character avatar. Moreover, to help users know about dangers or flood areas, we colored the entire space of the three-dimensional building model according to its potential to cause floods. We also mapped the Geospatial Information Authority of Japan tiles onto the three-dimensional city model to improve its authenticity. Hence, it allows users to learn about dangerous areas during disasters by simply walking through a highly realistic three-dimensional city model. Furthermore, to incorporate a gamification element, we implemented a quiz function to educate users about disaster prevention and mitigation. In the evaluation experiment, we assessed the operability, effectiveness, relevance, and applicability of the proposed virtual hazard map. The results indicate that it is necessary to improve the operability of virtual hazard maps.

References

1. Ministry of Education, Culture, Sports, Science and Technology Japan, Survey on the status of initiatives related to plans for promoting school safety (actual results in FY2018) (2020). [https://www.mext.go.jp/content/20200330-mxt_kyosei02-000006167_3.pdf](https://www.mext.go.jp/content/20200330-mxt_kyousei02-000006167_3.pdf). Last viewed August 2024
2. Ministry of Education, Culture, Sports, Science and Technology Japan, FY2020 School Safety Comprehensive Support Project (2020). <https://anzenkyouiku.mext.go.jp/mextshiryou/index.html>. Last viewed August 2024

3. Sato, S., Fujioka, T.: A science teaching material for fostering the ability to select actions in a natural disaster: hazard assessment and appropriate evacuation for disasters caused by heavy rain. *J. Res. Sci. Educ.* **61**(2), 287–297 (2020). (in Japanese)
4. Kikuchi, S., Makanae, K.: Development and evaluation of learning effect of the serious game for evacuation training in facilities. *J. Jpn. Soc. Civ. Eng. Ser. F3 Civ. Eng. Inform.* **71**(2), 64–71 (2015). (in Japanese)
5. Shimbo, T., Terayama, K., Koshino, M., Okino, K., Araki, K., Yoshita, R.: Development of “Disaster Prevention Sugoroku,” an application for disaster prevention education using VR content, and its educational effects. *J. Jpn. Soc. Civ. Eng. Ser. H Eng. Educ. Pract.* **78**(1), 1–9 (2022). (in Japanese)
6. Murakoshi, S., Mitsushita, K., Koyama, M.: Is the risk of natural hazards properly understood from hazard maps? : An examination of map literacy. *Map J. Jpn. Cartograph. Assoc.* **58**(4), 1–16 (2020). (in Japanese)
7. Harada, N., Fujimoto, M., Satofuka, Y., Mizuyama, T., Matsui, T., Takei, C.: Advances in hazard mapping using the metaverse - the iHazard map project -. *Intell. Inform. Infrastruct.* **4**(2), 102–113 (2023). (in Japanese)
8. Ministry of Land, Infrastructure, Transport and Tourism Japan, PLATEAU | 3D city model development and open data project across Japan led by the Ministry of Land, Infrastructure, Transport and Tourism Japan (2024). <https://www.mlit.go.jp/plateau/>. Last viewed August 2024
9. Geospatial Information Authority of Japan, About Geospatial Information Authority of Japan Tiles (2024). <https://maps.gsi.go.jp/development/siyou.html>. Last viewed August 2024
10. Ministry of Land, Infrastructure, Transport and Tourism Japan, PLATEAU VIEW 2.0 (2024). <https://plateauview.mlit.go.jp/>. Last viewed August 2024
11. Geospatial Information Authority of Japan, Digital Japan Basic Map (Ortho Image) (2024). <https://www.gsi.go.jp/gazochosa/gazochosa40001.html>. Last viewed August 2024



Earthquake Virtual Reality Simulation System for Appropriate Evacuation Actions

Koichi Nishino and Tomoyuki Ishida^(✉)

Fukuoka Institute of Technology, Fukuoka 811-0295, Fukuoka, Japan
s20b1043@bene.fit.ac.jp, t-ishida@fit.ac.jp

Abstract. In this study, we developed a virtual reality (VR) earthquake evacuation training system that allows users to wear a head-mounted display and navigate freely within a virtual environment using a controller. Our proposed system provides users with a simulation of evacuation procedures during an earthquake. In the VR environment, users receive guidance about the recommended actions before, during, and after an earthquake and during the evacuation process. By selecting and performing these actions, users engage in evacuation drills that closely simulate real-life scenarios. The aim of our proposed system is to enhance disaster prevention awareness and evacuation skills. We evaluated the system with 30 participants to assess operability, functionality, effectiveness, and sense of presence, achieving high ratings across all aspects.

1 Introduction

The probability of a magnitude 8 or 9 class earthquake (Nankai Trough mega earthquake) occurring within the next 30 years is estimated to be between 70% and 80%. Such an event is expected to trigger a large tsunami exceeding 10 m, affecting a wide area along the Pacific coast, from the Kanto region to the Kyushu region [1].

Given these risks, precautionary measures must be taken to ensure a rapid response to earthquakes whenever they occur. Key actions include strengthening the earthquake resistance of buildings and infrastructure, securing furniture and home appliances, reviewing local hazard maps, ensuring clear evacuation routes and safe locations, developing action plans, and participating in disaster prevention drills. Therefore, regular evacuation drills are extremely important. However, effective evacuation drills require advanced preparation, including planning, space, and qualified instructors. Additionally, the impact of the new coronavirus infection, which began in 2019, has led to a decrease in the number of organizations conducting earthquake disaster drills, training sessions, and participant involvement [2].

Thus, we propose an earthquake evacuation training system to facilitate evacuation drills at any time and place, enhancing disaster preparedness and evacuation proficiency.

2 Related Works

Nakamoto et al. [3] emphasized the value of repeatedly educating people and developed a virtual reality (VR)-based disaster prevention education system in which users repeatedly experienced disasters and practiced response measures. Their system employed

three-dimensional VR images to depict disaster scenarios and simulate countermeasures. Subsequently, Nakamoto et al. [3] attempted to enhance users' disaster prevention awareness through repeated exposure to simulated disaster scenarios depicting various causes of damage.

Yamamoto et al. [4] sought to increase awareness of earthquake threats by developing an application that allows users to virtually experience earthquakes through VR goggles. The development of this application allowed users to simulate earthquake experiences anywhere, unlike previous experiences that were limited to specialized earthquake experience rooms.

Matsushita et al. [5] developed a monitoring system that simultaneously records floor responses and indoor conditions during earthquakes to effectively understand indoor conditions. Their research demonstrated that highly accurate simulations can be achieved by measuring the friction coefficient between furniture and flooring materials well in advance.

Kobayashi et al. [6] designed a VR system for earthquake and fire evacuation studies, aiming to understand evacuation behavior characteristics during earthquakes and fires, particularly exploring how distance influences evacuation decisions. Their study established experimental methodologies to study user evacuation behaviors under varying distances from fire outbreak points.

3 Earthquake VR Simulation System Configuration

This system consists of a user agent, a simulation space control function, and an earthquake evacuation training simulation environment. Figure 1 illustrates the system configuration.

- User agent

The user agent interacts with the earthquake evacuation training simulation environment using the head-mounted display (HMD). Users can select options on the screen and navigate through the simulation space using a controller attached to the HMD. The simulation space control function processes this operational information and updates the earthquake evacuation training simulation accordingly.

- Simulation space control function

The simulation space control function displays the earthquake evacuation training simulation environment to the user and integrates the user's operational inputs into the simulation. Information about the training simulation space is conveyed to the user via the HMD, while the user controls operations (such as walking and interacting with objects) using the controller. The simulation also presents the recommended actions for disaster scenarios, allowing users to select whether to perform the recommended actions during a disaster.

- Earthquake evacuation training simulation space

The earthquake evacuation training simulation space includes the start screen, room space, and evacuation route space leading to the evacuation site. From the start screen, the user selects the earthquake seismic intensity they wish to experience. This choice allows the user to simulate shaking corresponding to the selected seismic intensity within the room space. After the earthquake subsides, the user can simulate an evacuation site using a walk-through function.

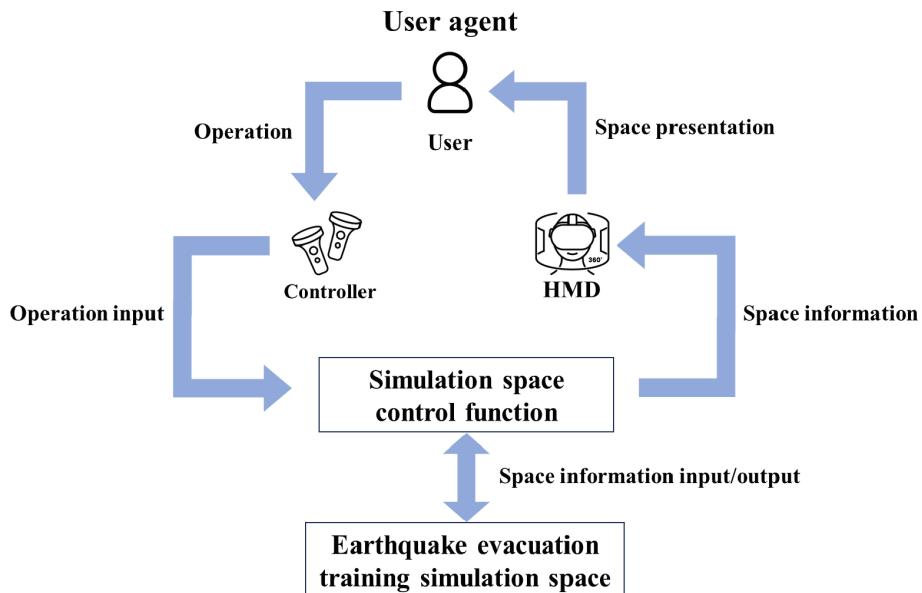


Fig. 1. Configuration of the earthquake virtual reality simulation system.

4 Earthquake VR Simulation System

Figure 2 displays the start screen, where users select the seismic intensity of the earthquake that they wish to experience: seismic intensity 4, seismic intensity 5, or seismic intensity 6. Users then experience shaking at the chosen seismic intensity.

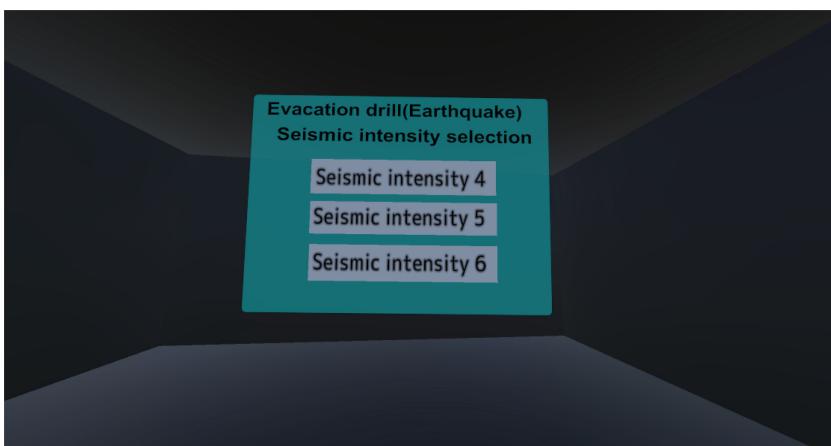


Fig. 2. Start screen of the earthquake virtual reality simulation system.

Table 1 lists the parameters (mass and static friction coefficient) of the objects installed within the room space. These object parameter settings remain constant across all seismic intensities in the system. Table 2 provides information on the elapsed time from entering the room space until the earthquake commences, along with examples of training actions. An earthquake early warning sounds 10 s after entering the room, followed by shaking lasting 15 s after the shaking subsides, users receive instructions to initiate evacuation action.

Table 1. Object parameter settings.

Object	Mass (kg)	Static friction coefficient
Chest (large)	100	0.8
Chest (small)	60	0.6
Desk (large)	40	0.55
Desk (small)	10	0.1
Sofa	70	0.6
Wallpaper	0.5	—
Television	20	0.1
Accessories	1	0.1

Table 2. Earthquake simulation time schedule.

Elapsed time (s)	Contents	Training behavior example
10	Earthquake early warning	Take appropriate response actions (e.g., hide under a desk)
20	Occurrence of shaking due to earthquake	Continue response actions
45	Presentation of evacuation action start guide	Start evacuation actions (evacuate outside, etc.)

After the earthquake early warning sounds, users are presented with recommended actions, as illustrated in Fig. 3, aimed at guiding them to protect themselves during the earthquake. When shaking occurs, objects like furniture in the room space topple over, as shown in Fig. 4. A seismic layer provided by Unity is positioned below the room objects to simulate earthquake shaking. This simulation involves adjusting gravitational forces and setting up collision detection among objects to replicate the shaking and falling of objects during an earthquake. The intensity of shaking for each seismic level in this study was varied by adjusting the shaking width and velocity of the seismic layer.

Once the earthquake shaking subsides, users are guided through recommended evacuation actions and transitions to an evacuation preparation scene, as shown in Fig. 5. In this scene, users receive instructions on actions to prevent secondary disasters within the

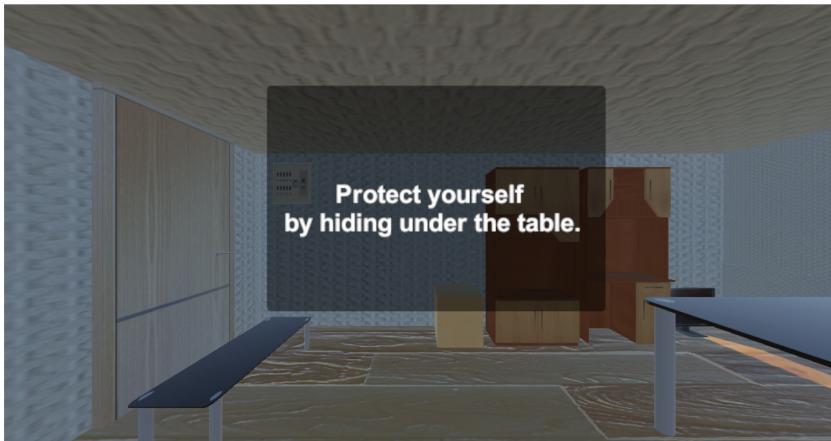


Fig. 3. Presentation of recommended actions after earthquake early warning.



Fig. 4. The state of the room space after the earthquake shaking subsides.

room before heading to the evacuation site. Users are given the choice to follow these recommended actions during a disaster. The following three recommended actions are provided to users, including turning off the circuit breaker (Fig. 6):

- Do you take evacuation goods?
- Do you turn off your circuit breaker?
- Do you turn off your stove?

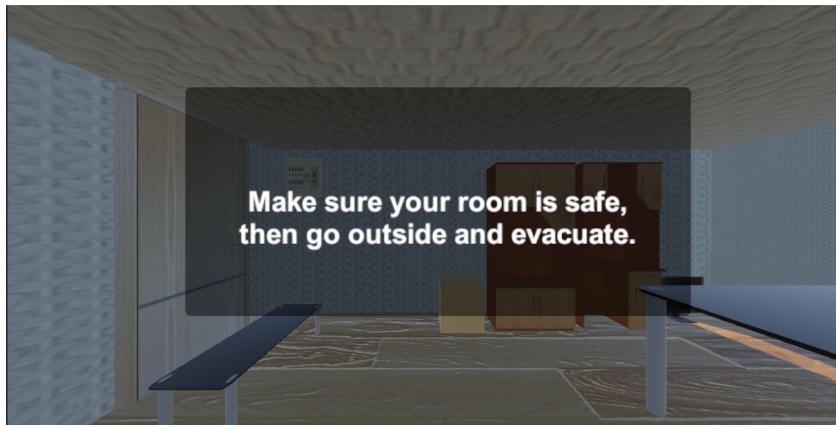


Fig. 5. Presentation of recommended actions after the earthquake subsides.

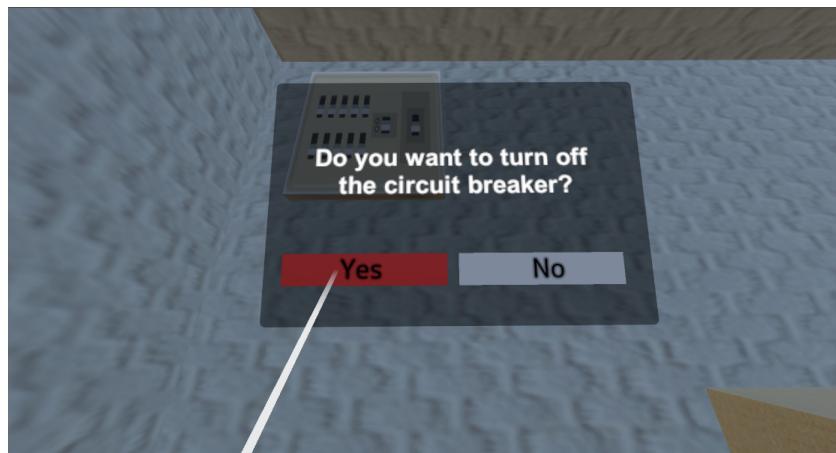


Fig. 6. Recommended action to turn off the circuit breaker on the selection screen.

In the evacuation route space leading to the evacuation site, users can simulate the evacuation process using a walk-through function toward designated refuge areas, such as parks. The path to the evacuation site is marked by signboard landmarks placed within the evacuation route space. During evacuation, a message window provides users with information on evacuation precautions, route selection, recommended actions, and other relevant details. Figure 7 illustrates how these evacuation precautions are presented to the user.



Fig. 7. Evacuation precaution.

5 Earthquake VR Simulation System Evaluation

To assess the operability, functionality, effectiveness, and sense of presence of the earthquake VR simulation system, we conducted an evaluation experiment involving 30 participants. During this experiment, participants experienced the earthquake VR simulation system firsthand and provided evaluations.

Figure 8 displays the operability evaluation results, showing that 73% of the participants rated it as “easy,” and 27% rated it as “somewhat easy,” confirming high operability of VR simulation system.

Figure 9 presents the functionality evaluation results of the VR simulation system, wherein 60% of the participants reported being “satisfied” and 33% reported being “somewhat satisfied,” confirming its high functionality.

Figure 10 displays the effectiveness evaluation results of the VR simulation system, with 73% of the participants finding it “effective” and 27% “somewhat effective,” indicating high effectiveness of this system.

Figure 11 displays the results of sense of presence evaluation in the VR simulation system. The evaluation indicated that 83% of the participants rated their sense of realism as “high,” while 17% rated it as “somewhat high,” confirming a strong user experience of presence using the system.

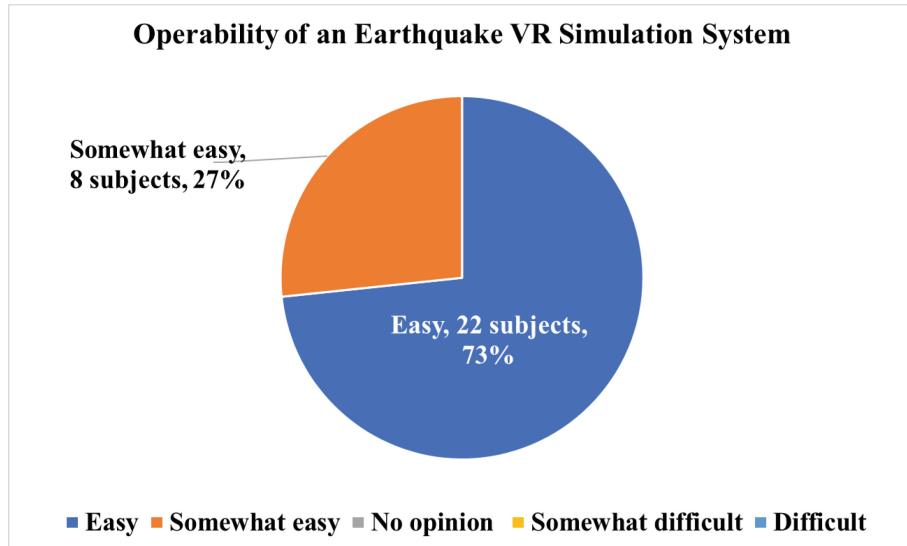


Fig. 8. Operability of the earthquake virtual reality (VR) simulation system ($n = 30$).

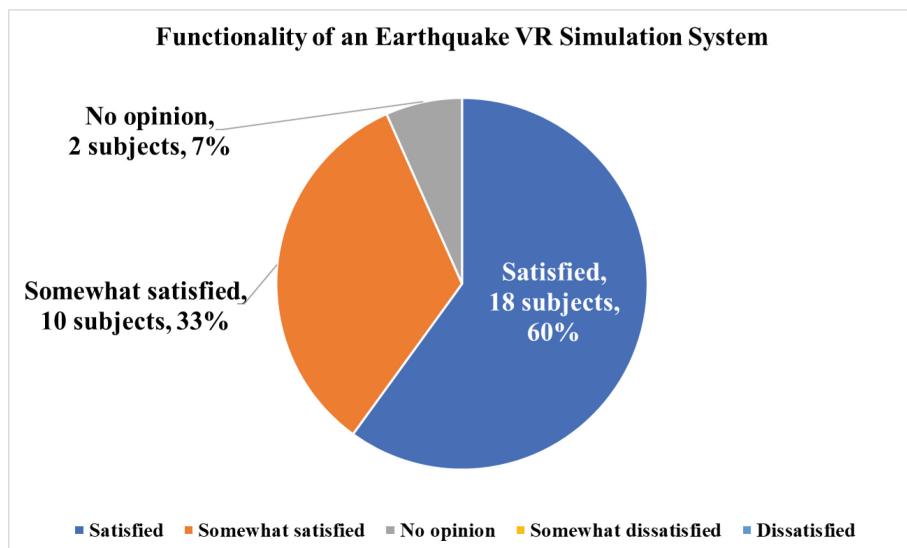


Fig. 9. Functionality of the earthquake virtual reality (VR) simulation system ($n = 30$).

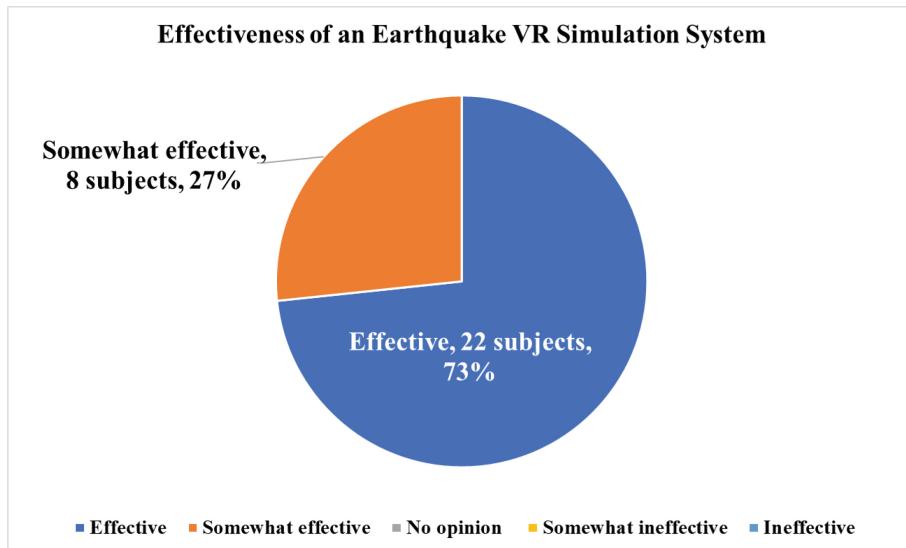


Fig. 10. Effectiveness of the earthquake virtual reality (VR) simulation system ($n = 30$).

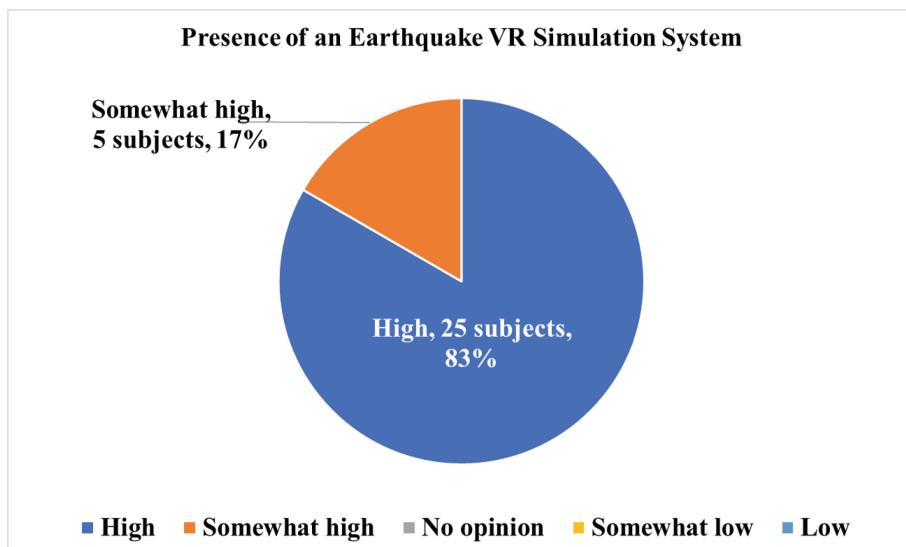


Fig. 11. Sense of presence of the earthquake virtual reality (VR) simulation system ($n = 30$).

6 Conclusion

In this study, we outlined the development and evaluation of an earthquake VR simulation system designed to promote effective evacuation behavior during disasters. The system features a user agent, a simulation space control function, and an earthquake

evacuation training environment, enabling users to engage in highly realistic evacuation training experience. Users can experience shaking corresponding to varying magnitudes of the seismic intensity and learn recommended actions following an earthquake, as well as evacuation procedures leading to designated sites. To access the operability, functionality, effectiveness, and sense of presence of this system, we conducted an evaluation experiment involving 30 participants. The results indicated high ratings across all evaluated aspects, confirming the system's effectiveness in enhancing disaster preparedness and evacuation skills.

References

1. Japan Meteorological Agency: About the Nankai Trough Earthquake (2023). <https://www.data.jma.go.jp/svd/eqev/data/ntrq/index.html>. Last viewed August 2024
2. Cabinet Office Japan: White Paper on Disaster Management 2022 (2022). https://www.bousai.go.jp/kaigirep/hakusho/r04/honbun/3b_6s_52_00.html. Last viewed August 2024
3. Nakamoto, S., Tanioka, R., Yoshino, T.: Proposal of disaster preparedness education system by repeated disaster experiences and countermeasures using virtual reality. In: Proceedings of the 2017 Information Processing Society of Japan Kansai Branch Conference, pp. 1–6 (2017). (in Japanese)
4. Yamamoto, T., Mizuno, S.: A self earthquake simulator on a smartphone. In: Proceedings of the Multimedia, Distributed, Cooperative, and Mobile Symposium 2017, pp. 592–597 (2017). (in Japanese)
5. Matsushita, T., Kurata, K., Tobita, J., Fukuwa, N., Yoshizawa, M., Nagae, T.: Development of monitoring system and simulation of indoor situation during earthquake based on shaking table experiment. AIJ J. Technol. Des. **19**(43), 871–874 (2013). (in Japanese)
6. Kobayashi, D., Kato, T., Kawahara, D., Shimura, T., Eda, T.: Evacuation behavior characteristic during post-earthquake urban fire spreading with use of virtual reality. J. Soc. Saf. Sci. **31**, 59–68 (2017). (in Japanese)



Mixed Reality-Based Japanese Calligraphy Learning System: Development and Evaluation

Riko Oohashi and Tomoyuki Ishida^(✉)

Fukuoka Institute of Technology, Fukuoka 811-0295, Fukuoka, Japan
s20b2009@bene.fit.ac.jp, t-ishida@fit.ac.jp

Abstract. In this study, we developed a Japanese calligraphy learning support system using mixed reality technology. The proposed system allows learners to practice eight kanji writing skills in an immersive environment. By using our proposed system, learners can concentrate on practicing their weaker kanji skills even in the absence of a calligraphy instructor. We conducted a questionnaire survey involving 30 participants to assess the operability, visibility, functionality, and effectiveness of the proposed system. The results revealed the need to improve the operability of the proposed system. In future work, we will aim to design a user-friendly interface suitable for novice users, particularly those using head-mounted displays.

1 Introduction

Calligraphy, an ancient Japanese tradition, involves the art of writing characters on paper using traditional calligraphy tools and ink. During the Edo period, temple schools became popular educational institutions that focused on “writing–practice” to improve calligraphy skills. In the Muromachi period, temples gathered children from common families and taught them, leading to the widespread use of the term calligraphy around this time [1].

However, the Agency for Cultural Affairs has highlighted several notable challenges: the aging population of calligraphy instructors and practitioners, the decline in the number of young participants, and decreased opportunities to practice calligraphy [1]. Furthermore, the advancement of information and communication technology in modern societies has reduced the significance of handwriting. Consequently, many people feel a sense of urgency about the potential loss of calligraphy culture and recognize the need to preserve it as an essential cultural heritage [2]. Over the past 30 years, the number of people practicing calligraphy has steadily declined, dropping from a peak of 7.5 million in 1992 to around 2 million by early 2020, as per a survey conducted between January and February 2020 [3].

There exist numerous techniques within the realm of calligraphy. The term “Eiji Happo” refers to the eight basic techniques embodied in the single character “ei.” By comprehensively understanding and diligently practicing these “Eiji Happo” techniques, learners can master brush movements [4]. To achieve progress and improvements in these techniques, learners must identify their weaknesses and areas for improvement. Suzuki

[5] emphasized the importance of receiving feedback from third parties to enable learners to quickly identify challenges and progress points. This feedback corrects incorrect practices and facilitates smooth learning processes.

Conversely, Japan's science and technology policy aims to create a human-centered society by achieving economic development and addressing social challenges through an integrated system that merges cyberspace (virtual space) with physical space (real space). This vision is encapsulated in the framework of "Society 5.0" [6]. One of the key concepts in this framework is digital transformation. Regarding educational innovation, it is important to stay abreast with trends in learning support technology, including augmented reality (AR), virtual reality (VR), and mixed reality (MR), in addition to technologies used in conventional learning support [7].

Therefore, in this study, we have developed a "Japanese calligraphy learning support system" using MR technology.

2 Related Works

Muranaka et al. [8] developed a calligraphy mastering support system that generates VR calligraphy videos using a frame image display computer. It also generates learning models using a system that integrates virtual and real images.

Fujitsuka et al. [9] developed an AR-based penmanship learning support system. This system allows learners to study brushstrokes by observing the teacher's brush movements displayed on a camera-equipped head-mounted display (HMD).

Okamura et al. [10] designed a system that displays calligraphy technical data on a computer screen, including calligraphy pressure, speed, and brush movement data.

Henmi et al. [11] developed a virtual calligraphy system that realistically conveys the sensation of a teacher's brush strokes, including the force and position information, in a VR environment.

Miyaji [12] constructed a model image and a work database using an image scanner, and developed a system that supports the practice of calligraphy evaluation methods, including brush strokes and character shapes.

3 Japanese Calligraphy Learning Support System Configuration

Figure 1 depicts the configuration of the proposed Japanese calligraphy learning support system. This system consists of three components: the learner, evaluation scene, and practice scene.

• Learner

The learner utilizes the Microsoft HoloLens 2 [13]. Upon starting the Japanese calligraphy learning support system, a menu screen is displayed. The learner views the explanation of Eiji Happo from the menu and proceeds to the writing screen upon pressing any button. On this screen, learner performs calligraphy using real calligraphy tools. After completing the calligraphy, the learner presses the "Evaluation" button, transitioning to the evaluation scene. Here, the learner recognizes the AR marker,

which causes the evaluation character to be displayed superimposed on the character the learner actually wrote. Additionally, pressing the “Practice” button within the evaluation scene transitions to the practice scene, where the learner can practice weak techniques. This process enables the learner to self-evaluate their calligraphy in the evaluation scene and then repeatedly practice weak techniques in the practice scene.

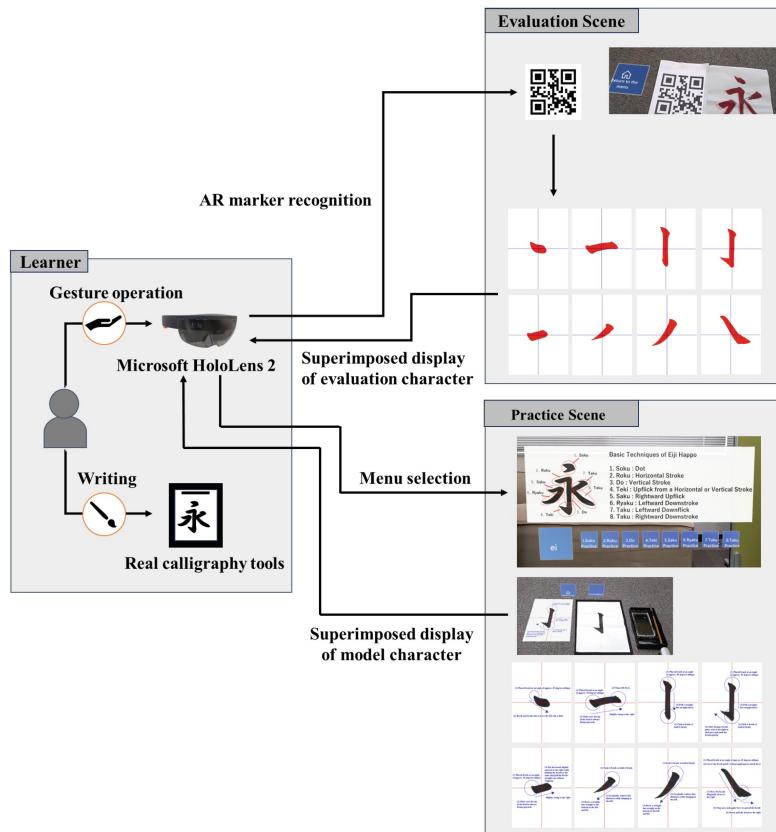


Fig. 1. Configuration of the Japanese calligraphy learning support system.

