

Álvaro Rocha · Hojjat Adeli ·
Gintautas Dzemyda ·
Fernando Moreira ·
Ana Maria Ramalho Correia *Editors*

Trends and Applications in Information Systems and Technologies

Volume 1

Advances in Intelligent Systems and Computing

Volume 1365

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Advisory Editors

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India

Rafael Bello Perez, Faculty of Mathematics, Physics and Computing,
Universidad Central de Las Villas, Santa Clara, Cuba

Emilio S. Corchado, University of Salamanca, Salamanca, Spain

Hani Hagras, School of Computer Science and Electronic Engineering,
University of Essex, Colchester, UK

László T. Kóczy, Department of Automation, Széchenyi István University,
Gyor, Hungary

Vladik Kreinovich, Department of Computer Science, University of Texas
at El Paso, El Paso, TX, USA

Chin-Teng Lin, Department of Electrical Engineering, National Chiao
Tung University, Hsinchu, Taiwan

Jie Lu, Faculty of Engineering and Information Technology,
University of Technology Sydney, Sydney, NSW, Australia

Patricia Melin, Graduate Program of Computer Science, Tijuana Institute
of Technology, Tijuana, Mexico

Nadia Nedjah, Department of Electronics Engineering, University of Rio de Janeiro,
Rio de Janeiro, Brazil

Ngoc Thanh Nguyen, Faculty of Computer Science and Management,
Wrocław University of Technology, Wrocław, Poland

Jun Wang, Department of Mechanical and Automation Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).

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

More information about this series at <http://www.springer.com/series/11156>

Álvaro Rocha · Hojjat Adeli ·
Gintautas Dzemyda · Fernando Moreira ·
Ana Maria Ramalho Correia
Editors

Trends and Applications in Information Systems and Technologies

Volume 1



Springer

Editors

Álvaro Rocha
ISEG
University of Lisbon
Lisbon, Portugal

Hojjat Adeli
College of Engineering
The Ohio State University
Columbus, OH, USA

Gintautas Dzemyda
Institute of Data Science
and Digital Technologies
Vilnius University
Vilnius, Lithuania

Fernando Moreira
DCT
Universidade Portucalense
Porto, Portugal

Ana Maria Ramalho Correia
Department of Information Sciences
University of Sheffield
Lisbon, Portugal

ISSN 2194-5357

ISSN 2194-5365 (electronic)

Advances in Intelligent Systems and Computing

ISBN 978-3-030-72656-0

ISBN 978-3-030-72657-7 (eBook)

<https://doi.org/10.1007/978-3-030-72657-7>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021

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

Preface

This book contains a selection of papers accepted for presentation and discussion at the 2021 World Conference on Information Systems and Technologies (WorldCIST'21). This conference had the scientific support of the University of Azores, Information and Technology Management Association (ITMA), IEEE Systems, Man, and Cybernetics Society (IEEE SMC), Iberian Association for Information Systems and Technologies (AISTI), and Global Institute for IT Management (GIIM). It took place online at Hangra do Heroismo city, Terceira Island, Azores, Portugal, March 30–31 to April 1–2, 2021.

The World Conference on Information Systems and Technologies (WorldCIST) is a global forum for researchers and practitioners to present and discuss recent results and innovations, current trends, professional experiences, and challenges of modern information systems and technologies research, technological development, and applications. One of its main aims is to strengthen the drive toward a holistic symbiosis between academy, society, and industry. WorldCIST'21 built on the successes of WorldCIST'13 held at Olhão, Algarve, Portugal; WorldCIST'14 held at Funchal, Madeira, Portugal; WorldCIST'15 held at São Miguel, Azores, Portugal; WorldCIST'16 held at Recife, Pernambuco, Brazil; WorldCIST'17 held at Porto Santo, Madeira, Portugal; WorldCIST'18 held at Naples, Italy; WorldCIST'19 held at La Toja, Spain; and WorldCIST'20, which took place online at Budva, Montenegro.

The Program Committee of WorldCIST'21 was composed of a multidisciplinary group of 309 experts and those who are intimately concerned with information systems and technologies. They have had the responsibility for evaluating, in a ‘blind review’ process, the papers received for each of the main themes proposed for the conference: A) information and knowledge management; B) organizational models and information systems; C) software and systems modeling; D) software systems, architectures, applications and tools; E) multimedia systems and applications; F) computer networks, mobility and pervasive systems; G) intelligent and decision support systems; H) big data analytics and Applications; I) human-computer interaction; J) ethics, computers and security; K) health informatics;

L) information technologies in education; M) information technologies in radio-communications; N) technologies for biomedical applications.

The conference also included workshop sessions taking place in parallel with the conference ones. Workshop sessions covered themes such as healthcare information systems interoperability, security and efficiency; user expression and sentiment analysis; gamification application and technologies; code quality and security; amalgamating artificial intelligence and business innovation; innovation and digital transformation for rural development; automatic detection of fake news in social media; open learning and inclusive education through information and communication technology; digital technologies and teaching innovations in COVID-19 times; devops and software engineering; pervasive information systems; advancing eHealth through software engineering fundamentals; blockchain and distributed ledger technology (DLT) in business; innovation and intelligence in educational technology, evolutionary computing for health care; ICT for auditing and accounting; and leveraging customer behavior using advanced data analytics and machine learning techniques.

WorldCIST'21 received about 400 contributions from 51 countries around the world. The papers accepted for oral presentation and discussion at the conference are published by Springer (this book) in four volumes and will be submitted for indexing by WoS, EI-Compendex, Scopus, DBLP, and/or Google Scholar, among others. Extended versions of selected best papers will be published in special or regular issues of relevant journals, mainly JCR/SCI/SSCI and Scopus/EI-Compendex indexed journals.

We acknowledge all of those that contributed to the staging of WorldCIST'21 (authors, committees, workshop organizers, and sponsors). We deeply appreciate their involvement and support that was crucial for the success of WorldCIST'21.

March 2021

Álvaro Rocha
Hojjat Adeli
Gintautas Dzemyda
Fernando Moreira

Contents

Intelligent and Decision Support Systems

Latent Impacts on Digital Technologies Sustainability Assessment and Development	3
Egils Ginters and Emils Dimitrovs	
Using Kinect to Detect Gait Movement in Alzheimer Patients	14
David Castillo-Salazar, Laura Lanzarini, Cesar Guevara, and Héctor Gómez Alvarado	
The Analysis of Triples of Triangular Norms for the Subject Area of Passenger Transport Logistics	29
Zbigniew Suraj, Oksana Olar, and Yurii Bloshko	
Real-Time Stair Detection Using Multi-stage Ground Estimation Based on KMeans and RANSAC	39
Yuchen Li, Lina Yang, and Patrick Shen-Pei Wang	
Artificial Intelligence Based Strategy for Vessel Decision Support System	49
Andrius Daranda and Gintautas Dzemyda	
Improved Multi-scale Fusion of Attention Network for Hyperspectral Image Classification	59
Fengqi Zhang, Lina Yang, Hailong Su, and Patrick Shen-Pei Wang	
Predictive Models in the Assessment of Tax Fraud Evidences	69
Fabiola Cristina Venturini and Ricardo Mattos Chaim	
Mobile Manipulator Robot Control Through Virtual Hardware in the Loop	80
Byron S. Jorque, Jéssica D. Mollocana, Jessica S. Ortiz, and Víctor H. Andaluz	

Optimizing Regularized Multiple Linear Regression Using Hyperparameter Tuning for Crime Rate Performance Prediction	92
Alexandra Vultureanu-Albișăi and Costin Bădică	
Modelling a Deep Learning Framework for Recognition of Human Actions on Video	104
Flávio Santos, Dalila Durães, Francisco Marcondes, Marco Gomes, Filipe Gonçalves, Joaquim Fonseca, Jochen Wingbermuehle, José Machado, and Paulo Novais	
Torque Control of a Robotic Manipulator Joint Using LQG and LMI-Based Strategies with LTR	113
José N. N. Júnior, Gabriel F. Machado, Darielson A. Souza, Josias G. Batista, Ismael S. Bezerra, Antônio B. S. Júnior, Fabrício G. Nogueira, and Bismark C. Torrico	
Forecasting the Retirement Age: A Bayesian Model Ensemble Approach	123
Jorge M. Bravo and Mercedes Ayuso	
Using Bayesian Dialysis and Tetrads to Detect the Persistent Characteristics of Fraud	136
Ignacio González García and Alfonso Mateos	
Benchmark of Encoders of Nominal Features for Regression	146
Diogo Seca and João Mendes-Moreira	
Optimizing Model Training in Interactive Learning Scenarios	156
Davide Carneiro, Miguel Guimarães, Mariana Carvalho, and Paulo Novais	
Early Prediction of student's Performance in Higher Education: A Case Study	166
Mónica V. Martins, Daniel Toledo, Jorge Machado, Luís M. T. Baptista, and Valentim Realinho	
Object Detection in Rural Roads Through SSD and YOLO Framework	176
Luis Barba-Guaman, Jose Eugenio Naranjo, Anthony Ortiz, and Juan Guillermo Pinzon Gonzalez	
Study of MRI-Based Biomarkers on Patients with Cerebral Amyloid Angiopathy Using Artificial Intelligence	186
Fátima Solange Silva, Tiago Gil Oliveira, and Victor Alves	
One-Pixel Attacks Against Medical Imaging: A Conceptual Framework	197
Tuomo Sipola and Tero Kokkonen	

Big Data Analytics and Applications

- Implementation of Big Data Analytics Tool in a Higher Education Institution** 207

Tiago Franco, P. Alves, T. Pedrosa, M. J. Varanda Pereira, and J. Canão

- Big Data in Policing: Profiling, Patterns, and Out of the Box Thinking** 217

Sónia M. A. Morgado and Sérgio Felgueiras

- Research Trends in Customer Churn Prediction: A Data Mining Approach** 227

Zhang Tianyuan and Sérgio Moro

- Roll Padding and WaveNet for Multivariate Time Series in Human Activity Recognition** 238

Rui Gonçalves, Fernando Lobo Pereira, Vitor Miguel Ribeiro, and Ana Paula Rocha

- Automatic Classifier of Scientific Contents** 249

Samuel Machado and Jorge Oliveira e Sá

- Crowdsourced Data Stream Mining for Tourism Recommendation** 260

Fátima Leal, Bruno Veloso, Benedita Malheiro, and Juan C. Burguillo

- W-core Transformer Model for Chinese Word Segmentation** 270

Hai Lin, Lina Yang, and Patrick Shen-Pei Wang

Computer Networks, Mobility and Pervasive Systems

- Implementation and Evaluation of WBBR in ns-3 for Multipath Networks** 283

Thalia Mijas-Abad, Patricia Ludeña-González, Francisco Sandoval, and Rommel Torres

- Trends, Challenges and Opportunities for IoT in Smallholder Agriculture Sector: An Evaluation from the Perspective of Good Practices** 293

C. Alexandra Espinosa, Jhon Pineda, Oscar Ortega, Astrid Jaime Author, Román Sarmiento, and George Washington Archibald Taylor

- Optical Wireless Communication Applications and Progress to Ubiquitous Optical Networks** 302

Simona Riurean, Monica Leba, Andreea Ionica, and Álvaro Rocha

- The Perspective of Cyclists on Current Practices with Digital Tools and Envisioned Services for Urban Cycling** 313

Inês Fortes, Diana Pinto, Joana Vieira, Ricardo Pessoa, and Rui José

Enhancing the Motorcycling Experience with Social Applications: A Study of User Needs	323
Inês Fortes, Diana Pinto, Emanuel Sousa, Vera Vilas-Boas, and Rui José	
Maximizing Sensors Trust Through Support Vector Machine	333
Sami J. Habib and Paulvanna N. Marimuthu	
Spreading Factor Analysis for LoRa Networks: A Supervised Learning Approach	344
Christos Bouras, Apostolos Gkamas, Spyridon Aniceto Katsampiris Salgado, and Nikolaos Papachristos	
 Ethics, Computers and Security	
Web Guard	357
Mohamed Haoud, Raid Djehiche, and Lalia Saoudi	
Filters that Fight Back Revisited: Conceptualization and Future Agenda	365
Sampsaa Rauti and Samuli Laato	
Malware Security Evasion Techniques: An Original Keylogger Implementation	375
Álvaro Arribas Royo, Manuel Sánchez Rubio, Walter Fuertes, Mauro Callejas Cuervo, Carlos Andrés Estrada, and Theofilos Toulkeridis	
Legal Ethical Implications in the Exercise of Communication and Information Technologies (ICT) in Telemedicine and e-Law in Medellín - Colombia	385
José Antonio García Pereáñez and David Alberto García Arango	
A Proposal for Artificial Moral Pedagogical Agents	396
Paulo Roberto Córdova, Rosa Maria Vicari, Carlos Brusius, and Helder Coelho	
 Human-Computer Interaction	
State of the Art of Human-Computer Interaction (HCI) Master's Programs 2020	405
Gabriel M. Ramirez V., Yenny A. Méndez, Antoni Granollers, Andrés F. Millán, Claudio C. Gonzalez, and Fernando Moreira	
On Bridging the Gap Between Far Eastern Cultures and the User Interface	415
Antoine Bossard	
Promotion of Social Participation in Smart City Developments: Six Technologies for Potential Use in Living Labs	425
Marciele Bernardes, Francisco Andrade, Paulo Novais, Herbert Kimura, and Jorge Fernandes	

Electroencephalography as an Alternative for Evaluating User eXperience in Interactive Systems	435
Sandra Cano, Jonathan Soto, Laura Acosta, Victor Peñefor, and Fernando Moreira	
Heuristic Evaluation Method Applied to the Usability Assessment of Smart Homes Applications	445
Ana Isabel Martins, Ana Carolina Oliveira Lima, and Nelson Pacheco Rocha	
Smart Glasses User Experience in STEM Students: A Systematic Mapping Study	455
Ronny Santana, Gustavo Rossi, Gonzalo Gabriel Méndez, Andrés Rodríguez, and Viviana Cajas	
Multimodal Assistive Technology for the Support of Students with Multiple Disabilities	468
Valentim Realinho, Luís Baptista, Rafael Dias, Daniel Marmelo, Paulo Páscoa, and João Mourato	
Hunter-Gatherer Approach to Math Education - Everyday Mathematics in a San Community and Implications on Technology Design	478
Samuli Laato, Shemunyenge T. Hamukwaya, Laszlo Major, and Shindume L. Hamukwaya	
Emotions and Intelligent Tutors	488
Rámon Toala, Dalila Durães, and Paulo Novais	
Health Informatics	
Factors affecting the Usage of e-Health Services in Kuwait	499
Issam A. R. Moghrabi and Manal H. A-Farsi	
Data Mining Approach to Classify Cases of Lung Cancer	511
Eduarda Vieira, Diana Ferreira, Cristiana Neto, António Abelha, and José Machado	
Mobile Burnout Estimation Device - An Agile Driven Pathway	522
Raluca Dovleac, Marius Risteu, Andreea Cristina Ionica, and Monica Leba	
Development of Adaptive Software for Individuals with Hearing Loss	532
Pedro Giuliano Farina, Cibelle Albuquerque de la Higuera Amato, and Valéria Farinazzo Martins	
Ensemble Regression for Blood Glucose Prediction	544
Mohamed Zaim Wadghiri, Ali Idri, and Touria El Idrissi	

Virtual Reality in the Treatment of Acrophobia	555
Vanessa Maravalhas, António Marques, Sara de Sousa, Pedro Monteiro, and Raquel Simões de Almeida	
FOMO Among Polish Adolescents. Fear Of Missing Out as a Diagnostic and Educational Challenge	565
Łukasz Tomczyk	
Elderly Monitoring – An EPS@ISEP 2020 Project	575
Julian Priebe, Klaudia Swiatek, Margarida Vidinha, Maria-Roxana Vaduva, Mihkel Tiits, Tiberius-George Sorescu, Benedita Malheiro, Cristina Ribeiro, Jorge Justo, Manuel F. Silva, Paulo Ferreira, and Pedro Guedes	
Emotion Recognition in Children with Autism Spectrum Disorder Using Convolutional Neural Networks	585
Rodolfo Pávez, Jaime Díaz, Jeferson Arango-López, Danay Ahumada, Carolina Méndez, and Fernando Moreira	
Author Index	597

Intelligent and Decision Support Systems



Latent Impacts on Digital Technologies Sustainability Assessment and Development

Egils Ginters^{1,2(✉)} and Emils Dimitrovs¹

¹ Riga Technical University, Riga 1658, Latvia

egils.ginters@rtu.lv, socotech@socotech.net

² Sociotechnical Systems OU, Sakala Street 7-2, 10141 Tallinn, Estonia

Abstract. Technological developments ensure the well-being of society, but unreasoned introduction can be detrimental to both society and the environment. It is important for technology authors and investors to know whether the technology will be accepted, adopted and used to reap the benefits of the financial resources invested. The article deals with hidden impacts and unanticipated effects that can influence the sustainability development of technology and assessment simulation. Bayesian networks are proposed for risk evaluation of unanticipated effects.