• Evaluation scene

After performing calligraphy on the writing screen, the learner presses the “Evaluation” button to transition to the evaluation scene. Here, the AR marker used to superimpose the evaluation character is recognized. Once Microsoft HoloLens 2 has identified the AR marker, the evaluation character is displayed superimposed on the

learner's written character. Simultaneously, the "Practice" button appears, allowing the learner to transition to the practice scene to practice weak techniques.

- **Practice scene**

In the evaluation scene, upon selecting the "Practice" button, the scene transitions to the practice scene. This scene includes basic technique information on Eiji Happo, a "Find weak techniques" button, and eight "Select weak techniques" buttons. The basic techniques provide a detailed description of the eight techniques. When the learner presses the "Find weak techniques" button, a model character is displayed superimposed. The learner then performs the character onto actual paper using a calligraphy brush while observing the model character, allowing for comparison between the learner's character and the model character. Furthermore, selecting any of the eight "Select weak techniques" buttons displays a model character containing writing instructions and advice for the chosen technique, superimposed on the screen.

4 Japanese Calligraphy Learning Support System

- **Menu screen**

After launching the application on HoloLens 2, the menu screen appears on the display, as shown in Fig. 2. The menu screen includes explanations of the fundamental techniques of Eiji Happo, a button to find weak techniques, and another to select weak techniques.

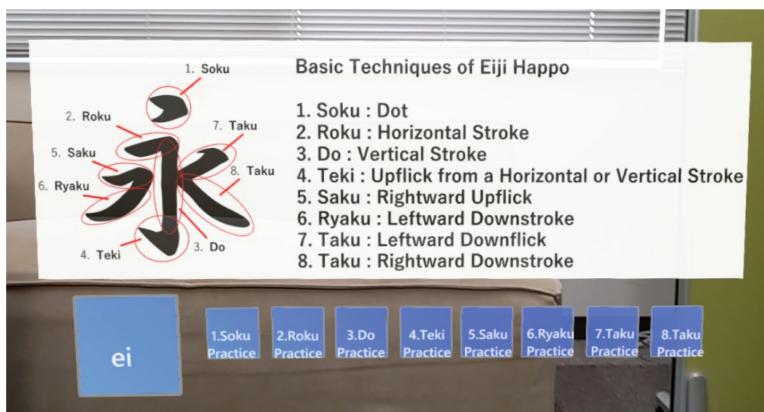


Fig. 2. Menu screen of the Japanese calligraphy learning support system.

- **Find weak techniques**

When the learner presses the "Find weak techniques" button, the screen transitions to the writing screen where kanji calligraphy is performed, as depicted in Fig. 3.

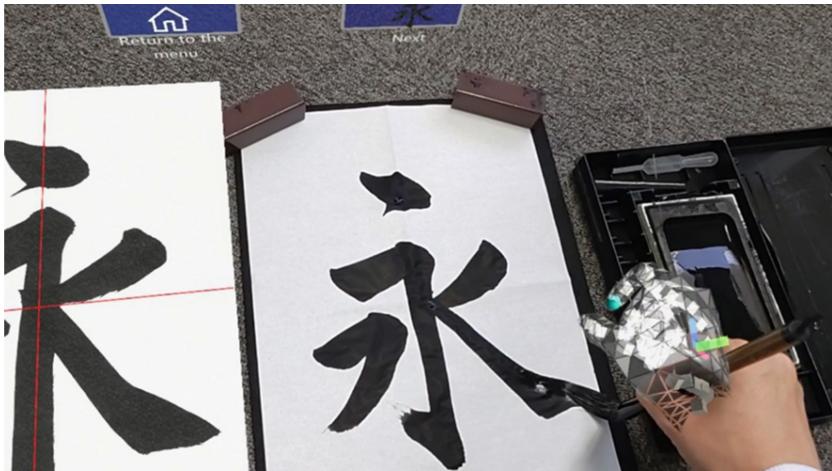


Fig. 3. Scene showing Kanji calligraphy.

After completing the kanji, clicking the “Next” button shifts the screen to the evaluation scene. Here, the registered AR marker is positioned to the left of the paper. Upon recognition by HoloLens 2, the evaluation character appears on the display, as depicted in Fig. 4. The learner can identify their own weak techniques by comparing their character with the evaluation character. After identifying the weak techniques, the learner can return to the menu screen and practice only those weak techniques.



Fig. 4. Scene showing Kanji evaluation.

- Practice for weak techniques

In the practice scene, when the learner selects a weak technique item from the menu screen, the interface transitions to a dedicated writing screen for that specific technique, as depicted in Fig. 5. The learner practices the weak technique while observing the model character and advice.



Fig. 5. Scene showing weak technique calligraphy.

Upon completing the practice, clicking the “Next” button shifts the scene to the evaluation scene. Here, the registered AR marker is positioned to the left of the paper. Once recognized by the HoloLens 2, the evaluation character is superimposed on HoloLens 2 display, as shown in Fig. 6.



Fig. 6. Scene showing weak technique evaluation.

5 Japanese Calligraphy Learning Support System Evaluation

In this study, we conducted a questionnaire survey involving 30 participants to evaluate the operability, visibility, functionality, and effectiveness of the Japanese calligraphy learning support system.

The evaluation result for the operability of the Japanese calligraphy learning support system is shown in Fig. 7. The survey revealed that 77% of the participants found the system “easy” or “somewhat easy” to operate, while 23% expressed “no opinion” or found it “somewhat difficult.” Notably, those who indicated “no opinion” or “somewhat difficult” were individuals who had experienced an HMD for the first time. Consequently, their unfamiliarity with HMD may have hindered their ability to gauge distances accurately when operating the menus, thereby impacting their assessment of the system’s operability.

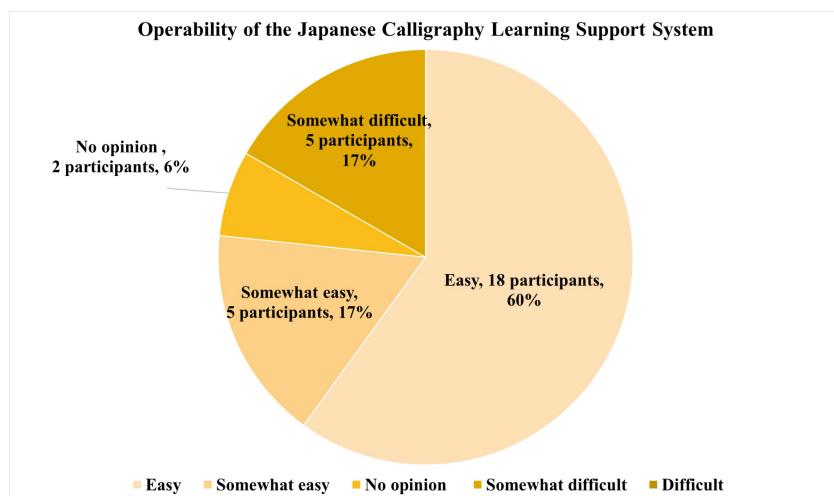


Fig. 7. Operability of the Japanese calligraphy learning support system ($n = 30$).

The evaluation result for the visibility of the Japanese calligraphy learning support system is presented in Fig. 8. According to the data, 93% of the participants rated the visibility as “high” or “somewhat high,” confirming its high visibility. However, 7% of the participants rated it as “somewhat low.” This lower rating can be attributed to misalignment in the superimposed display of the model character and the time it took for the evaluation character to be displayed in a stable position.

The evaluation results for the functionality of the Japanese calligraphy learning support system is depicted in Fig. 9. According to the data, 87% of the participants indicated that they were “satisfied” or “somewhat satisfied,” confirming the system’s high functionality. However, 13% of the participants expressed “no opinion” or were “somewhat dissatisfied.” Those who provided these responses lacked prior experience with calligraphy, suggesting that beginners found it difficult to accurately understand the model advice.

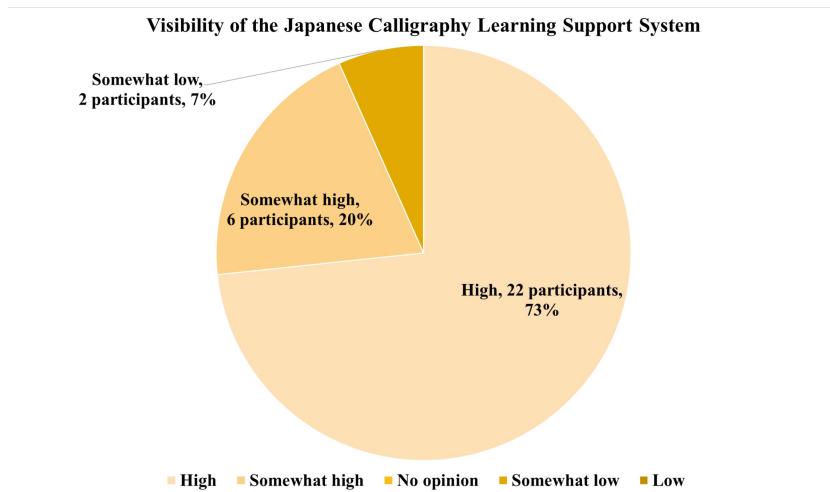


Fig. 8. Visibility of the Japanese calligraphy learning support system ($n = 30$).

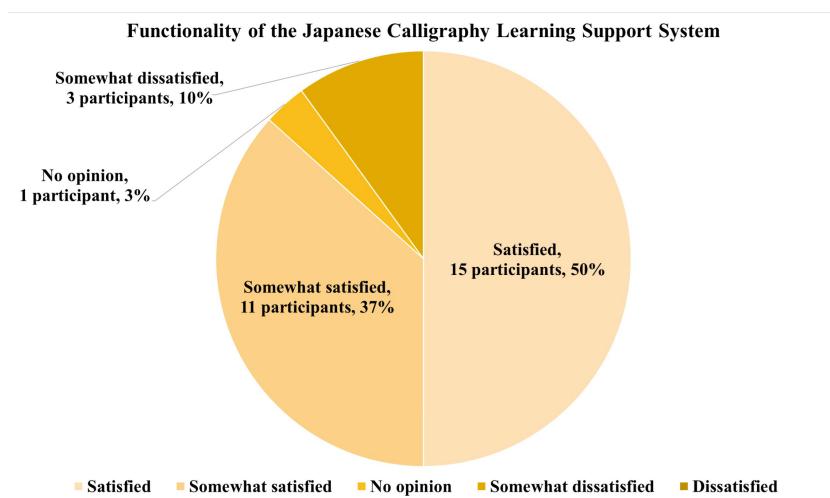


Fig. 9. Functionality of the Japanese calligraphy learning support system ($n = 30$).

The evaluation result for the effectiveness of the Japanese calligraphy learning support system is depicted in Fig. 10. According to the survey, 90% of the participants rated the system as “effective” or “somewhat effective,” thereby confirming its high effectiveness.

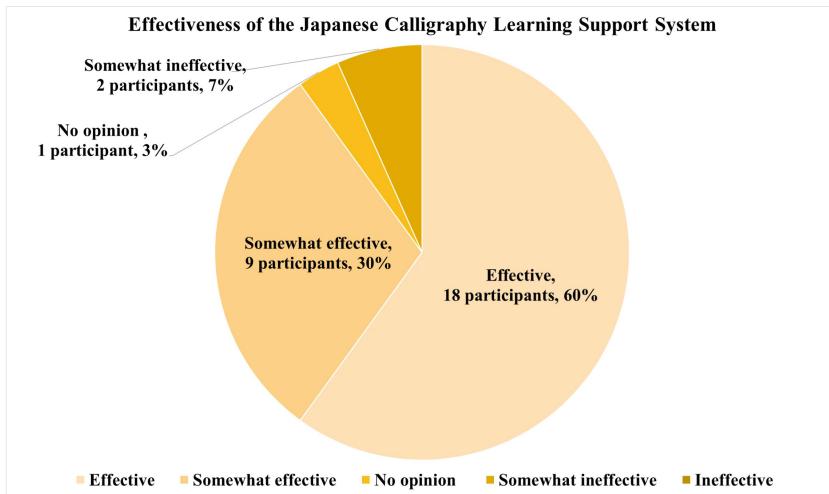


Fig. 10. Effectiveness of the Japanese calligraphy learning support system ($n = 30$).

6 Conclusion

In this study, we developed and evaluated the “Japanese calligraphy learning support system” that utilizes MR technology. This system comprises the calligraphy learner, evaluation scene, and practice scene. In the practice scene, the learner can practice calligraphy by referring to a model character superimposed in real space. Additionally, in the evaluation scene, the learner can identify weak techniques by superimposing the evaluation character onto the character they wrote. We conducted a questionnaire survey involving 30 participants to evaluate the operability, visibility, functionality, and effectiveness of the system. While the results indicated high ratings across many items, there is still room for improvement in terms of operability.

References

1. Headquarters for Vitalizing Regional Cultures, Agency for Cultural Affairs, 2020 Life Culture Survey Research Project (Calligraphy) Report (2020). https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/seikatsubunka_chosa/pdf/93014801_05.pdf. Last viewed August 2024
2. Japanese Calligraphy UNESCO Registration Promotion Council, Basic research report on calligraphy culture (2019). <http://www.shosan.jp/images/152.pdf>. Last viewed August 2024
3. Japan Productivity center, Leisure White Paper 2020 (2020). <https://www.jpc-net.jp/research/detail/004580.html>. Last viewed August 2024
4. Shoun Gakuen, Basics of calligraphy techniques. What is Eiji Happou? (2022). <https://shoun-e-nippon.co.jp/blog/7>. Last viewed August 2024
5. Suzuki, M.: Effects of a rubric: values of a test, motivation for learning, and learning strategies. *Jpn. J. Educ. Young Child.* **59**(2), 131–143 (2011). (in Japanese)
6. Cabinet Office, Government of Japan, Society 5.0 (2024). https://www8.cao.go.jp/cstp/english/society5_0/index.html. Last viewed August 2024

7. Takaoka, R., Mitsuhashi, H., Setozaki, N., Funaoi: Trends and prospects for technologies that enable digital transformation in primary and secondary education. *Jpn. J. Educ. Technol.* **45**(3), 283–294 (2021). (in Japanese)
8. Muranaka, N., Yamamoto, T., Imanishi, S.: A calligraphy mastering support system using virtual reality technology and its learning effects. *Trans. Inst. Electr. Eng. Jpn. A Publ. Fundam. Mater. Soc.* **123**(12), 1206–1216 (2003). (in Japanese)
9. Fujitsuka, T., Iwakura, J., Yamashita, S., Arai, H.: A penmanship learning support system using augmented reality. In: Proceedings of the 2014 IEICE General Conference, p. 163 (2014). (in Japanese)
10. Okamura, Y., Nagasaki, N., Nakamura, M.: Calligraphy teaching materials which added skill information and its effect. *J. JSEI* **19**(2), 3–8 (2003). (in Japanese)
11. Hemmi, K., Yoshikawa, T.: Virtual lesson and its application to virtual calligraphy system. *Trans. Virtual Real. Soc. Jpn.* **3**(1), 13–19 (1998). (in Japanese)
12. Miyaji, I.: The support system for learning the evaluation method of brush-written works using the analytic hierarchy process. *Bull. Inf. Process. Center Okayama Univ. Sci.* **17**, 1–11 (1996). (in Japanese)
13. Microsoft, Microsoft HoloLens Mixed Reality Technology for Business (2024). <https://www.microsoft.com/en-us/hololens>. Last viewed August 2024



Finding Representative Frames from Surveillance Video for Visualizing Viewer Behavior

Kaoru Sugita^(✉)

Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
sugita@fit.ac.jp

Abstract. In today's educational institutions, PCs and tablets have been introduced, and online lectures and e-learning platforms are being utilized. In these learning environments, live video streaming and recorded videos replace face-to-face classes, while exercises and exams are conducted online. However, some users (students) do other works and do not watch the playing video. In this paper, we introduce a method for finding representative frames of students' behaviors from recorded video, aiming to clearly visualize their actions during online learning.

1 Introduction

In today's educational institutions, PCs and tablets have been introduced, and online lectures and e-learning platforms are being utilized. In these learning environments, live video streaming and recorded videos replace face-to-face classes, while exercises and exams are conducted online. However, some users (students) do other works without watching the playing video. Additionally, e-learning systems often face the well-known issue of low completion rates. When using online lectures and e-learning platforms, these problems make it difficult to observe the actual learning activities. Therefore, it is necessary to monitor the learning activity during online education.

There are many studies on e-Learning systems. These studies have primarily focused on platforms designed to replicate the characteristics of traditional education in an electronic system [1]. In [2], it is presented the significance of e-learning systems. While in [3], a tracking system is designed to analyze real-time information about attentiveness of a student during an ongoing e-learning class using a simple web camera. In this study, attentiveness has been classified into three states: attentive, sleepy and disappeared. Also, an automatic student modelling approach has been introduced to identify learning styles in learning management systems [4]. This approach uses data about students' behavior during learning to gather insights about their learning styles and to provide courses and materials that fit each style.

In our previous works, we have developed and evaluated a monitoring system for measuring the gaze time during viewing a display. In these works, we found that it was difficult to determine whether a learner was taking notes or performing other tasks (i.e. lazy behavior) during viewing the content [5]. Therefore, we have introduced video

frame segmentation using a video frame correlation matrix for learner monitoring [6] and evaluated its performance [7]. Using this approach, we were able to obtain video scenes that accurately reflected student behavior, enabling us to distinguish between serious learning activities and idling. However, this approach requires significant time to review each video scene and monitor students.

In this paper, we introduce a method for finding representative frames of students' behaviors from the recorded video, aiming to clearly visualize their actions during online learning.

This paper is organized as follows: In Sect. 2, we introduce the method for visualizing viewer behaviors from surveillance video. Section 3 presents the implemented method. In Sect. 4 are illustrated some example results. Finally, conclusions and future work are given in Sect. 5.

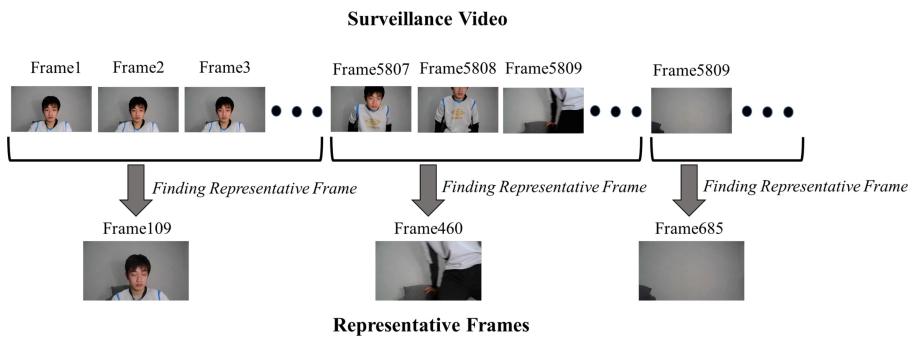


Fig. 1. Finding representative frames from surveillance video.

2 Visualizing Viewer Behaviors from Surveillance Video

In order to enable a quick overview of student behaviors during online learning, we introduce a method for finding representative frames from a video recording of a student as shown in Fig. 1. This method consists of two processes: video segmentation and representative frame extraction.

In the video scene segmentation, the video is segmented into scenes based on correlation values [6, 7] as shown in Fig. 2. The correlation values are calculated for all frames recorded in the video and put on a video frame correlation matrix. The correlation value increases for smaller differences between frames and decreases in case of larger differences. These values can be used as a criterion for segmenting the video.

The representative frame extraction involves finding a representative frame from each segmented scene as shown in Fig. 3. In this study, the representative frame is selected based on a median correlation value between frames in the scene.

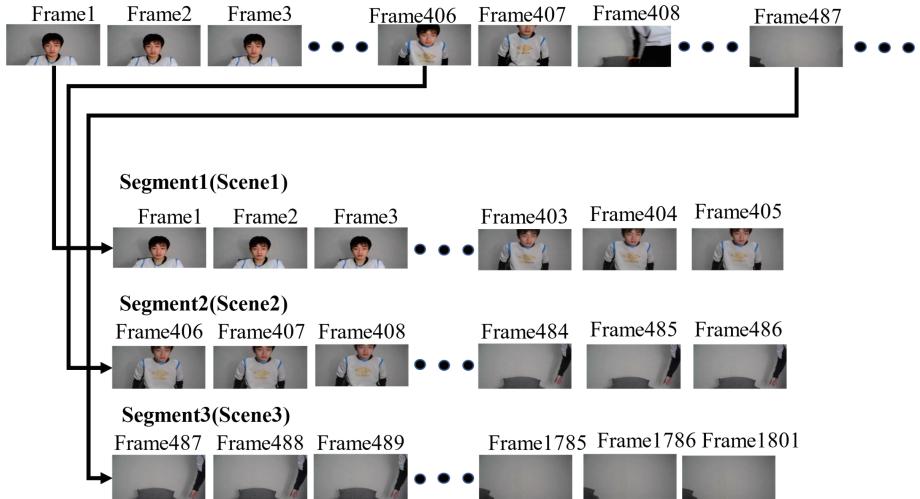


Fig. 2. Video scene segmentation.

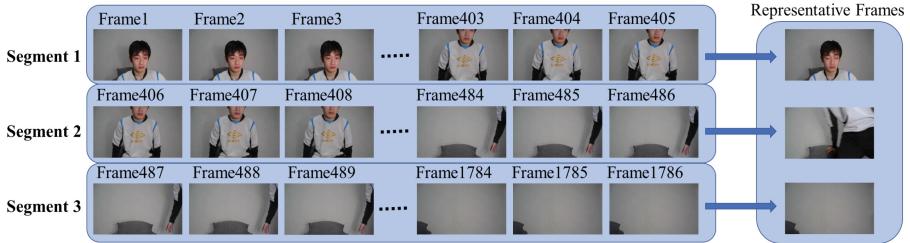


Fig. 3. Finding representative frame.

3 Implementation

We developed a Windows application using the environment described in Tables 1 and 2. The application calculates correlation values from a video file and saves them to a CSV file as a correlation matrix. The correlation value is calculated by ZNCC (Zero Mean Normalized Cross Correlation) supported by OpenCV 4.6.0. Figure 4 shows the application displaying a video and processing the calculation of correlation values.

4 Example Results

We founded representative frames of students' behavior from a video, as shown in Fig. 5. The video settings are as follows:

Resolution: HD quality

Frame Rate: 30 FPS

Number of Frames: 1,786

Playback Time: Approximately 1 min

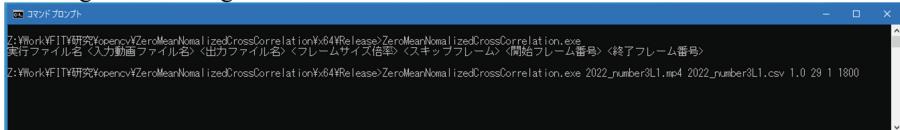
Table 1. PC environment.

Device	Product
CPU	Ryzen 9 5900X
Memory	80 GB DDR4-3200
GPU	GeForce RTX 3080
SSD	2 TB
OS	Windows 10 Pro 22H2

Table 2. Software environment.

Software	Product
IDE	Visual Studio Professional 2017
Language	Visual C++ 2017
Library	OpenCV 4.6.0

Starting a Scene Segmentation



Processing a Scene Segmentation

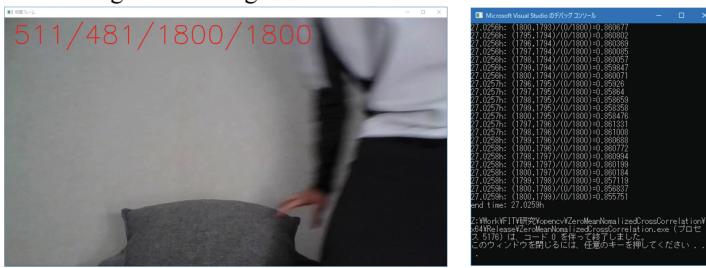


Fig. 4. Implementation.

In the video, the student stands up in frame 406 and leaves the seat in frame 487. In the video scene segmentation, we used the difference between adjacent correlation values to divide scenes according to the following criteria:

- (1) From the frame where the difference in correlation values exceeds a threshold to the frame where this difference becomes almost 0.
 - (2) From the frame where the difference in correlation values becomes almost 0 to the frame where it exceeds the threshold.

The thresholds were set from 0.1 to 0.3 incremented by 0.05. From the results, we confirmed that by finding representative frames from student video recordings, we can find their actions during online learning.

Threshold	Representative Frames						
0.1	Frame118	Frame244	Frame262	Frame298	Frame361	Frame406	Frame1207
0.15	Frame181	Frame451	Frame514	Frame550	Frame1558		
0.2	Frame109	Frame451	Frame514	Frame550	Frame1558		
0.25	Frame109	Frame478	Frame685				
0.3	Frame109	Frame460	Frame685				

Fig. 5. Example results.

5 Conclusions

In this paper, we introduced a method for finding representative frames of students' behaviors from recorded video, aiming to clearly visualize their actions during online learning.

We have shown the Windows application calculating the video frame correlation matrix between all video frames. Furthermore, we segmented the video into scenes based on correlation values in the video frame correlation matrix and extracted representative frames from each segmented scene. From the results, we confirmed that by finding representative frames from student video recordings, we can find their actions during online learning.

Currently, we are improving the representative frames to better reflect student actions, making it easier to understand students' behavior from these frames. In the future work, we would like to carry out evaluations for various types of students.

References

- Thakkar, S.R., Joshi, H.D.: E-learning systems: a review. In: Proceedings of IEEE Seventh International Conference on Technology for Education (T4E-2015), pp. 37–40 (2015). <https://doi.org/10.1109/T4E.2015.6>
- Alsadhan, A.O., Shafi, M.M.: What is the importance of different E-learning tools and systems?: An implementer's point of view. In: Proceedings of International Conference on Multimedia Computing and Systems (ICMCS-2014), pp. 686–689 (2014). <https://doi.org/10.1109/ICMCS.2014.6911364>

3. Narayanan, S.A., Prasanth, M., Mohan, P., Kaimal, M.R., Bijlani, K.: Attention analysis in e-learning environment using a simple web camera. In: Proceedings of IEEE International Conference on Technology Enhanced Education (ICTEE-2012), pp. 1–4 (2012). <https://doi.org/10.1109/ICTEE.2012.6208618>
4. Graf, S., Kinshuk., Liu, T.-C.: Identifying learning styles in learning management systems by using indications from students' behaviour. In: Proceedings of Eighth IEEE International Conference on Advanced Learning Technologies (2008), pp. 482–486 (2008). <https://doi.org/10.1109/ICALT.2008.84>
5. Takegawa, C., Sugita, K., Uchida, N.: Evaluation of re-watching support functions on educational content using learner monitoring system. In: Proceedings of 83-th National Convention of IPSJ (2021), pp. 143–144 (2021). (In Japanese)
6. Sugita, K.: A video scene segmentation approach for learner monitoring. In: Proceedings of International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2023), Lecture Notes on Data Engineering and Communications Technologies, vol. 189, pp. 214–221. Springer (2023). https://doi.org/10.1007/978-3-031-46970-1_20
7. Sugita K.: Performance evaluation of video scene segmentation approach to visualize learner behaviors. In: Proceedings of AROB 29-th 2024, International Society of Artificial Life and Robotics OS11–7, pp. 1109–1112 (2024)



A Comparative Sensitivity Analysis of Loss Functions in Machine Learning-Based Weather Forecasting

Aaron Van Poecke^{1(✉)}, Lukas Meuris¹, Matteo Cisneros¹, Michiel Van Ginderachter², Peter Hellinckx¹, and Hossein Tabari^{1,2,3}

¹ Modeling for Sustainability (M4S), Department of Electronics and ICT, University of Antwerp, Antwerp, Belgium
aaron.vanpoecke@uantwerpen.be

² Department of Meteorological and Climatological Research, Royal Meteorological Institute of Belgium, Brussels, Belgium

³ Environment and Health, United Nations University Institute for Water, Richmond Hill, ON, Canada

Abstract. Over the last year it has become apparent that advanced machine learning models are capable to compete with and even outperform conventional numerical weather prediction models. One of these models is GraphCast, developed by Google, and able to produce deterministic forecasts of hundreds of weather variables under one minute at state-of-the-art accuracy. These skills were learned by GraphCast during an extensive training at the heart of which lies the minimization of a loss function. Given this key role, understanding the model's sensitivity to the loss function is crucial. In this paper we present a comparative analysis of GraphCast's performance when trained on different loss functions, where we retrain GraphCast employing the mean absolute error and the log-cosh function next to the benchmark mean squared error. We assess the overall impact of different loss functions by calculating various error metrics and demonstrate the influence this choice has regarding the accuracy of weather forecasts across various regions and lead times.

1 Introduction

Accurate medium-range weather forecasts are vital for the functioning of our society as a whole, given that they lie at the basis of crucial decision-making processes across various domains, ranging from agriculture to public safety and from transportation to renewable energy [1–4]. Realizing these accurate forecasts remains, however, a daunting challenge due to the chaotic nature of the atmosphere [5]. Numerical weather prediction (NWP) models still form the backbone of operational weather forecasting today [6], encapsulating the evolution of the atmosphere in physics-based partial differential equations. These models are, however, subject to computational restraints and are prone to errors due to imperfect initial conditions and incomplete parametrizations [7]. These

drawbacks of physics-based weather forecasting might be largely overcome in the future with the advent of Machine Learning (ML) based weather forecasting. Driven by the fast surge in computational power [8], weather forecasting, like every major scientific field, has been revolutionized in recent years which, coupled with the availability of big data sets, led to high performing data-driven models [9]. In 2020, [5] asked the open question “*Can deep learning beat numerical weather prediction?*”, a question which found a seemingly positive answer in the three years that followed. Purely data-driven models like GraphCast by Google [10], FourCastNet by Nvidia [11] and PanguWeather by Huawei [12] have shown recently that they can compete with or even outperform state-of-the-art NWP models, all while needing orders of magnitude less calculation time. GraphCast, in particular, reached higher accuracy than the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF), globally considered as the most accurate NWP model, for over 90 % of the more than 1000 meteorological variables it forecasted [10]. GraphCast, however, was trained on nearly 40 years of ERA5 reanalysis data provided by the ECMWF [13], emphasizing its dependence on the accuracy of NWP models, which effectively means that, at least for the coming decades, we will most likely see a synergy between purely data-driven and physics-based models, instead of a replacement of the latter by the former. GraphCast will be the subject of the remainder of this paper, although the analysis could also be carried out for similar data-driven weather models.

The loss function, which quantifies the prediction error and is minimized during training, is located at the heart of large machine learning-based models like GraphCast [14]. Depending on the circumstances, different loss functions might lead to faster and better convergence as compared to others. The loss function employed during the training of GraphCast is the Mean Squared Error (MSE) or L2 loss, a popular choice for regression problems. Despite often resulting in satisfying results, this function also has its drawbacks, of which sensitivity to outliers and risks of missing minima during backpropagation are examples [15]. It is therefore relevant to analyze the effect of different loss functions on the training process of GraphCast and its subsequent predictive power regarding different variables, lead times and world regions, an analysis which, to the best of our knowledge, is currently missing in the literature.

In this work, the sensitivity of GraphCast to various loss functions is assessed. Due to computational complexity, we retrain a downsized version of the model, where we employ a different loss function during each training session. Next, we analyze how the loss function affects performance across multiple variables by calculating different performance metrics, which are compared across different lead times and world regions. The remainder of this paper is structured as follows: Sect. 2 elaborates on the methodology, where first the general structure of GraphCast is explained before diving deeper into the loss functions employed in this work. The results of our analysis are presented and discussed in Sect. 3, while we conclude and propose possible directions of future research in Sect. 4.

2 Methodology

2.1 GraphCast

GraphCast is a ML-based medium-range weather forecasting model released by Google in November 2023, which can produce accurate weather forecasts up to 10 days in less than a minute, utilizing only modest computational power. The main features of the algorithm are described here, but for more details we refer to the original paper [10]. The general structure of the model is represented by Fig. 1. GraphCast forecasts in an autoregressive way, in the sense that it takes as input the current and previous (with a six hour gap) state of the atmosphere and consequently feeds itself its own predictions to forecast up to ten days in advance, in timesteps of six hours. The underlying structures of the model are Graph Neural Networks (GNNs) present in the encoding, processing and decoding phase, as explained in detail by [16]. The complete GraphCast model is described by 36.7 million parameters which put out a global weather state on a 0.25° latitude/longitude grid, corresponding to a resolution of roughly 28×28 km.

2.2 Loss Functions

In addition to the benchmark MSE loss, utilized during the original development of GraphCast, we retrain the model on two loss functions: the Mean Absolute Error (MAE) or L_1 loss and the log-cosh loss. We briefly describe each of these functions in the following subsections.

Mean Squared Error

The MSE or L_2 loss can be formulated as follows:

$$L_2(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (1)$$

which is the ‘operational’ loss function of GraphCast. Here y represent the true values, \hat{y} the model’s estimates and n is the number of observations. This notation will be used throughout the remainder of this paper. The square in Eq. (1) implies a larger penalty for outliers and differentiability at $x = 0$. For small values of the loss function, decrease is more gradual implying a more certain convergence to the minima [17]. For large values, however, the square may result in large leaps when performing back propagation, jeopardizing chances of converging to the minimum [15].

Mean Average Error

The MAE or L_1 loss is given by:

$$L_1(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (2)$$

Despite its simplicity, Eq. (2) presents a relatively robust loss function. Its main advantages, following [15], are computational cheapness and insensitivity to outliers, while non-differentiability at $x = 0$ and steepness of the function, which

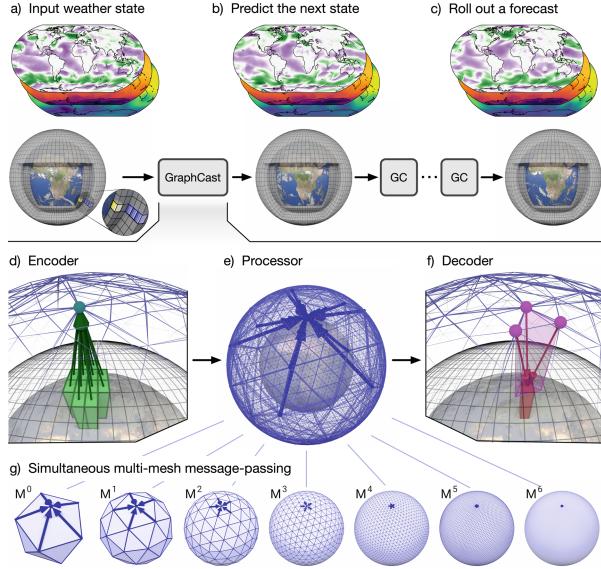


Fig. 1. A schematic overview of the GraphCast model [10].

may result in missing minima during back propagation, are its main disadvantages.

Log-Cosh

The expression for the log-cosh is given by:

$$\log \cosh = \sum_{i=1}^n \log (\cosh(\hat{y}_i - y_i)). \quad (3)$$

Similar to the L_2 loss, this function is differentiable, but not as sensitive to outliers. It essentially combines the advantages of linear and quadratic scoring functions, albeit at a higher computational cost [15].

2.3 Training GraphCast with Different Loss Functions

Training the original, complete GraphCast model with only one loss function took four weeks employing 32 Cloud TPU v4 devices utilizing batch parallelism [10]. Given a combination of time constraints and efficient use of computational resources, we decided to run a downscaled version of GraphCast when carrying out this research. This approach is justified given that we are mostly interested in relative difference between the performance of GraphCast when trained on different loss functions, and less in the absolute accuracy of the forecasts. Therefore, we removed the finest resolution layer M_6 (see Fig. 1), reduced the messaging passing steps from 16 to 8, decreased the latent feature size from 212 to 128

Table 1. The variables employed as input during the training of our downsized version of GraphCast. Atmospheric variables are available on 11 different pressure levels: 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925 and 1000 hPa. Forecasts of the variables in bold are the subject of further verification.

Parameter name	Short name	Units	Level
2 m Temperature	T2M	K	surface
10 m U wind component	U10M	ms^{-1}	surface
10 m V wind component	V10M	ms^{-1}	surface
Mean sea level pressure	MSL	hPa	surface
Total precipitation 6h	TP6H	mm	surface
Temperature	T	K	Atmospheric
Geopotential	Z	$\text{m}^2 \text{s}^{-2}$	Atmospheric
Specific humidity	Q	g kg^{-1}	Atmospheric
Vertical wind velocity	W	ms^{-1}	Atmospheric

and downsized the pressure levels from 37 to 13. In accordance with the original GraphCast model, training was carried out employing the reanalysis dataset ERA5 as ground truth [13], which is widely considered as being the most accurate global weather and climate dataset [18]. Table 1 lists the meteorological variables included during training, where bold abbreviations indicate the variables focused on during the verification process. The training period spanned from 1980 until 2019, while 2020 served as test year. During training, the loss function consisted of a weighted average of all variables, where more weight was given to surface variables [10] and with AdamW [19] serving as optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.95$ and a weight decay of 0.1. During a first phase, 58000 gradient descent steps were carried out with one autoregressive step, where the learning rate decreased from 1×10^{-3} towards 0, while the second phase consisted of 11000 gradient descent updates with autoregressive steps every 1000 updates ranging from 2 to 12 and a fixed learning rate of 1×10^{-7} . Training took around 12 h on one Nvidia GeForce RTX 4090 GPU with 16.384 CUDA cores and 24 GB RAM.

3 Results and Discussion

In order to assess the performance of the downsized GraphCast model trained on different loss functions, we verify the forecasts of three variables: temperature at 2 m (T2M), the wind velocity in the east-west direction (U10M) and the total precipitation in the last six hours (TP6H). These variables are of major interest for multiple applications including extreme weather event forecasting, renewable energy applications and the agricultural sector. As error metrics, we firstly calculate the Root Mean Squared Error (RMSE) to obtain a general impression of the average magnitude of the error in function of the lead time for each model

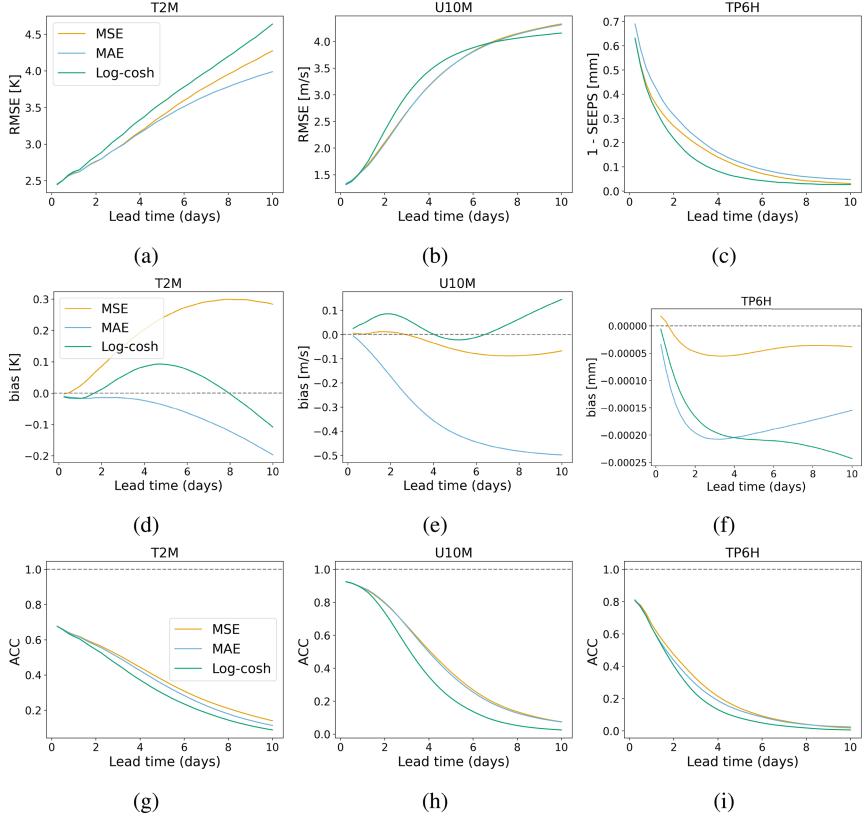


Fig. 2. Error metrics for all three loss functions with varying lead time. RMSE values for T2M (a), U10M (b), and 1 - SEEPS for TP6H (c). Bias for T2M (d), U10 (e), and TP6H (f). ACC values for T2M (g), U10 (h), and TP6H (i).

[20]. For TP6H, however, we employed, instead of the RMSE, the Stable Equitable Error in Probability Space (SEEPS), a metric tailored for the verification of precipitation forecasts and designed to penalize wrongful prediction of heavy precipitation [21]. Next, we assess the bias, which reflects the average difference between the forecasts and the ground truth, to provide the direction of said error, i.e., whether the model tends to over- or underestimate the meteorological variables. Lastly, we calculate the Anomaly Correction Coefficient (ACC) to assess the capability of each of the models to reproduce anomalies present in the dataset, where -1 and 1 translate to perfect disagreement and agreement respectively, whereas 0 means that there is no linear correlation present. The results of the verification are depicted in Fig. 2. A first important thing to note is that none of the three loss functions lead to an overall best performance, meaning the most accurate forecasts over all variables and all lead times. Concerning the RMSE, shown in Fig. 2a, 2b and 2c, the model trained on the MAE loss function

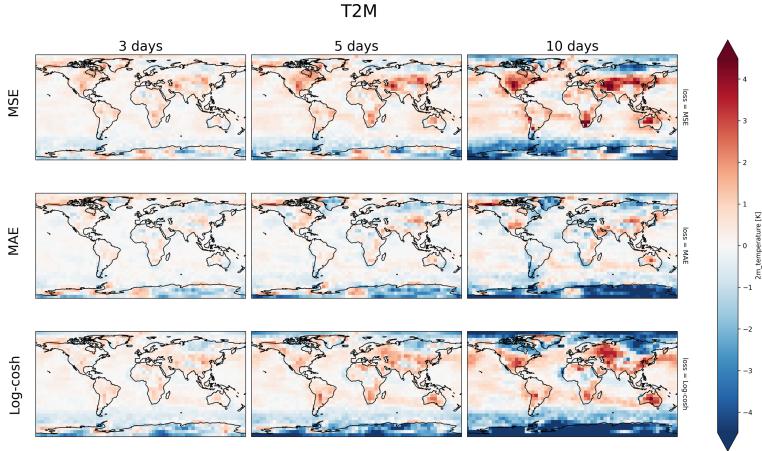


Fig. 3. Global bias of the T2M forecasts of the downsized GraphCast model when trained on MSE (top), MAE (center) and log-cosh (bottom), for three different lead times: 3 days (left), 5 days (middle) and 10 days (right).

outperforms the other ones for T2M and TP6H, which might be explained by its insensitivity towards outliers, i.e., extreme temperatures and rainfall in this case, as compared to the MSE and log-cosh. Overall, for RMSE and SEEPS, the model trained on log-cosh performs the worst. Interestingly, the MSE does not lead to the lowest RMSE values for T2M and U10M, reflecting the fact that the model is trained to minimize a weighted sum of MSE values of all variables, and not just the ones verified here. Figure 2g, 2h and 2i present the results regarding the ACC, where MSE and MAE result in very similar performance across all variables, while the log-cosh loss function scores worse than expected from other sources in the literature, where the loss function is presented as a viable alternative to the MAE and MSE [22]. Future research will unveil whether this is an artefact of the loss function itself, i.e., due to its computational complexity or possible training instability, or is linked to the downscaling of the model. Lastly, concerning the overall bias, the log-cosh performs best for T2M and U10M, but has a relatively large bias for precipitation. The MAE, on the other hand, has a negative bias for all variables, whereas the MSE leads to a positive bias for temperature and negligible bias for U10M and TP6H. This large positive bias becomes apparent when examining Fig. 3, where for large lead times the model trained on the RMSE largely overestimates the temperature over Asia, North-America and South-Africa as compared to the other two loss functions. Similarly, the underestimation of temperature around the South pole is less present for the MSE than it is for MAE and log-cosh, leading to the strong positive bias visible in Fig. 2d. It becomes clear that, for all three loss functions, a positive bias is exhibited in warmer regions and a negative bias in colder regions. This tendency is most pronouncedly negative in the model trained with log-cosh loss and most pronouncedly positive in the model trained with MSE. An explanation

for this could be the presence of large amounts of heat extremes in the training data. A model trained with MSE is penalized disproportionately for incorrect predictions, leading it to place redundant emphasis on these heat extremes. In contrast, a model trained with the log-cosh function is less susceptible to this phenomenon. Additionally, we should note that one year of testing data is possibly insufficient to thoroughly assess bias statistics [9], which is e.g., exemplified by all three models underestimating the exceptional 2020 heat wave in Siberia. Overall, the log-cosh function seems to lead to least accurate results, whereas the best performing method alternates between MSE and MAE across the different variables and various error metrics.

4 Conclusions and Future Prospects

In this work, we carried out a comparative sensitivity analysis of loss functions for GraphCast, a state-of-the-art ML-based weather forecasting model developed by Google. A downsized version of the model was retrained employing three different loss functions: the mean squared error, which served as benchmark, the mean absolute error and the log-cosh function. Next, forecasts of temperature, wind velocity and precipitation as generated by these models were compared and assessed across different error metrics, namely the root mean squared error, the bias and the anomaly correction coefficient. It became clear that the overall GraphCast performance varies significantly across variables and lead times when trained on different loss functions. The mean absolute error and root mean squared error produced the most accurate results, with the log-cosh trailing behind, although this last one is often put forward as a sound choice for training neural networks [23]. In future work, we will retrain the complete GraphCast model to investigate to what degree our conclusions can be extrapolated. Additionally, we will expand the set of loss functions by adding elastic-net regularization [24]. Furthermore, the goal is to carry out a more thorough analysis by enlarging the testing data to cover multiple years and assessing additional meteorological variables. Lastly, we plan to research the influence of these various loss functions on the capability of GraphCast to forecast extreme weather events across various world regions.

References

1. Lazo, J.K., Morss, R.E., Demuth, J.L.: 300 billion served: sources, perceptions, uses, and values of weather forecasts. *Bull. Am. Meteorol. Soc.* **90**(6), 785–798 (2009)
2. Nurmi, P., Perrels, A., Nurmi, V.: Expected impacts and value of improvements in weather forecasting on the road transport sector. *Meteorol. Appl.* **20**(2), 217–223 (2013)
3. Mohanty, U.C., et al.: A great escape from the Bay of Bengal “super sapphire-phailin” tropical cyclone: a case of improved weather forecast and societal response for disaster mitigation. *Earth Interact.* **19**(17), 1–11 (2015)

4. Van Poecke, A., Tabari, H., Hellinckx, P.: Unveiling the backbone of the renewable energy forecasting process: exploring direct and indirect methods and their applications. *Energy Rep.* **11**, 544–557 (2024)
5. Schultz, M.G., et al.: Can deep learning beat numerical weather prediction? *Philos. Trans. R. Soc. A* **379**(2194), 20200097 (2021)
6. Rabier, F.: Longer Ranges (2024). URL <https://www.ecmwf.int/en/newsletter/179>. Accessed 7 May 2024
7. Bouallègue, Z.B., Weyn, J.A., Clare, M.C.A., Dramsch, J., Dueben, P., Chantry, M.: Improving medium-range ensemble weather forecasts with hierarchical ensemble transformers. *Artif. Intell. Earth Syst.* **3**(1), e230027 (2024)
8. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
9. Rasp, S., et al.: Weatherbench 2: a Benchmark for the Next Generation of Data-Driven Global Weather Models (2023). [arXiv:2308.15560](https://arxiv.org/abs/2308.15560)
10. Lam, R., et al.: Graphcast: Learning Skillful Medium-Range Global Weather Forecasting (2022). [arXiv:2212.12794](https://arxiv.org/abs/2212.12794)
11. Pathak, J., et al.: Fourcastnet: a Global Data-Driven High-Resolution Weather Model Using Adaptive Fourier Neural Operators (2022). [arXiv:2202.11214](https://arxiv.org/abs/2202.11214)
12. Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q.: Pangu-Weather: a 3D High-Resolution Model for Fast and Accurate Global Weather Forecast (2022). [arXiv:2211.02556](https://arxiv.org/abs/2211.02556)
13. Hersbach, H., et al.: The era5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**(730), 1999–2049 (2020)
14. Wang, Q., Ma, Y., Zhao, K., Tian, Y.: A Comprehensive survey of loss functions in machine learning. *Ann. Data Sci.* **9**, 1–26 (2020)
15. Jadon, A., Patil, A., Jadon, S.: A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting (2022). [arXiv:2211.02989](https://arxiv.org/abs/2211.02989)
16. Keisler, R.: Forecasting Global Weather with Graph Neural Networks (2022). [arXiv:2202.07575](https://arxiv.org/abs/2202.07575)
17. Allen, D.M.: Mean square error of prediction as a criterion for selecting variables. *Technometrics* **13**(3), 469–475 (1971)
18. Horton, P.: Analogue methods and era5: benefits and pitfalls. *Int. J. Climatol.* **42**(7), 4078–4096 (2022)
19. Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization (2017). [arXiv:1711.05101](https://arxiv.org/abs/1711.05101)
20. Chai, T., Draxler, R.R., et al.: Root mean square error (RMSE) or mean absolute error (MAE). *Geosci. Model Dev. Dis.* **7**(1), 1525–1534 (2014)
21. Rodwell, M.J., Richardson, D.S., Hewson, T.D., Haiden, T.: A new equitable score suitable for verifying precipitation in numerical weather prediction. *Q. J. R. Meteorol. Soc.* **136**(650), 1344–1363 (2010)
22. Liu, Y., Zou, C., Chen, Q., Zhao, J., Caowei, W.: Optimization of critical parameters of deep learning for electrical resistivity tomography to identifying hydrate. *Energies* **15**(13), 4765 (2022)
23. Chen, P., Chen, G., Zhang, S.: Log Hyperbolic Cosine Loss Improves Variational Auto-encoder (2019)
24. De Mol, C., De Vito, E., Rosasco, L.: Elastic-net regularization in learning theory. *J. Complex.* **25**(2), 201–230 (2009)



Autonomous Shipping in Complex Situations

Matteo Cisneros^(✉), Oliver Rommens, Renzo Massobrio, and Peter Hellinckx

Faculty of Applied Engineering, Department of Electronics-ICT, University of Antwerp,
Antwerp, Belgium

matteo.cisneros@uantwerpen.be

Abstract. Recent advancements in artificial intelligence (AI) have had a significant impact on various sectors, especially maritime navigation. Since human behavior plays a major role in maritime accidents, there is a growing need for autonomous solutions to enhance safety. Additionally, the maritime industry faces a shortage of skilled captains, further necessitating the development of autonomous systems. Consequently, this study focuses on applying deep reinforcement learning (DRL) techniques to develop Maritime Autonomous Surface Ships (MASS), particularly addressing ship collision avoidance in compliance with the International Regulations for Preventing Collisions at Sea (COLREGs). We introduce a collision avoidance system within a custom Unity simulation environment, designed to handle static obstacles and dynamic encounter scenarios. Our approach utilizes Proximal Policy Optimization (PPO) to train an autonomous agent through a comprehensive state space and continuous rudder control. A detailed reward structure guides the agent towards safe and efficient navigation. The effectiveness of this methodology is validated through experiments in path-following, overtaking, head-on, and crossing give-way scenarios. Results show that the PPO algorithm enables COLREG-compliant maneuvers and successful navigation in these scenarios. This research advances autonomous maritime navigation, demonstrating DRL's potential to enhance safety and efficiency in maritime operations.

1 Introduction

Over the last decade, advancements in artificial intelligence (AI), particularly in reinforcement learning (RL), have become essential in modern innovations. A notable application in the maritime sector is the rise of Maritime Autonomous Surface Ships (MASS) recognized by the International Maritime Organization (IMO). MASS ranges from vessels with automated processes supervised by sailors to fully autonomous ships [1,2].

The development of MASS is driven by the need to enhance maritime safety. Research by the European Maritime Safety Agency (EMSA) indicates that human actions were responsible for 80.7% of maritime incidents from 2014 to 2022 [3]. Economically, MASS offers potential improvements in logistics efficiency, fuel optimization, and reduced operational costs [4].

In open sea environments, adherence to the International Regulations for Preventing Collisions at Sea (COLREGs) set by the IMO is crucial. These regulations ensure

safe navigation and collision avoidance by providing guidelines for specific encounter situations, such as head-on, crossing, and overtaking, and defining the roles of stand-on and give-way vessels [5]. Understanding and adhering to these guidelines is essential for safe navigation and compliance with international maritime policies.

Traditional collision avoidance methods in maritime navigation have relied on geometrical principles like the ship domain and closest point of approach (CPA) [6], as well as algorithms such as the artificial potential field (APF) [7] and velocity obstacle (VO) methods [8]. However, these methods often struggle with local minima issues and may not always ensure compliance with COLREGs. Recently, deep reinforcement learning (DRL) has emerged as a promising solution, leveraging environmental interaction to learn optimal strategies without explicit labeled data [9].

Our goal is to solve the collision avoidance problem in compliance with COLREGs. We create a custom simulation environment within the 3D dynamic framework Unity [10], replicating COLREGs scenarios. We implement a dual-layered collision avoidance approach. The first layer focuses on navigation, path-following, and avoiding static obstacles in an open sea environment. The second layer addresses COLREG encounter scenarios, allowing our agents to dynamically adapt to various situations. The primary challenges include navigating around vessels moving at different velocities and directions, avoiding static obstacles while maintaining the planned path, and ensuring regulatory compliance during all maneuvers.

The remainder of this paper is organized as follows: Sect. 2 presents the methodology, Sect. 3 describes the experiments and analyzes the results, and Sect. 4 concludes with a discussion of future work.

2 Methodology

2.1 COLREGs Compliance

To navigate safely, our DRL model must comply with the COLREGs. We use a first-person perspective, referring to our vessel as “Own Ship” (OS) and other vessels as “Target Ship” (TS).

Understanding that the COLREGs prioritize safety and collision avoidance, we incorporate the Goodwin Ship Domain’s principles [6] along with the Distance Closest Point of Approach (DCPA) collision risk index. This allows for a quantitative evaluation of risk, serving as a proxy for COLREGs compliance and also to calculate collision risk areas and define a safe area around the OS.

Translating human-centric COLREGs concepts into machine-readable formats is a significant challenge. Terms like “large enough” and “substantial action” must be accurately interpreted and replicated in a quantifiable manner.

Our study focuses on *Part B* of the COLREGs, known as the “Steering and Sailing Rules” [5], which govern encounter scenarios between vessels, including crossing, head-on, and overtaking, detailed in *rules 13 to 17* and illustrated in Fig. 1.

- *Rule 13:* For overtaking scenarios, when one vessel approaches another from behind at a higher speed and intends to pass, it must maintain a safe distance and may pass on either the starboard or port side.

- *Rule 14:* For head-on situations, when two vessels are heading directly towards each other, both should alter their course to starboard to pass on each other's port side.
- *Rule 15:* For crossing scenarios, if the OS encounters the TS on its starboard side, the OS is the give-way vessel. *Rule 16* clarifies that a give-way vessel must take substantial action to avoid a collision.
- *Rule 17:* If the OS encounters the TS on its port side, the stand-on vessel must maintain its course and speed, adjusting only if necessary to avoid a collision.

Our research does not cover the stand-on scenario, as the OS is not supposed to take any actions, which does not add value to our collision avoidance algorithm.

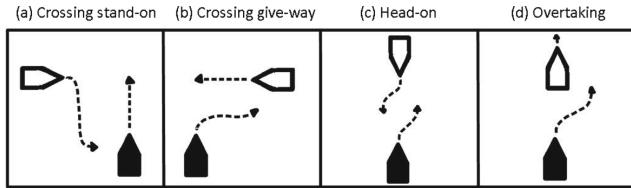


Fig. 1. Encounter scenarios with OS in black and TS in white.

2.2 Global Path Planning

We use the A* algorithm [11] for path planning to guide our ship to the goal by placing intermediate checkpoints for the ship to follow. The environment is divided into a grid where each cell represents a possible position for the ship. A* then searches for the shortest path by examining nodes from the starting point towards the objective, keeping a priority queue of nodes to be assessed using a cost function. The Euclidean distance function will be used as our cost function, $f(n)$, and is defined in Eq. 1.

$$f(n) = g(n) + h(n) \quad (1)$$

In Eq. 1, $g(n)$ represents the cost from the start node to the current node n and $h(n)$ represents a heuristic estimate of the cost from n to the goal node. The Euclidean function $h(n)$ plays a critical role in optimizing the search process.

2.3 Proximal Policy Optimization

PPO [12] is an actor-critic-based DRL algorithm designed to achieve stable and efficient policy updates. Within the actor-critic framework, the actor determines the agent's actions based on the current policy, while the critic evaluates the actions by estimating the advantage \hat{A}_t , providing feedback to the actor for continuous improvement. We estimate \hat{A}_t using the generalized advantage estimation (GAE) [13], which helps to reduce the variance while maintaining low bias, thereby improving the stability of policy updates. PPO approximates the policy π using a neural network with parameters θ and optimizes it by maximizing the clipped surrogate objective, as shown in Eq. 2.

$$L^{CLIP}(\theta) = \mathbb{E}_t [\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)] \quad (2)$$

The ratio function in Eq. 3 calculates the divergence between the old and current policies. This ratio compares the probability of the current action under the new policy to that under the old policy, ensuring updates remain within a safe range to avoid overly drastic changes.

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} \quad (3)$$

The final PPO loss function, shown in Eq. 4, combines the clipped surrogate objective with a value loss and an entropy bonus. The value loss L^{value} represents the squared error between predicted and target values, and the entropy bonus L^{entropy} encourages exploration by maintaining policy randomness. Constants c_1 and c_2 balance these components, promoting stable and efficient learning.

$$L^{\text{PPO}} = L^{\text{clip}} + c_1 L^{\text{value}} - c_2 L^{\text{entropy}} \quad (4)$$

2.4 State Space

The state space is the complete description of the environment, based on which the agent selects its actions. We assume the state space is fully observable by our agent. The encounter scenario is based on the bearing of the target ship which divides the OS's surrounding area into four quadrants representing different encounter scenarios.

To achieve COLREGs-compliant navigation and effective path-following, we developed a state space consisting of S_{tOS} (own ship), S_{tTS} (target ship), S_{tPF} (path-following), and S_{tOS} (static obstacles).

$$S_{tOS} = [\mathbf{v}_{OS}, \psi_{OS}] \quad (5)$$

In Eq. 5, \mathbf{v}_{OS} represents the OS's velocity and ψ_{OS} represents its course.

$$S_{tTS} = [d_{TS}, \beta_{TS}, \mathbf{v}_{TS}, \psi_{TS}] \quad (6)$$

In Eq. 6, d_{TS} is the distance to the TS, β_{TS} is the bearing to the TS with respect to the OS, \mathbf{v}_{TS} represents the TS's velocity, and ψ_{TS} denotes the TS's course.

$$S_{tP} = [\mathbf{u}_{PF}, \mathbf{f}_{PF}] \quad (7)$$

In Eq. 7, \mathbf{u}_{PF} is the vector to the next checkpoint and \mathbf{f}_{PF} indicates the forward direction to the next checkpoint.

The state for static obstacle avoidance, S_{tOS} , uses a Ray Perception Sensor in Unity, emitting rays to detect static obstacles. Each ray measures the distance to the nearest obstacle, forming an observation vector that informs the agent about its environment, similar to LiDAR systems.

$$S_t = [S_{tOS}, S_{tTS}, S_{tPF}, S_{tSO}] \quad (8)$$

Equation 8 defines a comprehensive state space that enables the agent to perform COLREGs-compliant navigation, effective path-following, and static obstacle avoidance.

2.5 Action Space

The OS uses continuous rudder angle control for smooth direction changes, including maintaining course, turning right, or turning left. Speed modifications are excluded due to the operational limitations of large vessels at sea. These vessels typically have fixed-pitch propellers and protection systems that prevent abrupt engine speed changes. Additionally, their significant inertia results in minimal speed reduction when propeller rotations are decreased. Therefore, for large vessels at maximum speed in the open sea, adjusting the rudder for course changes is more feasible and effective for collision avoidance than altering speed.

2.6 Reward Functions

To speed up the learning process, we use curriculum learning, where the agent starts with simpler tasks and gradually progresses to more difficult ones. Initially, a greater distance between the agent and the checkpoint is considered successful. As the agent consistently completes this task, the allowed distance is gradually decreased. This approach helps the agent learn more effectively, leading to faster convergence. We define the following rewards to guide the agent in making optimal decisions:

- **Goal Reward:** To account for necessary deviations from the path due to collision avoidance, the goal reward R_{goal} guides the OS to subsequent checkpoints and the final destination. R_{goal} is defined in Eq. 9.

$$R_{\text{goal}} = \begin{cases} r_{\text{check}} & \text{if } \|P_t - P_{\text{check}}\| < \gamma_0 \\ r_{\text{finish}} & \text{if } \|P_t - P_{\text{finish}}\| < \gamma_0 \end{cases} \quad (9)$$

Here, P_t is the agent's current position, P_{check} the next checkpoint, and P_{finish} the final checkpoint. Rewards r_{check} and r_{finish} encourage reaching checkpoints, with r_{finish} being higher to emphasize the final goal.

- **Advance reward:** To assess progress toward each checkpoint we define the advance reward denoted R_{adv} and defined in Eq. 10.

$$R_{\text{adv}} = \begin{cases} -r_{\text{adv}} & \text{if } \|P_{\text{check}} - P_t\| > \|P_{\text{check}} - P_{t-1}\| \\ r_{\text{adv}} & \text{if } \|P_{\text{check}} - P_t\| < \|P_{\text{check}} - P_{t-1}\| \end{cases} \quad (10)$$

P_{check} is the checkpoint's position, and P_t and P_{t-1} are the current and previous positions of the agent. Positive rewards are given for moving closer to the checkpoint, while penalties apply for moving away.

- **Proximity penalty:** In order to guide the vessel away from vessels and obstacles a reward function denoted $R_{\text{proximity}}$ is defined in Eq. 11.

$$R_{\text{proximity}} = \begin{cases} -\frac{r_{\text{prox}}}{d_{TS}} & \text{if } d_{TS} < \gamma \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Here, d_{TS} is the distance to the target vessel or obstacle, r_{prox} is the penalty, and γ is the threshold distance. The penalty increases as the agent gets closer to the TS.

- **Special cases:** For overtaking scenarios, existing rewards suffice as the overtaking vessel needs to maintain a safe distance without a specific side requirement as

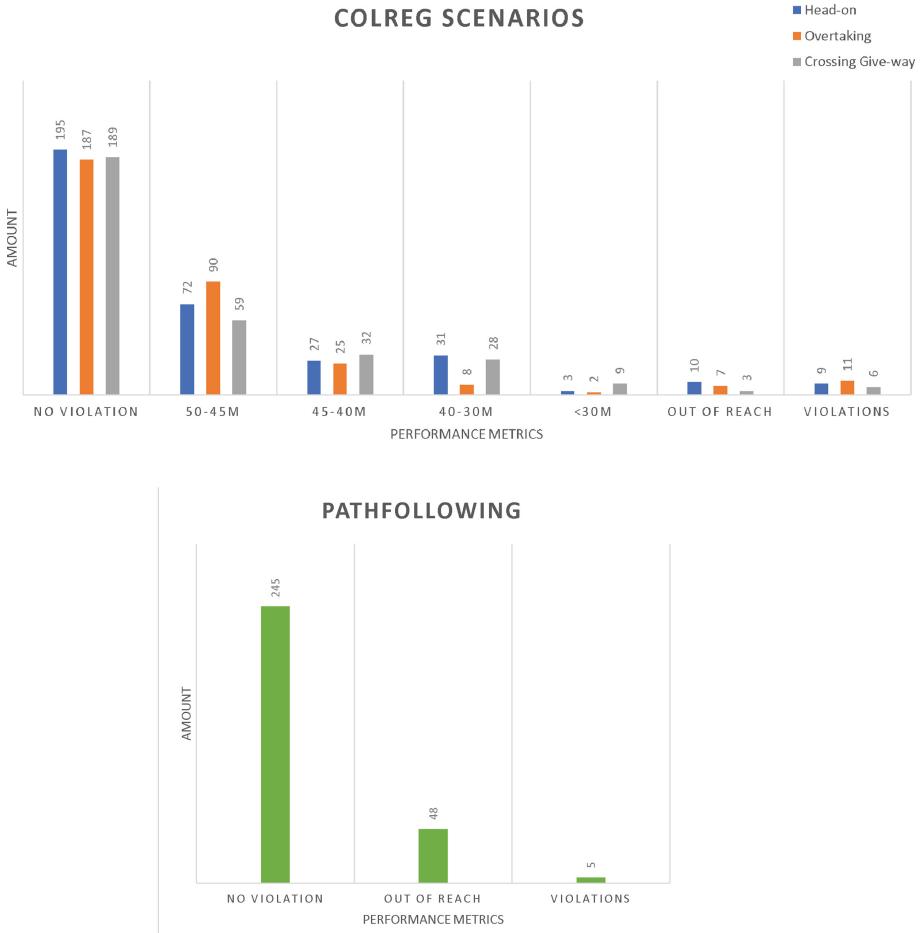


Fig. 2. DRL performance of approximately 300 simulations on different COLREGs scenarios and path-following tasks. Metrics: 50–45 m, 45–40 m, 40–30 m, and <30 m define the closest proximity the OS had during the simulation; out of reach means that the OS went too far from the A* path; Violation define a collision between ships or a forbidden manoeuvre.

described in COLREGs' rule 13. For head-on and crossing scenarios, the reward function R_{SPcases} in Eq. 12 applies.

$$R_{\text{head-on}} = \begin{cases} r_{\text{right}} & \text{if } 247.5^\circ < \beta_{\text{CPA}} < 355^\circ \\ -r_{\text{left}} & \text{if } 5^\circ < \beta_{\text{CPA}} < 112.5^\circ \end{cases} \quad (12)$$

β_{CPA} represents the target ship's relative direction at the closest point of approach. Rewards and penalties encourage passing port-to-port, reducing collision risk per maritime regulations. In crossing scenarios, the give-way vessel uses a similar reward function to avoid crossing ahead and reward passing astern.

3 Experiments and Results

We used a custom Unity environment to validate the effectiveness of the proposed PPO algorithm for collision avoidance and COLREGs-compliant path-following. Four scenarios were designed: path-following, head-on, overtaking, and crossing give-way. The OS was evaluated on COLREGs compliance and path-following, with metrics shown in Fig. 2. Violations included wrong-side crossings and collisions. We established three domain zones: safe (over 100 m), maneuvering (50–100 m), and caution (under 50 m). The caution zone was further divided into 45–50 m, 40–45 m, 30–40 m, and under 30 m, with increasing penalties as the OS approached the TS, reflecting collision risk.

For path-following, curriculum learning was used, starting with checkpoints 40 m away and reducing to 25 m in 5-m increments. In COLREGs scenarios, checkpoints within 80 m were hit to account for a 50-m TS domain, with a 30-m window. Curriculum learning for these scenarios started at 100 m, reducing to 80 m in 5-m increments.

To expose the agent to varied scenarios, we introduced randomness in each encounter. The OS spawned at the first checkpoint with a -5 to 5 m y-coordinate deviation. Each scenario had the TS offset relative to the OS, as shown in Fig. 2. In the crossing scenario, TS orientation ranged from 60 to 120°. The TS speed was randomly set for each run but was always slower than the OS.