Keywords: Sustainability assessment · Digital technology · Unanticipated risks simulation

1 Introduction

The daily life of a society is determined by technologies, and practically all technologies are digital today. Minor changes to the software can lead to fundamentally different functionalities and new technology. These are significant differences and benefits of digital technology development. Technology offers society a necessary and appropriate service that enhances and/or ensures the well-being of individuals. One of the integrated parameters that determines intelligent customer choice is the sustainability of the technology. From the customer's point of view, it could be treated as the ability of technology to successfully provide quality service at a reasonable price for the required period. This means that the adopted technology does not conflict with other technologies, nor does it cause harm to the environment and the user. However, the main sustainability parameter for a technology author or investor is profit, but the social and environmental aspects of technology use are secondary. The investors are interested in the customers using the adopted technology for as long as possible while developing a new and improved technology that is compatible with the one already used and ensures income continuity. The government is interested in the taxes paid and public satisfaction, so social and environmental factors dominate the assessment.

The sustainability of the technology can be evaluated using life cycle assessment (LCA) methods based on three whales or three bottom line (TBL) called Business, People and Environment. The set of methodologies provides a diverse assessment of the impact of technology on the above segments. Currently, more than a hundred different

sustainability assessment methodologies are mentioned [1], which are adapted to the object to be assessed.

In most cases, these methods are used for large projects assessment and assume *a priori* that the audience will use the technology. Sustainability assessment are based on statistics, micro-analytical and system dynamics simulation, structural equation modelling, public surveys/measurements and politico-economic conjuncture.

However, for authors of new technologies, especially startups working in private business, one of the most important aspects is technology acceptance in society [2]. If the technology will not be accepted, then it will not be adopted and used. Therefore, engineers and business angels are primarily interested in acceptance issues, but their attitude towards sustainability is still not serious [3]. Therefore, there are few integrated sustainability assessment methodologies that include both acceptance and sustainability assessment.

Acceptance forecasting still dominates in digital technology assessment and is based on extensive surveys of potential audiences using such as Technology Assessment Model (TAM), Unified Theory of Acceptance and Use Technology (UTAUT) and others. The main problems of these methods are the lack of interactivity and high workload. This is especially important in the digital society, where the life cycle of ordinary technology is shortening every year. Therefore, the developers of acceptance assessment methodologies try to reduce the proportion of potential audience surveys in the assessment by using sophisticated solutions such as Rogers diffusion theory and others [4].

Some sustainability forecasting methods offer a static assessment in the form of a sustainability index, which makes it possible to compare the sustainability of different projects/technologies but does not provide an enough understanding of what this index means. To enhance perception and transparency, the authors of Integrated Acceptance and Sustainability Assessment Model (IASAM), use sustainability index measurements in units that are comprehensible to the audience, such as *skypes* [4]. It offers not only a peer review of new projects/technologies, but also a comparison of the sustainability of individual projects against the Skype technology reference line. It also allows forecasting of sustainability index development based on system dynamics simulation.

Assessing IASAM according to the classic TBL scheme of sustainability impacts, 70% of the assessment is related to Business, 25% - to People, but only about 5% of the sustainability assessment forecast respects the impact on the Environment. Although digital technology is not an oil and gas industry, it is in any case a serious shortcoming of the IASAM approach. However, it should be noted that the impact on environment is mainly related to latent and unanticipated factors.

Nowadays, sustainability assessment methods generally do not respect the hidden values, incentives and other latent and stochastic factors that influence the use and further development of the technology.

This means that the forecast of the possible development of the sustainability may differ significantly over the life cycle of the technology and may shatter the initial expectations.

The aim of the article and the related research question is to clarify the hidden impacts and factors that technology sustainability assessment methodologies must respect in order to provide a more accurate forecast. The results and conclusions of the work are

based on the author's diverse experience in various information technology projects, holistic approach and detailed analysis of the previous projects consequences using UTAUT and IASAM methodologies. The audience of the article can be the developers of sustainability assessment methodologies.

2 Uncertainty Factors Influencing Sustainability Development

Traditionally, technology sustainability assessment has been based on specific visible and measurable characteristics that predict the impact of technology on TBL segments. It is assessed whether the investment will be profitable, whether the society will be satisfied, and the environment will not be harmed. However, in the longer term, additional impacts may emerge that could significantly change the initial sustainability assessment.

If we abstract from practically unpredictable risks, then the attributes of digital technology itself are important. Unfortunately, there is still no generally accepted paradigm for this set of attributes. In the LIASAM project launched in 2020, the authors identified six interrelated attributes of digital technology (A_T): Performance, Complexity, Uncertainty, Evolutionism, Pervasiveness and Reliability. The set of attributes determines the sustainability of each digital technology Sus_T^i :

$$Sus_T^i = f(A_T^i) \quad (1)$$

The riskiest attribute is Uncertainty, which at the same time significantly affects Reliability. The effect of the Uncertainty attribute on the People and Environment in the TBL model can be critical. Unfortunately, the assessment of Uncertainty $A_T^i(U)$ is challenging as the factors are usually latent and hidden:

$$A_T^i(U) = \langle DI, UE, AF, SD, UU, UC, DU, I \rangle \quad (2)$$

where $\{DI\}$ - Determined and systematic impacts, $\{UE\}$ - Unforeseen stochastic effects, $\{AF\}$ - Age dynamics factor, $\{SD\}$ - Technology self-development, $\{UU\}$ - Unexpected use, $\{UC\}$ - Unanticipated consequences, $\{DU\}$ - Dual use technologies, $\{I\}$ - Incentives.

Uncertainty influencing factors can be determined, stochastic, external, or embedded in the technology itself.

Determined and Systematic Impacts (DI). Events and impacts that are clearly predictable and will occur over a longer period.

It is clear, that ambient temperatures are rising unevenly and heterogeneously. The same time resources of minerals, including fossil fuels, is declining. The above factors promote further development of green technologies. Of course, only if these technologies are really green.

The intellectual capacity of technologies is increasing day by day because human labour becomes more and more expensive. For example, the Artificial Intelligence (AI) software market forecast from 2018 (USD 10.1 billions) promises a significant increase in 2020 to USD 22.59 billion, but if current trends continue, the AI software market in 2025 may exceed USD 126 billion [5].

Similarly, there is no doubt about the growth of the wireless technology market as wires become more expensive, but ether splitting is just a political decision.

Unforeseen Stochastic Effects (UE). Events and effects that are not expected during the initial assessment of sustainability and are independent of technology evaluated.

For example, the Covid-19 pandemic imposed significant restrictions on the movement of people but did not reduce the need for social and labor contacts. As a result, remote access technologies experienced unexpectedly rapid development. The number of users of the Zoom communications platform increased 30 times, while market capitalization reached USD 48.78 billion. The Zoom platform sustainability development experienced an exponential leap [6]. It is expected that the market for other assistive, augmented and virtual reality technologies and 3D printing will also grow significantly. The question remains whether the technology sustainability assessment model can predict such a grandiose impact. This could be very difficult, however the assessment of smaller stochastic impacts [7] could be possible.

Age Dynamics Factor (AF). The effect of age group changes on technology acceptance, adoption and use.

Digital technology acceptance research conducted in 2020 showed a significant impact of the age factor [8]. A group of several impacts related to the age of the user was identified: welfare, demand, complexity solving ability and busyness. At a young age, user well-being depends on parental funding, but the best opportunities to invest in technology are achieved in middle age. However, in retirement age, these opportunities are declining again. In turn, the trend of desire and need for technology is different. Initially this desire is growing very rapidly, but at the stage when the potential user reaches the peak of his well-being, the desire to use technologies decreases. However, with the deteriorating of health and physical condition, the current needs for the use of technologies are growing again. The complexity solving ability trend is like the welfare curve, however, it is more shifted to the user's youth period. Upon entering a routine, complexity solving ability gradually decreases. This factor is very important because digital technologies are changing very fast, which inevitably creates feedback and impact on the development of society. New concepts and digital habits are emerging, even changes in thinking influenced by societal technologies. Authors of new technologies usually create technologies for their generation. They may not even be able to create new digital technologies that are suitable for either the previous or the next generation, because the authors have different understandings. The complexity solving ability of an individual is very important especially for the previous generation, because the ideas embedded in the new technology are not taken for granted. An important factor is busyness. Has the user enough time to learn how to use the new technology? If there is enough time in the early youth and retirement years, then at the beginning of the career a user is busy.

The resulting digital technologies acceptance curve [8] reflected a nonlinear relationship between the human age and technologies acceptance, adoption and use. The observed changes in acceptance could reach more than 60% of the highest measurement value during user's lifetime. This means that the assessment of the sustainability of the digital technology must anticipate changes in the age dynamics of the potential audience along the life cycle of this technology.

Technology Self-Development (SD). Specific attributes of technology that determine the possibilities of its further development and successful improvement.

It is determined by different factors. The existence of talented and persistent designers, for example, because funding without an idea can only provide quantitative results. An important factor is the openness of the technology (open source, standardized interfaces and communication protocols) and the modular design (Arduino), which, as in the LEGO game, can transform and improve technology with minor investments. One more requirement is the intellectual capacity of the technology, which is determined by built-in self-diagnostics and reconfiguration options. A critical moment is approaching when users will say “Stop” to big software business sharks, because the user is tired of the ever-increasing and pointless cost of repairing’s and upgrading’s. Even if these fixes are free like MS Windows, then the number of updates and patches is increasing but the user must pay for the traffic and delay of work. To put it mildly, from the end of 2015 to the autumn of 2020, there have been 10 different versions of MS Windows 10, with a total of more than 837 different changes (263 builds and 574 patches) that have desperately tried to creep in the user’s computer [9].

It is expected that soon technologies incorporate AI cognition capabilities and be able to generate new functionalities without assistance of providers that would be a significant contribution to digital technologies sustainability development.

Unexpected Use (UU). A situation when the purpose of technology development is different from the further application of the technology.

One of the typical examples that has changed the habits of the society is the audio signal encoding format - MP3, which got its name in 1995. In 1982 for doctoral student Karlheinz Brandenburg [10], who was studying at Ilmenau Technical University, his supervisor formulated interesting topic of the research. The essence of the task was to develop compressed encoding of audio signals in order to reduce data transmission traffic on ISDN telephony channels. The scientist was very persistent, and the new technology emerged that revolutionized data storage, portability, mobility and streaming. MP3 destroyed the traditional music records industry and created the conditions for convenient P2P record sharing.

When starting to develop the technology, such application results did not appear to anyone even in a dream. Unexpected use factors are like external stochastic effects, nevertheless they are more dependent on the technology itself.

Unanticipated Consequences (UC). Positive or negative additional effects of technology that were hidden during the development and implementation of the technology but become significant during the technology life cycle.

Modern systems are complex, so all possible effects are difficult to predict [11]. Unfortunately, these side effects are sometimes deliberately ignored.

Mankind is developing green technologies, such as electric vehicles (EV), with good intentions. A fight has been announced to reduce pollution and against the greenhouse effect. Carmakers are rapidly abandoning diesel engines, and countries are subsidizing the use of EV in cities. Bloomberg [12] has estimated that the EV market is growing by around 60% each year, reducing the cost of batteries and making them more energy intensive. The forecast revealed that in 2040 around 35% of all new cars will be electric.

The new EVs do not spoil the atmosphere, their quiet movement reduces noise in the city. Unfortunately, these positive effects have their downsides. Electricity must be

generated and stored. Energy and heat production industry is one of the most environmentally harmful sectors related with CO₂ emissions in the EU28. A situation may arise when green vehicles are shifting emissions off roads onto power plants only.

Lithium, cobalt, nickel, manganese and other elements are required to produce EV batteries. Between 2017 and 2027, EV-initiated demand for lithium-ion batteries is expected to grow by about 11% each year, while the total market size in 2027 will reach about USD 41.1 billion [13]. Lithium is extracted from two sources: brines and hard rock mining. The largest deposits are in Australia, Chile, Argentina, Zimbabwe, Tibet and China. Lithium mining emits large amounts of emissions and has serious footprint. The mining process of brines requires extensive amounts of water, which causes aquifer depletion and makes the environment a desert. About 1900 tons of water are needed to produce 1 ton of lithium [14].

The Fraunhofer Institute for Building Physics estimated that the energy consumption for EV production is twice that of a conventional one. The main reason is the production of batteries and the disposal of waste batteries [15]. Batteries are not rubber tires, the ridges of which mainly create an unpleasant aesthetic effect. The disposal of used batteries is expensive and toxic to the environment. One can try to trick households into using second hand EV batteries to store electricity from solar panels, thus putting future problems in the hands of happy new owners. Of course, battery treatment and re-use is possible because EV batteries are recyclable. However, the cell recycling is very complicated because the electrolyte is flammable, explosive and toxic. Unfortunately, in the case of EV, a significant proportion of used batteries are either sent to landfills or stockpiled and stored. Respecting that more than 1 million EVs had already been sold by 2017, "...researchers in the United Kingdom calculated that 250,000 metric tons, or half a million cubic meters, of unprocessed battery pack waste will result when these vehicles reach the end of their lives in about 15 to 20 years - enough to fill 67 Olympic swimming pools" [16]. This is a large amount of toxic waste and someone will have to pay for disposal someday. This could be a classic example when initially good and green EV technology may not be realistically sustainable over time due to various completely unanticipated consequences.

Dual Use Technologies (DU). A situation when the technology is designed to achieve the basic goal, but the additional benefit is clearly visible and planned.

A typical example is cogeneration or combined heat and power (CHP), where the main task is to produce heat for heating systems in larger buildings where a heat carrier, such as water, circulates. The water is usually heated by burning fossil fuels or renewable resources such as wood chips, biogas, etc. However, when the temperature is high enough, the water turns into steam, which can spin an electricity generator connected to the steam turbine. It can directly generate alternating current, which does not require additional storage in the batteries and modification. Electricity generation is an added value of cogeneration technology, which significantly increases the sustainability. If heat and electricity are produced separately, the efficiency of heat production is about 55%, but when both benefits are obtained at the same time, the efficiency reaches 87% [17].

Incentives (I). The characteristics of the technology that contribute to the expansion of its market, offering ever new options.

The 1980s saw the emergence of the first cellular mobile phones that copied analogues telecommunications capabilities, but a decade later the first GSM mobile phones emerged. Installation and maintenance costs were reduced because no wires were needed. It began the irreversible wireless telephony victory march. Initially in GSM telephony an option of alphanumeric message (SMS) sending inherited from pagers has been added just for experimental purposes. However, cellular technology had good opportunities for self-development: openness and modularity. Protocols and batteries evolved, graphic screens appeared, and buttoned cell phones were replaced by smart phones with touch screens. Today the current development and competition of smart phones is based on incentives. With each new smart phone the user gets additional options: movie channels streaming, versatile games, AI based schedule planning, social media and networks access, bank accounts management, cloud storage of data, language translation, navigation, objects recognition, labels scanning and so on. PRNewswire [18] data show that the global smartphone market compound annual growth rate (CAGR) is 7.05%. It is expected that the market will increase from USD 179,972.89 million in 2018 to USD 290,098.28 million by the end of 2025.

This is an example when a user being purposefully encouraged to continue using the technology and purchasing more and more new gadgets and services, paying for it with a loss of privacy. Perhaps that the society has only a short step to mobile phone implants, because the human body is a wonderful battery to operate the phone 24 h a day.

3 Uncertainty Modelling in Sustainability Assessment

The above suggests that the group of uncertainty factors may be broader, but the set $A_T^i(U)$ can be mentioned as one of the most significant factors which rise risk $R(p, t)$ with a possible probability p at time t that influences sustainability development of assessed digital technology:

$$A_T^i(U) \Rightarrow R_T^i(p, t) \quad (3)$$

A major problem is fuzzy and latent effects proportion assessment in sustainability forecasts. The result will be subjective, as new technologies do not have historical data and must rely on comparisons with similar technologies, which are not accurate enough. In addition, the factors have a high stochastic component, so even the presence of timeline data does not promise anything good. However, efforts may be made to reduce subjectivity through risk assessment.

One possible solution is Bayesian networks (BNs), which are widely used in AI and expert systems. Unlike neural networks, Bayesian direct acyclic graphs show real relationships and links between events rather than basing them on the execution of certain rules. The approach allows the causes of events to be explained because it provides reasoning in the opposite direction from the result to the start cause, which is usually not allowed by rule-based theories. Random variables are usually described by Boolean values, where the possible probability of each value is calculated. The probability of occurrence of each subsequent child (A) event in the network is influenced by

the probability of occurrence of the related parent (B) events. The use of BNs provides quantification of risks and risk handling under conditions of uncertainty.