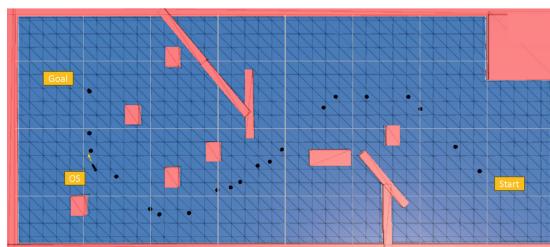


Fig. 3. Own ship (OS) following the A* path from start to goal while avoiding colliding with environment

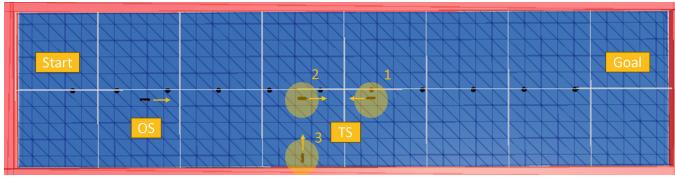


Fig. 4. Initialization of different COLREGs situation between own ship (OS) and target ship (TS): (1) Head-On, (2) Overtaking, (3) Crossing Give-Way

In the path-following scenario, illustrated in Fig. 3, the agent navigated a predefined path while avoiding static obstacles. The A* algorithm calculated the shortest path with a 50-m buffer from obstacles. Out of 293 runs, the agent successfully completed 245 runs. Failures included 48 instances of deviating too far from checkpoints due to sharp corners and 5 collisions with static obstacles.

The agent's compliance with the COLREGs was evaluated for the head-on, overtaking, and crossing give-way scenarios, illustrated in Fig. 4 as scenarios 1, 2, and 3, respectively. The results of our experiment on the compliance with COLREGs are presented as follows:

- **Head-On:** The OS altered its course to starboard to avoid a head-on collision, successfully executing the maneuver in 61.3% of 318 runs. Approach distances included 22.6% within 45–50 m, 8.5% within 40–45 m, 9.7% within 30–40 m, and 0.9% within less than 30 m. Failures included 3.1% out of reach alerts and 2.8% COLREGs violations, including 2.2% collisions.
- **Overtaking:** The OS encountered a slower-moving TS and had to safely overtake while adhering to COLREGs. The OS successfully overtook the TS in 61.3% of 305 runs. The approach distances were as follows: 29.5% within 45–50 m, 8.2% within 40–45 m, and 0.7% within less than 30 m. Failures included 2.3% out of reach alerts and 3.6% collisions with the outer barrier.
- **Crossing Give-Way:** The OS gave way to a TS crossing its path from the starboard side, completing the course in 62.0% of 305 runs. Approach distances included 19.3% within 45–50 m, 10.5% within 40–45 m, 9.2% within 30–40 m, and 2.9% within less than 30 m. Failures included 1.0% out of reach alerts and 2.0% collisions, including 1.3% with the TS.

4 Conclusion and Future Work

This paper proposes a collision avoidance system for autonomous marine navigation using DRL techniques to comply with COLREGs. The A* algorithm is used for global path planning by setting intermediate checkpoints, ensuring the vessel can follow a predetermined course. PPO is employed within a customized Unity simulation environment to enhance the safety and efficiency of MASS. The results demonstrate the OS's ability to navigate safely and comply with COLREGs during path-following, overtaking, head-on, and crossing give-way scenarios. However, improvements are needed in handling close-range encounters and maintaining safe distances.

Future research could explore Inverse RL for generating reward functions based on expert demonstrations, addressing the challenge of defining accurate rewards for the ambiguous COLREGs. Furthermore, incorporating a more complex and precise physics model into the simulation can provide more authentic scenarios, improving the management of inertia and momentum for large vessels, and resulting in more realistic decision-making processes. Additionally, expanding from a single-agent to a multi-agent framework can enable more complex interactions and collaborative behaviors among multiple independent ships, ensuring collision avoidance and COLREGs compliance in a complex environment.

References

1. IMO. Autonomous shipping (2022)
2. Munim, Z.H., Haralambides, H.: Advances in maritime autonomous surface ships (MASS) in merchant shipping. *Marit. Econ. Logistics* **24**(2), 181–188 (2022)
3. EMSA. Annual overview of marine casualties and incidents. EMSA Lisbon (2023)
4. Tsvetkova, A., Hellström, M.: Creating value through autonomous shipping: an ecosystem perspective. *Marit. Econ. Logistics* **24**(2), 255–277 (2022)
5. International Maritime Organization. COLREG: Convention on the International Regulations for Preventing Collisions at Sea, 1972. International Maritime Organization (2002)
6. Goodwin, E.M.: A statistical study of ship domains. *J. Navig.* **28**(3), 328–344 (1975). <https://doi.org/10.1017/S0373463300041230>
7. Lyu, H., Yin, Y.: Colregs-constrained real-time path planning for autonomous ships using modified artificial potential fields. *J. Navig.* **72**(3), 588–608 (2019). <https://doi.org/10.1017/S0373463318000796>
8. Kuwata, Y., Wolf, M.T., Zarzhitsky, D., Huntsberger, T.L.: Safe maritime autonomous navigation with colregs, using velocity obstacles. *IEEE J. Oceanic Eng.* **39**(1), 110–119 (2014). <https://doi.org/10.1109/JOE.2013.2254214>
9. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (2018)
10. Juliani, A., et al.: Unity: a general platform for intelligent agents. [arXiv:1809.02627](https://arxiv.org/abs/1809.02627) (2018)
11. Foead, D., Ghifari, A., Kusuma, M.B., Hanafiah, N., Gunawan, E.: A systematic literature review of a* pathfinding. *Proc. Comput. Sci.* **179**, 507–514 (2021). <https://doi.org/10.1016/j.procs.2021.01.034>. ISSN 1877-0509. 5th International Conference on Computer Science and Computational Intelligence 2020
12. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms (2017)
13. Schulman, J., Moritz, P., Levine, S., Jordan, M., Abbeel, P.: High-dimensional continuous control using generalized advantage estimation. [arXiv:1506.02438](https://arxiv.org/abs/1506.02438) (2015)



Transfer Learning for Traffic State Predictions in Small and Medium-Sized Cities

Mohammadmahdi Rahimiasl¹✉, Ynte Vanderhoydonc¹, Siegfried Mercelis¹, Laure De Cock², Thomas Kusmirczak², and Tamara De Swert²

¹ IDLab - Faculty of Applied Engineering, University of Antwerp - imec, Antwerp, Belgium

{Mohammadmahdi.Rahimiasl,Ynte.Vanderhoydonc,
Siegfried.Mercelis}@uantwerpen.be

² AI & Algorithms, imec, Leuven, Belgium

{Laure.Decock,Thomas.Kusmirczak,Tamara.Deswert}@imec.be

Abstract. In Flanders, 80% of the population resides in small and medium-sized cities, which face significant challenges in data-driven traffic management due to a lack of data. Although deep learning models surpass traditional methods for traffic state prediction, they require extensive data, often unavailable in these smaller cities. Transfer learning offers a potential solution by utilizing models trained in data-rich environments to enhance predictions in data-scarce regions. This ongoing research investigates the application of transfer learning for traffic state prediction in small and medium-sized cities. Preliminary tests in Greater Manchester and Lisbon under the TANGENT H2020 project have shown promising results. The next phase involves applying this technique to the Flemish city of Mechelen as part of the CitCom.ai project. The goal is to improve traffic state prediction in data-limited environments, thereby contributing to more efficient urban traffic management.

1 Introduction

The EU is distinguished by its many small and medium-sized cities and towns. Almost half of all EU cities have populations between 50.000 and 100.000, inhabitants, and there are over 8.000 towns with populations ranging from 5.000 to 50.000. Together, these towns and cities are home to nearly 30% of the EU's population. In Flanders, this figure rises to over 80%. This means that the majority of Flemish citizens is scattered amongst 285 cities and communities, resulting in significant challenges for mobility data management, such as resource scarcity and data deficiencies.

A lack of mobility data limits the possibilities for a data-driven and proactive traffic management in small and medium-sized cities. As urban transportation systems evolve, accurately predicting traffic states, including active modes such as walking and cycling, and crowd dynamics, is crucial for efficient traffic management and proactive congestion response. Data-driven approaches can

capture features in traffic data to forecast future states, and advancements in deep learning have led to models that surpass traditional statistical methods in performance. However, these approaches require a lot of data, which is often lacking in small and medium-sized cities.

This research introduces a novel approach for traffic state prediction in small and medium-sized cities, using transfer learning algorithms. Transfer learning can expedite the development of deep learning models in regions with limited historical traffic data, but its effectiveness for traffic forecasting requires further investigation [1,2]. Transfer learning is not only a promising solution to the data scarcity of small and medium-sized cities, but it can also save significant effort, time, and resources needed to develop new prediction models, especially deep learning models. The transferability of deep learning models trained on data-rich cities to those with limited data is the key focus of this research.

This research will be conducted for the CitCom.ai project, in the medium-sized Flemish city of Mechelen, and will draw from the case study of Greater Manchester that was conducted for the TANGENT H2020 project.

Section 2 provides a background on traffic state predictions using deep learning models and transfer learning. Section 3 presents our results for two case studies of Greater Manchester and Lisbon, together with a possible path towards future research directions. Finally, Sect. 4 provides a short conclusion.

2 Background

2.1 Traffic State Predictions

Data-driven approaches capture features in traffic data, which can be used to predict the future traffic state. With the advancement of deep learning models in various fields [3], researchers have developed multiple models for traffic supply forecasting, which surpassed traditional methods, e.g., statistical methods in performance. One of the early models DCRNN (Diffusion Convolutional Recurrent Neural Network) [4] presents a forecasting approach by integrating graph convolutional and recurrent neural networks which improved the state-of-the-art significantly. Meanwhile, other models are developed and improved the DCRNN model by introducing many different components like attention mechanisms for spatial and temporal aspects of the time series data [5–7].

The process of traffic state prediction is described in Fig. 1, and the different steps of the process are detailed below:

1. Input

Training a traffic state prediction model via data-driven approaches requires historical traffic data. The historical observations are used to train a model that predicts future traffic conditions. Historical data (e.g., time series of flows, speeds) coming from available traffic counters, floating vehicle data, etc., should preferably represent one year worth of data to incorporate all seasonal effects and is used for the training of the deep learning model. Real-time data, with a similar format, is used to provide traffic supply predictions in real-time based on the trained model.

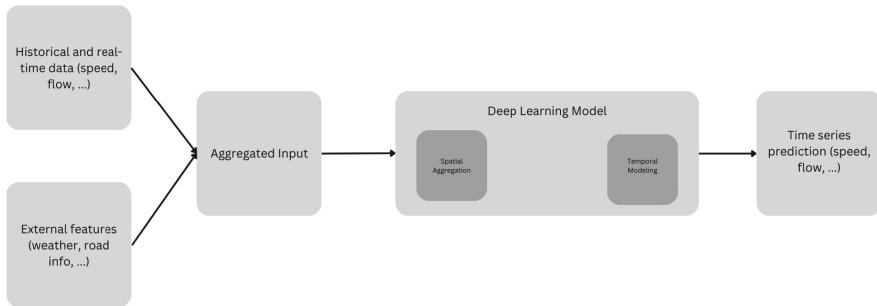


Fig. 1. Process of traffic supply forecasting with deep learning

Optionally, external features that influence traffic (e.g., weather data, road information like road type, number of lanes) can be collected, to be combined with historical and real-time traffic data. Note that the input of weather data should be the same granularity as the traffic data itself. Furthermore, it is required to have historical and real-time weather data, and a weather forecast.

2. Deep learning model

Deep learning modelling is performed on the collected input data. For testing, a traffic dataset is split in terms of training, validating, and testing. 70% of the historical data is used to train the deep learning model, 10% is used for validation, and 20% for testing purposes. Validation data helps in fine-tuning the model and prevents overfitting while testing data assesses its performance on previously unseen data and ensures its generalizability. For each dataset, model's hyperparameters should be optimized on the validation set to get the best results. The modelling task can be separated into two subtasks: spatial aggregation and temporal forecasting. Spatial aggregation is the task of aggregating the data from neighbouring nodes for the node of interest, and temporal forecasting is the task of predicting the future values of traffic states using the current values.

The model used in the preliminary tests in the remainder of this research, refers to a patent regarding traffic prediction, filed by imec and published with number US2024/0054321. It outperforms other methods in training time and has strong transfer learning capabilities. The model uses a convolution module to model the spatial-temporal relationships of the traffic data and an encoder-decoder architecture. The convolution module is applied to the time dimension which also considers the neighborhood and generates relation-based traffic data. The relation-based traffic data is then fed to an encoder to generate the fixed length vector latent representations. Afterwards, these representations are fed to the decoder to generate the predictions. The encoder and the decoder are both implemented using Recurrent Neural Networks (RNNs) like Long Short-Term Memory (LSTM) architecture.

3. Output

The output of these models are time series predictions for the future time steps.

2.2 Transfer Learning

Transfer learning encompasses a range of machine learning techniques designed to transfer knowledge from a source domain, which is abundant in data, to a target domain, which is limited in data. This approach has proven highly successful in areas such as text classification and product recommendation. [8] acknowledges that the transfer learning approach has high potential for smart city applications, but existing algorithms cannot be blindly copied from other domains, because of the spatiotemporal patterns that exist in the urban domain. It defines three types of transfer learning for the smart city domain:

1. Cross-modality, which transfers between different data types. For example, social media data can be used as a proxy for crowd dynamics.
2. Cross-city, which transfers between different cities. For example, crowd dynamics of one city can be used as a proxy for the crowd dynamics of another city.
3. Cross-modality and cross-city, which transfers between modalities and between cities.

The overall procedure for transfer learning is shown in Fig. 2, a spatio-temporal deep learning model is first trained on the source dataset, with the objective to minimize the loss value. This procedure is called pretraining. In the pretraining step, the weights of the deep learning model are initialized randomly using initialization methods such as Kaiming or He initialization. After the pre-training step, the weights of the model are used as initiate weights for training the same model using the target dataset. It might be needed to change some parts of the model, usually at the beginning or end of the model, based on the architecture of the model and training data shapes. Those altered weights are initiated randomly.

To address the data scarcity in small and medium-sized cities, transfer learning methods have been proposed for cross-city prediction models. For example, RegionTrans [9] concentrates on fine-tuning a source model, while [10] aims to selectively learn from the source city to better facilitate target fine-tuning. [11] proposed Graph Neural Networks for transfer learning in the urban environment, as the graph representation effectively captures the spatiotemporal patterns of the urban environment. The resulting transfer learning model with graph partitioning for traffic speed prediction (TEEPEE) outperformed traditional traffic prediction models and transfer learning models without graph partitioning.

Although transfer learning has proven to be a promising technique for traffic predictions in small and medium-sized cities, few efforts have been made to date to combine transfer learning, and deep learning for traffic state predictions in real-world settings, with real-world data. To address this gap, we aim to develop

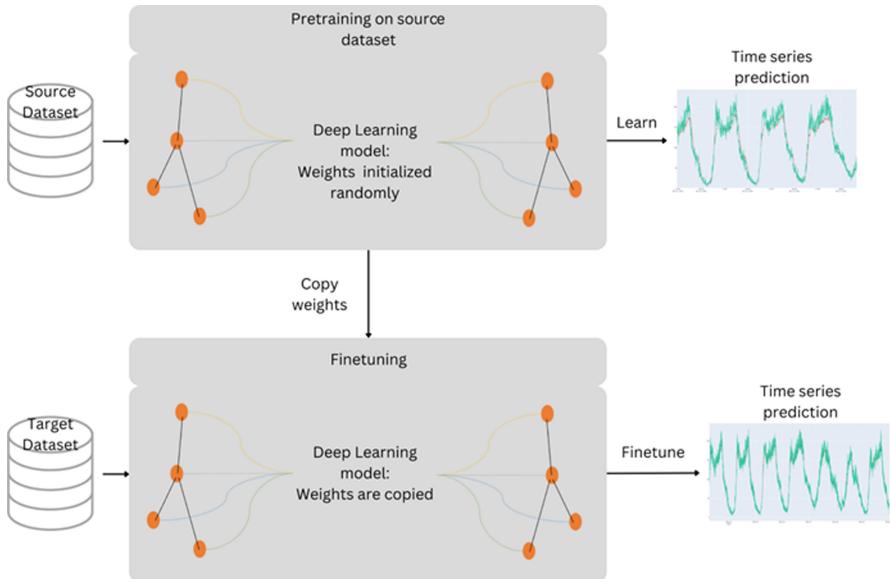


Fig. 2. Process of traffic supply forecasting with transfer learning

and test this method on real-world datasets, coming from several Flemish cities for the CitCom.ai project. In a first phase, we conducted a series of tests for one of the case studies within the TANGENT project, specifically focusing on Greater Manchester and Lisbon.

3 Transfer Learning for Traffic State Predictions in Real-World Settings

3.1 Case Study of Greater Manchester

One of the use cases in the TANGENT project concentrates on the area of Greater Manchester. For this use case, a dataset of traffic counters is available, including speed bins and traffic flows (locations shown in Fig. 3). This data has been made available, with a resolution of 5 min and a duration of 13 months (1st January 2022 to 31st January 2023).

To test the possibilities of transfer learning for traffic state prediction, two additional open datasets were collected, from two different cities:

1. Los Angeles (US)

The METR-LA dataset [12], including counts on the highway of Los Angeles, is commonly used in literature for benchmarking methods on traffic state prediction.

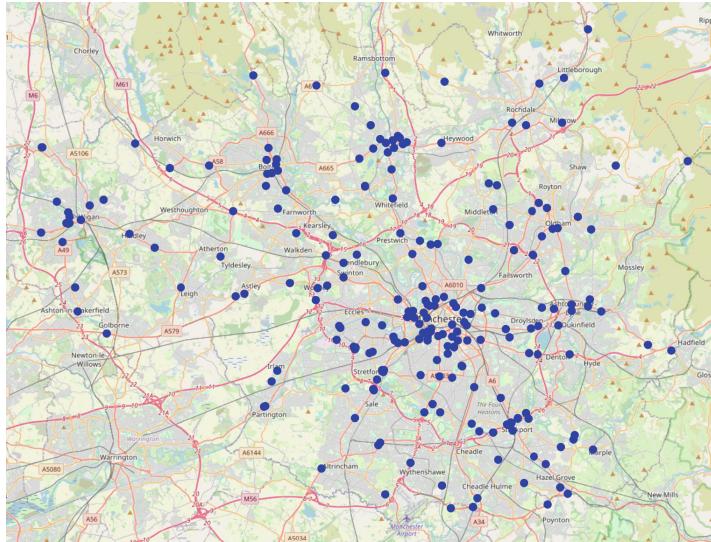


Fig. 3. Location of loop detectors in Greater Manchester

2. Luzern (Switzerland)

The UTD19 dataset [13] is a large-scale traffic dataset from detectors on urban roads in 40 cities worldwide. Luzern is one of them, and contains traffic flows, from loop detectors, in three-minute time stamps.

Both datasets are very different, as the METR-LA dataset, only contains speed data on highways, while UTD19 contains urban data. As a result, the METR-LA dataset is smooth and cleaned, while the UTD19 urban traffic flows change very abruptly.

The three datasets allowed for some explorative tests on transfer learning for traffic state predictions. Two different sets of tests were performed: one to predict traffic counts (see Table 1), and one to predict traffic speeds (see Table 2)). For counts and speed two models were trained: a classic traffic state prediction model that was trained on historical data from Manchester and made predictions for Manchester; a transfer learning model that was trained on historical data from either LA or Luzern and made predictions for Manchester. The preliminary results are presented in the tables below, including details regarding the amount of data used in both the source and target city, and error measurements in preliminary testing. In tables, Mean Absolute Error (MAE) measures the average of the absolute differences between prediction and ground truths. Mean Absolute Percentage Error (MAPE) measures the average percentages errors between predictions and ground truths. Symmetric Mean Absolute Percentage Error (SMAPE) is a symmetric variation of MAPE and is bounded between 0% and 200% and is useful when the ground truths has small values. Finally Root Mean Squared Error (RMSE) Measures the square root of the average of squared differences between predictions and ground truths.

Table 1. Comparison fully trained and transferred model on counts of target city Greater Manchester

	Fully trained traffic state prediction (count) model	Transfer learning from Luzern to Manchester
Source city	Greater Manchester	Luzern
Amount of data of source city used (for training)	9 months	8.4 months
Target city	Greater Manchester	Greater Manchester
Amount of data of target city used (for training/finetuning)	9 months	23 days
MAE	6.806 cars/5 min	10.674 cars/5 min
MAPE	0.4496	0.9752
SMAPE	0.3406	0.5632
RMSE	52.906	31.8771

These results are promising, as the difference in mean absolute errors between the classic and transfer learning methods is not substantial. Bear in mind that with transfer learning the applied models have seen less historical data of the cities (23 days compared to several months in these tests), in this case Greater Manchester, while they are still able to get comparable results to the classic method. The difference in mean absolute errors is the smallest for speed, which might be caused by the statistical distribution of this data. Even, in extremis, when starting from a trained model on a highway dataset (METR-LA) we still achieve promising results with transfer learning for traffic speeds in the urban area of Manchester.

Table 2. Comparison fully trained and transferred model on speeds of target city Greater Manchester

	Fully trained traffic state prediction (speed)	Transfer learning from Los Angeles to Manchester
Source city	Greater Manchester	Los Angeles
Amount of data of source city used (for training)	9 months	3 months
Target city	Greater Manchester	Greater Manchester
Amount of data of target city used (for training/finetuning)	9 months	23 days
MAE	1.929 mph	1.928 mph
MAPE	0.1667	0.1556
SMAPE	0.1881	0.1965
RMSE	5.198	6.7326

3.2 Case Study of Lisbon

Another use case in the TANGENT project focuses on Lisbon. For this case study, we utilize a dataset containing traffic speed and count data on 8 loop detectors from A5 motorway. The dataset has a 5 min resolution and covers a period of 1 year 9 months, from January 1st 2022, to September 30th 2023. Similar to the Greater Manchester case study, we apply transfer learning techniques using the METR-LA dataset for traffic speed and the UTD19 dataset for traffic count. The preliminary results are presented in Tables 3 and 4, which follow a similar structure as the Greater Manchester case study.

Table 3. Comparison fully trained and transferred model on counts of target city Lisbon

	Fully trained traffic state prediction (count)	Transfer learning from Luzern to Lisbon
Source city	Lisbon	Luzern
Amount of data of source city used (for training)	14.7 months	8.4 months
Target city	Lisbon	Lisbon
Amount of data of target city used (for training/finetuning)	14.7 months	1.6 months
MAE	19.961 cars/5 min	44.561 cars/5 min
MAPE	0.1636	0.7584
SMAPE	0.1412	0.3569
RMSE	31.626	58.594

The preliminary results show that the transfer learning model for traffic speed is working a bit better than the fully trained prediction models. However, it performs poorly with the traffic count predictions model in Lisbon. We observed that the mean daily count for Luzern is 10.958 cars per day, while it is 49.751 cars per day for Lisbon. The mean daily count for Manchester is 10.818 cars per day which can explain superior accuracy in Manchester. This is due to the fact that the Lisbon dataset only contains locations on a motorway, while Luzern and Manchester contain a diverse set of various road types. Section 3.3 explains future research directions how we tend to tackle this problem.

Table 4. Comparison fully trained and transferred model on speeds of target city Lisbon

	Fully trained traffic state prediction (speed)	Transfer learning from Los Angeles to Lisbon
Source city	Lisbon	Los Angeles
Amount of data of source city used (for training)	14.7 months	3 months
Target city	Lisbon	Lisbon
Amount of data of target city used (for training/finetuning)	14.7 months	1.6 months
MAE	5.941 kph	5.556 kph
MAPE	0.1338	0.1550
SMAPE	0.0950	0.0889
RMSE	11.723	10.437

3.3 Future Research Towards CitCom.ai

The preliminary tests performed in the TANGENT project showed promising results, which will be further investigated in CitCom.ai. Many Flemish cities have now taken steps towards a data-driven policy. The city of Bruges, for example, has collected a lot of traffic count data during the VLOED projects. Also, the city of Genk has a strong focus on data, and captures a lot of real-time data such as parking occupancy in car parks and on the street with a scanning vehicle. The city of Mechelen, as a medium-sized city, also aspires a digital transformation but currently lacks the data for traffic state prediction. The preliminary results of the Manchester use case in TANGENT show that the data of Bruges and Genk could be used for traffic state predictions in Mechelen, through transfer learning.

Where the focus in TANGENT was on motorized traffic, the focus in CitCom.ai will be on active modes. Incorporating active modes such as walking and cycling into traffic prediction models presents unique challenges and opportunities. Active modes contribute significantly to urban mobility but have different patterns and requirements compared to motorized traffic. Accurate prediction of these modes can enhance traffic management by providing a holistic view of urban movement, thereby improving infrastructure planning and safety measures. The integration of crowd dynamics further enriches the predictive capabilities, addressing the complexities of human movement in dense urban areas. The following research questions will be tackled in the CitCom.ai project:

- How to incorporate data from multiple cities?
- Which external features can improve the predictive model (e.g., weather, public transport data, cashless payments, etc.)?

- How can we maintain the balance between feature relevance and model generalizability? How about generalizability that might get lost by taking too many details in external features into account?
- As the tests regarding transfer learning in the TANGENT project were preliminary, a complete validation with e.g., bin statistics, visualizations, etc. was not included. To further investigate transfer learning, these insights are however required to give a better perspective towards small and medium-sized cities.
- Referring to the issues presented with transferring to Lisbon in Sect. 3.2, how to adapt the methodology to work with scenarios where the source dataset has a different distribution than the target dataset? We will investigate domain adaptation methods such as feature alignment or domain-invariant feature learning, as they have shown promising results in previous research.

These research questions resulted from limitations we faced during the analysis of the preliminary results with different distributions, e.g., how to deal with various datasets, how to deal with challenges related to data compatibility, particularly in reference to external features, etc.

4 Conclusion

In summary, research conducted as part of the TANGENT project highlights the feasibility, challenges and benefits of using transfer learning to predict traffic states in data-scarce regions. This promotes more efficient and scalable traffic management solutions, accommodating diverse urban transportation modes and dynamics, and enhancing the overall mobility experience in cities. The method will be further explored for the city of Mechelen in the CitCom.ai project where we will tackle the challenges that we addressed in this paper together with challenges that are common in transfer learning.

Acknowledgements. The open datasets METR-LA and UTD19 (utd19.ethz.ch) have played an important role in the success of the research. This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant No 955273 (TANGENT), from the Digital Europe programme under grant No 101100728 (CITCOM.AI) and from the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” programme.

References

1. Rahmani, S., Baghbani, A., Bouguila, N., Patterson, Z.: Graph neural networks for intelligent transportation systems: a survey. *IEEE Trans. Intell. Transp. Syst.* **24**(8), 8846–8885 (2023). <https://doi.org/10.1109/TITS.2023.3257759>
2. Manibardo, E.L., Lana, I., Del Ser, J.: Deep learning for road traffic forecasting: does it make a difference?. *IEEE Trans. Intell. Transp. Syst.* (2021). <https://doi.org/10.1109/TITS.2021.3083957>.

3. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015). <https://doi.org/10.1038/nature14539>
4. Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: data-driven traffic forecasting. In: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, pp. 1–16 (2018)
5. Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C.: Connecting the dots: multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, in KDD 2020, pp. 753–763. Association for Computing Machinery, New York (2020). <https://doi.org/10.1145/3394486.3403118>.
6. Shang, C., Chen, J., Bi, J.: Discrete graph structure learning for forecasting multiple time series (2021)
7. Bogaerts, T., Bosmans, S., Casteels, W., Hellinckx, P.: Traffic prediction US20240054321A1
8. Wang, L., Guo, B., Yang, Q.: Smart city development with urban transfer learning. *Comput.* (Long Beach Calif) **51**(12), 32–41 (2018). <https://doi.org/10.1109/MC.2018.2880015>
9. Wang, L., Geng, X., Ma, X., Liu, F., Yang, Q.: Cross-city transfer learning for deep spatio-temporal prediction. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence, pp. 1893–1899 (2019)
10. Jin, Y., Chen, K., Yang, Q.: Selective cross-city transfer learning for traffic prediction via source city region re-weighting. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 731–741. ACM, New York (2022). <https://doi.org/10.1145/3534678.3539250>.
11. Huang, Y., Song, X., Zhang, S., Yu, J.J.Q.: Transfer learning in traffic prediction with graph neural networks. In: IEEE Conference on Intelligent Transportation Systems, Proceedings, pp. 3732–3737. ITSC, Institute of Electrical and Electronics Engineers Inc. (2021). <https://doi.org/10.1109/ITSC48978.2021.9564890>.
12. “liyaguang/DCRNN: Implementation of Diffusion Convolutional Recurrent Neural Network in Tensorflow (github.com).”
13. Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W.: Understanding traffic capacity of urban networks. *Sci. Rep.* **9**(1), 16283 (2019). <https://doi.org/10.1038/s41598-019-51539-5>



Evaluating the Impact of Suboptimal HVAC Systems on Control Strategies

Pieter Jan Houben¹(✉), Stef Jacobs², Renzo Massobrio¹, Hossein Tabari¹,
Ivan Verhaert², and Peter Hellinckx¹

¹ M4S, Faculty of Applied Engineering - Electronics-ICT, University of Antwerp,
Groenenborgerlaan 171, 2020 Antwerp, Belgium
[{pieterjan.houben,renzo.massobrio,hossein.tabari,
peter.hellinckx}@uantwerpen.be](mailto:{pieterjan.houben,renzo.massobrio,hossein.tabari,peter.hellinckx}@uantwerpen.be)

² EMIB, Faculty of Applied Engineering - Electromechanical Engineering
Technology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium
{stef.jacobs,Ivan.Verhaert}@uantwerpen.be

Abstract. To find optimal strategies to control heating, ventilation and air-conditioning systems (HVAC), significant progress has been made using data-driven methods like model predictive control and (deep) reinforcement learning, which use simulation models to obtain optimal control strategies. These models simulate the system as if all components behave optimally, although in reality most HVAC systems operate in suboptimal conditions due to faults and degeneration of components. As a result, the generated control strategies are unaware of the suboptimal operation of the system. This discrepancy between the simulated behaviour and the actual behaviour of the system is called the sim-to-real gap. This paper aims to examine the impact of faults in a system on the performance of control methods that were tuned assuming the system operates optimally. A simple case study, a space heating network controlled by a PI controller, is simulated using a physics-based model. In this system, faults are introduced in the form of corrosion at the valves. The system is simulated for different levels of corrosion and different settings for the controller. Lastly, these simulations are used to assess the impact of faults on the optimal settings for the controller. They show that a setting that minimizes energy use for a system working optimally does not minimize energy use for a system that works suboptimally.

1 Introduction

Buildings are responsible for 40% of the global energy use, of which 36% is accounted for by heating, ventilation, and air-conditioning (HVAC) [11]. Therefore, finding optimal strategies to manage HVAC systems in buildings is key in the transition to a more sustainable society. Significant progress has been made in the field of model predictive control (MPC) and (deep) reinforcement learning (DRL) to generate optimal control strategies while maintaining a certain level of comfort [5, 9, 17]. These methods use simulated environments to find control

strategies that optimize the energy use of the system. As a consequence, their performance relies heavily on the accuracy of the underlying simulation models.

Current practice is to use physics-based modelling, in which thermodynamic equations are simulated using time-step methods. To keep calculations feasible, simplifications are necessary, causing the model to simulate the system as if all of its components operate optimally [14]. This causes a discrepancy between the simulated behaviour and the actual behaviour of the HVAC system, called the sim-to-real gap [3]. Moreover, after a while, degradation in the components will reinforce the sim-to-real gap [18]. In recent research on MPC in HVAC, closing the sim-to-real gap was addressed as a future research direction to improve performance and avoid specific tuning for different HVAC systems [2].

This paper aims to examine the impact of the sim-to-real gap on the control strategy of the system by simulating corrosion in a space-heating system and comparing the performance of different settings of the controller for different levels of corrosion.

In Sect. 2, the case study is introduced, as well as the simulation environment and the method of introducing corrosion in the system. Section 3 describes the settings of the simulations that were carried out and discusses the results, showing that different levels of corrosion require other settings to minimize average energy use (AEU). In Sect. 4, it is concluded that a system operating suboptimally does affect the optimal settings for the control algorithm, suggesting that to optimize performance of DRL or MPC for HVAC control, it is necessary to close the sim-to-real gap.

2 Methodology

2.1 Case Study

This paper aims to compare the performance of control strategies for HVAC systems using a standard heating setup as shown in Fig. 1. A geothermal heat pump is connected to a storage tank to provide heat for a dwelling with underfloor heating with design ingoing and outgoing temperature of $35^{\circ}\text{C}/30^{\circ}\text{C}$. The heat pump has a heat power of 2343.5 W and the storage is designed to buffer 1 h of operation of the heat pump. To control the supply temperature, mixing valves are used together with a pump.

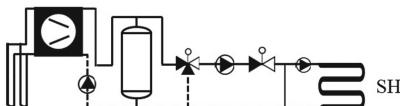


Fig. 1. Setup for the case study. A dwelling is equipped with underfloor heating with a geothermal heatpump and a storage tank.

2.2 Simulation Environment

Because the aim is to compare two control strategies, the setup is simulated in Python, to make sure all circumstances are the same when operating the system. The simulation environment is based on the work of Van Riet [13] and was also used in the context of collective heating systems in [7]. There, the transient thermal behaviour is represented by simulating the temperature y according to Eq. (1) in timesteps of 10 s. More specifically, an explicit solution is simulated where a and b are constant over each timestep, in other words a zero-order hold solution.

$$\frac{dy(t)}{dt} = -a(t)y(t) + b(t) \quad (1)$$

2.3 Occupant Behaviour Profiles

The occupant behaviour profiles are statistical representations of the occupancy of the dwelling. This affects the use of domestic hot water, internal heat gains due to the presence of people in the dwelling and zone temperature set point schedules. For this use case, domestic hot water consumption is not considered. The stochastic profile generator developed within the Instal2020 project was used to generate occupant behaviour [15]. This is based on a survey on 700 dwellings in Belgium. In total there are nine different family types, ranging from 1 to 4 inhabitants [10]. For this case study, an elderly couple was chosen as profile.