One of the software tools that can be used for risk modelling and latent effects risk probability assessment is AgenaRisk, which includes sophisticated algorithms to create complex system uncertainty models. AgenaRisk [19] is based on the creation of probability tables and visual risk maps that represent causal relationships and support both diagnostic and predictive reasoning about uncertainty. This approach allows representation of expert judgment using subjective probability and construction of hybrid models containing discrete and continuous uncertain variables. However, if sustainability researchers are not afraid of math and programming, then BNs modelling capabilities are also available in PyMC and other Python based applications.

Eight years ago, a group of authors developed the first version of the IASAM methodology [20], which included an acceptance phase as an integral component of sustainability assessment. The IASAM sustainability index is measured in *skypes* to ensure intuitive project evaluation capabilities. The IASAM index has a breakdown in four groups with 0.25 points deviation. First group is problematic, but upper group technology is reasonable for investments. Technology sustainability development $Sus_T^i(t)$ is specified as the interaction of four streams:

$$Sus_T^i(t) = Sus_T^i(t - dt) + \left(Accept.T^i + Manag.T^i + Quality_T^i + Domain_T^i \right) * dt \quad (4)$$

Flow dynamics are determined by sets of relevant factors: $Acceptance_T^i$ flow (12 factors), $Mangement_T^i$ (22 factors), $Quality_T^i$ (18 factors) and $Domain_T^i$ (7 factors).

Uncertainty risk assessment options provide an opportunity to improve the IASAM model and determine that:

$$\Delta Sus_T^i(t) = Sus_T^i(t - 1) + R_T^i(p, t) \quad (5)$$

As a result, sustainability development assessment can be implemented as a closed loop approach based on observations feedback (see Fig. 1). The application of feedback allows to perform estimation error correction $\Delta Sus_T^i(t)$ during the life cycle of the technology.

The results of risk assessment scenarios are displayed in the form of a boxplot, adding risk colour codes: high, relevant, low, minor. Used visual trend analytics facilitates risk monitoring capabilities.

The LIASAM risk assessment demonstrator is developed and verified according to the model described above. Unanticipated effects analysis and their share of the common risk assessment is still ongoing. Therefore, the demonstrator is designed as an open and modular structure in order to ensure risk assessment network links and weights configuration options. Delphi methods and comparison with similar risk assessment methods will be used to validate the LIASAM approach.

The advantages of the proposed approach are correction possibilities and higher accuracy of the sustainability development forecast respecting unanticipated impacts of digital technology.

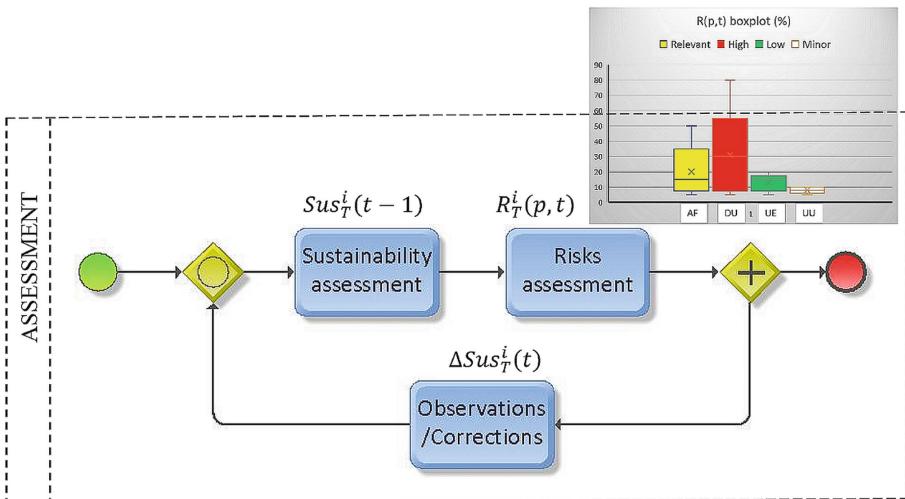


Fig. 1. BPMN2 swimlane of sustainability assessment simulation loop control system.

4 Conclusion

The sustainability assessment of the digital technologies or IT projects is difficult, as it depends of many cross related and stochastic factors.

No less important is the assessment of the dynamics of sustainability development along the life cycle of technology. A more accurate sustainability forecast allows technology authors and investors to plan financial flows in a timely manner. In turn, comparing the sustainability of projects is particularly important for policy makers and crafters who spend taxpayers' money. Comparable quantitative forecasts reduce the potential risks of corruption in public procurement.

The authors of the article analysed the essence one of the attributes of digital technology - Uncertainty, which significantly affects the People and Environment part in the sustainability forecast. A set of unanticipated factors affecting this attribute was described. Unanticipated factors are mostly stochastic and heterogeneous, and their effects are latent. In addition, these factors are hidden during the initial sustainability forecasting. Therefore, it is very important to assess the risks of unanticipated effects from the beginning, as well as to continue monitoring them during the use of digital technology.

The LIASAM project concept is based on a modified Integrated Acceptance and Sustainability Assessment Model (IASAM) and probabilistic Bayesian acyclic graphs use. Using the risk assessment network, the impact of unanticipated factors and the sustainability adjustment are calculated, which refines the forecast of the sustainability of digital technology. Unlike most other sustainability forecasting methodologies, LIASAM provides an interactive and quantitative self-assessment option. The results of the research can be important for both technology authors and investors, because this approach ensures an assessment of the sustainability even before the technology development begins.

Further work will be related with the validation of the proposed concept, using data of already implemented projects and introduced technologies. Machine learning algorithms will be developed to assess the weight of the factors that determine each specific risk.

Acknowledgements. The article publication is funded by LZP-2020/2–0397 “Latent Impacts on Digital Technologies Sustainability Development Assessment (LIASAM)”.

References

1. Saurat, M., Ritthoff, M., Smith, L.: SAMT. Deliverable D1.1 Overview of existing sustainability assessment methods and tools, and of relevant standards. <https://www.spire2030.eu/SAMT>. Accessed 17 July 2020
2. Piera, M.A., Buil, R., Ginters, E.: State space analysis for model plausibility validation in multi agent system simulation of urban policies. In: Proceedings of 25th European Modeling and Simulation Symposium (EMSS 2013), pp. 504–509 (2013)
3. Ginters, E.: New trends towards digital technology sustainability assessment. In: 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), pp. 184–189. IEEE, London (2020)
4. Ginters, E., Aizstraute, D.: Technologies sustainability modelling. In: Rocha, A., Reis, C. (eds.) Trends and Advances in Information Systems and Technologies (WorldCIST 2018). Advances in Intelligent Systems and Computing, vol. 746, pp. 659–668. Springer, Cham (2018)
5. Shanhong, L.: Artificial intelligence software market revenue worldwide 2018–2025. <https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/>. Accessed 21 July 2020
6. Ghosh, I.: Zoom is now worth more than the world's 7 biggest airlines. <https://www.visualcapitalist.com/zoom-boom-biggest-airlines/>. Accessed 05 Sept 2020
7. Ben Abdelaziz, F., Colapinto, C., La Torre, D., Liuzzi, D.: A stochastic dynamic multiobjective model for sustainable decision making. Ann. Oper. Res. 1–18 (2018)
8. Ginters, E.: Digital technologies acceptance/adoption modelling respecting age factor. In: Rocha et al. (eds.) Trends and Innovations in Information Systems and Technologies. AISC 1160, vol. 2, pp. 621–630. Springer, Cham (2020)
9. Windows 10 version history. https://en.wikipedia.org/wiki/Windows_10_version_history. Accessed 05 Sept 2020
10. Rose, J., Ganz, J.: The MP3: a history of innovation and betrayal. <https://www.npr.org/sections/therecord/2011/03/23/134622940/the-mp3-a-history-of-innovation-and-betrayal?t=1599074174353>. Accessed 05 Sept 2020
11. Mulder, K.F.: Impact of new technologies: how to assess the intended and unintended effects of new technologies? In: Kauffman, J., Lee, K.M. (eds.) Handbook of Sustainable Engineering. Springer, Dordrecht (2013)
12. Randall, T.: Here's how electric cars will cause the next oil crisis. <https://www.iene.eu/heres-how-electric-cars-will-cause-the-next-oil-crisis-p3240.html>. Accessed 05 Sept 2020
13. Consumer goods and services. Market research report. Lithium-ion battery market. <https://www.transparencymarketresearch.com/lithium-ion-battery-market.html>. Accessed 05 Sept 2020
14. Hebl, B.: The mining process requires extensive amounts of water, which can cause aquifer depletion. <https://www.ocmal.org/lithium-firms-depleting-vital-water-supplies-in-chile-analysis-suggests/>. Accessed 05 Sept 2020

15. How eco-friendly are electric cars. <https://www.dw.com/en/how-eco-friendly-are-electric-cars/a-19441437>. Accessed 05 Sept 2020
16. Hunt, K.: The rapid rise of electric vehicles could lead to a mountain of battery waste. <https://edition.cnn.com/2019/11/06/business/electric-vehicles-battery-waste-scn/index.html>. Accessed 05 Sept 2020
17. Why cogeneration is more efficient than conventional coal power plants? <https://wiki.energytransition.org/infographics/>. Accessed 05 Sept 2020
18. PRNewswire. <https://www.prnewswire.com/news-releases/the-global-smartphone-market-is-expected-to-grow-from-usd-179-972-89-million-in-2018-to-usd-290-098-28-million-by-the-end-of-2025-at-a-compound-annual-growth-rate-cagr-of-7-05-301054105.html>. Accessed 05 Sept 2020
19. Fenton, N., Neil, M.: Risk Assessment and Decision Analysis with Bayesian Networks. 2nd edn. Chapman and Hall/CRC (2018)
20. Ginters, E.: Sustainable physical structure design of AR solution for golf applications. In: Rocha, A., Reis, C. (eds.) New Knowledge in Information Systems and Technologies (WorldCIST 2019). Advances in Intelligent Systems and Computing, vol. 931, pp. 675–685. Springer, Cham (2019)



Using Kinect to Detect Gait Movement in Alzheimer Patients

David Castillo-Salazar^{1,2(✉)}, Laura Lanzarini^{2(✉)}, Cesar Guevara^{3(✉)},
and Héctor Gómez Alvarado^{4(✉)}

¹ SISAu Research Group, Universidad Tecnológica Indoamérica, Ambato, Ecuador
davidcastillo@uti.edu.ec

² LIDI Computer Research Institute (CICPBA Center), Facultad de Informática,
Universidad Nacional de La Plata, La Plata, Argentina
{david.castillos,laura1}@lidi.info.unlp.edu.ar

³ Mechatronics and Interactive Systems - MIST Research Center, Universidad Indoamérica,
Quito, Ecuador
cesarguevara@uti.edu.ec

⁴ Centro de Posgrado, Universidad Técnica de Ambato, Ambato, Ecuador
hf.gomez@uta.edu.ec

Abstract. During the aging process, the elderly can experience a progressive and definitive deterioration in their gait, especially when they have neurological disorders such as Alzheimer's disease. Effective treatment requires accurately assessing these issues in mechanical stability, the muscular-skeletal system, and postural reflexes. For Alzheimer patients in particular, gait analysis represents an important method for determining stability and treatment, which is the key objective of this investigation. Thus, this article describes the creation of a dataset on the walking gait, focusing on the distance covered by the patients and the angle of their legs as registered by a Kinect device. All patients were examined at a recognized center for elderly care in the canton of Ambato, Ecuador. We worked with a population of 30 Alzheimer patients whose ages ranged between 75 and 89 years old. The retrieved numerical data were processed with Diffused Logic, which, when based on a series of rules, can determine the instability and stability of a patient with a neurological illness. As a result, it was possible to create a dataset that included numerical values of the walking distance for each patient. This information will be important to future health care research, especially for physiotherapists and pose estimation.

Keywords: Alzheimer · Fuzzy · Kinect · Walking gait

1 Introduction

The ability to walk is an important characteristic in human beings. These movements enable one to move the upper part of the body, like the arms and the head, and to carry out daily activities. In [1], ADI (Alzheimer's Disease International) and the WHO (World Health Organization) define the elderly as those who are 60 years old or older

and sensitive in relation to the motor skills that they use on a daily basis. This definition implies that the elderly experience displacement when walking while trying to maintain their posture—thus leading to the intervention of the bones, articulations, tendons, muscles, and the brain. As the article written by [2] explains, Kinect is able to detect the human skeleton in motion, recognize it, and position it on a plane. Technologies like Deep Learning have enabled us to work with Depth YOLO and Decision Maker to obtain adjusted skeletons to reality. The results show that Kinect offers low coverage of skeleton information (61.73%). However, Depth YOLO improves this performance by 72.36%, because it does not need the whole body to extract the position of the articulations. At 80.60%, Decision Maker enables one to further increase the detection of the extracted skeletal structure and show body recognition coverage.

The article written by [3] presents a technological approach to fall prevention using RGBD devices as experimental systems (ES) that capture the movements of patients who live at home and are only intermittently visited by a physiotherapist. For this reason, Microsoft Kinect is a reliable and appropriate device to conduct the Functional Reach Test (FRT), which is used to therapeutically measure equilibrium—thereby helping notably with diagnosis. The results show an absolute average difference of 2.84 cm (± 2.62) via a Student's t-test, indicating that a significant statistical difference does not exist.

Author [4] described the cognitive and physical activities of elderly people with the aim of carrying out an analysis of their walking patterns using cluster algorithms based on a dataset collected by Kinect. This focus is connected to the CAC brand (Controller Application Communication) application, which can transmit skeletal packages and RGB information to the physical system. The study's results show that the DBScan grouping algorithm can be applied with success to the location data collected by Kinect.

Author [5] details the following procedures for the Kinect sensor, which should be followed during authentication: 1) Inputting words using speech recognition, 2) introducing passwords using the position of any part of the human body, 3) initiating keys using the posture of the human structure, 4) longitudinally measuring the skeleton, and 5) validating access using physical keys, such as a printed card with the required information. These procedures incorporate three categories of keys using physical and biometric data, making the Kinect system more reliable, less complex, and able to detect both the voice of the patient and the distance of their movements, among other data. The results show the importance of shoulder measurements, with a minimum threshold of 9 cm needed to identify the patient and to determine which biometric method is the most appropriate.

The article written by [6] emphasizes the sequence of an individual's gait as a relevant characteristic to biometrics. A preliminary study was focused on biometrics based on the way a person walks, particularly five articulations of the axial skeleton that can differentiate two people while they move. The results are presented in coordinates (x and z). An analysis of variance (ANOVA) produced values lower than 0.05 for the neck and the head, indicating the validity of the Kinect. However, for the vertebral column, the middle column, and the shoulder blade, the figures were approximately 0.05, 0.048, and 0.047, respectively.

The research carried out by [7] shows the importance of monitoring and providing medical care to the elderly. New vision technologies using computers and image processing techniques have been relevant to this type of research, including three-dimensional (3D) depth sensors such as Microsoft Kinect, which can trace skeletons that work with articulations and contextualize the bodies in the temporal space of the activities that are carried out.

The mathematical focus of this study is based on the euclidean distances by Minkowski and Cosinos regarding distances between 3D articulations. This dataset was processed through an algorithm based on extremely random trees using Microsoft MSR, 3D Action, and MSR Daily Activity 3D. The results demonstrate that datasets can contribute to the efficiency of monitoring systems in various contexts, such as when it is possible to apply three series of movements: Action 1 with an average of 89.99%; Action 2 with 92.18%; and the overlapped Action with 80.92%. The biggest of these averages supports different activities within an established context.

The research carried out by [8] highlights the importance of Kinect to pose estimation procedures and monitoring human movements in various contexts, such as in the field of physiatry. With this technology, one can evaluate displacement in patients of various ages. The results show that virtual rehabilitation in synchronization with other methods are adequate for these purposes. Specifically, the precision of the biometric estimates, the reliability of the position of the articulations, and the angles among the segments of the skeleton show a Pearson coefficient of 0.93 for the right leg, 0.90 for the left leg, and 0.99 for the back—thus demonstrating a strong correlation to the angle of the back. The articulation was 99% for a threshold of 5° and 100% for threshold values of 7.5° , 10° , and 15° . These findings confirm the effectiveness of this system under certain conditions.

An article written by [9] also studied the elderly and their performance while carrying out daily physical tasks within a pre-determined context with the purpose of discover anomalies in their activities, showing that new technologies facilitate the evaluation and quality of this research. For example, computerized vision systems can be utilized to detect skeletal neuromuscular disturbances captured by Asus Xtion PRO, the Microsoft Kinect Depth camera. The study's probability model classified the tests as "normal" and "abnormal". Regarding the quality of the displacement, a linear regression analysis was used. The evaluated performances corresponded to the simplest tasks, such as sitting down, stopping, walking, and walking up and down stairs. The results have generated a new field of research in automatic medical monitoring processes in the home, where it is shown that the average plus the standard deviation has a variation from 0.56 to 0.79 for all the executed actions.

Research paper [10] discusses postural stability in Alzheimer patients who develop a mental disease that progresses in an unalterable way. In this study, the patients had an illness level classified from a medical point of view as being slight to moderate and that had reduced their static and functional balance. The patients also experienced other dysfunctions, such as inattention while carrying out two activities at the same time and a loss of visual information, which are key factors in postural instability. The results show that the Alzheimer patients had worse scores on a performance-based mobility assessment than healthy patients ($P = 0.01$). Moreover, the sub-analysis did not show

significant differences between the healthy participants and those with light Alzheimer's ($P > 0.05$).

In this article, the authors propose a preliminary study into the angle, distance, speed of the walking gait among other parts of the body of Alzheimer patients. The process is focused on three stages: the first stage is to install software to communicate with the Kinect device; the second is to detect the patient's skeleton using Kinect; the third is to capture the walking gait to obtain the data that are required for the research. In the following sections, these procedures are further detailed.

This article is structured as follows. Section 2 explains the methods and materials used in this work. Section 3 describes the process for obtaining the data and the proposed model for the detection of the skeleton of Alzheimer patients. Section 4 presents the results of this study. Finally, Sect. 5 provides the conclusions of the study and describes future lines of investigation.

2 Methods and Materials

In the following section, the materials and techniques used to develop an algorithm to detect a walking gait are described.

2.1 Methods

The method used was based on the skeleton detection and articulation coordinate systems for SDK Kinect, which are described in [7]. Microsoft Kinect, which can focus on skeletons by describing the temporal-spatial aspects of a sequence of human activities, was used.

Author [11] mentions the relationship between computer vision and the identification of a person's gait. These technologies can be used in several technical or scientific fields to identify individuals by the way they walk and to obtain results according to the researcher's needs.

The skeletal system in Kinect shows 20 positions in 3D space: one for each articulation of the human body, as seen in (see Fig. 1). The points of interest are the mid hip, knee, left foot, and right foot, respectively, for determine the values related to the angle, speed, and distance of the patient's walking gait. In [12], the author describes the structure of local articulations and the locations of the parts of the body in those 20 positions. Only 13 are selected, which include the head, the middle of the shoulder, mid hip, right shoulder, left shoulder, right elbow, left elbow, right hand, left hand, right knee, left knee, right foot, and left foot.

The study carried out by [13] used Kinect to measure the gait and the static body during passive activities and determined novel characteristics with respect to gait movement and distance. Which are comparable with anthropometric qualities.

In the proposed study, the Kinect device and its representation in code lines are used to measure certain points of interest on the skeletal in the AnkleLeft (left ankle) and AnkleRight (right ankle), allowing one to determine the coordinates x and y. In turn, their distances can be calculated by a function that has two input parameters (left foot, right foot), returning a figure in millimeters. Similarly, the angle can be calculated with

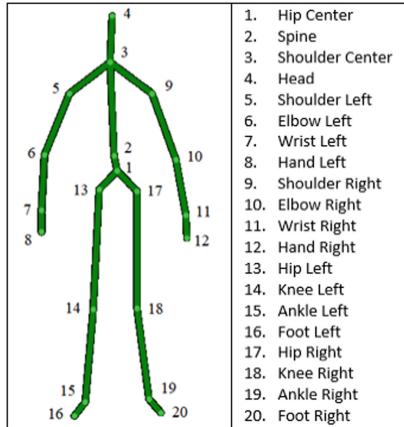


Fig. 1. The Kinect skeletal structure

three input parameters (right knee, left knee, and hip), with the resulting value expressed as a grade. The obtained results are then stored in a text file with referential information, such as the date, hour, angle, and distance.

This data was collected in elderly care centers in an area dedicated for this activity. The Kinect software generated a text file that signaled the beginning of the data processing stages: the collection of data and the preparation of the data with the information that one would work with, such as angle and distance. Once the data is clean, one can process the data in Matlab Fuzzy Logic Toolbox—thereby obtaining the desired results for analysis.

The criteria to measure the distance between the individual and the Kinect device is represented by the vector of Euclidean distance, as shown in Eq. (1). In [14], the authors refer to Kinect measurements of Euclidean distance patterns that are dependent on the time involved in the movements. These values allow one to calculate simple structures and even the positioning of body parts.

$$d = \sqrt{x^2 + y^2} \quad (1)$$

The longitude of a vector in a 2D space is calculated with Eq. (2). This characteristic vector represents the distance between one articulation and another. In [15], the Kinect device is used to apply the plane detection technique, which calculates a step using the Euclidean distance of the ankle to the independent floor and of the individual to the Kinect device.

$$d = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}, \quad (2)$$

where x_a , y_a and z_a represent the coordinates x, y, and z of the “A” point, respectively. The characteristics, such as the longitude of a step, can be calculated using the Eq. (2) [11].

The distance traveled is shown in Eq. (3), where the horizontal longitude is determined between the left ankle and the right ankle. In [16], it is shown that the gait cycle

completes after three steps, beginning with 1) the position of the foot at rest, 2) then continuing as the right foot moves forward to rest, 3) and finishing when the left foot moves forward to rest.

$$d = (X_{\text{Ankle Left}} - X_{\text{Ankle Right}}) \quad (3)$$

Table 1 shows the angles of the legs and referential step distances as the Alzheimer patients walked. In [17], the grades of the inferior extremities of the body are supported, while in [18] one justifies the cycle of human walking patterns. To determine the risk of an elderly person falling, it is important to know and to evaluate the patient's equilibrium and walking pattern with the Tinetti scale [19] which is for gait and balance assessment.

Table 1. Bands to determine normal walking in Alzheimer patients

	Unstable		Stable		Unstable	
	Min	Max	Min	Max	Min	Max
Angulo de las Piernas	0	24	25	30	31	38
Distancia de Paso	0	24	12	36	24	50

Based on the data in Table 1, where the minimum and maximum ranges of an individual's steps are presented, we established the values of the stepping distance that are described in Table 2, namely those that were obtained with the Kinect device.

Equation (4) shows the function of ownership or membership, which refers to the part of the theory of a series where the grade in which each element of a given universe belongs to this group is indicated. In other words, the function of ownership of series A in universe X will be as follows: $\mu_A(x) \rightarrow [0,1]$, where $\mu_A(x) = r$. If r is the grade in which x belongs to A and the value is 0, there is no ownership; if the value is 1, there is ownership [20].

$$f(x; a, b, c) = (\max(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0)) \quad (4)$$

A diffused pattern is defined by the function of ownership or membership denominated as $U_{A(x)}$, which indicates the grade in which variable x is included in the concept represented by the label $A(a \leq U_{A(x)} \leq 1)$. Specifically, we are using the function of Eq. (5).

$$U_{A(x)} = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{m-a} & \text{si } a < x \leq m \\ \frac{b-x}{b-m} & \text{si } m < x \leq b \\ 0 & \text{si } x \geq b \end{cases} \quad (5)$$

Similarly, the triangular characteristic function $(a; a, m, b)$ shown in (see Fig. 2) was used for its mathematical simplicity and manageability. It is defined by limits inferior

to a , an inferior limit b , and an m value where $a < m < b$ the $[a, b]$ is the level of confidence of the diffused triangular number (a, m, b) .

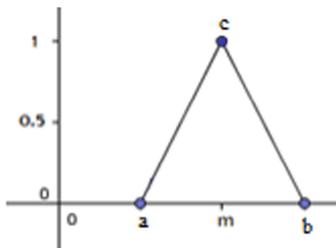


Fig. 2. The triangular function

This information allows one to describe the development of a diffused model, which is defined by the diffused control or Mamdani inference [21] and based on the following series of steps: the fuzzification of the input variables, the evaluation of the rules, the aggregation of the outputs of the rules, and the de-fuzzification of the linguistic variables of the inputs (grade, distance, angle, hip, head, shoulder) and outputs (Unstable/Stable). These values are represented by the diffused groups.

With regard to fuzzification, the values of the input variables were 45.35 for distance, 8.03 for the angle, 18 for the waist, 11 for the head, 17 for the shoulder. These variables were also given three labels: low, moderate, and high.

During the evaluation of the rules, one takes into consideration the previously inputted values. This step was applied to the nine proposed diffused rules. As there are various previous examples, one should use the AND operating system, as shown in Table 2. To evaluate the operating system on a habitual basis, one uses the minimum standard T-norm $U_{A \cap B} = \min [U_A(x), U_B(x)]$.

Table 2. Rules using the AND operating system

Rule 1	If (leg angle is low degrees) AND (distance steps are medium) and (head is high) then (unstable/stable is stable)
...	...
Rule 15	If (Head high) AND (Low shoulder) and (Middle hip) Then (Unstable/Stable is Unstable_Posterior)

Aggregating is the process of unifying the outputs of all the rules implicit in the model, including the functions of the pertinence of the distance, angle, waist, head, and shoulder of all the subsequent variables combined, which are used to obtain a unique diffused series as a stable/unstable output variable.

De-fuzzification is the final step, during which one takes an input from the previously obtained diffused series to give an output value. Further details are determined in the

results section. Furthermore, in (see Fig. 3), the relationships of the functions that were applied with Matlab Fuzzy Logic Toolbox are presented.

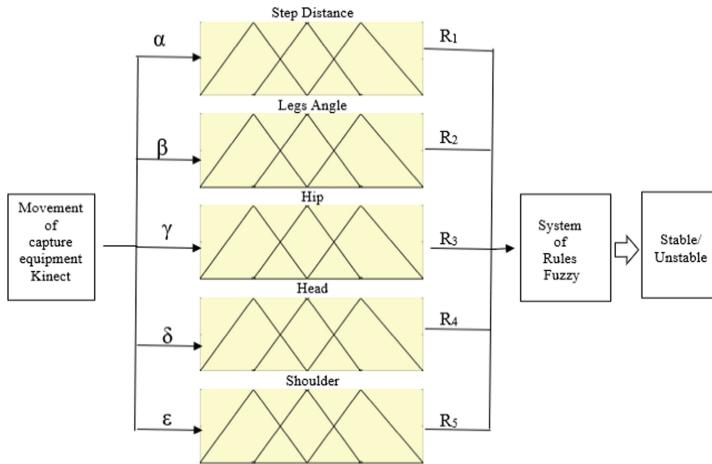


Fig. 3. Simulink implementation of the gait detector with Kinect

During the application of this diffused control, various tests were carried out with diverse Alzheimer patients. The findings are presented in the results section.

In Table 3, the data from the investigation is presented, including where one can observe the data obtained with the Kinect device. In [17], details of the ranges of movements of the articulations are described. In [22], the values of each trajectory are identified (flexion angle), along with the angular speed for several sub-phases of the cycle of the human gait. To calculate the averages of the articular angles in [23], a speedometer was used, which allowed one to estimate the angle of the hip, the knees in static posture, and the dynamic angles that were obtained with the segmentation of the articulation in Kinect.

In [24], the exercises applied to the Alzheimer patients regarding their balance, strength, and coordination during daily activities, including getting up, walking, and sitting down, are listed. Other factors, such as speed and temporal space variables, are described in [25] and [26], respectively. This information includes the number of steps at a specific pre-determined time that an elderly person makes when he/she walks at a spontaneous speed. This process is called step cadence, which describes the minimum and maximum values for the patients who were between 60 and 80 years old. In [27], important approaches related to the longitude and width, step cadence, longitude of the stride, stepping time, stride, support, balance, and stepping speed of the elderly with range limits are described. as shown in Table 3.

2.2 Materials

The present study contains data for 30 Alzheimer patients, including 15 men and 15 women whose ages ranged from 75 to 86 years old among the men and 75 to 89 years old

among the women. The progressive illness of the patients impacted their step. According to [28], the gait step is a basic cognitive action that is altered during the first stages of neuro-degenerative dementia. These patients were selected by clinical psychologists because they fulfilled the neurological criteria to carry out the stepping procedure.

The procedure required specific equipment, including a portable device and an external Kinect application for Windows. In addition, we used Visual Studio 2019 and.NET languages, such as C#, that are compatible with Microsoft Kinect.

In the following section presents the development of the model to the research proposal.

3 Development of the Model

The proposed model has a practical focus: to capture the human skeleton, recognize it, and position it in a digital plane to generate the skeletal form of Alzheimer patients.

A succession of activities are required to detect the human body. First, the Kinect device, while managed by Kinect computer software, establishes the patient's location coordinates on the digital plane (x and y) to an approximate longitude of 3 m. Then, during the recognition process, the skeleton form is visualized on the human structure. Generating the stepping movements can subsequently provide the distance, angle, and location points of the steps, as shown in (see Fig. 4). In [29], the distance between the Kinect and the participant was 2.5 m, the height of the Kinect sensor was 0.7 meters, and the angle of the Kinect was 0°.

The axis for coordinate x was selected because of its orientation to the coronal plane of the body [30] and because of the sagittal orientation of the body. In [31], a walking pattern is generated in the sagittal plane that possesses characteristics of normal human walking, with the body parts synchronized in relation to speed, stride for stature, elevation of the pelvis, quantified articulation, angle in a certain moment, and free space in the heel.

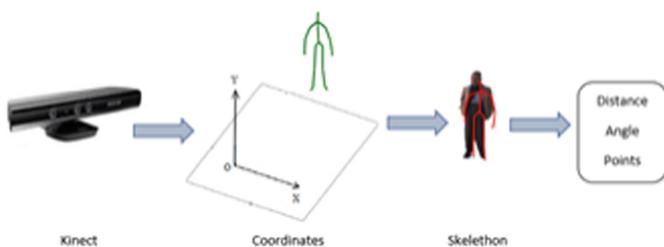


Fig. 4. The skeleton data process

With regard to the ranking and selection of attributes (column), we used functions implemented by Weka with the aim of selecting a subset of values to mediate the quality and the relevance of the data. This process included CfsSubsetEval or CFS, an algorithm for the selection of attributes that describes the ranking of a sub-series of agreement attributes with their correlation based on the estimate of a heuristic evaluation function

applied with a BestFirst in-depth search that uses a retroprocessing limit. The results show 0.993 of the related subsets. In [32] and [33], several applicable techniques in Weka and other tools are identified.

The relevant characteristic to register data in the Kinect device was the lateral pose of the individual during the initial step of the right foot towards the device, which was straight ahead of the patient.

The model identified the base in the inferior structure of the body, which determined the coordinates of an initial point in a (x, y) and a final point in b (x, y), thereby obtaining a distance that generated an angle in c (x, y), as shown in (see Fig. 2).

The present study allows us to meditate on the kinematics [34] responsible for these body movements in connection with the longitude, time, and angle variables. In addition, through biomechanics [35], this study can analyze the displacements of the body in the support phases and the oscillation in the spine, hip, knee, ankle, foot, arms, and trunk, among other aspects of the movement. When individuals vary the positions of their feet, the whole posture of the body changes. For this reason, the variables of the inferior part of the body, including the hip, knee, and foot, have been selected.

The foot angle, which affects the positions of the hip, right knee, and left knee in the Kinect skeletal form, demonstrates where the Euclidean distance is applied in Eq. (1) and determines the points that will be applied to Eq. (6), which produces the inverse cosine of the elements of x [14, 36]. The real values are in an interval of $[0, \pi]$ in the radians in [29]. A similar calculation for the distances using the cosine function and longitude.

$$\text{angulo}_{\text{radianes}} = \text{acos}(x) \quad (6)$$

$$\alpha = \text{angulo}_{\text{radianes}} * \left(\frac{180}{\pi} \right)$$

The distances between the articulations of the tips of the right and left feet in the Kinect skeletal form are part of Eq. (6), the same one that identifies the coordinates of the points x and y.

In the following section presents the results of this research.

4 Results

Using the Kinect device, we gathered data on the walking gait of 30 Alzheimer patients who had common neurological characteristics, such as anxiety and depression, and were 75 to 89 years of age on average. When processing this data in the simulator, the distance traveled on foot and the angle of the formation of the legs demonstrated that the value of uncertainty regarding the patient's stability was between 0 and 1, with the highest value referring to maximum stability, as shown in Table 3. Thus, the applied pattern generates data in an automatic way based on these established rules.

Fuzzy logic was applied because many data series exist in real life where the limits are diffused. If we consider a stable step to be 12 to 35 cm, a person who has a stable step of 36 cm is probably stable in this view, even though he or she would not be stable using normal logic. Therefore, diffused logic is more appropriate for understanding these

Table 3. Data of the walking gait (Unstable/Stable)

Id	Gender	Step distance	Angle	Hip	Head	Shoulder	Unstable/Stable
1	F	26.47	29.70	19	10	18	0.5000
2	F	18.02	16.45	19	11	17	0.8467
...
29	M	18.53	15.44	18	10	8	0.8366
30	M	21.47	19.07	19	9	8	0.5000

transitions, as it allows one to study movement in a more gradual way and is open to rules such as “if in 90%, not in 10%, while in Boolean logic, you can only say yes and no”.