2.4 Weather Profiles

The weather files are integrated into the simulation and contain data on outdoor temperatures and solar radiation for each cardinal direction. They are based on weather data from Belgium, taking the average of data between 2001 and 2020 [8]. To ensure a high space heating demand, the first week of the month of January is selected.

2.5 Control and Corrosion

2.5.1 Control

The setpoint for the supply to the storage tank is constant at 37 °C. The temperature setpoint in the dwelling is 21 °C during the day and 19 °C during the night. After the storage tank there is a mixing valve to control the supply temperature and finally a passive mixing controlled by a pump for the underfloor heating. The valve opening θ is determined by the temperatures of the supply flow coming from the storage tank T_1 and the return flow T_2 and the supply temperature setpoint T_{SP} as in Eq. (2).

$$\theta = \frac{T_{SP} - T_2}{T_1 - T_2} \quad (2)$$

Finally, the pump is controlled using a PI controller. This is a standard type of controller that adjusts the system not only based on the current error between the set point and process value, but also on the integral of past errors [6]. In short, for a temperature set point $T_{SP}(t^*)$ and inside temperature $T(t^*)$ at time t^* , the reaction of the system is given by Eq. (3).

$$u(t^*) = K_p(T_{SP}(t^*) - T(t^*)) + K_I \int_{0 \leq t \leq t^*} T_{SP}(t) - T(t) dt \quad (3)$$

where K_p , also referred to as gain, and K_I are constants characterising the controller. Equation (3) is often reformulated as in Eq. (4), where the parameter T_n is used to define the weight for the integral term as $K_I = \frac{K_p}{T_n}$. Although T_n is often referred to as the replay time, it is dimensionless, as is K_p .

$$u(t^*) = K_p \left((T_{SP}(t^*) - T(t^*)) + \frac{1}{T_n} \int_{0 \leq t \leq t^*} T_{SP}(t) - T(t) dt \right) \quad (4)$$

Tuning the control algorithm of the system means setting the values for K_p and T_n .

2.5.2 Corrosion

To simulate a system working suboptimally, corrosion at the valves is introduced into the simulation model. In [12] and [16], corrosion in valves was found to reduce the flow rate. Since the simulation was done over the course of one week and corrosion develops over the scale of years, it is a reasonable assumption that the flow rate through the valves is reduced by a constant factor throughout the whole simulation [4]. Thus, for a certain level of corrosion c , the mass flow rate through the valve, \dot{m} , is given by:

$$\dot{m} = \dot{m}_{ideal}(1 - c), \quad (5)$$

where \dot{m}_{ideal} is the mass flow rate through the valve in ideal conditions. The goal of this paper is to examine the impact of corrosion on the optimal values for K_p and T_n .

3 Experiments and Discussion

3.1 Setup

The goal of this paper is to tune the PI controller of the system for various levels of corrosion. Based on the findings in [4], the corrosion levels for which simulations were carried out were

$$c = 0, 0.1, 0.2, 0.3.$$

Average energy use is the main KPI to assess the performance of the controller. Based upon the recommendations of [1], and taking the simulation time into account, the simulations were carried out for

$$K_p = 0.5, 1, 2, 3$$

$$T_n = \{ 0.5 \cdot k \mid k \in [1, 28] \}.$$

To ensure that the system would be operating thoroughly during simulation, an elderly couple occupying the dwelling in the first week of January was chosen as occupation model and simulation moment.

3.2 Results

3.2.1 Average Energy Use

Figure 2 shows for each level of corrosion the average energy use of the system over the course of the simulation. Figures 2(a) and 2(b) show that for higher levels of corrosion, $K_p = 0.5$ gives the best results in terms of energy use for all values of T_n . In Figs. 2(c) and 2(d), a value of $K_p = 0.5$ does not perform the best over all values of T_n , but the minimum average energy use is not worse than the minimum average energy use for other values of K_p , suggesting that $K_p = 0.5$ is a good value for all values of corrosion.

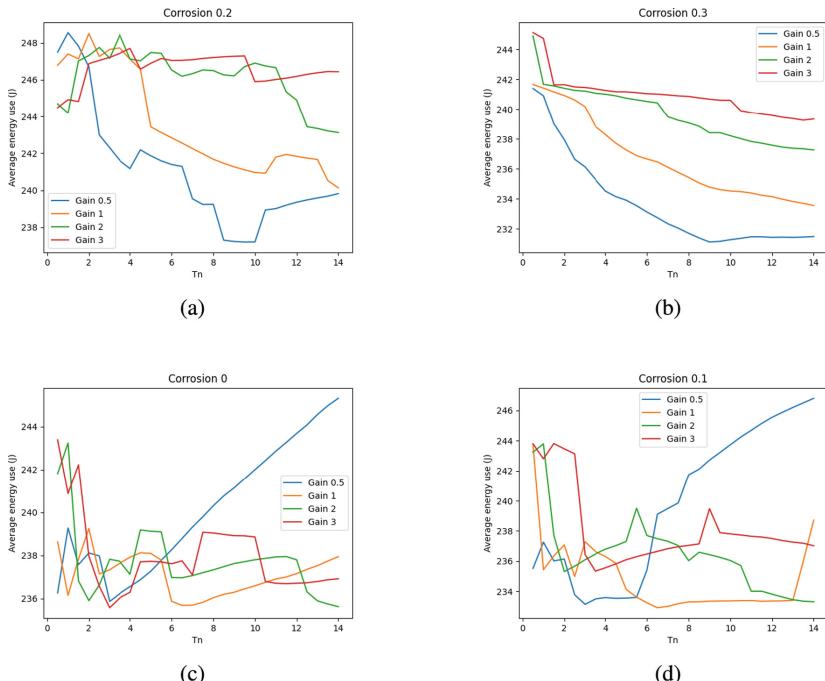


Fig. 2. Average energy use for different levels of corrosion and different setting for the controller.

Figure 3 shows the average energy use for $K_p = 0.5$ for each level of corrosion. For low levels of corrosion, the optimal value for T_n lies around $T_n = 3$, whereas for higher levels of corrosion it shifts more to a value around $T_n = 9$. Moreover, the optimal value of $T_n = 3$ for the ideal system results in a much higher energy use for systems with higher levels of corrosion, showing that a control strategy that is optimal for a system that operates optimally is not necessarily optimal for a system that works under suboptimal circumstances.

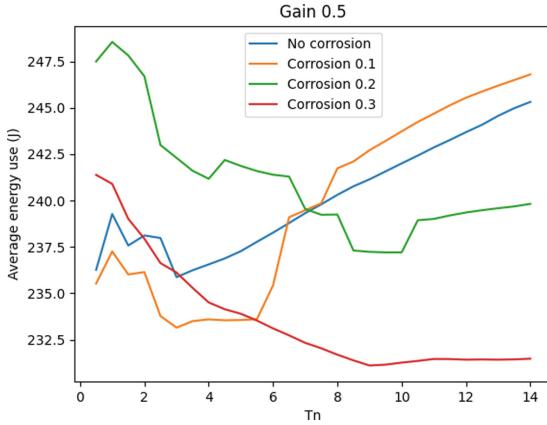


Fig. 3. Average energy use vs T_n for $K_p = 0.5$.

3.2.2 Room Temperature Lack

In Sect. 3.2.1, the average energy use was used as the only KPI for the performance of the system. With an eye for sustainability, this is of course reasonable, but since the system's purpose is to heat a room to a certain temperature, the performance on reaching that set point cannot be ignored. The room temperature lack, given in Eq. (6), provides a measure for this:

$$RTL(t^*) = \int_{0 \leq t \leq t^*} (T_{SP} - 0.5 - T)_+ dt. \quad (6)$$

Here, the difference between the temperature set point T_{SP} and the inside temperature T is integrated, but only when T is lower than T_{SP} and tolerating a difference of 0.5°C , since the system is only meant to heat the dwelling. Therefore it differs from the integral term in (4).

Figure 4 shows the room temperature lack per day versus T_n for different levels of corrosion. It can be interpreted as the amount of hours per day the temperature is exactly 1°C below the setpoint. Obviously, corrosion has a negative impact on the room temperature lack. Also, a lower value for T_n drastically improves the performance of the system. In general, there is no significant difference in behaviour for different levels of corrosion. However, the room temperature

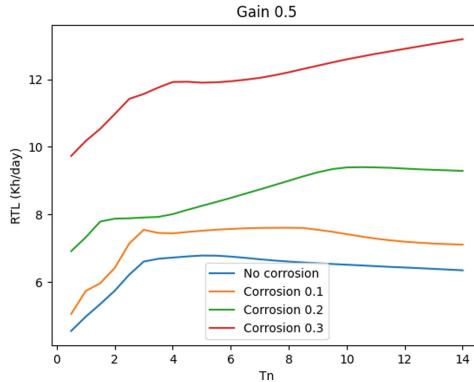


Fig. 4. Room temperature lack per day for different values of corrosion

lack, although different, is strongly related to the integral term of the PI controller. The importance of this term is directly determined by a factor $\frac{1}{T_n}$, as can be seen in Eq. (4). Therefore, it is no surprise that the behaviour is the same for different levels of corrosion.

3.2.3 Combining AEU and RTL

Figure 5 shows the sum of the AEU and RTL for different levels of corrosion. In order for the sum to be meaningful, the results were normalised using min-max normalisation. As in Sect. 3.2.1, a different behaviour can be observed for different levels of corrosion. Due to the RTL term, a lower value of T_n is generally preferred, but for higher levels of corrosion, the AEU term causes the value of T_n to shift more to the right, again showing that the optimal setting for the optimally operating system is not optimal for a system operating under suboptimal conditions.

3.2.4 Note on Versatility of a PI Controller

Looking at Figs. 2 and 3, it can be noted that the energy gained by changing the settings of the controller, although not negligible, is not significantly large. This is because a PI controller is not very versatile in its policies, due to its simplicity. However, the aim of this paper is not to reduce the energy use of HVAC systems by tuning a PI controller, but to examine the impact of a fault, in this case study corrosion, on the optimal settings of a control algorithm. The fact that for an versatile control model such as a PI controller a significant impact can be observed, suggests that for versatile methods such as DRL and MPC, the effect will be even more significant.

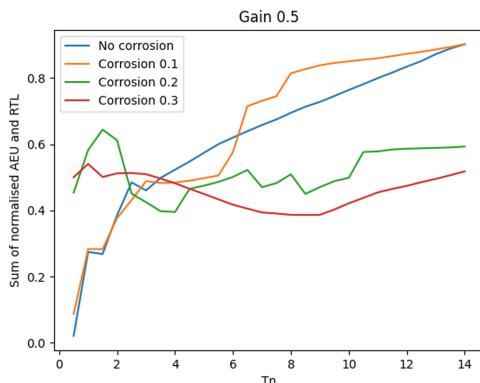


Fig. 5. Sum of normalised AEU and RTL vs T_n for different levels of corrosion.

4 Conclusion

This study investigates the impact of corrosion on the optimal tuning parameters of a PI controller that operates an HVAC system. The results demonstrate that the optimal settings for a PI controller differ substantially between systems operating under ideal conditions and those operating under suboptimal conditions. Specifically, for systems with higher corrosion levels, the tuning parameter T_n needs to be adjusted to minimize energy use and maintain efficiency.

The study underscores the limitations of the PI controller's versatility, suggesting that while significant impacts are observed, more advanced control strategies such as Deep Reinforcement Learning (DRL) and Model Predictive Control (MPC) could potentially adapt more effectively to suboptimal system conditions. This can only be achieved by closing the sim-to-real gap for the underlying simulation model.

In conclusion, as more advanced and flexible control strategies show improved performance for HVAC systems, calibrating them on simulations that grasp the imperfections of the system will improve their performance even further. Future research should therefore aim to develop methodologies to close the sim-to-real gap for these simulation models. This approach will be crucial for the sustainable and efficient operation of HVAC systems in real-world scenarios.

References

1. Åström, K., Hägglund, T.: PID Controllers: Theory, Design, and Tuning. ISA - The Instrumentation, Systems and Automation Society (1995)
2. Bamdad, K., Mohammadzadeh, N., Cholette, M., Perera, S.: Model predictive control for energy optimization of HVAC systems using energyplus and ACO algorithm. *Buildings* **13**(12), 3084 (2023)
3. Chong, A., Gu, Y., Jia, H.: Calibrating building energy simulation models: a review of the basics to guide future work. *Energy Build.* **253**, 111533 (2021)

4. Cole, I., Marney, D.: The science of pipe corrosion: a review of the literature on the corrosion of ferrous metals in soils. *Corros. Sci.* **56**, 5–16 (2012)
5. Du, Y., et al.: Multi-task deep reinforcement learning for intelligent multi-zone residential HVAC control. *Electr. Power Syst. Res.* **192**, 106959 (2021)
6. Franklin, G., Powell, J., Emami-Naeini, A.: *Feedback Control of Dynamic Systems* (1994)
7. Jacobs, S., et al.: Grouped charging of decentralised storage to efficiently control collective heating systems: limitations and opportunities. *Energies* **16**(8) (2023)
8. Klein, S., et al.: Mathematical reference: Type 54 (hourly weather data generator) (2009)
9. Li, J., Zhang, W., Gao, G., Wen, Y., Jin, G., Christopoulos, G.: Toward intelligent multizone thermal control with multiagent deep reinforcement learning. *IEEE Internet Things J.* **8**(14), 11150–11162 (2021)
10. Schutter, J., Verhaert, I., De Pauw, M.: A methodology to generate realistic random behavior profiles for space heating and domestic hot water simulations (2018)
11. UNEP. Global status report for buildings and construction: Towards a zero-emission, efficient and resilient buildings and construction sector. Technical report (2021)
12. Valves, C.: *Flow of Fluids Through Valves, Fittings Pipe: Technical Paper 410 Metric Version*. Vervante (1998)
13. Van Riet, F.: *Hydronic Design of Hybrid Thermal Production Systems in Buildings*. University of Antwerp (2019)
14. Vera-Piazzini, O., Scarpa, M.: Building energy model calibration: a review of the state of the art in approaches, methods, and tools. *J. Build. Eng.* **86**, 108287 (2024)
15. VLAIO. Instal project: integraal ontwerp van installaties voor sanitair en verwarming (dutch). VIS 13589, 2014–2018
16. Winston, R.R., Herbert, U.: Corrosion and corrosion control (2006)
17. Yu, L., et al.: Multi-agent deep reinforcement learning for HVAC control in commercial buildings. *IEEE Trans. Smart Grid* **12**(1), 407–419 (2021)
18. Zhao, Y., Li, T., Zhang, X., Zhang, C.: Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future. *Renew. Sustain. Energy Rev.* **109**, 85–101 (2019)



AI for Anticipating Human Behavior

Jeoffrey Canters, Pieter Jan Houben^(✉), Renzo Massobrio, and Peter Hellinckx

University of Antwerp, Antwerp, Belgium

jeoffreycanters@icloud.com,

{pieterjan.houben,renzo.massobrio,peter.hellinckx}@uantwerpen.be

Abstract. Automated systems are increasingly integrated into our daily lives, streamlining various tasks and enhancing convenience. Despite their careful design to improve our everyday experiences, problems still occur, often due to human interaction with these systems. This paper addresses the challenges posed by human impact on autonomous systems, aiming to predict and mitigate errors caused by such interactions. By incorporating human behavior into the training process, we hypothesize that the trained agent's ability to anticipate and withstand these behaviors will improve. Leveraging artificial intelligence (AI) and reinforcement learning (RL) in particular, a controller is developed for automated processes designed to anticipate human impact and minimize errors. We differentiate irrational human behavior into two categories: short-term irrationality and long-term irrationality. This research focuses on the short-term irrational behaviors as a manageable subset. To address the lack of data on irrational human behavior, we define an irrational model within a straightforward environment to evaluate RL's ability to recognize and anticipate such behavior. The environment used is the card game UNO, because of its simple rules and emphasis on player interaction. Results reveal a notable 0.4% improvement in win rate for the anticipating controller compared to the rational controller when playing against the human agent. Additionally, the anticipating controller achieves a 51% win rate against the rational controller, demonstrating its ability to match performance in a rational context.

1 Introduction

Automated processes are a common aspect of everyone's daily lives. These systems are currently designed within an idealized environment, not taking irrational human decisions into account. Autonomous vehicles for example, cannot predict that an abrupt maneuver might irritate other drivers if it has only been trained in an environment where all vehicles follow the rules. Irritated drivers may react unpredictably, so it would be beneficial for the autonomous vehicles to be prepared for such scenarios. This is something we want to address by integrating irrational human behavior into the training process, expecting that this will enable the controller to better navigate and adapt to human behavior, while maintaining the same level of performance as the current controllers. This behavior is influenced by various factors such as stress and emotions [4, 11, 13].

We differentiate irrational human behavior into two categories: short-term irrationality and long-term irrationality. This study focuses on short-term irrational behavior as a manageable subset. To evaluate our hypothesis, we chose the card game UNO as our test environment due to its clear rules and emphasis on player interaction. Its simplicity and focus on interpersonal dynamics provide an ideal setting to explore human decisions and the ability to anticipate these decisions. To create an anticipating controller, reinforcement learning (RL) [12] is employed. Due to a lack of data on UNO-games playes by human players, an irrational playing agent is built using a rule-based strategy. Using RL and self-play, an agent is trained to play the game in a setting without irrational decisions. It is therefore labeled as a rational agent. The rational and irrational model then form the building blocks of a human model. Lastly, this human model is used to train the anticipating model, thereby incorporating human irrational behavior into the training process.

Section 2 provides an overview of the methodology. Section 3 explains the model's application and experimental setup. Section 4 presents the results and analysis derived from the conducted experiments. Finally, Sect. 5 offers conclusions drawn from the study and proposes avenues for future research.

2 Methodology

The general setup for this research involves an environment in which irrational and rational decisions are made and a RL controller that learns to handle these irrational decisions by encountering them during the training process. RL, a mathematical framework used by agents to optimize cumulative rewards through environmental interactions [12], forms the basis of this approach. The irrational and rational decisions are simulated by introducing other agents into the environment who are trained to take these decisions. These agents are fundamentally rational but are trained to occasionally make irrational decisions, simulating the emotional reactions seen in humans when interacting with systems. The main objective of this research is to evaluate this methodology using a straightforward use case, namely UNO, as elaborated further in Sect. 3.

3 Case Study

3.1 UNO as Case Study

To evaluate our hypothesis, we chose the card game UNO as our test environment due to its clear rules and emphasis on player interaction. Its simplicity and focus on interpersonal dynamics provide an ideal setting to explore human influences and the ability to anticipate these influences. The goal is to be the first player to discard all cards. The deck includes numbered cards (0–9 in red, blue, green, and yellow), action cards (skip, reverse, draw 2 in the same colors), and wild cards (change color and draw 4).

The predefined UNO environment of the RLCard [15] framework is used. RLCard provides card game environments, pre-implemented agents, training algorithms, and evaluation tools. A pre-implemented RLCard agent used in this research is a rule-based (RB) model. This model follows a simple strategy. First, if the agent possesses a “Wild Draw 4” or “Change color” card, it immediately plays it, selecting the most common color in its hand. Otherwise, the agent randomly chooses an action from its legal options.

3.2 Rational Agent

The rational agent is trained with the deep Q-network (DQN) algorithm due to its capability to manage large and intricate state and action spaces [8,9]. In this research, three different opponents are used during training. The first opponent consistently makes random moves. This serves as a baseline, enabling the agent to adapt and develop its own tactics in response to unpredictable gameplay. The second opponent is the pre-implemented rule-based model mentioned in Sect. 3 that performs very well against various other agents and rule-based models. The final opponent is itself, a concept commonly referred to as self-play [3,10]. This method worked well in the past for other applications in games and doubles the experience since the opponent is also training.

For each training process, the same reward function and state space are utilized. The reward function is straightforward: the agent earns a point every time it wins and incurs a penalty point for every loss. The state space includes the target card, whose turn it is, the cards in your hand, and the previously played cards.

Additionally, a rule-based model is developed to evaluate the agents’ performance against a weak opponent, which avoids playing action or wild cards and selects the least frequent color when playing a “color change” card. This comparison helps assess the agents’ effectiveness against a less competitive player.

3.3 Irrational Agent

Due to a lack of data on UNO players’ behaviour, a RB model was developed that simulates short-term irrational behavior, which is representative due to the simplicity of the game. Emotions, particularly anger, significantly impact short-term irrationality in games, where frustration and irritation are common. We model anger within UNO to approximate real-life irrational behavior.

According to [6], individuals experiencing anger tend to favor high-risk, high-reward options in decision-making, reflecting the irrational agent’s propensity for employing such strategies. Anger-driven individuals typically exhibit overconfidence in their abilities and perceive a heightened sense of control over the situation [7]. Additionally, angry individuals tend to seek revenge on other individuals [2]. In UNO, this manifests through actions like aggressively playing Draw 4 or Reverse cards to disrupt opponents’ progress and gain a competitive edge.

The developed RB model employs a straightforward strategy: it plays an action or wild card when available, and otherwise selects a random legal action.

We have implemented a scalability feature for the irrationality of this agent, known as the irrationality factor (η). This parameter determines the probability of the agent making an irrational action. When the irrationality factor approaches 1, the agent is highly likely to make irrational decisions. Conversely, when it approaches 0, such decisions are less probable. This parameter was introduced to allow for variability in the level of irrationality exhibited by the irrational agent.

3.4 Human Agent

The human agent is a RB model that switches between the rational agent and irrational agent. When to switch depends on specific scenarios where a human player becomes angry or frustrated. According to cognitive consistency theory, anger arises from discrepancies between expectations or desires and actual experiences, often prompting aggressive responses to restore consistency [1,5,14]. In UNO, players typically aim to win using rational strategies, but unexpected outcomes can lead to frustration and irrational behavior. Thus, the specific scenarios triggering anger include:

- A player receives a Draw 4 card
- A player has to skip two or more turns consecutively due to receiving Skip cards
- A player has one card left but is then obliged to draw extra cards
- The difference in the number of cards between players exceeds five

When anger is triggered, the human agent will utilize the irrational agent to make irrational decisions. However, the degree of irrationality varies depending on the situation. The irrationality factor, as discussed earlier and initialized by equation (1), modulates this behavior. Specifically, when a player has fewer cards than their opponent, they are less likely to make irrational decisions compared to when they have more cards.

$$\eta = \begin{cases} \min\left(1, \frac{1}{2} + \frac{\Delta}{20}\right) & \text{if } \Delta > 0 \\ \max\left(0, \frac{1}{2} + \frac{\Delta}{20}\right) & \text{if } \Delta \leq 0 \end{cases} \quad (1)$$

where:

- η represents the irrationality factor
- Δ represents the card difference, calculated as your cards minus your opponent's

After triggering anger and initializing an irrationality factor, it is necessary to determine how long this human agent behaves irrationally. To manage this, we implemented a cooldown system where the irrationality factor decreases by 0.1 each round until it falls below zero. Once it drops below zero, the rational agent resumes decision-making duties.

3.5 Anticipating Agent

The anticipating agent, like the rational agent, is trained using the DQN algorithm. This agent also employs the same reward function, state space, and hyperparameters to ensure fair comparison and consistency in evaluation.

4 Discussion

Table 1 shows the win rates of the three trained agents: Agent A (trained through self-play), Agent B (trained against a rule-based agent) and Agent C (trained against a random agent). The table presents the win rates over 1000000 games against a rule-based agent, random agent and bad agent. The agents' win rates do not differ significantly, although the self-play trained agent slightly outperforms the others due to it receiving twice as much training data. The poorest performing agent is the one trained against a random agent, as it has been trained against a player with no strategy. The agents achieve a win rate of approximately 52% when matched against a highly skilled opponent, and about 59% when facing an opponent employing the least effective strategy possible. There is a 7% difference in win rates between playing against skilled versus unskilled opponents. When comparing proficient agents among themselves, it becomes evident that they have to be analysed on a different scale.

Table 1. Win rates of three different agents when competing against a rule-based agent and a random agent

Agent	WR against RB agent	WR against random agent	WR against bad agent
DQN Agent A	0.5268	0.5756	0.5930
DQN Agent B	0.5245	0.5714	0.5886
DQN Agent C	0.5209	0.5683	0.5858

Figure 1 illustrates the win rate distribution for various agents against different opponents. It shows that a trained agent can achieve a maximum win rate of 60% against an opponent employing the worst possible strategy. This indicates that even with intentional poor play, one can still win 40% of the games due to luck. All win rates fall between 40% and 60%, representing a narrow margin in which improvements can be made. This narrow margin suggests that a different scale should be used. Instead of analyzing and comparing agents on a scale of 100%, they should be evaluated on a scale of 20%.

The anticipating agent was trained against the human agent using the same reward function and hyperparameters. However, the rational agent, benefiting from self-play training, received double the training data. To ensure a fair comparison, the anticipating agent underwent 100,000 episodes of training, twice that of the rational agent. Figure 2 shows the reward progression during its training.



Fig. 1. Winrate distribution of different agents versus different opponents. (Color figure online)

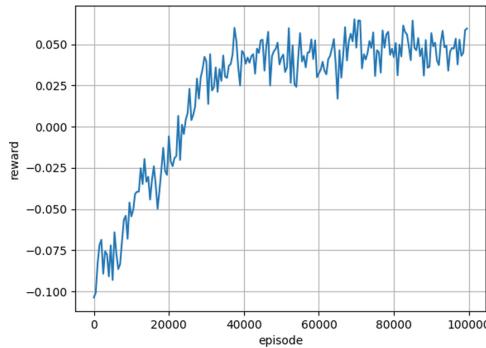


Fig. 2. Reward progression over 100,000 episodes

Table 2 shows the win rates over 1000000 games of the rational agent and the anticipating agent against both the rational and human agents. The anticipating agent shows approximately a 0.4% higher win rate against both irrational and human opponents compared to the rational agent. This 0.4% performance boost comes down to an improvement of 2% when evaluated over a scale of 20%. This indicates that when comparing proficient agents, there is very little room for improvement. Thus, while the improvement appears marginal, it represents notable progress. Against the rational agent, the anticipating agent achieves a win rate of approximately 51%, demonstrating its effectiveness in anticipating and countering human behavior while maintaining rational gameplay.

Table 2. Win rates of three different agents when competing against a human agent and an irrational agent

Agent	WR against human agent	WR against rational agent
Rational Agent	0,5221	0,5002
Anticipating agent	0,5262	0,5081

5 Conclusions

Automated systems, while beneficial, can encounter errors due to human interaction. This research addresses this challenge by focusing on human influence within a simplified environment using the card game UNO. We modeled both irrational human behavior and rational decision-making to develop a human controller capable of mimicking both. Training a new controller against this model aimed to enhance its resilience to human influences and match the performance of rational controllers. Our findings show a 0.4% performance improvement in the anticipating agent compared to the rational agent, indicating its better management of human-like irrational behavior. These results underscore the importance of integrating human behavior into training processes to bolster resilience. Despite the modest performance gain, which reflects the significant role of luck in UNO, these results represent a promising step toward integrating irrational behavior into the training process.

Future studies can build upon these initial experiments by applying them in environments with readily available human influence data. This would provide deeper insights into how the anticipating controller performs when confronted with real human behavior. Additionally, exploring a variety of algorithms could help identify whether there are more effective methods for anticipating irrational behavior. The overarching goal is to translate these research findings into practical applications for critical real-world scenarios, such as autonomous vehicles.

References

1. Archer, J.: The organization of aggression and fear in vertebrates
2. Barber, L., Maltby, J., Macaskill, A.: Angry memories and thoughts of revenge: the relationship between forgiveness and anger rumination. *Pers. Individ. Differ.* **39**, 253–262 (2005)
3. Chen, Z., Deng, Y., Yuan, H., Ji, K., Gu, Q.: Self-play fine-tuning converts weak language models to strong language models (2024)
4. Duque, A., Cano-López, I., Puig-Pérez, S.: Effects of psychological stress and cortisol on decision making and modulating factors: a systematic review. *Eur. J. Neurosci.* **56** (2022)
5. Hebb, D.: The organization of behavior a neuropsychological theory
6. Leith, K.P., Baumeister, R.F.: Why do bad moods increase self-defeating behavior? Emotion, risk taking, and self-regulation. *J. Pers. Soc. Psychol.* **71**, 1250–1267 (1996)

7. Litvak, P.M., Lerner, J.S., Tiedens, L.Z., Shonk, K.: Fuel in the Fire: How Anger Impacts Judgment and Decision-Making, pp. 287–287–310. Springer, New York (2010)
8. Mnih, V., et al.: Playing Atari with deep reinforcement learning (2013)
9. Mnih, V., et al.: Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015)
10. Silver, D., et al.: Mastering chess and shogi by self-play with a general reinforcement learning algorithm (2017)
11. Starcke, K., Brand, M.: Decision making under stress: a selective review (2012)
12. Sutton, R.S.: Introduction: The Challenge of Reinforcement Learning, pp. 1–3. Springer, Boston (1992)
13. van der Pligt, J.: Decision making, psychology of. In: International Encyclopedia of the Social Behavioral Sciences, pp. 3309–3315 (2001)
14. van Kampen, H.S.: The principle of consistency and the cause and function of behaviour (2019)
15. Zha, D., et al.: RLCard: a toolkit for reinforcement learning in card games (2019)



Mamdani Type-1 Non-singleton Fuzzy Logic System (T1 NSFLS) for a Quality Control Process Based on Industrial Image Processing

Pascual Noradino Montes-Dorantes^{1,2}, Adriana Mexicano-Santoyo^{1(✉)},
Jesús C. Carmona-Frausto¹, and Gerardo Maximiliano Mendez³

¹ Tecnológico Nacional de México, Instituto Tecnológico de Ciudad Victoria, 87010 Victoria City, México

{pascual.md, adriana.ms, jesus.cf}@cdvictoria.tecnm.mx

² Tecnológico Nacional de México, Instituto Tecnológico de Saltillo, 25280 Saltillo, México

³ Tecnológico Nacional de México, Instituto Tecnológico de Nuevo León, 67170 Cd. Guadalupe, N. L, México

gerardo.m@nuevoleon.tecnm.mx

Abstract. This paper presents an application of the Mamdani type-1 non-singleton fuzzy logic system (T1 NSFLS) for a quality control process based on industrial image processing. The proposed application is used in a cutting process to obtain support plates for picture framing, the width and high of the plate are needed to obtain the quality parameters. In this system, the uncertainty is filtered by the inputs that are treated as fuzzy numbers (FN) instead of a crisp number, the inputs whose values came from the image sensor. This process happens in the feed forward and the adjustment of the error is made in the backward pass of the fuzzy system. The FN are used to find the dispersion of the corrupted measurements via the standard deviation to obtain an adjusted crisp value to make production decisions about the quality of the product. The results show that the proposed model obtains a precision of 91% when the uncertainties are in the range of one standard deviation.

1 Introduction

This paper focuses on the modelling of a non-singleton (NS) fuzzy system based on Mamdani model, the system works with a crisp consequent and the solution is obtained by the fuzzy basis function (FBF) and their product with the firing level of every rule. The NS presents a way to deal with the uncertainty by an adjustment provided in the fuzzification, which is once again adjusted in the training. The literature only presents three types of membership functions (MF's) for T1 NSFLS. Since the uncertainty can happen on both sides of the function the MF needs to be symmetrical. The noise of measurements should be equivalent in all points of the MF as mentions [1], the MF used in this case being (1) as recommended [2]. The main difference in the type of fuzzification is that in singleton fuzzification, Fig. 1(a) the input is represented by a single point of

the universe of disclosure. On the other hand, in the NS the input is treated as a fuzzy number (fuzzy set), see Fig. 1(b),

$$\mu_{x_i} = e^{-\left((x_i - x'_i)^2 / 2\sigma^2\right)} \quad (1)$$

where: x_i represents the center of fuzzy set, x'_i represents the input, σ and c represents the spread of the fuzzy sets (FS) and i represents the number of variables.

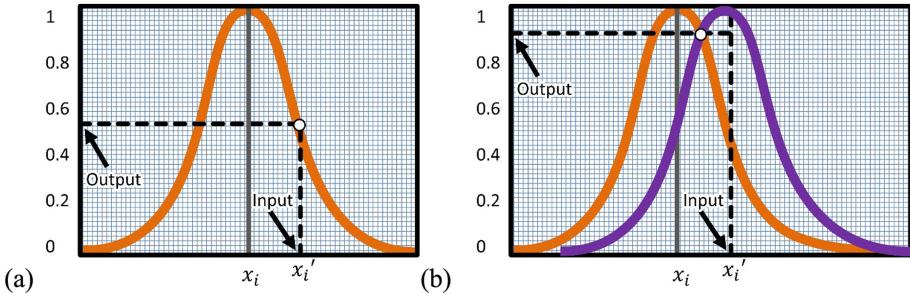


Fig. 1. Fuzzification: Singleton model (a), non-singleton model (b).

In the evaluation process of a fuzzy system, there are the FBF's that achieve the output fuzzy values for every rule by combinations of FS after the implication (T-norm or T-conorm). These FBF's cannot process the uncertainty, which is processed in the training phase via iterations; that uncertainty comes from the input data that can be treated partially with the use of FN as inputs. The final evaluation of a FBF gives the fuzzy output of the overall system for a test data pair.

A literature survey shows the existence of a couple theoretical propositions [2, 3] and some successful cases of real-world applications of this method [4–15]. The T1 NSFLS did not have the same level of development of its singleton counterpart (T1 SFLS). On the other hand, for this model, the inputs are represented as type-1 non-singleton FN, these inputs have a deviation from the real value or noise in the measurements which corrupts the data and their membership values. The literature presents some cases of application of the type-1 (T1) NS model as well, such as: the use of the NS inputs [13, 15–20]; the NS model is also used for tuning [21, 22]; comparing [8]; classifying [22]; predicting [4]; simulating [23]; designing the interval type-2 model [24] and the use of the interval type-1 (IT1) to handle the uncertainties precedent to the inputs [25]; it is also used to predict, or as a reference kernel [26], signal processing [10], finally in literature the NS appears as a mathematical condition. And in the other hand, the interest for this technology has not been much in the past since 1997, when it was created, however in recent times some interest exists, as mentioned in [27] in a few papers that include the models in type-1 e.g. [28] but, these are applied in type-2 [29–31] and type-3 [32–34] with type-1 inputs among others.