The data in the column (attribute) “unstable/stable” in Table 3 had a weight from 0 to 0.5 if unstable and from 0.51 onwards if stable. As a result, the variables were defusified in order to obtain a quantifiable value in [37]. A clear value was assigned to the aggregated result of the system. This value should also be selected in such a way that it represents the result of the evaluation. Furthermore, the values of the distance, origin, and angle of the gait were validated by a physiotherapist so that a comparison with the data obtained by the Kinect could be carried out. This process allowed the author to calculate the results of the confusion matrix, which were as follows: True Positive ($VP = 11$), True Negative ($VN = 16$), Positive Negative ($FN = 2$), and False Positive ($FP = 1$). In [38], a confusion matrix was used. In [39] the gait pattern was analyzed using a virtual support machine algorithm to determine the precision and applicability of the confusion matrix.

Within the group of 30 people, the results included 12 positive detection tests of the illness. Among these positive tests, 11 were true positives. This finding means that the illness was present in 11 patients, with 1 false positive for a person that did not suffer from the illness.

Therefore, within this study, the probability of having the illness and being registered as positive in the analysis was approximately 91.66% ($11/12 = 0.916$). The probability of obtaining a false positive was approximately 8.33% Eq. (7). These data obtained from the patients show the sensitivity value in 84.61% and the specificity in 94.11%, which indicates that the sick patients were properly diagnosed. In [40], the author used metrics to calculate the confusion matrix results that are shown in Eq. (8).

$$Positivo = \frac{Verdadero Positivo}{(Verdadero Positivo + Falso Positivo)} * 100 \quad (7)$$

$$Negativo = \frac{Falso Positivo}{(Verdadero Positivo + Falso Positivo)} * 100$$

$$Sensibilidad = \frac{Verdadero Positivo}{(Verdadero Positivo + Falso Negativo)} * 100 \quad (8)$$

$$Especificidad = \frac{Verdadero Negativo}{(Verdadero Negativo + Falso Positivo)} * 100$$

The threshold base was calculating using the average total of the data for the distance and angle of the men and women's gait, as shown in Eq. (9). For these values, the law of homogeneity was applied [41, 42] with a quantity of ± 2 , reflecting the principle of mathematical consistency that postulates it is possible to add or subtract physical magnitudes of the same nature, as shown in Eq. (10) and obtain the threshold, as shown in Table 4.

$$\sum_{i=1}^n x_i \quad (9)$$

$$h1 = \sum_{i=1}^n x_i + 2; h2 \sum_{i=1}^n x_i - 2 \quad (10)$$

Table 4. Threshold data

Gender	Law of homogeneity	Total distance		Angle	
F	Promedio	25.968		17.738	
F	± 2	23.968	27.968	15.738	19.738
M	Promedio	31.074		27.728	
M	± 2	29.074	33.074	25.728	27.728

In the following section, we present the conclusion for this research.

5 Conclusion

Previous studies on the walking gait of Alzheimer patients have been highly significant for this research because they have established the criteria for the goniometric, biomechanical, and postural measurements relevant to physiotherapy, among other important references for the superior and inferior parts of the human body.

It is important to understand the characteristics of an Alzheimer patient's step in the sagittal plane, where some parts of the human body intervene during walking.

Using the Kinect device allows one to capture the walking gait of Alzheimer patients within a controlled environment and register patterns of movement in a computer.

The diffused logic system has allowed us to simulate these movements and obtain data about real patients. In addition, the diffused control system has allowed us to combine input variables. For example, the distance and foot angles were defined as a diffused series bound by rules that produced output values for the stability and instability of the patients.

In classical logic, computers compile data in chains of ones and zeros. In other words, they process information based on Boolean Logic. In diffused logic, information about the real world is analyzed by assigning different grades of ownership to the elements of a group, such as between true (1) and false (0) manipulating concepts, like stability or instability, that refer to a person when balancing or losing balance while walking.

We worked on a data series presented in tabular form, where each column contained a specific variable that represented the distance, foot angle, and stability or instability of an Alzheimer patient. These variables can be used as formal criteria to determine the treatment of Alzheimer patients.

In future research, we intend to use pose estimation techniques on common activities for Alzheimer patients to generate behavior alerts.

References

1. Evans-lacko, S., et al.: Actitudes hacia la demencia Informe Mundial sobre el Alzheimer 2019 (2019)
2. Cebrián, H.M.: Esqueletos de Kinect mediante el algoritmo YOLO (2018)
3. Ayed, I., Moyà-alcover, B., Martínez-bueso, P., Varona, J., Ghazel, A., Jaume-i-capó, A.: el equilibrio: el test de alcance funcional con Microsoft Kinect. *Rev. Iberoam. Autom. Inf. Ind.* **14**(1), 115–120 (2017)
4. Konstantinidis, E.I., Bamidis, P.D.: Density based clustering on indoor kinect location tracking: a new way to exploit active and healthy aging living lab datasets. In: 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering, BIBE 2015 (2015)
5. Mountadi, F., García-Vázquez, L.A.: Autenticación multifactor con el uso de un sensor kinect. *Iteckne* **13**(1), 23–35 (2016)
6. Nadhif, M.H., Hadiputra, A.P., Whulanza, Y., Supriadi, S.: Gait analysis for biometric surveillances using kinectTM: a study case of axial skeletal movements. In: 2019 16th International Conference on Quality in Research QIR 2019: International Symposium on Electrical and Computer Engineering, pp. 1–4 (2019)
7. Hbali, Y., Hbali, S., Ballihi, L., Sadgal, M.: Skeleton-based human activity recognition for elderly monitoring systems. *IET Comput. Vis.* **12**(1), 16–26 (2018)
8. Bonenfant, M., et al.: A computer vision system for virtual rehabilitation. In: Proceedings - 2017 14th Conference on Computer and Robot Vision, CRV 2017, vol. 2018-Janua, pp. 269–276 (2018)
9. Elkholy, A., Hussein, M.E., Gomaa, W., Damen, D., Saba, E.: Efficient and robust skeleton-based quality assessment and abnormality detection in human action performance. *IEEE J. Biomed. Health Inf.* **24**(1), 280–291 (2020)
10. Mesbah, N., Perry, M., Hill, K.D., Kaur, M., Hale, L.: Postural stability in older adults with alzheimer disease. *Phys. Ther.* **97**(3), 290–309 (2017)
11. Gowtham Bhargavas, W., Harshavardhan, K., Mohan, G.C., Nikhil Sharma, A., Prathap, C.: Human identification using gait recognition. In: Proceedings of 2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017, vol. 2018-Janua, pp. 1510–1513 (2018)
12. Du, L., Chen, H., Mei, S., Wang, Q.: Real-time human action recognition using individual body part locations and local joints structure, pp. 293–298 (2016)
13. Yang, K., et al.: National laboratory for parallel and distributed processing. College of Computer, National University of Defense Technology, no. May (2016)
14. Tabaghi, P., Dokmani, I., Vetterli, M.: Kinetic euclidean distance matrices. **2017**(November 2017), 1–15 (2019)
15. Amini, A.: An improved technique for increasing the accuracy of joint-to-ground distance tracking in kinect v2 for foot-off and foot contact detection. *J. Med. Eng. Technol.* **43**(1), 8–18 (2019)
16. Rahman, M.W., Zohra, F.T., Gavrilova, M.L.: Rank level fusion for kinect gait and face biometrie identification. In: 2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings, vol. 2018-Janua, pp. 1–7 (2018)

17. El, C.: Capítulo 3. El cuerpo humano Extremidades inferiores, pp. 35–46
18. Miodonska, Z., et al.: Biomedical signal processing and control Inertial data-based gait metrics correspondence to Tinetti test and Berg balance scale assessments. *Biomed. Signal Process. Control* **44**, 38–47 (2018)
19. López, S.A., Larrea, L.C., Ferrer, C.N., Labanda, R.M.: Análisis de las caídas en una residencia de ancianos y de la influencia del entorno Analysis of the falls environmental influence, vol. 27, no. 1, pp. 2–7 (2016)
20. Martín, D., Jiménez, J., Álvarez, F., Carrasco, L.: A novel approach for movement evolution tracking in Parkinson's disease using data analysis and fuzzy logic, pp. 455–461 (2018)
21. Wang, G., Gao, T., Sun, G.: Analytic expression of Mamdani fuzzy system constructed by fuzzy similarity degree and its output algorithm. *J. Intell. Fuzzy Syst.* **37**, 3593–3603 (2019)
22. Rai, J.K., Tewari, R.P., Chandra, D.: Trajectory planning for all sub phases of gait cycle for human-like walking. *Int. J. Eng. Syst. Modell. Simul.* **1**(4), 206–210 (2009)
23. Carneiro, S., Silva, J., Madureira, J., Moreira, D., Guimarães, V., Allen, R.A.: Inertial sensors for assessment of joint angles, pp. 1–4 (2016)
24. Sala, C.A.: Seis semanas de ejercicio físico mejoran la capacidad funcional y la composición corporal en pacientes con Alzheimer Six weeks of physical exercise improve functional capacity and Introducción El Alzheimer se define como una patología neurodegenerativa progresiva, que afecta a, pp. 156–166 (2020)
25. People, H.E.: Velocidad de marcha del adulto mayor funcionalmente saludable Gait Speed in Functionally and Healthy Elder People Velocidade da marcha do idoso funcionalmente saudável, vol. 5, no. 2, pp. 93–101 (2018)
26. Barral, N.C., Aparicio, V.R.: GeroInfo geroinfo, vol. 13, no. 1, pp. 1–23 (2018)
27. Herrero, A.: Estudio de los parámetros espaciales de la marcha en la población anciana española y su asociación con resultados adversos de salud, p. 164 (2017)
28. Martín-Gonzalo, J.A., et al.: Permutation entropy and irreversibility in gait kinematic time series from patients with mild cognitive decline and early Alzheimer's dementia. *Entropy* **21**(9), 1–21 (2019)
29. Ahmed, M.H., Tahir Sabir, A., Maghdid, H.S.: Kinect-based human gait recognition using triangular gird feature. In: 1st International Conference on Advanced Research in Engineering Sciences, ARES 2018, pp. 1–6 (2018)
30. Katsumi, R., Mochizuki, T., Sato, T., Kobayashi, K., Watanabe, S.: Contribution of sex and body constitution to three-dimensional lower extremity alignment for healthy, elderly, non-obese humans in a Japanese population (2018)
31. Lee, S., Lee, J., Lee, D.G.: Walking pattern generation in sagittal plane possessing characteristics of human normal walking, pp. 1066–1072 (2016)
32. Informática, E., Nacional, U., Educación, D., Juan, C.: Comenzando con Weka: Filtrado y selección de subconjuntos de atributos basada en su relevancia descriptiva para la clase (2016)
33. Dwivedi, S.: Comprehensive Study of Data Analytics Tools (RapidMiner, Weka, R tool, Knime) (2016)
34. Maeda, H., Ikoma, K., Toyama, S., Taniguchi, D., Kido, M.: Gait & posture a kinematic and kinetic analysis of the hip and knee joints in patients with posterior tibialis tendon dysfunction; comparison with healthy age-matched controls. *Gait Posture* **66**(August), 228–235 (2018)
35. Pedrinolla, A., Venturelli, M., Fonte, C., Munari, D.: Exercise training on locomotion in patients with Alzheimer's disease: a feasibility study. *J. Alzheimer's Dis.* **61**, 1599–1609 (2018)
36. Amini, A., Emb, I., Banitsas, K., Hosseinzadeh, S.: A new technique for foot-off and foot contact detection in a gait cycle based on the knee joint angle using microsoft kinect v2, pp. 153–156 (2017)
37. Tóth-laufer, E.: Improvement possibilities of the maximum defuzzification methods, no. 2, pp. 339–344 (2019)

38. García-balboa, J.L., Alba-fernández, M.V., Ariza-lópez, F.J., Rodríguez-avi, J.: Homogeneity test for confusion matrices: a method and an example, pp. 1203–1205 (2018)
39. Savić, S.P., Prodanović, N., Devedžić, G.: Algorithm, pp. 8–11 (2020)
40. Balaji, E., Brindha, D., Balakrishnan, R.: Jou RNA IP. Appl. Soft Comput. J. 106494 (2020)
41. Objectives, L.: Basic Concepts and Definitions 1 (2017)
42. Tsai, S., Member, S., Jen, C.: H ∞ stabilization for polynomial fuzzy time-delay system: a sum-of-squares approach. **14**(8) (2018)



The Analysis of Triples of Triangular Norms for the Subject Area of Passenger Transport Logistics

Zbigniew Suraj¹, Oksana Olar², and Yurii Błoszko¹(✉)

¹ University of Rzeszów, Rzeszów, Poland
zbigniew.suraj@ur.edu.pl

² Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine
o.olar@chnu.edu.ua

Abstract. This paper presents the analysis of a group of triples of functions which imply different combinations of t-/s-norms in the weighted fuzzy Petri nets. The wFPN model was realized in the system PNeS for the subject area of passenger transport logistics. The experiment describes the realization of 15 different triples of functions associated in the range between a minimal triple (LtN, LtN, ZsN) and a classical one (ZtN, GtN, ZsN). It is expected to achieve a numerical rise of the resulting calculations in accordance to the change of the triples in the range from the minimal to the classical one. Additionally, the analysis of achieved decisions with the change of triples is considered for the given subject area.

Keywords: Decision-making system · Intelligent computational techniques · Weighted fuzzy Petri net · Knowledge representation · Transport logistic problem

1 Introduction

The improvement of decision-support systems requires the development, application and analysis of intelligent computational techniques in order to achieve the most concrete results for the given task [1–3]. The search for the best type of transport in the subject area of transport logistics can be considered with the application of a dynamic discrete mathematics. Therefore, the application of Petri nets can be a good starting point for such type of the problem [4, 5]. Yet, Petri nets have faced lots of modifications and improvements. The most significant is the conception of weighted fuzzy Petri nets (wFPN) which apply fuzzy coefficients which describe the strength of the connection between an input place and the transition [6]. 125 different combinations of t-/s-norms in a group of three functions under each transition can be tested in the wFPN [7, 8]. Therefore, it necessitates to provide a deep analysis of these triples and their application in the relevant subject areas.

As it was mentioned above, the subject area of passenger transport logistics can serve as a good example of wFPN disclosure. Until now, the structure of transport logistics was described in a four-level scheme represented with interconnected tables of type

“Object-property” [9–11]. The completeness of the structure allowed creation of the hierarchical wFPN [12]. There were tested two types of triples: optimized (ZtN, ZtN, LsN) and classical (ZtN, GtN, ZsN) [13]. The optimized triple of functions resulted in higher numerical values at output places comparing to the achieved results from the classical triple. In addition, the change of triples led to the change of target decisions, which influenced the hierarchy. Therefore, the deep analysis of triples of functions which are located in-between classical and optimized one in the research of the best type of transport (the initial wFPN model in the hierarchy) was conducted [13]. Two different approaches were presented and tested: a) the decisive majority of equivalent decisions achieved by different triples in-between increases the probabilistic accuracy of the classical or optimized triple; b) all numerical results achieved by every triple of functions for each object were calculated using the arithmetic mean value. Therefore, the influence of triples and their resulting numerical values were considered significantly in the process of decision making. Moreover, the difference between achieved results after application of those two approaches led to the necessity of applying additional factors for justifying the truth-probability of the achieved decisions (or considering some additional alternatives) [14].

Thus, this paper aims to continue the research in the subject area of passenger transport logistics aiming to find out the best type of transport. Different combinations of t-/s-norms which are located in-between minimal (LtN, LtN, ZsN) and classical (ZtN, GtN, ZsN) triples are considered in order to get a deeper understanding of achieved numerical values and associated decisions.

2 Description of the Task

The graphical description of the task aiming to achieve a decision on the best type of transport for the given input properties is presented in Fig. 1.

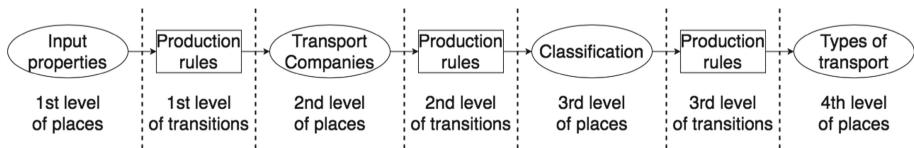


Fig. 1. Scheme of the process of finding the best type of transport.

Figure 1 can be considered as an introduction to the problem, which consists of four levels of places and three levels of transitions interconnected pairwise between each other. This graphical representation implies the simplified structure of the wFPN model to be created (Fig. 3).