Section 1 presents the introduction, Sect. 2 presents the theoretical background of T1 NSFLS, while Sect. 3 presents the method used to create and operate the system, Sect. 4 presents the experimental results, Sect. 5 presents the assemble of the proposal and Sect. 6 presents the conclusions.

2 Theoretical Backgrounds

Basically, in the T1 NSFLS before the fuzzification the process is the same as the T1 SFLS, the values of the firing rules are obtained by (2) in the Mendel model [1], which are decomposed in (3, 4) where (3) is the fuzzification of the X_k and represents the input i . The Eq. (4) represents the k^{th} rule,

$$\mu Q_k^l(x_k) \equiv \mu_{xk}(x_k) * \mu F_k^l(x_k) \quad (2)$$

$$\mu_{xk}(x_k) = e^{-\left((x_k - m_{xk})^2 / 2\sigma_x^2\right)} \quad (3)$$

$$\mu F_k^l(x_k) = e^{-\left((x_k - mF_k^l)^2 / (2\sigma_F^2 F_k^l)\right)} \quad (4)$$

where: $k = 1, \dots, p$, and $l = 1, \dots, M$; x_k is the input of k variable; and x_k is the input; m_{xk} is the mean of fuzzy set for the antecedent x_k and mF_k^l represents the mean of fuzzy set in the rule l . σ_x is the standard deviation of the x_k , and σ_F is the standard deviation of the set in the rule. In the case of T1 NSFLS an evaluation is added to adjust the values corrupted by noise, this needs to calculate the deviation of the real values of x_i , given by (5) and this is equivalent to (6), this deviation is given by (7). Once obtained the deviation called σ_k^l can be performed the necessary calculus to obtain the supremum value called: $\mu Q_k^l(x_k, max)$ that yields (8),

$$\delta_{xi} = |x_i - \hat{x}_i| \forall x' \in x \quad (5)$$

where: x_i represents a measurement provided by the sensor and \hat{x}_i represents the measurement of the specification.

$$\delta_{xi} = |\mu F_k^l(x_k) - \mu_{xk}(x_k)| \forall x' \in x \quad (6)$$

$$\sigma_{xi} = \sum_{i=1}^n \delta_{xi} = \delta_{xi}/n \forall x_i \in x \quad (7)$$

$$\mu Q_k^l(x_{k,max}) = e^{-\left(\left(m_{xk} - mF_k^l\right)^2 / \left(2\left(\sigma_k^2 - \sigma_F^2\right)\right)\right)} \quad (8)$$

With the value μQ_k^l can generate the secondary evaluation with the t-norm to adjust the values of the inputs and their MF's with (9). From this point the rest of the model is the same as a T1 SFLS.

$$\mu Bi = \mu Gi * \left[\prod_{i=1}^n \mu Q_i(\max(xi)) \right] \quad (9)$$

3 Methodology

The method for industrial image processing requires an adaptation process to generate the evaluation in a single layer matrix (gray scale image). This occurs because most of the digital cameras generate images in color with a red, green, and blue (RGB) codification, or a Yellow Ultraviolet (YUV) with other layer to provide an enhancement or an image of high definition. These codifications (RGB and YUV) produce an array of matrices that could not be processed due to the array. The separation of the channels is needed to process the image in this case. Then the grayscale image produces a single matrix with 8-bit resolution that provides 256 shades of gray. The method consists of the following steps:

1. Take pictures from 1 to n samples. This step is made to generate the thresholds of the specification.
2. Filter the image by a threshold. This step is made to saturate the pixels at the borders to fully activate it.
3. Count the activated pixels in the sample by summation.
4. Obtain the specification limits. Generate a summation for the width and other for the height of the sample to compare with the specification.
5. By interpolation, give the specification limits generated by step 3.
6. Take the specification limits by sampling (one iteration for every sample).
7. Obtain the crisp values for every variable with steps 5 and 6.
8. With the values obtained in step 7, make a MISO system to create the rule base for the expert system. These rules are obtained with the coefficients given by (10). The samples (Table 1) are treated to obtain only the limits to get the Table 2, that represents the rule base to the expert system using the knowledge of Table 1.
9. Generate the universe of discourse (UOD) with the permutations of the low and high states of every variable. E.g. For X_1 the low limit is 133 and the high limit is 143.
10. Use the Table 1 to generate the knowledge base to train the expert system.
11. Get the approximation to evaluate the process.

$$a_1x_1 @ a_2x_2 @ \dots @ a_nx_n = y \quad (10)$$

where: a_n , is a scalar (independent variable coefficient) and “@” is a mathematical operator.

Table 1. Sampling.

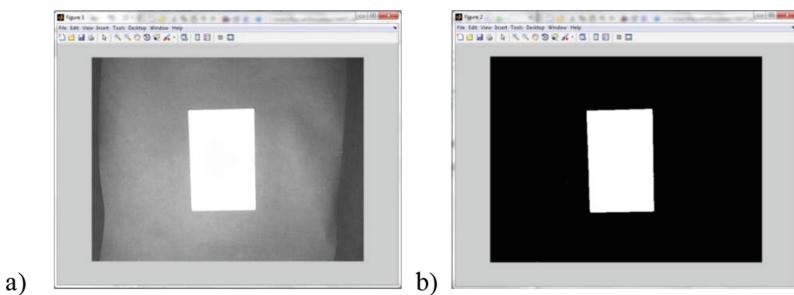
Samples	1	2	3	4	5	6	7	8	9	10	11	12	13
X1	208	214	218	218	210	209	205	218	206	211	211	208	207
X2	139	141	135	142	142	141	137	143	141	146	147	143	145
Goal	0.6	3.6	1.2	5.4	3.0	2.1	-1.5	6.0	1.2	5.7	6.3	3.0	3.9

Table 2. Rule base of the expert system obtained by CCD/2k [35].

Rule	1	2	3	4
X1	133	133	143	143
X2	198	218	198	218
Goal	-6.0	0	0	6.0

4 Experimental Test

The experimental results of industrial image processing are illustrated in Table 1. The modeling of the system uses the CCD/2^k method proposed in [35]. For the case study, a light cabinet is provided with a webcam and two sources of led light with a power of 550 lumens; each situated on both sides of the camera, and the objective sample is placed at 18" from the camera. All tests were performed in MatLab 2023b on a laptop with AMD Ryzen 5 5500U processor and Radeon Graphics @ 2.10 GHz. And 8 GB of RAM. There are two variables for input: height and width for dimensional inspection of the support plates, based on industrial image processing with T1 NSFLS Mamdani model. The height and width features were used to determine the product quality of the support plates. The specifications are width 80 ± 3 pixels and height 120 ± 5 pixels. A sample of the image used to process the information is shown in Fig. 2(a), and the filtered image is shown in Fig. 2(b).

**Fig. 2.** a) Sample image, b) Filtered image. Adapted from [36].

5 Application

The industrial image processing quality control system was tested with 16 samples (see Table 1). Where “X” is a matrix consisting of 16 data pairs. Furthermore, the level of quality of the samples is classified into two fuzzy subsets with two inputs (high and wide in low and high levels each). The images processed in the industrial image processing experiment listed in Table 1 are processed with the T1 NSFLS. The output is categorized by standard deviations presented in the sample versus the mean of the process,

corresponding to the categories “Low” & “High”. The T1 NSFLS had the following characteristics: Mamdani inference, product implication, centroid defuzzification, two inputs, one output, Gaussian MF’s and four rules.

6 Results

To evaluate the performance of the T1 NSFLS model, it is needed to run tests on the pure model T1 SFLS of this application. All models are designed with an adaptation of the T1 SFLS/CCD method proposed by [35]. The tests performed consist of two runs with 20 data pairs for training and 10 data pairs for a test in two different runs. The exposed results below, show that this process can be evaluated with the use of soft computing (SC) techniques. They also show that SC can be used for manufacturing quality assurance in the forms of measuring (dimensioning, shape checking), visual inspection (assembly, feature evaluation, presence, etc.). The singleton models are useful but require a lot of epochs for training and still have a significant error in some samples, as shown in Table 3. in there the Mean Square Error (MSE) estimator is used to see the performance of the model. Nevertheless, the uncertainty cannot be filtered with this model. If the training data is corrupted by noise as in this case, the output values cannot be adjusted by the model.

Table 3. MSE of prediction with T1 SFLS obtained by CCD/2k [35] with 500 epochs of training.

Samples	T1 SFLS (No noise)	T1 SFLS (95 dB white noise)	T1 SFLS (40 dB white noise)
1	0.00240100	0.00228484	0.37773316
2	0.01092025	0.01079521	0.00200704
3	0.00000656	0.00000902	0.23833924
4	0.00030976	0.00033856	0.18696976
5	0.00505521	0.00521284	0.20757136
6	0.00000739	0.00000739	1.22190916
7	0.00627264	0.00652864	1.84063489
8	0.00720801	0.00698896	0.10982596
9	0.00568516	0.00594441	0.01774224
10	0.00024025	0.00028561	0.02220100
MSE	0.00382318	0.00385432	0.42249338

Table 4 presents the approximation for a corrupted input data with uncertain values that have a rate of error in an interval of $\pm 5\%$. There is shown the prediction of the model after 500 epochs of training and we can see that the error presents cycles. With a change in the quantity of the number of epochs that train the system, it depends on the amount of uncertainty in the input. Table 4 shows that there is a process of adaptation tighter when this model uncertainty is higher. This is because the SC models show adaptation cycles as models appearing in the classical control systems [35] as can be seen in sample 3.

Table 4. MSE of prediction with T1 SFLS obtained by CCD/2k [35] with different quantity of epochs of training (40 dB white noise added).

	Epochs of training					
Samples	1	5	10	50	100	500
1	0.203401	0.06666724	0.664225	0.38713284	0.30713764	0.40208281
2	2.66440329	0.77387209	0.89283601	0.12013156	0.00893025	0.00025921
3	0.04558225	4.80223396	8.82268209	2.17975696	0.46036225	0.21104836
4	8.13276324	1.47671104	0.85618009	0.01926544	0.19909444	0.17969121
5	1.41729025	0.02900209	0.04774225	0.154449	0.14417209	0.23020804
6	2.601769	0.43401744	1.18657449	1.07205316	1.13379904	1.23832384
7	0.68508729	1.02353689	2.03490225	1.74662656	1.61239204	1.92238225
8	0.26884225	3.01265449	4.13674921	0.887364	0.1444	0.12996025
9	5.26610704	2.41989136	0.90687529	0.29387241	0.06240004	0.01375929
10	11.1395738	2.04833344	1.09998144	0.13645636	0.01954404	0.02108304
MSE	3.24248194	1.608692	2.06487481	0.69971083	0.40922318	0.43487983

Tables 5 and 6 show the approximation for T1 NSFLS and their enhanced capacity provided by the secondary evaluation. On the other hand, the adjustment of the process shows a fast convergence in the output values. The figure below shows the output prediction of the non-singleton system.

Table 5. MSE of prediction with T1 NSFLS obtained by CCD/2k [35] with different quantity of epochs of training (No noise added).

	Epochs of training					
Samples	1	5	10	50	100	500
1	0.42237001	0.18809569	0.00006724	0.00946729	0.01868689	0.00038809
2	2.62180864	0.73307844	0.83868964	0.08946081	0.02362369	0.00579121
3	0.01179396	4.260096000	7.97836516	1.23476544	0.04068289	0.00117649
4	9.02401600	3.12794596	2.59822161	0.39350529	0.00380689	0.00002025
5	1.13443801	0.01065024	0.01399489	0.00187489	0.01651225	0.00248004
6	3.60278361	0.00037636	0.08485569	0.00073441	0.00009604	0.00011236
7	0.16941456	0.09853321	0.03052009	0.01308736	0.02699449	0.00246016
8	0.39400729	2.26803600	3.05620324	0.31438449	0.00030276	0.00376996
9	4.81188096	1.07806689	0.11923209	0.02647129	0.00000841	0.00608400
10	11.6998203	3.16982416	2.19099204	0.38514436	0.00000961	0.00050176
MSE	3.38923333	1.4934703	1.69111417	0.24688956	0.01307239	0.00227843

Table 6. MSE of prediction with T1 NSFLS obtained by CCD/2k [35] with different quantity of noise at the inputs and 500 epochs of training.

Samples	T1 NSFLS (No noise)	T1 NSFLS (95 dB white noise)	T1 NSFLS (40 dB white noise)
1	0.00038809	0.00034596	0.40208281
2	0.00579121	0.00570025	0.00025921
3	0.00117649	0.00108241	0.21104836
4	2.025E-05	2.916E-05	0.17969121
5	0.00248004	0.00259081	0.23020804
6	0.00011236	0.00011025	1.23832384
7	0.00246016	0.00262144	1.92238225
8	0.00376996	0.00361201	0.12996025
9	0.006084	0.00635209	0.01375929
10	0.00050176	0.00056644	0.02108304
MSE	0.00227843	0.00230108	0.43487983

Sometimes big error rates occur because the system has an overtrain in some points at the transition phase of the probability distribution, in that case trial and error is needed to train the system and provide the adequate quantity of epochs to obtain the desired performance.

Tables 3, 4, 5 and 6 show that the process to obtain an adequate input for a particular process depends on the desired precision and the training should be made by trial and error to obtain a desired performance. Also, the cycles present an advantage that provides a way to reduce the complexity of the system, through the reduction to a single cycle which must be scaled for other states of the same system, this will execute the model itself.

7 Conclusions

The training phase contributes to adjust the noise and the uncertainty generated at the inputs. The nonsingleton fuzzification provides the necessary adjustment, provided by the secondary evaluation, diminishing the variation from 1.46% in T1 SFLS to 0.94% in T1 NSFLS, close to a single standard deviation on every side of the distribution, and qualifies as a superior quality assurance process validation.

The Mamdani model provides better performance in the non-singleton fuzzy systems form, in contrast to a trained singleton system that provides an SME of 0.4348 with a SNR of 40 dB. The fuzzy non-singleton system can reduce the error rate with training of 500 epochs, which consume less than half a second, that allows a system that works online. The experiments show that the proposal can be used in a real facility where an important level of the product's geometrical quality is needed. The product's geometry measured using a camera can be considered as a crucial tool in the production line to increase the number of produced parts under the required quality specifications.

References

1. Mendel, J.M.: Uncertain Rule-Based Fuzzy Systems. In Introduction and New Directions, 2nd ed.; Springer: Cham, Switzerland (2017)
2. Mouzouris, G.C., Mendel, J.M.: Non-singleton fuzzy logic systems: theory and application. *IEEE Trans. Fuzzy Syst.* **5**(1), 56–71 (1997)
3. Monzouris, G.C., Mendel, J.M.: Dynamic non -singleton fuzzy logic systems for nonlinear modeling. *IEEE Trans. Fuzzy Syst.* **5**(2), 199–208 (1997)
4. Akpolat, Z.H.: Non-Singleton Fuzzy Logic Control of a DC Motor. *J. Appl. Sci.* **5**(5), 887–891 (2005)
5. Chua, T.W., Tan, W.W.: Non-singleton genetic fuzzy logic system for arrhythmias classification. *Eng. Appl. Artif. Intell.* **24**(2), 251–259 (2011)
6. Loiola, M.B., Ribeiro, M.V., Romano, J.M.T.: A turbo equalizer using fuzzy filters. In: Machine Learning for Signal Processing, Proceedings of the 14th IEEE Signal Processing Society Workshop (2004)
7. Montes Dorantes, P.N., Hernández García, H.M., de la Rosa Elizondo, J., Méndez, G.M., Nieto González, J.P.: Sistemas difusos para monitoreo y control de metalurgia secundaria, Memorias del congreso internacional de metalurgia y materiales. Congreso **35**. 1(1), 354–363 (2013)
8. Mukerji, S.: Understanding the nonadditive probability decision model. *Econ. Theor.* **9**(1), 23–46 (1997)
9. Baranyi, P., Martinovics, A., Kovacs, S., Tikk, D., Yam, Y.: A general extension of fuzzy SVD rule base reduction using arbitrary inference algorithm. In: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, vol. 3, pp. 2785–2790 (1998)
10. Haddad, L., Lau, D.: On partial clones containing maximal clones. In: IEEE 43rd International Symposium on Multiple-Valued Logic, pp. 96–101 (2013)
11. Fontes, R.M.: Dynamic model identification with uncertain process variables using fuzzy inference system. In Computing Aided Chemical Engineering (31). 11th International symposium on Process Systems Engineering (PSE), pp. 955–959, (2012)
12. Serir, L., Ramasso, E., Zerhouni, N.: Evidential evolving Gustafson –Kessel algorithm for online data streams partitioning using belief function theory. *Int. J. Approx. Reasoning* **53**(5), 747–768 (2012)
13. Ng, G.S., Liu, F., Loh, T.F., Quek, C.: A novel brain-inspired neuro-fuzzy hybrid system for artificial ventilation modeling. *Expert Syst. Appl.* **39**(15), 11808–11817 (2012)
14. Liu, C., Li, P., Zhang, Y., Zhang, Y., Liu, C., Wei, S.: A construction method of personalized ECG template and its application in premature ventricular contraction recognition for ECG MobilePhones. In: Long, M. (ed.) IFMBE Proceedings, vol. 39, pp. 585–588. Springer Heidelberg (2012). https://doi.org/10.1007/978-3-642-29305-4_153
15. Liu, X., Li, S.: Cumulative distribution function estimation with fuzzy data: some estimators and further problems. In: Kruse, R., Berthold, M., Moewes, C., Gil, M., Grzegorzewski, P., Hryniiewicz, O. (eds.) Synergies of Soft Computing and Statistics for Intelligent Data Analysis, vol. 190, pp. 83–91. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-33042-1_10
16. Yuan, Y., Yuan, X., Li, H.: The probability distribution and fuzzy system based on bounded product implication. In: Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), vol. 2, pp. 786–790 (2010)
17. Zhou, C.: Belief functions on distributive lattices. *Artif. Intell.* **201**, 1–31 (2013)
18. Prokopowicz, P.: Flexible and simple methods of calculations on fuzzy numbers with the ordered fuzzy numbers model. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2013. LNCS, vol. 7894, pp. 365–375, Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38658-9_33

19. Hashim, M.N.: Induction motor modelling using fuzzy logic (Doctoral dissertation), Universiti Tun Hussein Onn Malaysia (2013)
20. Zhou, W.Z., Huan, J.: Research on method of anti-fuzzy for qualitative evaluation index of materiel support plan. *Adv. Mater. Res.* **765**, 3220–3224 (2013)
21. Pendharkar, P.: Fuzzy classification using the data envelopment analysis. *Knowl.-Based Syst.* **31**, 183–192 (2012)
22. Khalil, R., Sababheh, M.: A study of uniquely remotal sets. *J. Comput. Anal. Appl.* **13**(7), 1233–1239 (2011)
23. Ren, Y.Q., Duan, X.G., Li, H.X., Chen, C.P.: Multi-variable fuzzy logic control for a class of distributed parameter systems. *J. Process. Control.* **23**(3), 351–358 (2013)
24. Sayari, E., Yaghoobi, M.: Clustering of ECG signals based on fuzzy neural network with initial weights generated by genetic algorithm. *Majlesi J. Electr. Eng.* **8**(1), 1–9 (2013)
25. Sunberg, Z., Rogers, J.: A belief function distance metric for orderable sets. *Inf. Fusion* **14**(4), 361–373 (2013)
26. Starczewski, J.T.: (2012). Advanced Concepts in Fuzzy Logic and Systems with Membership Uncertainty, Studies in Fuzziness and Soft Computing, Vol. 284, Springer, Heidelberg (1997)
27. Chen, C., Zhao, Y., Wagner, C., Pekaslan, D., Garibaldi, J.M.: An extension of the FuzzyR toolbox for non-singleton fuzzy logic systems. In IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6 (2021)
28. Reyes D., Álvarez A., Rincón E.J., Valderrama J., Noradino P., Méndez G.M.: A PID Using a Non-singleton Fuzzy Logic System Type 1 to Control a Second-Order System. In: Melin P., Castillo O., Kacprzyk J., Reformat M., Melek W. (eds). *Fuzzy Logic in Intelligent System Design: theory and Applications. NAFIPS 2017. Advances in Intelligent Systems and Computing*, vol 648, Springer, Cham (2018)
29. Méndez, G.M., Montes Dorantes, P.N., Alcorta, M.A.: Dynamic adaptation of the PID's gains via Interval type-1 non-singleton type-2 fuzzy logic systems whose parameters are adapted using the backpropagation learning algorithm. *Soft. Comput.* **24**(1), 17–40 (2020)
30. Ontiveros-Robles, E., Castillo, O., Melin, P.: An approach for non-singleton generalized Type-2 fuzzy classifiers. *J. Intell. Fuzzy Syst.* **39**(5), 7203–7215 (2020)
31. Tong, W., Zhao, T., Duan, Q., Zhang, H., Mao, Y.: Non-singleton interval type-2 fuzzy PID control for high precision electro-optical tracking system. *ISA trans.* **120**, 258–270 (2022)
32. Méndez, G.M., López-Juárez, I., Alcorta García, M.A., Martínez-Peon, D.C., Montes-Dorantes, P.N.: The enhanced Wagner-Hagras OLS–BP hybrid algorithm for training IT3 NSFLS-1 for temperature prediction in HSM processes. *Mathematics* **11**(24), 4933 (2023)
33. Castorena, G.A.H., Méndez, G.M., López-Juárez, I., García, M.A.A., Martínez-Peon, D.C., Montes-Dorantes, P.N.: Parameter prediction with Novel enhanced Wagner Hagras interval Type-3 Takagi–Sugeno–Kang Fuzzy system with type-1 non-singleton inputs. *Mathematics* **12**(13), 1976 (2024)
34. Zhao, T., Yu, Q., Dian, S., Guo, R., Li, S.: Non-singleton general type-2 fuzzy control for a two-wheeled self-balancing robot. *Int. J. Fuzzy Syst.* **21**(6), 1724–1737 (2019)
35. Montes Dorantes, P.N., Praga-Alejo, R., Nieto González, J.P., Méndez, G.M.: Modelado de Sistemas Adaptativos de Inferencia Neuro-Difusa Usando Diseño Central Compuesto. *Avances en Inteligencia Artif. Res. Comput. Sci.* **62**, 259–269 (2013)
36. Montes Dorantes P.N., Nieto González J.P., Praga-Alejo R., Guajardo Cosio K.L., Méndez G.M.: Sistema inteligente para procesamiento de imágenes en control de calidad basado en el modelo difuso singleton tipo 1. *Res. Comput. Sci.* **74**, 117–130 (2014)



Weed Detection in a Sunflower Field Using Supervised Learning Techniques

A. Mexicano¹, J. C. Carmona^{1(✉)}, S. Cervantes², K. Bee¹, and P. N. Montes¹

¹ National Technological Institute/Technological Institute of Ciudad Victoria, Tamaulipas, Mexico

{adriana.ms, jesus.cf, m23380019}@cdvictoria.tecnm.mx

² Department of Computer Science and Engineering, Universidad de Guadalajara, 46600 Ameca, Mexico

salvador.cervantes7964@academicos.udg.mx

Abstract. Weed growth in crops represents a challenge for farmers as this can affect plant thriving due to nutrient stealing. This paper presents a tool for weed detection in a sunflower field using computer vision techniques. To this end, regions of interest were extracted from 25 multispectral images; 20 vegetation indices were used to characterize the classes Background, Weed and Sunflower. Afterwards, Correlation Analysis (CA), Principal Component Analysis (PCA), AutoEncoder (AE), CA-PCA and CA-AE techniques were applied to create 5 datasets to train the Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Naive Bayes (NB) classifiers. The classifier that obtained the best separation between the classes Background, Weed and Sunflower was the SVM classifier applied on the PCA set based on 3 principal components, with an Accuracy of 81.5%, Precision of 81%, Recall of 81%, F1 of 81% and Cohen Kappa of 72%.

1 Introduction

Precision Agriculture (PA) aims to innovate the usual techniques for the agricultural branch by collecting and examining crop information [1]. PA uses Geographic Information Systems (GIS), Remote Sensing (RS), Spatial statistics, Farm Management Information Systems (FMIS) and Variable-Rate Technology (VRT) as monitoring strategies with the purpose of providing relevant information to improve production and output, and to reduce pollution to the environment. One such example has been its focus on sunflower harvesting. The sunflower (*H. annuus* L.) is part of the Compositae (*Asteraceae*) group and is a native plant of North America [2]. One of the main uses of the sunflower is in the generation of oil due to its high content of vitamin E, which is used as an antioxidant method due to its a-, b- and γ-tocopherol content. However, there are impediments that can affect sunflower growth. These include diseases such as downy mildew (*Plasmopara halstedii*) and Phomopsis (*Diaporthe helianthi*); insects such as leafhoppers, European sunflower moth, sunflower seed weevil and sunflower stem weevil; as well as parasitic herbaceous plants such as Boomrape (*Orobanche cernua* Loefl.). In addition, it is also affected by stress conditions such as drought, wild birds that usually appear 3 weeks before the seed matures and the emergence of weeds. Weeds are

the anomaly of interest in this research, being a component of the environment that has shown to influence sunflower growth. Weeds are considered to be those plants that grow in an unwanted area and invade a crop [3]. This type of plant can be found in different environments due to its ecological adaptations and taxonomic classification. In addition, they can subsist in aquatic and ground environments and are sometimes aerial climbers. Damage that can be caused by weeds is categorized as direct and indirect losses. Direct losses consist of minimizing the quantity and quality of crop production. These losses are connected to the existence of weeds in row crops, vegetables, fruits, trees, ornamental plants, lawns, sports fields, pastures, rangelands and natural environments in general. Indirect losses, on the other hand, consist of those external factors that do not directly damage the profits obtained from the sales of the crop, but do have a cost for the population, traders and landowners. As an example, they tend to store insects, pests and cause health problems, leading to a decline in the value of the crops, generating costs for weed removal in non-agricultural areas, among other consequences. Therefore, alternatives are often implemented to counteract them [4] such as mechanical weed control like Tillage, Mowing and Flaming; chemical weed control like Herbicides; cultural weed control like Shading, Mulches and Animals; as well as weed control through Computer Vision, which is the basis of this research. Computer Vision has proven to be an opportunity gap for crops, being an example of its usefulness the work presented in [5], where potato plants that required more care due to water stress were identified; the automatic location of vegetation in karst areas such as caves with the help of an Unmanned Aerial Vehicle (UAV) [6]; the monitoring of aromatic crops [7]; the detection of weeds in canola fields using Maximum likelihood classification and deep convolutional neural networks [8]; the discrimination of weeds in crops by measuring simple morphological features of leaf shape and self-organizing neural network adapted to biological functioning for pattern detection [9]; and the weed detection with the help of the U-Net model with multispectral images taken by a UAV (unmanned aerial vehicle), the Green channel + Filtered-NIR + Normalized Difference Vegetation Index with the U-NET model [10]. In this paper, a methodology for weed detection in a sunflower field is proposed using multispectral images, vegetation indices, feature selection algorithms and classifiers such as SVM, KNN and NB.

2 Related Work

In the last decade, several research projects have been developed for the automatic detection of pests and weeds using UAVs. In 2020 Hamza et al. [8] developed a methodology to speed up the manual labeling of pixels in images of weedy crops through a 2-step procedure. In the first stage, the background and foreground were segmented by applying the Maximum likelihood classification. In the second stage, the pixels belonging to the weeds were manually labeled. The labeled elements were used to train semantic segmentation models to classify crop pixels: Background as one class and Weeds as a second class. The model that obtained the best results for weed detection was SegNeT based on ResNet-50, obtaining an average evaluation of 0.8288 in Intersection over Union (IoU) and 0.9869 in Frequency Weighted Intersection Over Union (FWIoU). The result obtained when classifying the Weed class was 0.6648 IoU and 0.9928 IoU

for the elements of the Background class. In 2021 Butte et al. [5] developed a project to analyze multispectral images of a Russet Burbank potato field in Bingham County, Idaho using neural networks. Their main goal was to show the ability of automated recognition to differentiate between healthy and water-stressed plants. The model used was Retina-UNetAG which achieved an average Dice Score coefficient of 0.74 obtained between healthy classes with 0.723 and stressed classes with 0.756. This coefficient assessed the overlap between actual classes and model predictions, specifically for the identification of healthy and stressed plants. In 2023 Mertkan et al. [11] designed an algorithm for detecting weeds. To this end, a repository based on multispectral images of a sunflower field was used. Images were captured in stages, starting from the emergence of the cotyledon up to the advanced growth of the sunflower stem. The U-Net model was tested on images of the crop at the late growth stage, when chemical treatments could already be applied. The algorithm uses the Green + NIR Filtered + NDVI Index channels as input data. The classes evaluated were Soil, Crop and Weed using the IoU measure, achieving an IoU of 0.990, 0.906 and 0.753 for each respective class. The purpose of this measure is to calculate how many pixels of the prediction match the pixels of the real mask. In 2024, Seiche et al. [10] presented a comparison study between a self-made multispectral camera and the MicaSense Altum tool for weed detection. To do this, images taken by a DJI Matrice 210 UAV were used with both cameras in a corn field. The pixel-based classification of weed and crop classes required a U-Net neural network. The evaluation metrics used were Recall, which measured the ability of the model to efficiently detect positive instances for the Weed Crop and Soil classes in the datasets; Precision, which measured the number of instances detected as positive by the model that are actually positive; and the F1 Score metric, which brings the above metrics together as a balanced average into a single score. Altum reached an F1 Score of 82%, while the self-made camera achieved 76%. In the case of the Recall, Altum achieved a 75% compared to 68% for the self-made camera. However, the self-made system achieved a Precision of 90%, making it an affordable option for weed detection. Furthermore, in addition to the approaches based on Deep Learning seen above, there are other projects that work with vegetation by applying techniques based on Supervised Learning, such as those used in [23]. For example, in 2020, Lan et al. [12] conducted a research to determine the feasibility of remote sensing for Huanglongbing (HLB) in citrus orchards. For this, a multispectral ADC-lite camera mounted on a DJI Matrice M100 UAV was used to obtain multispectral photographs, their own repository was built, 20 Vegetation Indices (VIs) were used, 5 datasets were developed by applying CA, PCA, AE, CA-PCA and CA-AE, and the classifiers SVM, KNN, Logistic Regression, NB, Ensemble Learning and a Neuronal Network were trained. The evaluation metrics used were Accuracy, Recall, Precision, Specificity and F1-Score to detect healthy citrus trees from those diseased with HLB. Precision and Specificity measured the classification correctness of healthy and diseased samples; Accuracy assessed the percentage of healthy and diseased samples that were correctly classified; Recall rated the ability of the models to identify HLB affected trees; F1-Score consisted of the combination of Recall and Precision. Cohen's Kappa coefficient determined the concurrence between the different classification algorithms. The best performing classifiers were Ensemble Learning and the Neural Network, with results of 100% and 97.28% respectively, in

their evaluation metrics for the detection of healthy and diseased HLB plants. In 2023, Pan et al. [6] collected multispectral image data to design a tool for vegetation detection in karst areas. To do this, they used Random Forest, SVM, Gradient Boosting Machine (GBM) and Deep Learning models to compare their efficiency in detection. Also, 16 VIs were used. The best model for vegetation detection in karst areas was GBM with a precision of 95.66%.

3 Materials

For the development of this sunflower and weed detection project in a sunflower field, a computer with a 3.42 GHz Intel Core i9 processor, a 64 GB RAM memory and a Windows 11 Pro OS was used. For the segmentation of the classes in the image sets, the *Region Of Interest (ROI) Tool* of the ENVI software version 5.3 was used. The project was implemented using Python programming language together with the following modules and libraries: OpenCV [13], Scikit-learn [14], Pandas [15], Keras [16], Tensorflow [17], Seaborn, Matplotlib [18], Numpy [19], Joblib [20] and Pickle [21]. The images used in this study were photographs of sunflower crops at an advanced stage of growth, where the plants were producing more leaves, their stalks were growing and were preparing for blooming. The repository was set up in 2016 by SAPIENZA UNIVERSITÀ DI ROMA [22]. The photographs were acquired in Jesi, Italy in the infrastructure of Assam, using a multispectral camera with *Red*, *Green*, *Blue* and *NIR* channels. The images are divided into four groups: RGB, NIR, GT: greyscale images and GT_COLOR: images with the representation of the classes within the RGB model.

4 Methodology

The proposed methodology consists of 6 stages: (A) Region of Interest (ROI) extraction; (B) Dataset augmentation; (C) Calculation of Vegetation Index; (D) Dataset construction; (E) Training dataset development; and (F) Classification.

- A) *ROI extraction*: In this stage, the Background, Weed and Sunflower classes were segmented using the ENVI tool. Figure 1 shows the result of extracting the regions of interest belonging to the Background, Sunflower and Weed classes from the RGB and NIR images. Figure 1a shows the *ROI* of the Background in RGB format; Fig. 1b depicts the *ROI* of the Background in NIR channel; Fig. 1c shows the *ROI* of the Sunflower in RGB format; Fig. 1d presents the *ROI* of the Sunflower in NIR channel; Fig. 1e shows the *ROI* of the Weed in RGB format; Fig. 1f presents the *ROI* of the Weed in NIR channel; Fig. 1g shows the original image in RGB format; Fig. 1h provides the original image in NIR channel.
- B) *Dataset augmentation*: To increase the amount of data, the transformations of rotation, shift, mirror, brightness and contrast enhancement were applied to the extracted ROIs, resulting in a total of 125 images.