First level of places includes properties for different passenger transport companies presented on the second level of places. Every level of places is presented in the knowledge tables of type “Object-property” [9, 10]. At the very beginning, the first level is considered as a list of properties (columns) in the table, while the second level – objects (rows). Production rules are used to establish a connection between properties

and objects. Transition can be interpreted as the application of the production rule both in the structure (Fig. 1) and wFPN model.

The example of production rule is as follows: IF r_{i1} AND (OR) ... AND (OR) r_{in} THEN d_j , where r_{ik} ($k = 1, \dots, n$) – property, and d_j – object. Herein, production rule was applied at the internal level – inside the table. The following steps imply the application of production rules at the outer level – between tables, where objects from the previous knowledge table become properties in the following table on the basis of which, new production rules are created.

Thereby, passenger transport companies which are presented on the second level of places and which are objects in the first knowledge table, - become properties in the following (second) knowledge table. These properties are connected to the classification of type of kinds of transport which is presented on the third level of places in Fig. 1. Production rules are created on the basis of the second knowledge table which imply a connection between properties and objects. In the same manner, a connection between third and fourth level of places in the last (third) knowledge table is established, where the last level of places is the final one (i.e. objects from the third knowledge table are the final decisions). The final level of objects describes target decisions of types of transport.

3 Mathematical Tool of the wFPN

Production rules which are created from the knowledge tables of type “Object-property” form a basis for creating wFPNs. Every wFPN consists of some number of places and transition connected pairwise with directed arcs. Every place describes property or object taken from the knowledge table, while production rules are represented with the corresponding transitions. In order to improve the accuracy of calculations, the FPN were modified with weights establishing the conception of wFPN [6]. Weights are only associated with arcs which connect input places with transitions and describe the strength of the corresponding connection. As follows, the input value (marking) of a place is multiplied by the value associated with the arc and the resulting value is set for the transition as the first input value.

Every transition includes the following list of mathematical tools: (1) truth degree betta $\beta(t)$: calculated by the following formula $\beta(t) = k/(k + 1)$, where k is a number of input places connected to this transition [11]; (2) threshold function gamma $\gamma(t)$: set by the experts in the corresponding subject area of research; (3) triple of operators (In , Out_1 , Out_2): every element of the triple is replaced with the corresponding t-/s-norm (with the respect to the logical operator used in the production rule).

In order to fire a transition, the following condition must be satisfied:

$$In(w_{i1} \cdot M(p_{i1}), w_{i2} \cdot M(p_{i2}), \dots, w_{ik} \cdot M(p_{ik})) \geq \gamma(t) > 0 \quad (1)$$

where: (a) In is an input operator instantiated with some t-/s-norm; (b) w_{ij} ($j = 1, \dots, k$) is a weight which is connected to the corresponding place; (c) $M(p_{ij})$ is a marking of a place.

When the condition is satisfied, the first operator In is replaced with some t-norm (in case of logical AND in the production rule) or s-norm (in case of logical OR in the production rule) which takes all markings from input places for calculations by the correspond function. In the same manner, the second operator Out_1 which is replaced

with some function, takes the numeric value calculated by the first operator *In* as the first input and betta $\beta(t)$ as the second one. The third operator *Out*₂ which is also replaced with some function, takes the resulting value of the second function of the triple as the first input and the markings of output places as the second one. In this manner, the result of the last function of the triple sets the final value for the output place.

The number of possible combinations of triples of functions is equal to 125 and can be graphically represented (Fig. 2) [13, 15].

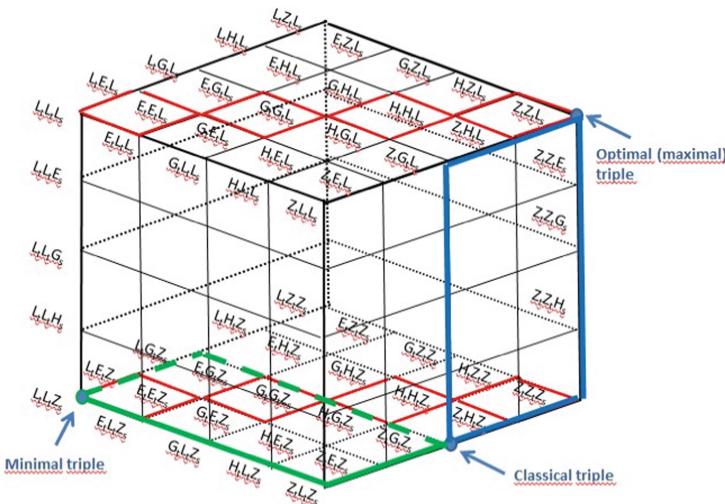


Fig. 2. Graphical representation of 125 combinations of triples of function (logical AND).

Figure 2 represents a cube with all-possible triples of function which are structurally ordered from the minimal triple (LtN , LtN , ZsN) which results in the lowest possible numerical result at the output to the maximal (optimal) triple (ZtN , ZtN , LsN) which results in the highest possible numerical result at the output. Additionally, there is highlighted the classical triple (ZtN , GtN , ZsN) which is located in-the-middle between minimal and maximal triples. The combinations of triples between classical and optimized are highlighted in blue forming a rectangle with 15 combinations of triples. In the same manner, the 15 combinations of triples which are in the range between minimal and classical triples are highlighted in green. The combinations shown in the blue rectangle have already been tested in the wFPN model (Fig. 3), the structure of which is presented in Fig. 1 [14]. The first function of the triple is always ZtN , since it is located on the plane of the cube which represents this function. Vertical lines of the blue rectangle represent 3 alternatives to choose from: HtN , GtN and ZtN . Horizontal lines of the blue rectangle represent 5 alternatives to choose from: ZsN , HsN , GsN , EsN and LsN . Thus, there was a possibility to test 15 combinations.

Yet, the specification of the wFPN model should be considered in the first place: the firing sequence is done in a step-by-step format (i.e. only the first level of places is filled with values while all other levels of places are empty, before calculations of the previous

level). It means that the last function of the triple loses the influence in the calculation process, because the second input value is equal to 0 (since all of output places are always empty). Therefore, the third function of the triple gets numerical value for the input only from the second function. Thus, the resulting value of the third function does not change and is equal to the result calculated by the second function of the triple in the wFPN model described in the step-by-step format.

As a result, the list of norms presented on the vertical lines of a blue rectangle is neglected because of the wFPN model structure (Fig. 1, 3). It leads to the reduction of the number of possible combinations from 15 to 3 which can be tested [13, 14]. Therefore, this paper suggests to test a green rectangle. This rectangle is supposed to give more combinations to be tested as well as a wider list of the achieved decisions. Additionally, it necessitates to analyze the advantages and disadvantages of applying additional combinations of triples for the given problem in the subject area of passenger transport logistics (Fig. 1).

The main benefit of the green rectangle lies in the fact that it is located on the different plane of the cube (Fig. 2) compared to the location of the blue rectangle. Therefore, the location of combinations of triples is changed. Since the wFPN model is made in a format of a step-by-step activation, the third element of the triple is again neglected. In this case, the difference is that it neglects only one function (ZsN) which is represented on the bottom plane of the cube which is different from the plane which includes a blue rectangle. Horizontal lines of the green rectangle describe the list of t-norms which can be set as the first element of the triple: LtN, EtN, GtN, HtN and ZtN. Vertical lines include a list of t-norms which can be set as the second element of the triple: LtN, EtN and GtN. As a result, it enables to test 15 different combinations of triples.

The mathematical description of functions presented in a green rectangle is as follows:

$$LtN(a, b) = \max(0, a+b-1) \text{ (Lukasiewicz t-norm)}, \quad (2)$$

$$EtN(a, b) = \frac{ab}{2 - (a + b - ab)} \text{ (Einstein t-norm)}, \quad (3)$$

$$GtN(a, b) = ab \text{ (Goguen t-norm)}, \quad (4)$$

$$HtN(a, b) = \begin{cases} 0 & \text{for } a = b = 0 \\ \frac{ab}{a+b-ab} & \text{otherwise} \end{cases} \text{ (Hamacher t-norm)}, \quad (5)$$

$$ZtN(a, b) = \min(a, b) \text{ (Zadeh t-norm)} \quad (6)$$

$$ZsN(a, b) = \max(a, b) \text{ (Zadeh's norm)}. \quad (7)$$

4 Net Representation of the Task

This paper presents the set of input values for wFPN model presented in Fig. 3: 0.9, 0.75, 0.8, 0.85, 0.9, 0.65, 0.5, 0.6, 0.8, 0.75.

Figure 3 represents wFPN model made according to the structure in the Fig. 1. It consists of four arrays of places and three arrays of transitions which were created on the basis of production rules extracted from three knowledge tables of type “Object–property”.

The results of application of different combinations of triples of functions from the green rectangle (Fig. 2) for three resulting objects in Fig. 3 are presented in the Table 1.

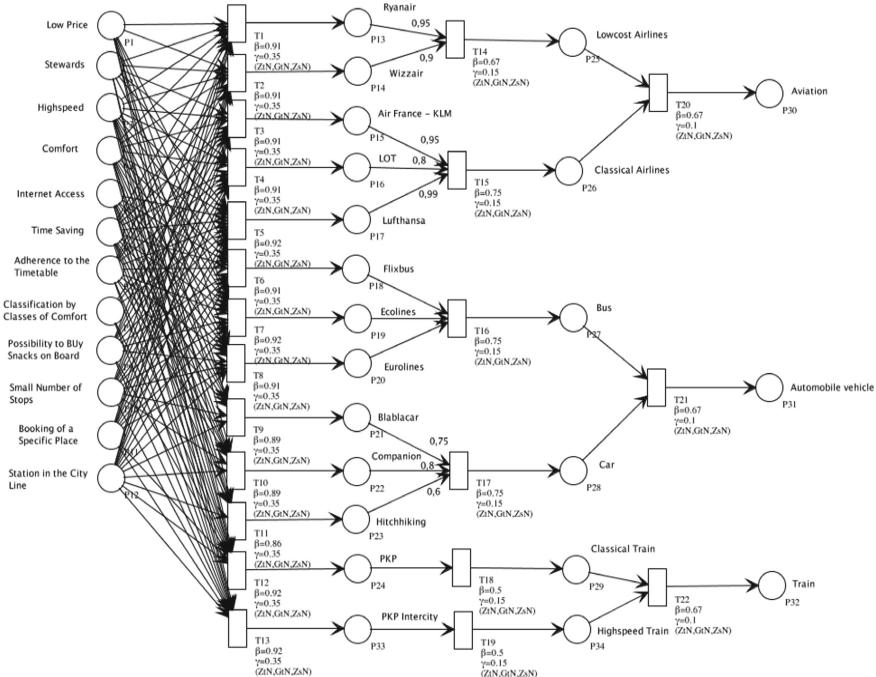


Fig. 3. wFPN model for the best type of transport.

As far as, wFPN model operates with the extremely low numerical values, Table 1 describes two cases.

Case 1. Gamma $\gamma(t)$ includes the values presented in Fig. 3 which are equal to 0.35 on the first level and to 0.15, 0.1 on the second and third levels correspondingly. In this case, only two last combinations of triples [(ZtN, EtN, ZsN) and (ZtN, GtN, ZsN)] marked in bold in the Table 1 resulted with some numerical values at the final output in the wFPN (Fig. 3). These two triples have finished calculations, because the condition in formula 1 was satisfied for all three levels of transitions. Also, triple (ZtN, LtN, ZsN) marked in italics achieved interesting results from the observational point of view: (1) all transitions were fired on the first level since the condition (1) was satisfied; (2) third level of transitions: transition which leads to the object “Aviation” has not been fired, since the condition (1) was not satisfied ($0.039 < 0.1$). Therefore, output value for the object “Aviation” was not received. Thus, the output for the “Aviation” object is marked as “Undefined”. The similar issue has been observed for the object “Car” in the process of firing the second level of transition for the third level of places (point 3); (3) second

level of transitions: transition which leads to the object “Car” was not fired since first the function (operator *In*) of the triple resulted in value equal to 0.141, which is lower than the gamma $\gamma(t)$ value which is equal to 0.15. Thus, the condition (1) was not satisfied and transition has not been fired. In accordance to this observation, resulting object of the wFPN model “Automobile vehicle” have lost one of input properties (“Car”) in the wFPN, the connection of which was previously established in the knowledge table; (4) second level of transitions: the remaining object “Bus” which became a property for the last level of transitions in the “Automobile vehicle” branch development could be used to finalize calculations on the third level of transitions to achieve final numerical result for the “Automobile vehicle” object and this branch of type of transport. The condition (1) for firing the transition on the third level is satisfied ($0.16 > 0.1$). Yet, the second function of the triple (LtN) zeroed the input values and therefore, the resulting object also received 0 value at the output. Thus, if there were any additional levels of transitions after this one, the development would not take place, since this branch is being cut out on the fourth level of places. The same phenomenon has been observed in the “Railway (Train)” branch on the second level (point 5); (5) second level of transitions: two transitions which result in objects “Classical Train” and “High-speed Train” previously satisfied the condition (1) ($0.4075 > 0.15$ and $0.42 > 0.15$ respectively). Yet, numerical values which were received at the output are equal to 0. It can be explained by the second function in the triple LtN, which zeroed input values after applying the formula. Therefore, the resulting object of the wFPN model “Train” lost both objects on the current level (which is considered as properties at the following level). In this way, the development of the situation for this branch was cut out on the third level of places. For this reason, the resulting output is marked as “Undefined”.

Table 1. Resulting values of 15 triples of function for thre output objects

Triples/Decisive objects	Aviation	Automobile vehicle	Train
LtN, LtN, ZsN	Undefined	Undefined	Undefined
LtN, EtN, ZsN	Undefined	Undefined	Undefined
LtN, GtN, ZsN	Undefined	Undefined	Undefined
EtN, LtN, ZsN	Undefined	Undefined	Undefined
EtN, EtN, ZsN	1,45237E–15	2,1184E–18	9,09742E–09
EtN, GtN, ZsN	4,8911E–15	7,91109E–18	3,17417E–08
GtN, LtN, ZsN	Undefined	Undefined	Undefined
GtN, EtN, ZsN	3,51698E–10	3,93223E–12	2,94159E–06
GtN, GtN, ZsN	1,17306E–09	1,44571E–11	1,02028E–05
HtN, LtN, ZsN	Undefined	Undefined	Undefined
HtN, EtN, ZsN	0,008693656	0,006622562	0,010795526
HtN, GtN, ZsN	0,015832969	0,011981633	0,022235616
ZtN, LtN, ZsN	Undefined (0)	0	Undefined
ZtN, EtN, ZsN	0,116475097	0,057340053	0,088145482
ZtN, GtN, ZsN	0,18291	0,09723375	0,1502475

Case 2. Gamma $\gamma(t)$ value is considered as the lowest possible value to make transitions fireable. Thus, there were achieved all 15 sets of results presented in the Table 1. The results for the last two triples has not been changed since gamma lowering did not influence the output results which were achieved with even higher threshold value of gamma. Yet, it allows to face changes in points 2 and 3, where the transition firing was blocked by the condition (1). Therefore, the next two paragraphs describe possible changes that could be observed in the process of application of the triple (ZtN, LtN, ZsN) for objects which did not receive their values due to the failure to comply the condition (1).

In the case 2, the second point is modified and the transition is considered to be fireable. Unfortunately, the result for object “Aviation” became equal to 0 (it is noted in brackets in the Table 1), since the second function of the triple (LtN) zeroed input values in the same manner as it has been observed in “Automobile vehicle” object on the same level (case 1, point 4).

In the case 2, the third point should also be modified. When gamma $\gamma(t)$ is lowered and condition (1) is satisfied then the transition which leads to the object “Car” is considered as fireable. This object underwent the same phenomenon described in the previous paragraph and in case 1, point 4: the transition was fired, but the object received 0 value because of the second function LtN. Thus, object “Car” was again excluded from the further calculations in the wFPN. Moreover, it can be considered equally: the object did not receive output value because the condition (1) was not satisfied, thus this object is considered as “Undefined” (or zero informative value) and the situation when the transition fired, but also resulted in 0 value because of calculations. In both cases, 0 value means that the object is empty and cannot be longer considered in the wFPN model.

From the given analysis of the results presented in Table 1, the following issues can be observed: (1) as it was expected, the correctness of the following sequence LtN \leq EtN \leq GtN \leq HtN \leq ZtN is confirmed by the achieved numerical values, where LtN achieved the lowest possible output and ZtN achieved the highest possible output correspondingly; (2) the decisions differ with the change of triples of functions. There is an argue that with the incensement of number of properties, the result for the object becomes lower compared to the same situation differing in fewer properties. It should be noted that fuzzy values for properties did not differ a lot and are in the close range for objects to be calculated. This situation can be observed in the number of places connected to the second level of transition in Fig. 3. Therefore, this aspect will be analyzed in more detail in the following papers; (3) numerous issues raised from the application of LtN function. As it can be observed in Table 1, every combination of triples which imply application of LtN function resulted in 0 value (at best) or the value became so low that at some level of transitions it did not meet the requirement for firing the transition (condition (1)) and it resulted in “Undefined” state (at worst). In both cases, application of LtN did not provide values that can be analyzed and compared with other results. Thus, it is suggested not to apply LtN in calculations for wFPN with a large number of level of transitions, because there is high risk that soon or later, the value will drop so low that the transition will not fire at all (failure to comply the condition 1); (4) in the case of using LtN, the following situation was noticed: with the increase in the number of inputs calculated according to the formula LtN $(a, b) = \max(0, a + b - 1)$, the risk

of obtaining the value equal to 0 at the output increases. The following calculation: $a + b - 1$ requires large input values (a and b). An increase in the number of low values compared to the number of high values proportionally increases the risk of obtaining resulting value less than zero in the calculation of $a + b - 1$. As soon as this occurs, LtN(a, b) is set to 0, which leads to the neglect of the object in further calculations, because its degree of truthfulness can no longer be taken into account; (5) the decisions achieved by triples (ZtN, GtN, ZsN) and (ZtN, EtN, ZsN) which were received with the initially set values for the threshold functions are installed in the following order by their resulting values (from the most proposed decision to the least suggested decisions): "Aviation", "Train", "Automobile vehicle". (ZtN, LtN, ZsN) is a triple which includes the sequence change of the decisions with the change of values set for the threshold functions. Thus, decisions which were achieved by other triples except for three above mentioned are set in the following order: "Train", "Aviation", "Automobile vehicle". Remark: triples which achieved "Undefined" state were excluded. Thus, the threshold function gamma $\gamma(t)$ plays a vital role in the decision-making process for the triples presented in a green rectangle (Fig. 2).

5 Conclusion

This paper presented the application of 15 triples of functions (t-/s-norms) in the range from the minimal (LtN, LtN, ZsN) to the classical one (ZtN, GtN, ZsN) for the subject area of passenger transport logistics. The most important observations presented here are: 1) the output results from the triples in the green rectangle are sensibly lower; 2) the value of gamma $\gamma(t)$ plays a vital role in the wFPN model operation and in the process of firing a transition; 3) with the change of triples from the minimal to the classical one, numerical values at the output were changing correspondingly; 4) the exclusions from the statement in point 3) are applied for triples which includes LtN function. As far as LtN function gives the lowest possible output result (which is often equal 0), the resulting objects were receiving 0 values at the output or the transition was not even fired at all. Thus, it is suggested not to apply LtN in the wFPN model which follows the strategy of a step-by-step activation; 5) in the process of application of triples from the green rectangle, the number of input places plays more notable role compared to the triples located in the blue rectangle; 6) there is a switch in the order of decisions achieved by triples with initially installed threshold function gamma $\gamma(t)$ and decisions which were achieved after lowering the threshold function; 7) the decisions are sensibly changing with the change of triples of function and the threshold function in a green rectangle. Thus, it is complicated to estimate the correctness of the achieved decisions at this moment.

According to the observations described above, there is a need of combining achieved results of triples presented in blue and green rectangles. Hence, it necessitates to generalize the achieved decisions and to provide a concrete suggestion on the application effectiveness of the chosen triple of function. These goals will be considered in the following papers with the corresponding analysis and conclusions.

References

1. Cortes, Z.A.J., Serna, M.D.A., Gomez, R.A.: Information systems applied to transport improvement. *Dyna* **80**(180), 77–86 (2013)
2. Díaz-Parra, O., Ruiz-Vanoye, J.A., Bernábe Loranca, B., Fuentes-Penna, A., Barrera-Cámara, R.A.: A survey of transportation problems. *J. Appl. Math.* **2014**, 1–17 (2014)
3. Suraj, Z., Hassaniene, A.E., Bandyopadhyay, S.: Weighted generalized fuzzy petri nets and rough sets for knowledge representation and reasoning. In: *Lecture Notes in Artificial Intelligence*, vol. 12179, pp. 61–77 (2020)
4. Cardoso, J., Camargo, H. (eds.): *Fuzziness in Petri Nets*. Physica-Verlag (1999)
5. Suraj, Z., Grochowalski, P.: Petri nets and PNes in modeling and analysis of concurrent systems. In: *Proceedings of International Workshop on CS&P*, pp. 1–12 (2017)
6. Chen, S.-M.: Weighted fuzzy reasoning using weighted fuzzy Petri nets. *IEEE Trans. Knowl. Data Eng.* **14**(2), 386–397 (2002)
7. Klement, E.P., Mesiar, R., Pap, E.: *Triangular Norms*. Kluwer (2000)
8. Suraj, Z.: A new class of fuzzy Petri nets for knowledge representation and reasoning. *Fundam. Inform.* **128**(1–2), 193–207 (2013)
9. Lyashkevyc, V., Olar, O., Lyashkevych, M.: Software ontology subject domain intelligence diagnostics of computer means. In: *Proceedings of IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Berlin, September 2013, pp. 12–14 (2013)
10. Lokazyuk V.: Software for creating knowledge base of intelligent systems of diagnosing process. In: *Advanced Computer System and Networks: Design and Application*, Lviv, pp. 140–145 (2009)
11. Suraj, Z., Olar, O., Bloshko, Y.: Conception of fuzzy petri net to solve transport logistics problems. In: *Current Research in Mathematical and Computer Sciences*, II, pp. 303–313. University of Warmia and Mazury Press, Olsztyn (2018)
12. Suraj, Z., Olar, O., Bloshko, Y.: Hierarchical weighted fuzzy Petri nets and their application for transport logistics problem. In: *Proceedings of International Conference on Intelligent Systems and Knowledge Engineering*, Cologne, pp. 404–411. World Scientific (2020)
13. Suraj, Z., Olar, O., Bloshko, Y.: The analysis of human oriented system of weighted fuzzy petri nets for passenger transport logistics problem. In: *Advances in Intelligent Systems and Computing*, vol. 1197, pp. 1580–1588 (2021)
14. Suraj, Z., Olar, O., Bloshko, Y.: Modeling of passenger transport logistics based on intelligent computational techniques. *Int. J. Comput. Intell. Syst.* (2021), Submitted
15. Suraj, Z.: Toward optimization of reasoning using generalized fuzzy petri nets. In: *Lecture Notes in Artificial Intelligence*, vol. 11103, pp. 294–308. Springer (2018)



Real-Time Stair Detection Using Multi-stage Ground Estimation Based on KMeans and RANSAC

Yuchen Li¹, Lina Yang¹⁽⁾, and Patrick Shen-Pei Wang²

¹ School of Computer, Electronics and Information, Guangxi University,
Nanning 530004, People's Republic of China

lnyang@gxu.edu.cn

² Computer and Information Science, Northeastern University,
Boston, MA 02115, USA
patwang@ieee.org

Abstract. Multiplane estimation from three-dimensional (3D) point clouds is a necessary step in the negative obstacle detection. In recent years, different Random Sample Consensus (RANSAC) based methods have been proposed for this purpose. In this paper, we propose a multi-stage algorithm based on RANSAC plane estimation and KMeans clustering, and apply it to the negative stairs detection. This method contains two steps: first, it clusters the point clouds and downsamples them; second, it estimates the planes by iteratively using RANSAC algorithm with the downsampled data. Finally, according to the relationship between regions to determine whether there is an obstacle in front of the autonomous vehicle. Our experimental results show that this method has satisfactory performance.

Keywords: Negative stair detection · Multiplane estimatin · Safety drive · KMeans clustering

1 Introduction

With the rapid development of sensing technology, the detection of positive obstacles has been greatly studied and improved. However, due to the hard measurement and uncertainty of negative obstacles, the detection of negative obstacles is still a big challenge for the researcher in the filed of autonomous driving [1]. Stairs (especially the down stairs), which are frequently occurrence in streets, are one kind of negative obstacles. Stairs can significantly bring security risk to the autonomous driving robot. To reach the goal of safe driving on the open road, the detection of stairs is a necessary.

In order to deal with such issue, many researcher have done extensive work on it. For example, the authors in [2] proposed to solve it within the framework of 3D computer vision, for the reason of non-measurement of LiDAR device. Taking the ground plane as reference, a remarkable difference between positive obstacles

and negative obstacles is that: Positive obstacles usually protrude above the ground, while negative obstacles concave below the ground. According to this observation, the authors in [3] proposed a height-length-density (HLD) terrain classifier to uniformly detect the negative (such as stairs) and positive obstacles, using three kinds of features: the changes of height and the intervening distance (both from the upper and lower surfaces), and the point density in the vicinity of each discretized cell. Although this method works well in the restricted scene, the refined features do not fit the complex environment, for example the scene of open road. To robustly and steadily detect the stairs, we seek to the unified methods that do not rely on any prior models.

An important character of stair is that: either the height of points gradually decrease when the stair can be sensed by the onboard LiDAR device, or there are not any returned points when the stair can not be seen by the LiDAR device. This phenomenon implies that the stair has partly shared the property of positive obstacles and negative obstacles. Therefore, the key of stair detection is the estimation of ground plane. In other word, ground plane estimation provides a bridge between the positive obstacles and negative obstacles. It has become one of the efficient tools in intelligent transport system area.

One of the most popular approaches to estimate the ground from a set of 3D points is Random Sample Consensus (RANSAC). This method was first proposed by Schnabel *et al.* in [4].

Researchers selected a random subset of input values, and then estimated a model from these values. The quality of the model was evaluated based on the overall data set and the process was applied iteratively until convergence or for a predefined number of iterations. Finally, selecting the best model which was most suitable for the whole point set.

As RANSAC is not limited by specific dimensionality [5], RANSAC has been widely studied and used in the task of plane estimation. For example, in the 3D scene reconstruction, Xu *et al.* [6] compared the performance of plane estimation between Hough Transform (HT) and RANSAC, and Yufan *et al.* [7] applied RANSAC to plane detection when the robot stereo matching. Muhovic *et al.* [8] focused on finding waterborne planes in the moving scenarios during autonomous navigation. Zhou *et al.* [9] proposed a unified framework for robust 3D supporting plane estimation using results from object shape detection and 3D plane estimation.

Multiple planes estimation was needed be applied in many cases of negative obstacle detection, thus researchers have continually improved RANSAC algorithm to effectively extract multi-planes. The approach in [10] was performed by using iterative RANSAC to extract planar and non-planar regions. In this paper researchers firstly selected a set of random points and used RANSAC to estimate a plane, then repeated this process with the rest of the points until the remaining points could not be used to fit the estimated plane. [11] proposed an efficient multi-planes extraction method based on scene structure priors as an improvement of traditional RANSAC method. In recent years, researchers tend to pre-cluster the data in advance and then estimate plane from result of cluster,

and the related experiment has shown better performance. Gallo and Manduchi proposed the CC-RANSAC algorithm [12]. This method assumed that there were steps, curbs or slopes on the ground, and then used eight connected neighbors components to pre-cluster the obtained data. These pre-clustering data would make RANSAC get better results. Saval-Calvo and Marcelo [5] estimated each plane in the pre-clustering step by adding vector normal information and scene knowledge and then added constraints to estimate the best plane in the step of RANSAC. The experiment showed better results when the signal-to-noise ratio is low.

After considering the RANSAC methods for plane estimation from point clouds, it is obvious that there are still challenging problems when high accuracy is required, mainly with noisy data. In addition, most of the current studies on negative obstacles detection are focused on the field scene in which the obstacles are complex and irregular.

In this paper, a multi-stage ground estimation method based on RANSAC and KMeans is proposed to detect the stairs. The input includes the original 3D point clouds data collected by ourselves, and using KMeans to cluster the obtained data. Next, we iteratively use RANSAC with added constraints to estimate planes. Finally, whether there is a negative obstacle in front of the autonomous vehicle is determined according to the geometric structure of the negative obstacle by processing the obtained planes.

The main contributions of this paper include:

1. Explore the technical application of the scene of negative stairs obstacles.
2. An effective and stable negative obstacle detection method.

The remainder of this paper is organised as follows. The Sect. 2 introduces the proposed method, in which Sect. 2.1 describes negative obstacle features, Sect. 2.2 and Sect. 2.3 introduce KMeans clustering algorithm and RANSAC multi-plane estimation algorithm respectively. Then we presents the experimental process and results in Sect. 3. Finally we give a conclusion in Sect. 4, including the description of the advantages of the method and the summary of the characteristics of the related scenarios.

2 Proposed Method

The method is divided into two steps: using obtained point clouds data as inputs, we first cluster them with KMeans algorithm (see Sect. 2.2); Applying RANSAC algorithm to estimate the planes and making decision of the region categories (see Sect. 2.3). According to the fitted plane equation, we calculate the relationship between the point cloud in each grid and the plane equation. Judging whether the area is a negative obstacle area according to the calculated results. Figure 1 shows the diagram of KMeans and RANSAC steps.

The hyphen arrow represents the real points passed between steps, while the thick arrow represents the constraint in the model. The algorithm has one inputs: 3D LiDAR system Obtained point clouds data. Based on the point and normal

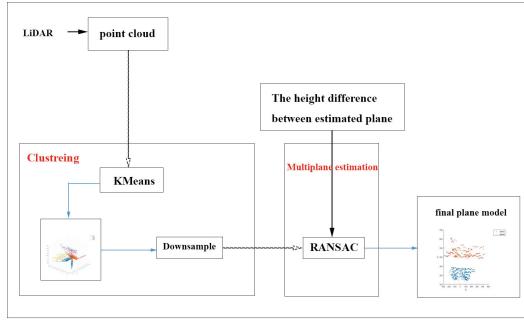


Fig. 1. General overview of the process

vector information, using traditional KMeans to cluster the points cloud, and the clustered data are downsampled to reduce the amount of data processing. Then the downsampled points data are sent to RANSAC to estimate the planes.

2.1 Feature of Negative Obstacle

Figure 2 shows the scene of stairs and the geometric structure of the negative stair in 2D side view respectively. In Fig. 2(b), the W represents the width of the negative obstacle, H represents the vertical distance from the position of LiDAR to the ground, and L represents the distance between the front end of the LiDAR and the side of the negative obstacle near the vehicle.

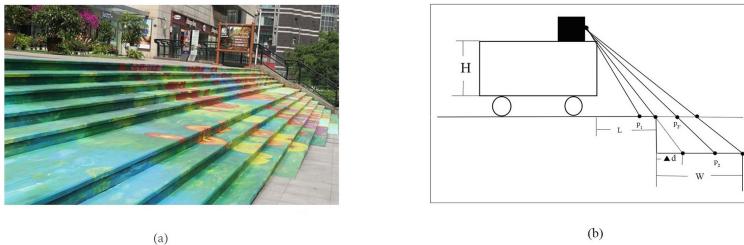


Fig. 2. (a): The scene of stair; (b): Typical geometric structure of negative obstacle

If there is a negative stair in front of the vehicle, the distance between the point belonging to the ground area and the point belonging to the stair area becomes larger when LiDAR scans the stair. In this way, the distance Lp_1p_2 is longer than $Lp_1p'_2$. And we can also see that existing a blind area Δd from the top view of the obtained LiDAR points cloud (Fig. 3(a)) due to the structural characteristics. According to Δd , we can use it to help determine whether existing stair in the interesting region.

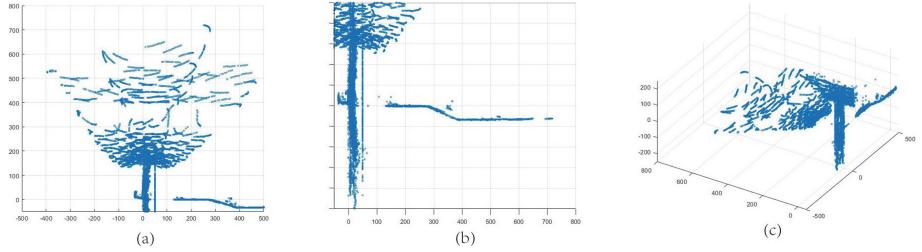


Fig. 3. Data describing the negative stair obstacle collected in front of the vehicle. (a): The top view; (b): The side view; (c) The oblique view

2.2 Point Cloud Clustering

Before carrying out the task of ground estimation, it is necessary to preprocess the point clouds data. As the LiDAR usually produces uneven point clouds data (for example, the points outside the interesting region), these points may make the results inaccurate. The solution is to pre-cluster the point clouds data and prune these noisy points according to the index information of the clustered points.