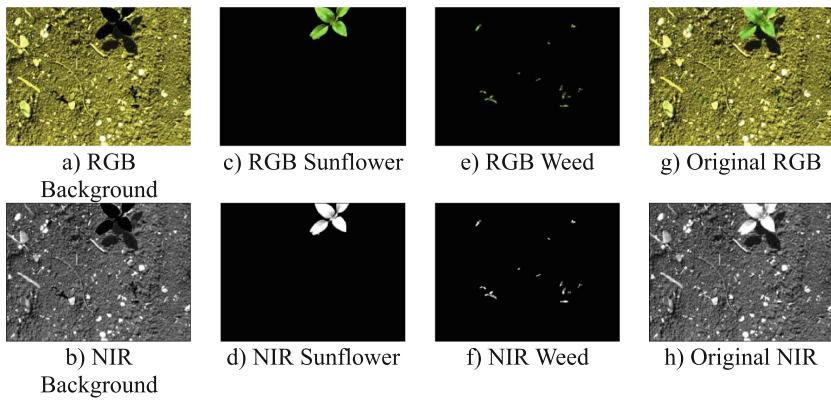


Fig. 1. ROI for each class in the different groups of images

C) *Calculation of vegetation indices (VIs):* The purpose of the VIs is to numerically represent the values that a plant or object possesses through the colors that compose it. In this study, the 20 VIs proposed in [12] have been applied. Among the selected indices are *Normalized Difference Vegetation Index (NDVI)*, *Structure Intensive Pigment (SIP)*, *Optimized Soil Adjusted Vegetation Index (OSAVI)* and *Ratio Vegetation Index (RVI)*, as shown in Table 1, see Eqs. 1 to 4.

Table 1. Extract of the equations used for the calculation of vegetation indices.

Index name	Acronyms	Equations	
Normalized Difference Vegetation Index	NDVI	$(NIR - RED)/(NIR + RED)$	(1)
Structure Intensive Pigment	SIP	$(NIR - GREEN)/(NIR + RED)$	(2)
Optimized Soil Adjusted Vegetation Index	OSAVI	$1 + L * \frac{NIR - RED}{NIR + RED + L}; L = 0.16$	(3)
Ratio Vegetation Index	RVI	NIR/RED	(4)

- D) *Dataset construction:* After calculating the VIs for each class (Sunflower, Background and Weeds), a dataset was generated where each record corresponds to a pixel and the columns correspond to its VIs and to the class each record belongs to.
- E) *Training dataset development:* Most studies of this type tend to use the VIs without index reduction. This is due to the fact that they only use 7 VIs. In this research, 5 datasets were generated after applying: a) Correlation Analysis (CA), b) Principal Component Analysis (PCA), c) AutoEncoder (AE), d) CA with PCA (CA-PCA) and e) CA with AE (CA-AE).

The CA was applied using Pearson's product moment correlation coefficient, implementing a linear transformation so that the range of the correlation was from 0 to 1, since the empirically established radius in the CA was from 0.96 to 1. Equation 5 shows the mathematical procedure of the analysis used, where η represents the correlation coefficient equation; X_i, X_j the comparison of vegetation indices; $cov(X_i, X_j)$ the covariance between X_i, X_j ; $var(X_i)$ the variance of X_i and $var(X_j)$ the variance of X_j [1].

$$\eta = \frac{cov(X_i, X_j)}{\sqrt{var(X_i) \cdot var(X_j)}} \quad (5)$$

$$cov(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))]$$

$$var(X_i) = E[(X_i - E(X_i))^2]$$

$$var(X_j) = E[(X_j - E(X_j))^2]$$

$$\eta \rightarrow \eta \times 0.5 + 0.5$$

After applying CA, it was observed that the NDVI index presented a high correlation with OSAVI, IPVI, MCARI1 and MTVI1, as the correlation values were within the established range of 0.96 to 1, therefore they were eliminated. 12 VIs were eliminated, leaving only 8 of them: NDVI, TVI, GDVI, G, CVI, MCARI1, Norm R and Norm G. After that, PCA was applied to significantly reduce the number of vegetation indices while retaining the relevant information from the original dataset. By applying PCA, three principal components were obtained: [0.80318816, 0.16290851, 0.01689905] giving a variance total of 0.9829957175640618.

AutoEncoder (AE) was applied, which consists of an *Encoder* and a *Decoder*, which are multilayer neural networks with dense layers. The *Encoder* consists of an input layer that receives the input dimensions (in this case, the 20 Vegetation Indices). Subsequently, the number of dimensions decreases due to fewer neurons in each layer. In the hidden layers of the *Encoder*, the *ReLU* activation function is used; this function converts negative values to 0 and keeps only positive values. Finally, in the latent space, three dimensions were obtained. The *Decoder* then reconstructs the input data again from the latent space, where the output layer reconstructs the 20 Vegetation Indices, using the *Linear* activation function. The *AE* is trained for 20 epochs with the objective of reducing as much as possible the error of reconstruction of the input information, because if the error is too high, it means that too much information is being lost and the desired error must lie within a range close to 0. The *AE* reduced the 20 indices to 3 dimensions. This analysis showed an 8.2753e–04 data loss in the reconstruction of the model with the latent layer neurons. Thus, this indicates that an adequate network was built to handle the data.

CA with PCA (CA-PCA) was applied with the purpose of further reducing the dimensions of the data through the order of variances of the 3 principal components; this would result in having fewer variables to analyse, enhancing the speed of the

training. The 3 principal components [0.70187342, 0.25556673, 0.01975199] of the remaining 8 *CA* indices were obtained, giving a sum of 0.9771921409345312 in variance.

Finally, CA with AE (CA-AE) was applied, reducing the 8 indices given by *CA* to only 3 dimensions, with a loss error of 0.0228.

- F) *Classification.* The classification algorithms used were Support Vector Machine with a Polynomial Kernel of Degree $d = 3$, Gaussian kernel with $\gamma = 0.0947$ and Linear kernel. Also, K-Nearest Neighbors was used with the Uniform Weight and Distance Weight models. And finally, Naive Bayes with the Gaussian, Multinomial and Bernoulli models. Equation 6 evaluates *Accuracy (%)*, which is the proportion of correct predictions among the total predictions made. Equation 7 evaluates *Recall (%)*, which is the ability of the classifier to correctly identify all positive instances. Equation 8 evaluates *Precision (%)*, which is the ability of the classifier to efficiently identify positive instances out of the total number of instances predicted as positive. Equation 9 evaluates the *F1 score*, which is a combination of *Precision* and *Recall* to analyze how well the classifier is performing. Equations 10 to 12 constitute the Cohen Kappa evaluation measure that assesses the degree of concordance in the classifications.

$$\text{Accuracy}(\%) = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \quad (6)$$

$$\text{Recall}(\%) = \frac{TP}{TP + FN} \times 100 \quad (7)$$

$$\text{Precision}(\%) = \frac{TP}{TP + FP} \times 100 \quad (8)$$

$$F1(\%) = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \quad (9)$$

$$P_0 = \frac{TN + TP}{TN + TP + FN + FP} \quad (10)$$

$$P_e = \frac{(TP + FP) * (TP + FN) + (FN + TN) * (FP + TN)}{(TP + FP + FN + TN)^2} \quad (11)$$

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \quad (12)$$

5 Results

Different tests were carried out by modifying the parameters of the classifiers in order to find the model that provided the best results.

Table 2. Evaluation metrics results for the best models of each classifier.

Metric	Gaussian Kernel PCA-SVM	CA-AE KNN with Uniform Weights	Gaussian CA-PCA NB
Accuracy	0.8150	0.7969	0.7611
Precision	0.8121	0.7926	0.7652
Recall	0.8145	0.7946	0.7604
F1	0.8102	0.7931	0.7466
Cohen Kappa	0.7225	0.6951	0.6418

5.1 Experimental Results

After having experimented with different configurations, the results of the best models tested were extracted, being PCA-SVM with the Gaussian Kernel the one that obtained the best performance, as can be seen on Table 2. Table 2 shows the evaluation metrics used and the respective model corresponding to each result.

Table 2 presents the obtained results of the Gaussian Kernel PCA-SVM, CA-AE KNN with Uniform Weights and CA-PCA NB models, through the evaluation metrics Accuracy, Precision, Recall, F1 score and Cohen Kappa. In it, it is noted that the Gaussian Kernel PCA-SVM model had better performance than the rest of the models, as it obtained the highest average of the evaluation metrics: Accuracy of 81.5%, Precision of 81%, Recall of 81%, F1 of 81% and Cohen Kappa of 72%. It can be concluded that these results are due to the fact that this approach used the 3 principal components with the Gaussian Kernel on the 20 base VIs, which unlike in other models, the base amount of VIs was smaller because they had first been applied CA and then PCA or AE depending on the model, thus losing data during the process.

5.2 Comparison with Other Approaches

In the project carried out by Mertkan et al. [11] results were obtained where the classes Crop, Weed and Soil could be identified. However, they used an approach based on Deep Learning, also using 3 repositories based on sunflower fields at different stages. Instead, in this study, only a small set of images from a single sunflower repository and a Supervised Learning-based Computer Vision approach were used. On the other hand, work has been done to detect weeds in crops using the You Only Look Once (YOLO) model [24, 25]. However, the potential of the developed tool is based on the analysis applied to each pixel of the image, avoiding taking elements that do not belong to the object as YOLO does [26] and also allows the creation of classes that differentiate the state of the plant through the VIs that are applied [12].

Table 3 shows the individual evaluations of the Kernel PCA-SVM model for the classes Background, Weed and Sunflower through the Precision, Recall and F1 metrics. By averaging the 3 metrics, the Background class presented a performance of 94%, the Weed class of 70% and the Sunflower class of 79%. In these results it can be noted that the Background class was easier to detect than the Sunflower and Weed classes,

Table 3. Class evaluation of the Gaussian Kernel PCA-SVM model.

Metric	Background	Weed	Sunflower
Precision	0.9122	0.7729	0.7513
Recall	0.9712	0.6405	0.8317
F1	0.9408	0.7005	0.7895

being the latter the worst. This is due to the amount of data there was for each class. Typically, Background covers most of an image, which, unlike Sunflower or Weeds, are only found in one region, causing a lack of variability in the data, because there were not a large number of images used where these two classes could be found at different points in the same image. To improve this situation, it would be necessary to extract more regions of interest, use the GridSearchCV from the Scikit-Learn library to find the right configuration for the classifiers and further testing.

6 Conclusions

Weed emergence is a problem that significantly affects the growth of any crop, generally causing losses in both production and profit for farmers. Therefore, applying techniques that apply *Computer Vision* will facilitate the detection of this anomaly over large areas. In this case, a supervised learning approach using Vegetation Indices and CA, PCA, AE, CA-PCA and CA-AE analyses has been proposed to generate the training datasets. The SVM classifier with Gaussian, Polynomial and Linear kernels were used; the KNN classifier was used with uniform weights and distance weights models; and the NB classifier was used with Gaussian, Multinomial and Bernoulli distributions to identify weeds. To validate the results, the set was divided into 80% for training and 20% for testing. The Precision obtained for each class using the Gaussian Kernel PCA-SVM model was of 0.91 for Background, 0.77 for Weed and 0.75 for Sunflower. These results can be considered acceptable since the classifier was able to detect the three classes through the VIs that were applied directly on multispectral images. However, as a future work, it is hoped to obtain better results through experimentation and the application of performance improvement techniques, in order to be a tool that has an impact and offers a different alternative than other projects already carried out.

References

1. Nandeha, N., Trivedi, A.: Precision and Sustainable Agriculture. *Frontiers of Agronomy*, vol. 1, pp. 20–34. Elite Publishing House, Rohini New Delhi (2023)
2. Kaya, Y., Jocic, S., Miladinovic, D.: Sunflower. In: Gupta, S. (ed.) *Technological Innovations in Major World Oil Crops*, vol. 1, pp. 85–129. Springer, New York (2012). https://doi.org/10.1007/978-1-4614-0356-2_4
3. Dille, J.: Weed Biology. *Encyclopedia of Applied Plant Sciences*, 2nd edn., vol. 3, pp. 469–472 (2017). <https://doi.org/10.1016/B978-0-12-394807-6.00026-5>

4. Hanson, B., Roncoroni, J., Hembree, K., Molinar, R., Elmore, C.: Weed Control in Orchards and Vineyards. Encyclopedia of Applied Plant Sciences, 2nd edn., vol. 3, pp. 479–484 (2017). <https://doi.org/10.1016/B978-0-12-394807-6.00032-0>
5. Butte, S., Vakanski, A., Duellman, K., Wang, H., Mirkouei, A.: Potato crop stress identification in aerial images using deep learning-based object detection. *Agron. J.* **113**, 3991–4002 (2021). <https://doi.org/10.1002/agj2.20841>
6. Pan, W., Wang, X., Sun, Y., Wang, J., Li, Y., Li, S.: Karst vegetation coverage detection using UAV multispectral vegetation indices and machine learning algorithm. *Plant Methods* **19** (2023). <https://doi.org/10.1186/s13007-023-00982-7>
7. Bahuguna, S., et al.: Unmanned aerial vehicle-based multispectral remote sensing for commercially important aromatic crops in India for its efficient monitoring and management. *J. Indian Soc. Remote Sens.* **50**, 397–407 (2022). <https://doi.org/10.1007/s12524-020-01302-5>
8. Hamza, M., Bais, A.: Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Inf. Process. Agric.* **7**, 535–545 (2020). <https://doi.org/10.1016/j.inpa.2019.12.002>
9. Aitkenhead, M., Dalgetty, I., Mullins, C., McDonald, A., Strachan, N.: Weed and crop discrimination using image analysis and artificial intelligence methods. *Comput. Electron. Agric.* **39**, 157–171 (2003). [https://doi.org/10.1016/S0168-1699\(03\)00076-0](https://doi.org/10.1016/S0168-1699(03)00076-0)
10. Seiche, A., Wittstruck, L., Jarmer, T.: Weed detection from unmanned aerial vehicle imagery using deep learning—a comparison between high-end and low-cost multispectral sensors. *Sensors* **24** (2024). <https://doi.org/10.3390/s24051544>
11. Mertkan, H., Miftahushudur, T., Grieve, B., Yin, H.: Segmentation of weeds and crops using multispectral imaging and CRF-enhanced U-Net. *Comput. Electron. Agric.* **211** (2023). <https://doi.org/10.1016/j.compag.2023.107956>
12. Lan, Y., et al.: Comparison of machine learning methods for citrus greening detection on UAV multispectral images. *Comput. Electron. Agric.* **171**, 105–234 (2020). <https://doi.org/10.1016/j.compag.2020.105234>
13. Kaehler, A., Bradski, G.: Learning OpenCV 3. O'Reilly Media, Inc., USA (2016)
14. Garreta, R., Moncecchi, G.: Learning Scikit-Learn: Machine Learning in Python. Packt Publishing Ltd., Birmingham B3 2PB, UK (2013)
15. Heydt, M.: Learning pandas. Packt Publishing Ltd., Birmingham B3 2PB, UK (2017)
16. Kalinowski, T., Allaire, J., Chollet, F.: keras3: R Interface to Keras (2024). <https://keras.posit.co/>. Accessed 10 Feb 2024
17. Gupta, P., Bagchi, A.: Essentials of Python for Artificial Intelligence and Machine Learning. Springer, Cham
18. Hunter, J., Dale, D.: The matplotlib user's guide (2007). https://www.jick.net/Manuals/Python/matplotlib-users_guide_0.90.0.pdf. Accessed 10 Feb 2024
19. Idris, I.: Numpy Beginner's Guide. Packt Publishing Ltd., Birmingham B3 2PB, UK (2015)
20. Kim, T., Cha, Y., Shin, B., Cha, B.: Survey and performance test of python-based libraries for parallel processing. In: The 9th International Conference on Smart Media and Applications, vol. 1, pp. 154–157 (2020)
21. Slaviero, M.: Sour Pickles (2011). https://sensepost.com/cms/resources/conferences/2011/sour_pickles/BH_US_11_Slaviero_Sour_Pickles.pdf. Accessed 10 Feb 2024
22. Sapienza Università di Roma: Sunflower Dataset (2016). <https://www.diag.uniroma1.it/~labrococo/fsd/sunflowerdatasets.html>. Accessed 15 Dec 2023
23. Gupta, P., Sehgai, N.: Introduction to Machine Learning in the Cloud with Python. Springer, Cham (2021). <https://doi.org/10.1007/978-3-030-71270-9>
24. Sunil, G.C., et al.: Field-based multispecies weed and crop detection using ground robots and advanced YOLO models: a data and model-centric approach. *Smart Agric. Technol.* **9** (2024). <https://doi.org/10.1016/j.atech.2024.100538>

25. Gbenga, O., Ashi, J., Guda, B.: Performance evaluation of YOLO v5 model for automatic crop and weed classification on UAV images. *Smart Agric. Technol.* **5** (2023). <https://doi.org/10.1016/j.atech.2023.100231>
26. Dhruw, D., Sori, A.K., Tingga, S., Singh, A.: Weed detection in soybean crop using YOLO algorithm. In: Sisodia, D.S., Garg, L., Pachori, R.B., Tanveer, M. (eds.) *Machine Intelligence Techniques for Data Analysis and Signal Processing*, pp. 777–787. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-0085-5_63



Behavior Tree as a Decision Planning Algorithm for Industrial Robot

Martina Hutter-Mironovova¹(✉), Benjamin Blumhofer², Christopher Schneider¹, and Achim Wagner³

¹ YASKAWA Europe GmbH, Yaskawastraße 1, 85391 Allershausen, Germany
{martina.hutter, christopher.schneider}@yaskawa.eu

² Technologie-Initiative SmartFactory KL e. V., Trippstadter Str. 122, 67663 Kaiserslautern, Germany
benjamin.blumhofer@smartfactory.de

³ German Research Center for Artificial Intelligence GmbH (DFKI), Trippstadter Str. 122, 67663 Kaiserslautern, Germany
achim.wagner@dfki.de

Abstract. In traditional industrial settings, robot execution planning is typically governed by pre-programmed instructions, with cell logic predominantly managed by PLC (Programmable Logic Controller) systems. However, the rapid advancements in artificial intelligence have unlocked new possibilities, enabling the deployment of robots in previously unautomated sectors. Enhanced machine vision, advanced data processing, and increased adaptability to dynamic environments are now within reach. These developments necessitate a re-evaluation of conventional approaches to robot and cell programming. This paper explores the implementation of behavior trees as an alternative execution planning algorithm, specifically applied to an industrial YASKAWA robot, demonstrating their potential to optimize performance and flexibility in complex industrial applications in dynamic environment.

1 Introduction

In a typical industrial application, robot's moves and interaction with other devices are executed by the robot program. Execution of the cell functionality is typically driven by a PLC control, either separated or integrated within the robot controller. This approach is based on the principle that the robot consistently performs the same movements and actions to meet a predefined cycle time and maintain an optimal workflow as defined by the programmer. However, it does not account for changes in the working process, variations in operating conditions, environmental shifts, or the need for flexibility within the work cell [1].

With decreasing amount of available human workforce and increasing demands of manufacturers, also areas of manufacturing or production, that were never considered for automation, are recently considered for deployment of robots. In comparison to fully automated solution, robots offer flexibility due to multi degrees of freedom and can be used in production, where conditions are changing. However, until now, such automation

has been challenging to achieve and has proven costly to implement and maintain. Adding new products and functionality required an expert in robot programming or machine vision and new learning for operators.

The recent advancements in artificial intelligence have enabled the deployment of robots in industries that have historically depended on human labor due to the necessity of cognitive abilities and creativity. These sectors include food processing and bakery operations, construction sites, healthcare facilities, biomedical and chemical industries, waste management systems, and other domains where automation was previously considered infeasible. Thanks to new development in artificial intelligence (AI), machine vision and data processing can be enhanced, and robots can be used in dynamic environments with adaptive capabilities. These use cases frequently demand more adaptive robot behavior, making it essential to explore new execution planning approaches for robots and cells to manage the increased complexity. Behavior trees offer modularity and scalability as well as clearer decision making with clear visualization. Such advantages offer a different approach in execution planning compared to traditional procedural programming approaches, where varying situations are typically solved using conditions and loops, which make system much more difficult to scale up.

This paper presents the implementation of behavior trees on an industrial YASKAWA robot from the NEXT generation and shows advantages of used execution planning algorithms.

2 State of the Art

Execution planning can be implemented through three primary approaches: traditional routine programming on a CPU, enhanced programming with an external PC or GPU, and programming utilizing IoT and big data technologies [2].

With a classical programming, robot job runs directly in the robot controller. This programming is typically used in a predefined environment with known robot paths and positions, which robot needs to reach. It does not involve changing environment or need for machine vision. These programs are easy to implement and debug, typically being written in a language specific to the manufacturer.

On the other hand, when the environment around the robot is changing or the robot task is varying, it is necessary to use additional sensors to achieve required performance of the system. Such sensors include cameras, lidars, radars, tactile sensors and force and torque sensors as most common ones. Usage of those devices enlarges the capabilities of the robot and if combined with methods of artificial intelligence, they can level up the device in its performance [3]. Methods of classical machine vision are boosted up with use of classification or other forms of machine learning algorithms [4]. However, such AI algorithms require more computation power and therefore, CPU of the robot controller is no longer capable of performing such calculations in reasonable time. Usually, such algorithms would need to run on an additional machine (industrial PC) connected to the robot controller. Adding another computer adds complexity to the hardware architecture and complicates robustness of the system.

Third method mentioned above relates usage of big data, such as databases of learned models or patterns. The data is stored in a cloud, which brings not only networking delays, but also feared loss of data or cyber security related issues.

Intelligent robots are supposed to work in less-structured environments together in collaboration with humans. Such workflows or tasks are composed of sub-tasks, that can be executed independently [5]. Example would be a grasping skill of the robot, based on the object and its position. Behavior trees simplify the composition of these sub-tasks in the whole application and the order in which the sub-tasks are executed is independent from their implementation, as sub-tasks can be designed, tested, and replaced independently. Such approach allows an easy composition of trees and creation of larger trees. In traditional programming, behaviors can be created as sequences of function calls, making it more difficult to implement, reuse, or maintain individual components, which can introduce bugs, make debugging more difficult and make the system less predictable [5]. Another huge benefit of behavior tree is parallelism, where behaviors can run simultaneously within the node. Implementation of parallelism in procedural programming often requires explicit threading or concurrent operations, which adds complexity and potential synchronization issues. Implementation of behavior trees also brings ease of use for non-skilled programmers. Robots can be deployed faster, therefore lowering costs for the companies [1]. Behavior tree should be designed in the way, that is utilizing modularity of the sub-tasks as much as possible. Behavior tree must fulfil all requirements, all nodes shall be executed in correct workflow and all behaviors must be covered. In case of very complex trees, this could be more difficult to perform [5]. More standardized approach to understanding of robot skills can simplify the implementation of the behavior tree and interconnectivity of individual nodes [6]. Each skill should be designed that it receives and provides information in a structured way for universal connection to other skills. Another aspect is to achieve adaptable movement of the robot, so it would simply move between positions without collisions with its environment. Working area of the robot can be monitored (with radar, lidar, camera, etc.) and collected data transformed into virtual environment. In such digital twin, robot paths can be automatically generated based on the obstacles or other equipment in the workcell [6].

For practical use, automatic generation of the BT architecture from classical workflow description would be enormous simplification for project development and can be considered as a further work [7]. Research in this field have been performed with promising results, especially with use of reinforcement learning methods [8], simulation and reality [9].

3 Approach to Execution Planning with Behavior Trees

Behavior trees started to be used in computer game industry for modelling of NPC characters' behavior (Non-player character). In the last few years, usage of this programming method spread also to robotics area as a more expressive tool to model behavior of autonomous agents [10].

Behavior tree (BT) is designed as a composition of behaviors, that are independent from each other. Execution (tick) of a BT starts from the root (node without a parent) and propagates through top to bottom and left to right. Execution ends at leaf, which has no other child nodes underneath. The rules are set how behaviors occur, and in which order they will be executed. BT follows traditional way of data structure, where execution starts at the root in fixed direction, so it does not look back or repeat itself.

Internal nodes are called control flow nodes and leaf nodes are called execution nodes. The control flow node must have at least one child and each node has one parent. Tick is a trigger to execute the node, which returns either running, failure or success. Control flow nodes are of four types: Sequence, Fallback, Parallel, Decorator and execution nodes are either Condition or Action [10]. Various node types are explained in the Table 1.

BT used in this example is BT.CPP, a C++ 17 library, that provides a framework for construction of a behavior tree. It is written in C++ language and tree can be defined using a scripting language based on XML. Created behavior tree can be visualized and edited via GUI called Groot 2 [11]. There are variety of different libraries in C++ or Python languages available for the user. List of those can be found in this publication [12] with description and comparison analysis.

Table 1. Explanation of nodes with labelling convention. [10]

Node	Labelling	Success	Failure
Sequence (AND)		All children must return success	At least one child return failure
Fallback (OR)		One child returns success	All children return failure
Parallel		When at least M child nodes succeed	When all child nodes fail
Decorator		According to user defined policy	According to user defined policy
Condition		If is true	If is false
Action		When completed	If not possible to complete

3.1 Hardware and Software Architecture

For the implementation purposes, new generation of intelligent robots MOTOMAN NEXT of YASKAWA is used. In addition to the robot controller CPU this robot is expanded with an integrated edge computer equipped with CPU and GPU. CPU of the robot controller takes care of standard robot functionalities, while GPU allows user to run expanded functions, such as robot control service, AI service, machine vision service, path planning and obstacle avoidance service and user defined skills, tasks or services. Architecture of the NEXT controller is illustrated in Fig. 1.

With this unique and novel architecture, it is easy to achieve optimal performance and develop custom applications, that can be deployed directly inside the robot controller without the need of additional PC. Services offer APIs for easy and direct implementation of functions into the user's own code, which can be then uploaded onto the ACU (Autonomous control unit) as a containerized application (Docker container) via graphical user interface.

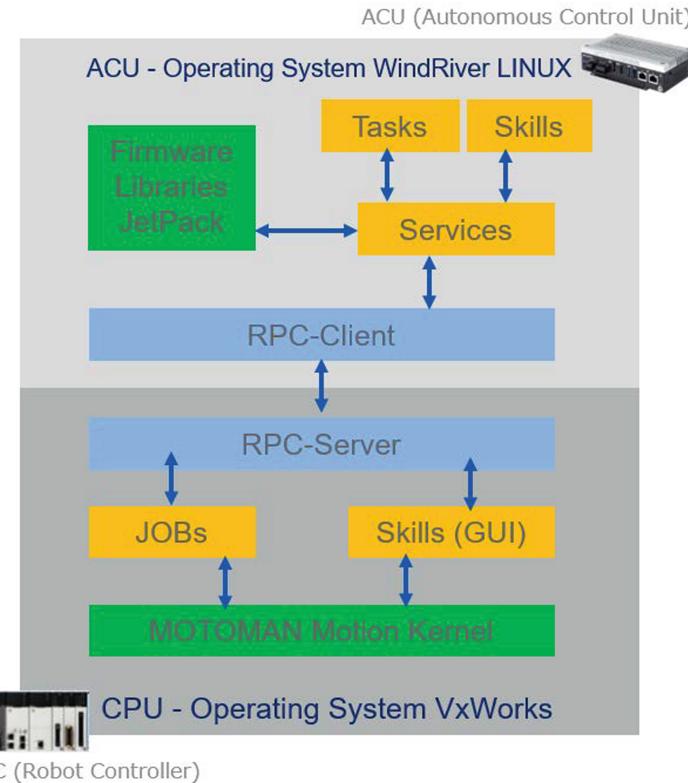


Fig. 1. Architecture of the NEXT controller. RC (Robot Controller) operating system is VxWorks, it hosts the motion kernel, which is core functionality for the robot functionality. On RC run JOBS (robot native programs) and skills user interfaces. RPC-Server serves for transferring tasks between CPU and ACU.

3.2 Description of the Robot Task

The demonstration of the pick skill designed by behavior tree is shown in Fig. 2. A Machine vision service provides the product type and 3D pose of the robot based on the recognized product type and its position in the cameras coordinate system. 3D pose is directly converted and written into the position variable in the RC as three positions and three rotations in robot coordinate system. Product type is written in variable in RC as well. Recognized products, their type and position are written in a table that is checked by the tick in decision tree. Based on the type of the product written in the table on i -th position during the i -th iteration, the gripping method is selected. Either a finger gripper or a vacuum suction cup gripper is used. This is simplified Pick Skill without error handling. Behavior tree can be either visualized in the graph form (Fig. 2) or written in XML format.

Behavior tree practically describes the workflow and actions performed by the robot. In its visual form, it is easy to understand and follow the job flow, as well as debug the relations between nodes. When the BT is created, it is necessary to pay attention to

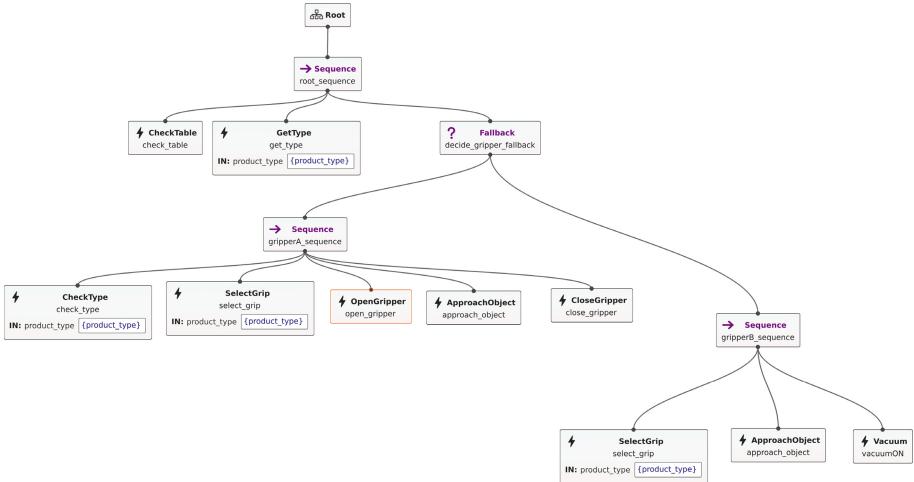


Fig. 2. Structure of the behavior tree. Individual nodes are represented as rectangles, arrows connect child nodes and parent nodes.

correct workflow and execution, because complex tasks can lead to complex structures of behavior trees. In such complexity, it might be difficult to cover all possible situations or behaviors [5]. However, one advantage of behavior trees is their ability to nest multiple trees within a single structure.

4 Implementation into the MOTOMAN NEXT

In the previous section, the structure of the behavior tree was established and executed in XML format (or visualized by Groot 2 software). In this section, an implementation of the behavior trees to the robot controller will be discussed.

Typically, behavior trees use the framework ROS (Robot Operating System) as the most common open source software development kit among companies or institutions in research and industry in robotics and automation providing an interface between the applications and the hardware. It is modular and reusable, allowing developers to create independent programs (called nodes), that can communicate with each other and be therefore re-used or shared freely among large community.

The NEXT controller from YASKAWA provides a comprehensive set of services (APIs) that offer a direct interface to the robot controller, eliminating the need for additional interface setup. This enables users to implement a library of functions and seamlessly exchange data between their applications and the robot controller, as well as across multiple applications. Users can manage variables (such as byte, integer, string, positions, etc.), initiate robot motions (linear, joint), execute robot JOBS (from the RC controller), and read or write system variables (such as servo on, start, hold, etc.). These services are fully compatible with one another, allowing for integrated functionality; for instance, the Machine Vision service can provide position data of detected objects, which the Robot Control service can process and send to the RC unit. Simultaneously,

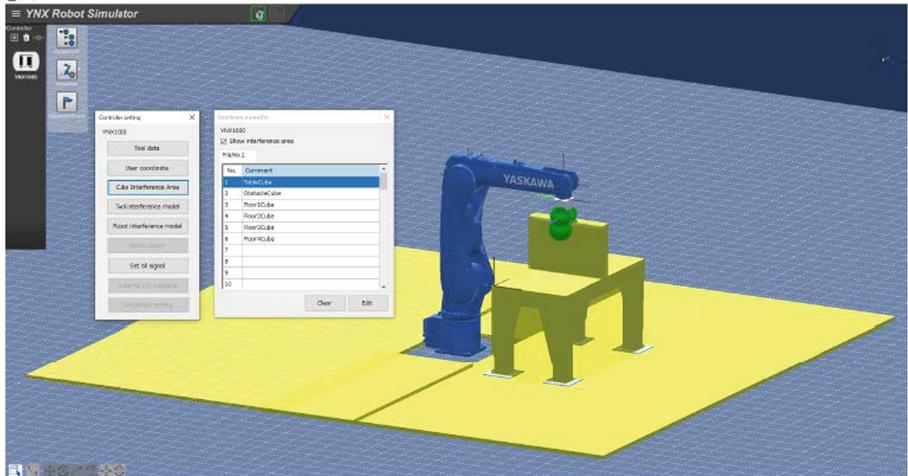


Fig. 3. Digital twin in the simulation environment YNX Robot Simulator and the surrounding environment with defined obstacles.

the Path Planning service creates an optimal trajectory for the robot between two points in space, taking into account environmental factors like obstacles, other machines, and parts. The robot follows this automatically generated collision-free path, and the path can be recalculated with each iteration of the program, enabling the robot to adapt to changing conditions in its environment (Fig. 3).

An example of such implementation of functions, that write to the robot controller variables (byte and position) and execute the robot motion are shown in the pseudo code. It shows one node implementation (this node executes a motion of the robot), creation of the tree from XML file and node registration to the tree structure. Every node has to be written separately as a function, that implements the node behavior. I.e., opening the gripper would set the variable that would trigger the external output (24 V signal) to the gripper.

```

// Import used libraries incl. APIs from YASKAWA
Import ...

// Define individual nodes and their functionality
class PickFruitBehaviourNode extends BehaviourNode {
...
virtual BehaviourNodeStatus execute() override {
    //YASKAWA functions for moving the robot,
    //writing variables, starting JOBS, etc.
    robot_controller_client_.robot_motion(...);
    robot_controller_client_.variable_byte_value_set(...);
    robot_controller_client_.job_start(...);
    return BehaviourNodeStatus();
}

// define the client for the communication with RC unit
RobotControllerClient robot_controller_client_;

}

function void main() {
// start the client for the communication with RC unit
robot_controller_client = new RobotControllerClient(...);
// generate BT from XML file
behaviour_tree = new BehaviourTree(...);
// register all nodes to BT
behaviour_tree.registerNode<PickFruitBehaviourNode>(robot
_controller_client);
behaviour_tree.execute();
}

```

The user application must be containerized and build within the ACU SDK environment. Once completed, the application can be uploaded and managed through the web-based interface of the ACU (Fig. 4).

The screenshot shows the ACU's graphical user interface. On the left is a vertical sidebar menu with tabs for Application, Skills, Path Planning, AI Model, File Explorer, Log, Settings, and logout. The main area has tabs for Application and System. Below these are two rows of application details:

App Group	App. Name	Description	Version	Status	Auto Start	
System app	RobotControlService	This is a service to communicate with robot controller and provide the APIs to access robot controller.	v1.1.0	Running	Enable	<button>Details</button> ≡
System app	PathPlanningService	This is the path planning service to generate collision free path. Please configure the environment setting before using this service.	v1.0.0	Running	Enable	<button>Details</button> ≡

Fig. 4. Graphical user interface in the ACU. Left hand side menu shows Services, Skills and Applications in the ACU, and logging and settings interfaces. Main screen displays list of all uploaded applications, their name, description, version number and status.

5 Conclusions and Further Work

This study demonstrates the effectiveness of integrating behavior trees with industrial robots, offering a flexible and adaptable approach to execution planning in dynamic manufacturing environments. Behavior trees provide a structured and intuitive method for managing complex tasks, and when combined with AI-driven machine vision, they significantly enhance the capabilities of industrial robots. The innovative architecture of the MOTOMAN NEXT robot, with its built-in ACU unit, facilitates the direct implementation of AI methods within the robot controller, eliminating the need for additional external computing resources. The APIs provided by YASKAWA further streamline the integration of native robot skills and services into custom user applications.

The current implementation serves as a foundation for further development, with plans to incorporate additional YASKAWA services, such as Machine Vision and Path Planning, to create a fully operational application. This will allow for a comprehensive evaluation of performance in real-world scenarios. Additionally, the behavior tree framework could be expanded into a user-friendly service, complete with pre-defined nodes for robot motion, gripper control, and other essential functions. This would greatly simplify the process for users to develop their own applications, enhancing the overall user experience and expanding the potential for innovation in industrial robotics.

References

1. Sidorenko, A., Rezapour, M., Wagner, A., Ruskowski, M.: Towards Using Behavior Trees in Industrial Automation Controllers (2024)
2. Naghib, A., Navimipour, N., Hosseinzadeh, M., Sharifi, A.: A comprehensive and systematic literature review on the big data management techniques in the internet of things, Springer, Wireless Networks (2022)
3. Li, Q., Wu, C., Yuan, Y., You, Y.: MSSP: A Versatile Multi-Scenario Adaptable Intelligent Robot Simulation Platform Based on LIDAR-Inertial Fusion (2024)
4. Peters, J., Tedrake, R., Roy, N., Morimoto, J.: Robot Learning (2017)
5. Colledanchise, M.: Behavior Trees in Robotics, Doctoral Thesis Stockholm, Sweden (2017)
6. Herrero, H., Moughlbay, A., Outon, J., Salle, D., Ipina, K.: Skill Based Robot Programming: Assembly, Vision and Workspace Monitoring Skill Interaction, Neurocomputing (2017)
7. Iovino, M., Smith, Ch.: Behavior Trees for Robust Task Level Control in Robotic Applications (2023)
8. Banerjee, B.: Autonomous Acquisition of Behavior Trees for Robot Control (2018)
9. French, K., Wu, S., Pan, T., Zhou, Z., Jenkins, O.Ch.: Learning Behavior Trees From Demonstration (2019)
10. Colledanchise, M., Ögren, P.: Behavior Trees in Robotics and AI: An Introduction, CRC Press, KTH (2022)
11. Faconti, D.: BehaviorTree.CPP [Software] (2019). GitHub: <https://github.com/BehaviorTree/BehaviorTree.CPP>. Accessed 21 Aug 2024
12. Ghzouli, R., Berger, T., Johnsen, E.B., Wasowski, A., Dragule, S.: Behavior Trees and State Machines in Robotics Applications (2022)



A Grey-Box Model for Real-Time Control and Monitoring

Ricardo Rodriguez-Jorge^{1,2(✉)}

¹ Ceit-Basque Research and Technology Alliance (BRTA), Manuel de Lardizábal 15,
20018 Donostia-San Sebastián, Spain
rrodriguezj@ceit.es

² Universidad de Navarra, Tecnun, Manuel de Lardizábal 13,
20018 Donostia-San Sebastián, Spain

Abstract. Grey-box models, which combine the explanatory power of first-principle models, with the ability to detect subtle patterns from data, are attracting great attention in various sectors. Simulating the dynamic process model with software sensors in parallel to the real plant, allows the automation system to observe all the modeled inner states of the plant even if they cannot be registered by sensors. In this work, software sensors serve to monitor the ammonia concentrations in the aerobic tanks of the plant line and in the anoxic tank. In addition, the software sensor tracks the maximum nitrification rate.

1 Introduction

Water is the most crucial resource for the survival of the most organisms. In the recent decades, the demand for and abuse of water resources have caused enormous pressure on the water supply [1, 2]. In contemporary industry, water is necessary. However, in the process of development, many companies use water inefficiently and even sometimes pollute it, for example by discharging wastewater unwisely. Solid wastewater treatment plants are facing growing limitations in terms of emissions and new rules for energy consumption and the conservation of resources [2]. The most critical elements in sustainable wastewater management are allowing universal access to clean water and basic sanitation services, especially wastewater treatment plants (WWTPs), which aim to improve living standards and treat emerging contaminants [3]. WWTPs are well known for their energy-intensive and costly operations due to the need for aeration and chemical additives in the system [3].

Urban and industrial waste waters are responsible for the disposal of various kinds of pollutants into other aquatic environments [2, 3]. Therefore, failure to operate a plant-specific wastewater infrastructure may create serious concerns regarding human health and environmental issues.

WWTPs are usually modeled via mechanistic models based on mathematical equations that describe the underlying physical, chemical, and biological processes. Activated sludge models (ASMs) constitute the core of biological wastewater

ater treatment process modeling and have been used for the dynamic simulation of water resource recovery facility (WRRF) operations for many years [4].

Traditional techniques are used to control WWTPs based on power consumption voltage regulators and input to alter the wastewater treatment procedure. However, control of certain efficiency strategies, such as a dynamic grey-box model with data-driven techniques, has recently been introduced as a solution for real-time prediction and control. An extended Kalman filter (EKF) is used to estimate the unknown model states of the WWTP model, as suggested in the BSM2 reference scenario. Unknown feed concentrations, plant-model mismatches, and large measurement errors are considered [5]. The EKF has usually been applied to state parameter estimation via models described by ordinary differential equations.

The aim of this paper is to develop an adaptive real-time dynamic model for advance control and development in water resource recovery facilities (WRRFs). This paper proposes a technical solution for advanced data management solutions full-scale WWTPs. In addition, the development of the model is proposed based on grey-box modeling and the extended Kalman filter (EKF). A grey-box model structure has been identified and validated under different scenarios. This model structure is then converted to an EKF to overcome the adaptivity issue.

The structure of this research paper is as follows. Section 2 illustrates the methodology applied for the development of the proposed system; in Sect. 3, the obtained results are reported and discussed; and finally, in Sect. 4, the conclusions are summarized.

2 Methodology

BSM2 implements a default control strategy based on proportional integral (PI) controllers, which maintain the dissolved oxygen concentration (So) of certain reactor tanks at the setpoint of ($2mg/L$) by modifying the oxygen transfer coefficient (K_{La}) [6]. The layout considered in the BSM2 scenario is shown in Fig. 1, where the water line and the sludge treatment modules are differentiated. In addition, different flow rates are considered in the BSM2 layout, where Q_{MLE} corresponds to the quantity of aerated flow from the last reactor tank to the first anoxic tank and where Q_{RAS} corresponds to the internal recycling flow of the sludge, which moves some sludges from the second clarifier to the first reactor tank.

In this work, a WWTP is improved when an EKF is applied as a state estimation mechanism. Accordingly, Fig. 2 gives an overview of the architecture used to develop the graphic interface for the software sensor based on the EKF model.

BSM2 implements a simulation protocol that considers the entire year of influent data. These data represent different types of weather: stormy, rainy, and dry. In addition, influent data also provide information about the nutrients present. Moreover, BSM2 must be calibrated before being able to simulate the real behavior of a WWTP facility. After validation, the grey-box model structure

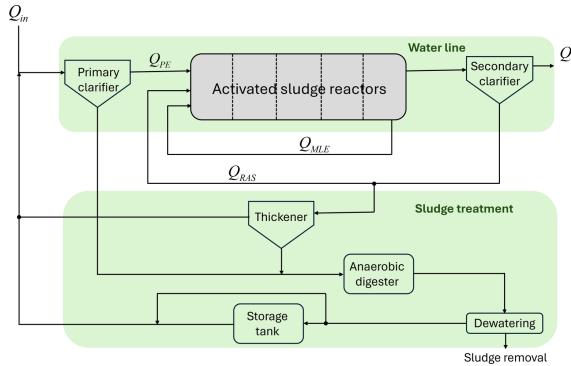


Fig. 1. BSM2 layout modeling of the architecture of a wastewater treatment plant [6]

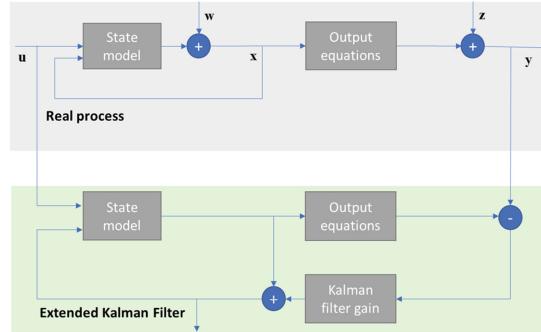


Fig. 2. Proposed methodology for the application of the extended Kalman filter

was used to develop the EKF. Simulated sensor data streams are fed into both the virtual plant and the EKF [6].

Therefore, the EKF operates as a state observer, and the online calculation of the internal states of the nonlinear dynamic system is performed [7–9]. WEST has been used as a virtual plant for generating data on the operational layout of Line 3 from the Tilburg plant. The realization pathway of the offline development of software sensors based on extended Kalman filter (EKF) models has been defined in two phases: (I) grey-box modeling, and (II) implementation of the EKF.

In phase I, input and output data under different scenarios are collected from WEST simulations, and a grey-box model structure is developed and validated. In phase (II), the grey-box model structure is converted into an EKF, which is an adaptive dynamic model, and it was developed in the Python programming language.

The virtual process that has been simulated thus far is an activated sludge process involving Line 3 from the Tilburg plant. This line is composed of an

anoxic zone, an aerobic zone with three sequential stages, and a mixed liquor recirculation from the end of the bioreactor to the anoxic zone [4,8,10–12].

2.1 Grey-Box Modeling

The development of a grey-box model usually begins with a general mathematical model that considers the time dependence of the reaction, since many environmental reactions do not occur instantaneously [13]. Equation 1 may be written to account for time-dependent transformation.

$$\text{Accumulation rate} = \text{input rate} - \text{output rate} \pm \text{transformation rate} \quad (1)$$

Time-dependent reactions are called *kinetic reactions*. The rate of transformation, or the reaction rate ρ , is used to describe the rate of formation or disappearance of a substance or chemical species. For biochemical reactions, one reactor, can be described by the mass balance for substrates and the biological reaction rate (see Eq. 1).

The mathematical model for biochemical reactors, as described by Eq. 1, is described in Eq. 2.

$$\frac{dC}{dt} = \frac{1}{V} \cdot (Q_{in} \cdot C_{in} - Q_{out} \cdot C_{out}) + \rho \quad (2)$$

where: $\frac{dC}{dt}$ represents the net concentration change rate of the substrate, V represents the reactor volume, Q_{in} and Q_{out} represent flows in and out of the reactor, respectively, C_{in} and C_{out} represent the substrate concentrations in and out of the reactor, respectively, and ρ represents the biochemical reaction rate.

2.1.1 Identification and Validation via Nonlinear Least Squares

The parameters in the grey-box model are estimated via nonlinear least squares [14,15]. The least-square method is defined as follows:

Given N data points, $[x_i, y(x_i)] = (x_i, y_i)$, the functional form of the approximating function is chosen to fit, $\tilde{y} = y(x)$, and minimize the sum of the squares of the deviations, $e_i = (y_i - \tilde{y}_i)$, where the vector \mathbf{x} denotes the $(m \times 1)$ dimension, as shown in Eq. 3.

$$\mathbf{x} = [x_0 \ x_1 \ \dots \ x_m]^T \quad (3)$$

To calculate weights via the least-squares method, the square error criterion (Eq. 4) between the neural outputs and the targets is considered [16,17].

$$\mathbf{Q} = \sum_{k=1}^N (y(k) - \tilde{y}(k))^2 \quad (4)$$

To calculate weights via the classical least-squares method, the set of Eqs. (see Eq. 5) is solved.

$$\frac{\partial \mathbf{Q}}{\partial \mathbf{w}_{\text{col}}} = \mathbf{0} \quad (5)$$

where **wcol** stands for the column-vector form of the weight matrix shown in Eq. 6, and **0** is a zero vector whose length is the total number of weights, and that represents the set of equations in a simplified notation, as shown in Eq. 7.

$$\mathbf{W} = \begin{bmatrix} w_{0,0} & w_{0,1} & \dots & w_{0,n} \\ 0 & w_{1,1} & \dots & w_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & w_{n,n} \end{bmatrix} \quad (6)$$

$$(\mathbf{y} - \mathbf{xrow} \cdot \mathbf{wcol})^T \cdot \frac{\partial(\mathbf{y} - \mathbf{xrow} \cdot \mathbf{wcol})}{\partial \mathbf{wcol}} = \mathbf{0} \quad (7)$$

Equation 7 can be rewritten in a matrix way as Eq. 8.

$$(\mathbf{y} - \mathbf{xrow} \cdot \mathbf{wcol})^T \cdot (-\mathbf{xrow}) = (-\mathbf{y}^T \cdot \mathbf{xrow} + \mathbf{wrow} \cdot \mathbf{xcol} \cdot \mathbf{xrow}) = \mathbf{0} \quad (8)$$

Thus, the least-squares method calculation of weights for a polynomial of *order* = 2 in a long-row-vector form is obtained as Eq. 9, or in a long-column-vector form as Eq. 10.

$$\mathbf{wrow} = \mathbf{y}^T \cdot \mathbf{xrow} \cdot (\mathbf{xcol} \cdot \mathbf{xrow})^{-1} \quad (9)$$

$$\mathbf{wcol} = (\mathbf{xcol} \cdot \mathbf{xrow})^{-1} \cdot \mathbf{xcol} \cdot \mathbf{y} \quad (10)$$

The previous equations are considered for direct calculation of weights via the least-squares method; they imply the existence of a unique solution, i.e., a unique global minimum [14, 15]. For the solution, the Levenberg-Marquardt (L-M) weight update algorithm is derived for a static polynomial function of *order* = 2, i.e., for a static quadratic neural unit [14, 15]. This weight-update algorithm, in its simplest form, is derived as Eq. 11.

$$\Delta \mathbf{wcol} = (\mathbf{xcol} \cdot \mathbf{xrow} + \frac{1}{\mu} \cdot \mathbf{I})^{-1} \cdot \mathbf{xcol} \cdot \mathbf{e} \quad (11)$$

where:

$\mathbf{xrow} = \mathbf{xcol}^T$ represents the Jacobian matrix, μ is the learning rate, and \mathbf{e} is the vector of neural output errors. All weights might appear to be linear; therefore the adaptation weights can be transformed into a state-space form. Therefore, the vector **wcol** has been introduced to perform the weight adaptation stability [14, 15].

2.1.2 Scenario Analysis

The anoxic zone and aerobic zones are represented as continuously stirred tank reactors (CSTRs) in series, each with a specific volume. The primary and secondary clarifiers were modeled as ideal separators. The primary recirculated activated sludge flow rate (Q_{RAS}), the internal recirculation flow rate (Q_{MLE}) and the primary effluent were set to specific flow rates in cubic meters per day. The initial ammonia loading was varied on a dynamic basis. Dissolved oxygen

levels (SO_2) at each aerobic reactor were first static loads, and later, testing was performed by varying the SO_2 level. The soluble ammonia signal was used as an input to the plant. The dissolved oxygen (SO_2) and soluble ammonia concentrations (S_{NH}) in the aerobic tanks were extracted from WEST for model fitting and evaluation.

Several scenarios have been considered to evaluate the ability of the model structure to estimate ammonia concentrations in the virtual plant model and to identify when model parameters need to be adjusted. The primary effluent flow (Q_{PE}) has been designed to have a daily pattern with small shifts, and the ammonia concentration (S_{NH_0}) input to the plant line has been designed with fluctuations.

2.2 Extended Kalman Filter

A general discrete nonlinear grey-box model with noise can be written in state-space form, as shown in Eq. 12.

$$\begin{aligned}\dot{\mathbf{x}}_k &= \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1} \\ \mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{y}_k\end{aligned}\quad (12)$$

where:

$\mathbf{x} \in \mathbb{R}^m$ is the state, $\mathbf{z}_k \in \mathbb{R}^n$ is the measurement, \mathbf{u}_k is the input at time step k , $\mathbf{w}_{k-1} \sim (\mathbf{0}, \mathbf{W})$ is the Gaussian process noise, $\mathbf{f}(\cdot)$ is the dynamic model function, and $\mathbf{h}(\cdot)$ is the measurement model function, $(\cdot)_k$ denotes variables at time step k .

The EKF is a recursive method for state estimation of nonlinear process models and measurement models (Eq. 12), it is an optimal state estimator in the sense that it minimizes the error variance. The estimation can be linearized around the current estimate using the partial derivatives of the process and measurement functions to compute estimates even in the face of nonlinear relationships.

In this case, the nonlinear function \mathbf{f} in the difference Eq. 12 relates the state at the previous time step $k-1$ to the state at the current time step k . It includes as parameters any driving function \mathbf{u}_{k-1} and the zero-mean process noise \mathbf{w}_{k-1} . The nonlinear function \mathbf{h} in the measurement equation from Eq. 12 relates the state \mathbf{x}_k to the measurement \mathbf{z}_k . The general steps for the EKF update equations are listed in the sequence of their implementation below (see Eq. 13–Eq. 17):

$$\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}^-, \mathbf{u}_{k-1}) \quad (13)$$

$$\mathbf{P}_k^- = \mathbf{F}_k \cdot \mathbf{P}_{k-1} \cdot \mathbf{F}_k^T + \mathbf{W} \quad (14)$$

$$\mathbf{K}_k = \mathbf{P}_k^- \cdot \mathbf{H}_k^T \cdot (\mathbf{H}_k \cdot \mathbf{P}_k^- \cdot \mathbf{H}_k^T + \mathbf{V})^{-1} \quad (15)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \cdot (\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-, \mathbf{u}_k)) \quad (16)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \cdot \mathbf{H}_k) \cdot \mathbf{P}_k^- \quad (17)$$

where:

- $\hat{\mathbf{x}}_k^-$ and $\hat{\mathbf{x}}_k$ denote a prior estimate and a posterior estimate of the state, respectively.
- \mathbf{F}_k and \mathbf{H}_k stand for the Jacobian matrix of partial derivatives of \mathbf{f} and \mathbf{h} with respect to \mathbf{x} which are evaluated with estimates (\mathbf{x}) and input \mathbf{u} at time step k , that is, (see Eq. 18 - Eq. 19):

$$\mathbf{F}_k = \frac{\partial}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}, \mathbf{u})|_{\mathbf{x}=\hat{\mathbf{x}}_k^-, \mathbf{u}=\hat{\mathbf{u}}_k} \quad (18)$$

$$\mathbf{H}_k = \frac{\partial}{\partial \mathbf{x}} \mathbf{h}(\mathbf{x}, \mathbf{u})|_{\mathbf{x}=\hat{\mathbf{x}}_k^-, \mathbf{u}=\hat{\mathbf{u}}_k} \quad (19)$$

- $\hat{\mathbf{x}}_k$ and \mathbf{I} are the Kalman filter gain and the identity matrix, respectively.
- \mathbf{P}_k^- and \mathbf{P}_k stand for the covariance matrices of a prior and a posterior estimation error, respectively.
- \mathbf{W} represents the covariance matrix of the process noise, and \mathbf{V} represents the covariance matrix of the measurement noise.

2.2.1 Tuning the Extended Kalman Filter

Both the mathematical structure of the internal states of nonlinear dynamic systems and the properties of the stochastic disturbances are known. The term “filter” indicates that the algorithm can reduce measurement noise and other disturbances of stochastic data.

In the EKF, the covariance rate and trade-offs between the plant line and sensor noise depend on the two tuning parameters, the process covariance matrix, and the measurement noise covariance matrix, which are the error covariances of the state and measurements. In practical applications, these are not known precisely; therefore, trial-and-error tuning is expected [18].

The estimation problem is extended to nonlinear process that results from a nonlinear state space model. The soft sensor mathematical structure can be defined as Eq. 20.

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_m \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_p \end{bmatrix} \quad (20)$$

where:

- \mathbf{x} is composed of the S_{NH} in the anoxic tank, S_{NH} concentrations in the i aerobic tank, and the r observed nitrification rate, which is calculated recursively.
- \mathbf{u} contains the input disturbance signals and the input controllable signals.

- \mathbf{p} represents the parameter vector composed of the volume of each reactor, the half saturation coefficient for the oxygen concentration in each aerobic reactor, the half saturation coefficient for the ammonia concentration, and the maximum ammonia change rate.

For the EKF algorithm, knowledge about the observed process must be provided a priori.

3 Results and Discussion

Reasonable estimations of the ammonia concentrations in the anoxic and aerobic tanks are observed as the estimation curves converge to the simulated “true” values in WEST. The noisy fluctuation is due to measurement noise propagating through the EKF. Fine-tuning or extra filtering may further reduce such noise. The covariance matrices are constantly being updated, which means that the error is growing, shrinking, and changing. By plotting the process covariance matrix, how the EKF work and whether it performs the way needed can be determined. Figure 3 presents the process covariance evolution in discrete time (k). This provides a clearer idea that the lowest covariance corresponds to the measured state variable. Therefore, it can be estimated with more precision.

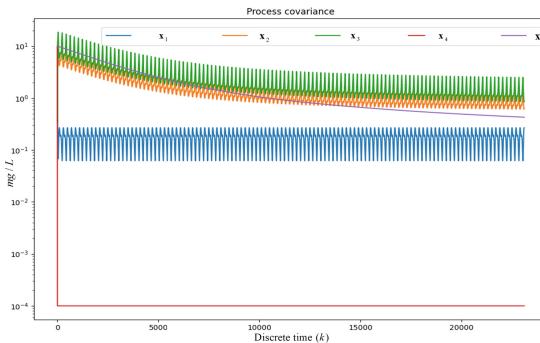


Fig. 3. Process covariance matrix

Moreover, the maximal nitrification rate, r , starts dropping from the first days, indicating tracking of the r value.

Figure 4 presents the Kalman gain values for each state variable during convergence for process covariance. The lower the Kalman gain is, the more the model's measurement fulfills the prediction.

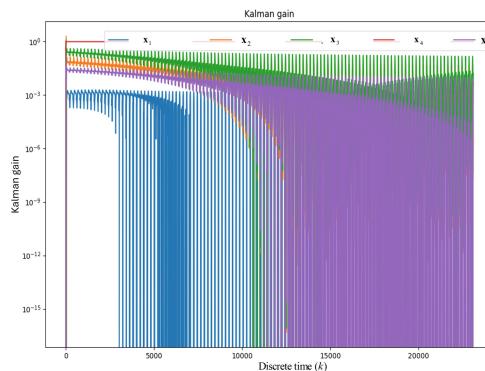


Fig. 4. Kalman gain values during convergence for state model covariance

4 Conclusions

In this paper, a method has been presented for the state estimation of a nonlinear system described by a class of differential-algebraic models via the EKF. The development of a grey-box model to adaptively estimate the EKF relies on an adequate grey-box model structure. The grey-box model is based on the mass balance and reaction rates governed by Monod kinetics. Although the grey-box model lacks many traditional model components (e.g., biomass concentration and COD) compared with state-of-art ASM models, the adaptive scheme can incorporate impacts from those neglected variables into the r values, ensuring the accuracy of the model. One benefit of this simplicity is the increased applicability to other processes, especially those that are not fully understood yet.

Future work will include the simulation of more scenarios for EKF calibration and performance evaluation metrics and implementation.

Acknowledgements. This work was funded by the European Union’s Horizon Europe Research and Innovation Program within the project DARROW (grant agreement 101070080). The views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

The author would also like to thank Eduardo Ayesa, Beñat Elduayen-Echave, Sofia Jaray Valdehierro, Itxaro Errandonea and Saioa Arrizabalaga from the Ceit-Basque Research and Technology Alliance (BRTA) and Universidad de Navarra, Tecnun, for insightful consultations and for providing me with the required data.

References

1. Yu, X., Chen, S., Zhang, X., Wu, H., Guo, Y., Guan, J.: Research progress of the artificial intelligence application in wastewater treatment during 2012–2022: a bibliometric analysis. *Water Sci. Technol.* **88**(7), 1750–1766 (2023). <https://doi.org/10.2166/wst.2023.296>

2. Hasan, M.A.: An emergent addition for the optimal systemization of wastewater utilization plants using artificial intelligence. *Water Sci. Technol.* **84**(10–11), 2805–2817 (2021). <https://doi.org/10.2166/wst.2021.203>
3. Heo, S., Nam, K., Tariq, S., Lim, J.Y., Park, J., Yoo, C.: A hybrid machine learning-based multi-objective supervisory control strategy of a full-scale wastewater treatment for cost-effective and sustainable operation under varying influent conditions. *J. Clean. Prod.* **291**, 125853 (2021). <https://doi.org/10.1016/j.jclepro.2021.125853>
4. Daneshgar, S., et al.: A full-scale operational digital twin for a water resource recovery facility-a case study of Eindhoven water resource recovery facility. *Water Environ. Res.* **96**(3), e11016 (2024). <https://doi.org/10.1002/wer.11016>
5. Busch, J., et al.: State estimation for large-scale wastewater treatment plants. *Water Res.* **47**(13), 4774–4787 (2013). <https://doi.org/10.1016/j.watres.2013.04.007>
6. Pisa, I., Santín, I., Morell, A., Vicario, J.L., Vilanova, R.: LSTM-based wastewater treatment plants operation strategies for effluent quality improvement. *IEEE Access* **7**, 159773–159786 (2019). <https://doi.org/10.1109/ACCESS.2019.2950852>
7. Yang, C., Seiler, P., Belia, E., Daigger, G.T.: An adaptive real-time grey-box model for advanced control and operations in WRRFs. *Water Sci. Technol.* **84**(9), 2353–2365 (2021). <https://doi.org/10.2166/wst.2021.408>
8. Schneider, M.Y., Carbajal, J.P., Furrer, V., Sterkele, B., Maurer, M., Villez, K.: Beyond signal quality: the value of unmaintained pH, dissolved oxygen, and oxidation-reduction potential sensors for remote performance monitoring of on-site sequencing batch reactors. *Water Res.* **161**, 639–651 (2019). <https://doi.org/10.1016/j.watres.2019.06.007>
9. Seshan, S., Vries, D., van Duren, M., van der Helm, A., Poinapen, J.: AI-based validation of wastewater treatment plant sensor data using an open data exchange architecture. In: IOP Conference Series: Earth and Environmental Science, vol. 1136, no. 1, p. 012055 (2023)
10. Rodríguez-Vidal, F.J., et al.: Monitoring the performance of wastewater treatment plants for organic matter removal using excitation-emission matrix fluorescence. *Microchemical J.* **175**, 107177 (2022). <https://doi.org/10.1016/j.microc.2022.107177>
11. Singh, N.K., Yadav, M., Singh, V., Padhiyar, H., Kumar, V., Bhatia, S.K., Show, P.L.: Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. *Bioresource Technol.* **369**, 128486 (2023). <https://doi.org/10.1016/j.biortech.2022.128486>
12. Ravi, N., Johnson, D.P.: Artificial intelligence based monitoring system for onsite septic systems failure. *Process Saf. Environ. Prot.* **148**, 1090–1097 (2021)
13. Davis, M.L., Masten, S.J.: Principles of Environmental Engineering and Science. McGraw-Hill Higher Education (2004). 9780072921861
14. Rodríguez Jorge, R.: Lung tumor motion prediction by neural networks. Ph.D. thesis. Czech Technical University in Prague, Czech Republic (2012)
15. Rodriguez-Jorge, R., Bila, J., Mizera-Pietraszko, J., Martínez-García, E.A.: Weight adaptation stability of linear and higher-order neural units for prediction applications. In: Choroś, K., Kopel, M., Kukla, E., Siemiński, A. (eds.) MISSI 2018. AISC, vol. 833, pp. 503–511. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-98678-4_50
16. Rodriguez-Jorge, R., Bila, J.: Cardiac arrhythmia prediction by adaptive analysis via Bluetooth. *MENDEL* **26**(2), 29–38 (2020). <https://doi.org/10.13164/mendel.2020.2.029>

17. Rodríguez-Jorge, R., Sánchez-Pérez, L., Bila, J., Škvor, J.: Rotating machinery fault diagnosis using a quadratic neural unit. *Int. J. Grid Util. Comput.* **13**(2–3), 309–319 (2022). <https://doi.org/10.1504/ijguc.2022.124403>
18. López Arenas, T., Pulis, A., Baratti, R.: On-line monitoring of a biological process for wastewater treatment. *Revista Mexicana de Ingeniería Química* **3**(1), 51–36 (2004). 1665-2738

Author Index

A

- Aloisio, Alessandro 209, 221
Al-Turjman, Fadi 67
Ampririt, Phudit 133
Armengol, Eva 25

B

- Babuc, Diogen 109
Barolli, Leonard 121, 133
Bee, K. 374
Bigiotti, Alessandro 258
Blumhofer, Benjamin 385

C

- Cacciagrano, Diletta 47, 67, 209
Canters, Jeoffrey 356
Carmona, J. C. 374
Carmona-Frausto, Jesús C. 364
Cervantes, S. 374
Chaudhary, Pradeep 47
Christopoulou, Maria 78
Ciampi, Mario 187
Cirillo, Egidia 1
Cisneros, Matteo 318, 327

D

- Damiano, Emanuele 187
De Cock, Laure 336
De Swert, Tamara 336
Debnath, Naren 142
Di Martino, Beniamino 25

F

- Fonisto, Mattia 1
Fortiș, Alexandra-Emilia 13, 109
Fortiș, Teodor-Florin 13, 245
Fukushima, Yumemi 282

G

- Ghobadi, Sajjad 270
Giacalone, Marco 1

H

- Hellinckx, Peter 318, 327, 347, 356
Higashi, Shunya 133
Hosoda, Yoshiki 176
Houben, Pieter Jan 347, 356
Hutter-Mironovova, Martina 385

I

- Ikeda, Makoto 133
Ishida, Tomoyuki 282, 292, 302

J

- Jaber, Aws 78
Jacobs, Stef 347

K

- Kamiya, Riku 99
Kathala, Krishna Chaitanya Rao 35
Kawakami, Tomoya 99
Kohana, Masaki 154, 165, 176
Koufos, Ioannis 78
Kushida, Takayuki 199
Kusmirczak, Thomas 336

M

- Massobrio, Renzo 327, 347, 356
Matsuo, Keita 133
Mazzante, Gianmarco 233
Mendez, Gerardo Maximiliano 364
Mercelis, Siegfried 336
Meuris, Lukas 318
Mexicano, A. 374
Mexicano-Santoyo, Adriana 364
Moccardi, Alberto 1
Montes, P. N. 374
Montes-Dorantes, Pascual Noradino 364
Mostarda, Leonardo 47, 233
Mukhopadhyay, Sajal 142
Mulayim, Oguz 25
Murthy, Shruthi Sreenivasa 35

N

- Naeem, Hamad [47](#)
Nakamura, Aino [154](#)
Nakamura, Shigenari [121](#)
Natwichai, Juggapong [59](#)
Navarra, Alfredo [233](#)
Nishino, Koichi [292](#)

O

- Oohashi, Riko [302](#)

P

- Pezzullo, Gennaro Junior [25](#)
Piselli, Francesco [270](#)
Puiu, Ionica-Larisa [245](#)

Q

- Qafzezi, Ermioni [133](#)

R

- Rahimiasl, Mohammadmahdi [336](#)
Rodriguez-Jorge, Ricardo [395](#)
Rommens, Oliver [327](#)

S

- Sakamoto, Shinji [121](#)
Salama, Ramiz [67](#)
Sato, Miyu [165](#)
Schneider, Christopher [385](#)
Sestili, Davide [233](#)
Shah, Purav [258](#)

Sicuranza, Mario [187](#)

Silvestri, Stefano [187](#)

Sugita, Kaoru [312](#)

Sutinaraphan, Sutipong [59](#)

T

- Tabari, Hossein [318, 347](#)
Takahashi, Futa [199](#)
Takizawa, Makoto [121](#)
Trestian, Ramona [258](#)
Tricomi, Giuseppe [187](#)

U

- Ullah, Farhan [47](#)
Ullah, Shamsher [47](#)

V

- Van Ginderachter, Michiel [318](#)
Van Poecke, Aaron [318](#)
Vanderhoydonc, Ynte [336](#)
Verhaert, Ivan [347](#)

W

- Wagner, Achim [385](#)

X

- Xhafa, Fatos [142](#)

Z

- Zhang, Guangli [35](#)
Zhao, Yue [47](#)