3D KMeans clustering algorithm. For the obtained point clouds data, the points in the same plane are usually continuous and adjacent. If there is a sudden change in the distribution of these data points, we usually determine whether the two points come from different planes in terms of the mean distance of points. In the process of general clustering, the selection of distance threshold Δt will have a direct impact on the clustering effect. If the selected threshold is too small, the LiDAR data belonging to the same plane may be divided into different clustering. In addition, the clustering process will also be too tedious and piecemeal, thus reducing the real-time of detection; on the contrary, if the selected threshold is too large, some small patches will be difficult to be detected resulting in error clustering. As the scanning beam of the LiDAR is fan-shaped, with the scanning points are gradually away from the vehicle the distance between the two adjacent points will continue to increase. Then the threshold Δt will increase accordingly. However, In the k-means algorithm, the distance measurement is based on the average distance from the intra-class points to the centroid, which avoids the misclassification problem caused by the change of threshold. The KMeans algorithm flowchart is shown in Fig. 4.

The k-means algorithm is needed to predefined the number of clusters k , and we set the k value to 2 according to the experiment. After the initial clustering, the obtained initial categories should be downsampled respectively. We choose the average grid method to downsample the point clouds. The advantage of this method is that the shape features of the point clouds are retained while the amount of point clouds data are reduced. Thus the accuracy of plane estimation step can be improved.

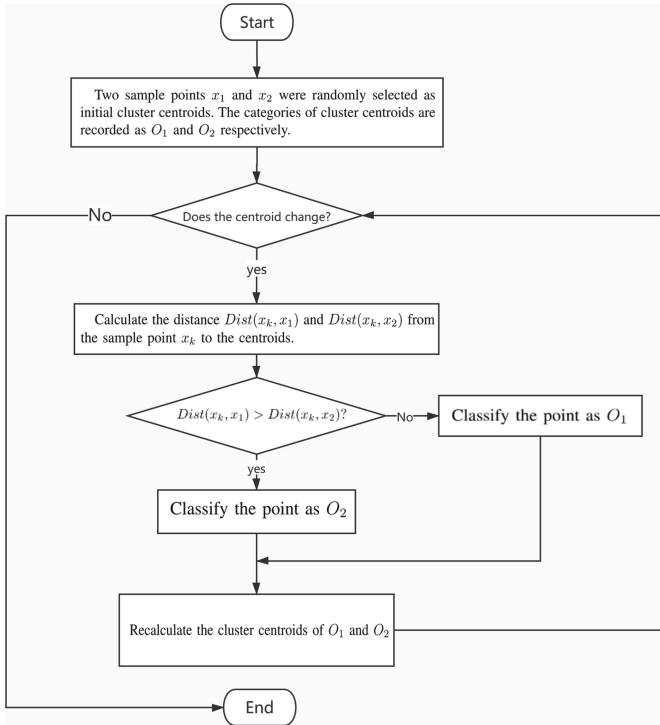


Fig. 4. KMeans algorithm flowchart

2.3 RANSAC Plane Estimation and Negative Obstacle Decision

The process in this stage is twofold. First, we perform an iterative RANSAC to estimate ground as follows: (1) using the heuristic filtering strategy to sieve the downsampled points that are near the front of the vehicle as a priori knowledge, and then to estimate the first plane with RANSAC; (2) checking the rest points that are near the front of the vehicle. If the inlier confidence supports the estimated plane, the current plane is determined as the credible plane. Then removing the points belonging the estimated plane; (3) step 1 and step 2 repeat until the second plane fits successfully (the points in inlier are sufficient). Otherwise, stop iterating and we judged that only exist one plane in detected area. The complete algorithm is briefly summarized in Algorithm 1.

Algorithm 1. Multiplane estimation

Input: downsampled point set $PC = \{pc_i | i = 1, \dots, p\}$

Output: plane P_1, P_2

1. $PC_{Prio} = \text{Heuristic}(PC)$
2. $[P_1, \text{inlier}] = \text{RANSAC}(PC_{Prio})$
3. for $i := < \text{Num}_{pc}$
4. for points which near vehicle not in inlier do
5. if pc_i fits P_1
6. add point to inlier
7. end for
8. if the number of points in inlier is $> d$
9. P_1 is the first estimated plane
10. $PC_{rest} = \text{Delete}(PC \setminus \{pc_i | i \in P_1\})$
11. $[P_2, \text{inlier}] = \text{RANSAC}(PC_{rest})$
12. for points which near the first plane not in inlier do
13. if pc_i fits P_2
14. add point to inlier
15. end for
16. if the number of points in inlier is $> d$
17. P_2 is the second estimated plane
18. else there is only one plane
19. end for
20. return P_1, P_2

Second, making negative obstacle decision according to the obtained plane equations. Here, α is the angle between the surface normal of point in the grid and that of the estimated plane. Calculating the relationship (including the distance from points the obtained plane and the angle α) between points in each grid and the obtained planes. We can get two decisions: (1) the vacant area is decided as negative obstacle region; (2) a few points exist in one region and the region is below the plane and $\alpha > 15^\circ$. Then, this region is decided as a ramp.

3 Experiment

3.1 Performance Assessment

The planes estimation result in this experiment is shown in Fig. 5. We use [12] defined metrics to evaluate the expected performance of RANSAC.

In RANSAC, setting point pc_i as inlier of the given plane P with $pc_i < \varepsilon$ for a given threshold ε . Given that $N_\varepsilon(P)$ is the set of inliers and $|I|$ represents cardinality of the set N . The measurement of the fitness f can be considered as $|N_\varepsilon(P)|$.

Here, $q(P)$ is the quality of a candidate plane P , and measured by number of points in plane P and in truth ground P_0 . we use it to describe the fitness between estimated plane and given plane. $q(P)$ needed to be normalized by using the number of points in inliers P_0 :

$$q(P) = |N(P) \cap N(P_0)| / |N(P_0)| \quad (1)$$

$q(P) = 0$ when P is far enough from the plane; $q(P) = 1$ when P coincides with P_0 .

Evaluating RANSAC's expected performance with the joint probability density function $d_{q,f}(q, f)$ of quality q and fitness f . Set f equals to $|N_\varepsilon(P)|$ for each plane P (here we assume that the space of candidate planes is continuous). q_N is random variable describing plane quality which is selected by algorithm after N iterations (each iteration corresponds to a random candidate plane). $\{f_n\}$ are the set of measured fitness values of the candidate planes, where each algorithm selects plane with $f(P) = \bar{f}, \bar{f} = \max \{f_i\}$. Then:

$$d_{qN}(q) = \int_{-\infty}^{\infty} d_{qN|\bar{f}}(q|f) f_{\bar{f}}(f) df \quad (2)$$

When $d_{qN|\bar{f}}(q|f) = d_{q|f}(q|f)$, it represents the quality of plane with highest fitness measure:

$$d_{\bar{f}} = N d_f(f) F_f^{N-1}(f) \quad (3)$$

Among them, $F_f(f)$ is the cumulative distribution function of f:

$$F_f(f) = \int_{-\infty}^f f_f(u) du \quad (4)$$

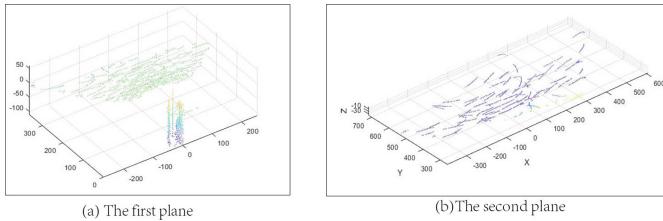


Fig. 5. Plane estimation results. Figure (a) is the first estimated plane; Fig. (b) is the second estimated plane.

3.2 Experimental Results

This subsection presents the experimental results of the multi-stage plane estimation method based on RANSAC and KMeans, and emphasises the application potential of this method in negative stairs obstacle detection. This exploration is important for safety issues in autonomous driving. Note that although we only introduce the detection results of negative stairs, these method can be applied

to other detection of structures containing planes. For example, small steps in open roads.

The result of the plane estimation is shown in Fig. 6, and the two estimated planes are represented by red and blue respectively. Figure 6(a) shows the top view for the results of plane estimation to the data in Fig. 3(a). We can see that the detection result is better. Figure 6(b) shows a side view corresponding to the data in Fig. 3(c). In this case, there is a vacant area between the two regions of planes. In this way, we can determine that currently existing a negative obstacle.

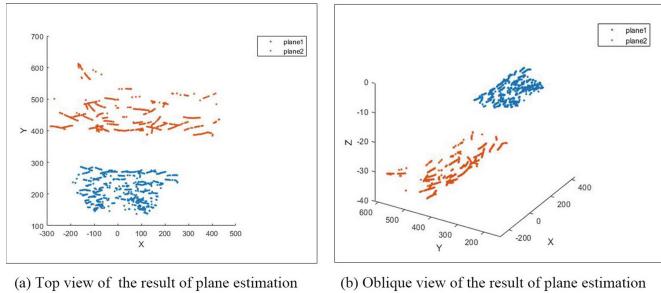


Fig. 6. Result of negative stair obstacle detection. The first plane is represented by blue and the second plane by red. Figure (a) is the top view of the region detection; Fig. (b) is the side view of the region detection

4 Conclusion

In this paper, we propose a multi-stage plane estimation method based on RANSAC and KMeans to explore the applicable of negative stairs obstacle detection. In this experiment, we firstly use KMeans to pre-cluster the obtained point clouds and downsample them in each obtained category to reduce the amount of data processing and to improve the efficiency in next step of detection. Second, RANSAC algorithm iteratively estimates the planes with the downsampled data, and we add the confidence of inlier to decide the plane. Finally, Calculating the geometrical relationship between regions according to the obtained planes and then making decision of negative stairs obstacle. The final experimental results show that this method can effectively detect stairs obstacles and timely make obstacle avoidance decisions to ensure the safety driving.

However, the approach in this paper is not really good at adaptive performance when the detection environment becomes complex. For example, the task of plane estimation according to inlier of confidence depends on the chosen distance threshold Δd . Δd may changes in different scene. In this way, it has a strong limitation for the scene. In the future work we will investigate to increase the adaptive performance of this approach by considering different constrains to improve the setting of inliers confidence in the step of RANSAC, so that the approach can be applied to other similar scenarios.

Acknowledgment. This work is financially supported by the Nature Science Foundation with No. 61862005, the Guangxi Nature Science Foundation with No. 2017GXNSFBA198226, the Scientific Research Foundation of Guangxi University with No. XGZ160483, the Higher Education Undergraduate Teaching Reform Project of Guangxi with No. 2017JGB108, and the project with No. DD3070051008.

References

1. Shang, E., An, X., Tao, W., Tingbo, H., Yuan, Q., He, H.: LiDAR based negative obstacle detection for field autonomous land vehicles. *J. Field Robot.* **33**(5), 591–617 (2016)
2. Gao, B., Xu, A., Pan, Y., Zhao, X., Yao, W., Zhao, H.: Off-road drivable area extraction using 3D LiDAR data. In: 2019 IEEE Intelligent Vehicles Symposium (IV) (2019)
3. Morton, R.D., Olson, E.: Positive and negative obstacle detection using the HLD classifier (2011)
4. Schnabel, R., Wahl, R., Klein, R.: Efficient RANSAC for point-cloud shape detection. *Comput. Graph. Forum* **26**(2), 214–226 (2007)
5. Saval-Calvo, M., Azorin-Lopez, J., Fuster-Guillo, A., Garcia-Rodriguez, J.: Three-dimensional planar model estimation using multi-constraint knowledge based on k-means and RANSAC. *Appl. Soft Comput.* **34**, 572–586 (2015)
6. Xu, G., Yuan, J., Li, X., Su, J.: Reconstruction method adopting laser plane generated from RANSAC and three dimensional reference. *MAPAN-J. Metrol. Soc. India* **33**, 307–319 (2018)
7. Zheng, Y., Liu, J., Liu, Z., Wang, T., Ahmad, R.: A primitive-based 3D reconstruction method for remanufacturing. *Int. J. Adv. Manuf. Technol.* **103**(9–12), 3667–3681 (2019)
8. Muhovic, J., Mandeljc, R., Bovcon, B., Kristan, M., Pers, J.: Obstacle tracking for unmanned surface vessels using 3-D point cloud. *IEEE J. Oceanic Eng.* **45**, 1–13 (2019)
9. Zhou, K., Richtsfeld, A., Varadarajan, K.M., Zillich, M., Vincze, M.: Combining plane estimation with shape detection for holistic scene understanding, pp. 736–747 (2011)
10. Qian, X., Ye, C.: NCC-RANSAC: a fast plane extraction method for 3-D range data segmentation. *IEEE Trans. Cybern.* **44**(12), 2771–2783 (2014)
11. Wang, W., Gao, W.: Efficient multi-plane extraction from massive 3D points for modeling large-scale urban scenes. *Vis. Comput.* **35**(5), 625–638 (2019)
12. Gallo, O., Manduchi, R., Rafii, A.: CC-RANSAC: fitting planes in the presence of multiple surfaces in range data. *Pattern Recogn. Lett.* **32**(3), 403–410 (2011)



Artificial Intelligence Based Strategy for Vessel Decision Support System

Andrius Daranda^(✉) and Gintautas Dzemyda

Vilnius University Institute of Data Science and Digital Technologies, Akademijos str. 4,
Vilnius, Lithuania
gintautas.dzemyda@mii.vu.lt

Abstract. Nowadays, it is vital to ensure safe vessel navigation. As in the old days, this responsibility lies on the marine navigator. The main issue to provide safety is to plan and predict the vessel maneuvering in the massive congestion. The enormous permanent stream of marine traffic data is tricky to process for vessel navigator. Thus, the main reason for the collision is human errors. So the autonomous navigation is the key to reduced major errors. However, the enormous size of the navigation data gathering from marine traffic causes a problem to create a vessel decision support system. We introduced the novel method of the decision support system to navigate safely in the high-density maritime traffic. The suggested method is based on Deep Reinforcement Learning. This technique could deal with the end-to-end approach for vessel navigation in marine traffic with considerable vessel congestions.

Keywords: Marine traffic · Reinforcement learning · Decision support system

1 Introduction

Nowadays, the marine traffic environment becomes very complicated and dangerous because the traffic operates on a large scale. The vessel collisions could lead to huge losses of human life and property. Therefore, the issue of ensuring safety at sea is the main goal for the mariners. This task could be solved by providing timely, essential navigational information. On the other side, this navigational information could not be enough to avoid the vessel collision. So, the safe trip with the vessel highly depends on the experience of navigators. At first, the navigators do it in the old fashion way by the visual watch keeping on the vessel bridge. In the technology century, navigators use a more sophisticated way to minimize the risk of collision. The most popular technology to avoid a collision is radars. Radars use electromagnetic waves to calculate the position of vessels or other obstacles. Otherwise, radars do not ensure safe navigation itself. The radar uses the Automatic Radar Plotting Aid (ARPA) [1] to reduce collision risk. ARPA is a great tool to calculate and predict dangerous vessel maneuvering. The main problem with using this calculation is that every second, the situation could change. So, ARPA is the most effective when all vessel speed and course are constants.

Autonomous vehicles have enormous potential in the maritime transport world. Nowadays, most maritime companies face hiring crew problems. Another significant issue is the vessel running costs. Most parts usually are for the crew salary. These costs cannot be saved by firing sailors because it could make a case for the vessel's safe voyage.

Moreover, the autonomous vessel technology could reduce fuel consumption and recruitment problems. Furthermore, it could increase operational times and improved lifestyles for the much lower number of seafarers, and increase maritime shipping capacity [2]. Otherwise, there is a significant skepticism on the capability of the autonomous vessel operations. Otherwise, many challenges (legal [3], technical [2]) could be solved by the autonomous vessel technology. Many projects are initiated to tackle autonomous shipping problems; for example, The Yara Birkeland electric, autonomous vessel [4], the MUNIN research project [5], the DIMECC 'One Sea' Consortium [6], and a start-up company is retrofitting old vessels to be autonomous.

Many complicated factors influence the assessment of the maneuvering situation. These factors could be geographical or hydro-meteorological conditions. However, the main reason for the vessel collision is human errors. The crew member must decide to navigate and maneuver safely. So, it very important to ensure the decision support system to maintain alertness.

In this paper, we disclosed a new possibility of applying artificial intelligence for the vessel decision support system (DSS). The DSS is acting through the agent who realizes the suggested decision. This agent learns to control the vessel in marine traffic.

2 Related Work

In recent years, research and industry initiatives have primarily aimed at developing and building autonomous vehicles. Land-based transport modes, especially autonomous cars, have been leading these initiatives. The marine industry is now following this trend, and autonomous vessels have become the subject of numerous research projects within organizations and companies. The technology of autonomous vessels is developing rapidly and tries remote control applications on the boats.

Many different methods were created to avoid the collision. These methods have been given in [1, 7, 8]. The collision avoids methods utilize very different approaches depending on the problem formulation - the Dynamic Games theory [9], the Dynamic Programming [10], the Maze Routing Algorithm [11], the Fast Marching Method [12], the Genetic Algorithm [13, 14], Cooperative Path Planning algorithm [15], the Artificial Potential Field [16], the Fuzzy Logic-based method [17], the Evolutionary Algorithms [18] and the Ant Colony Optimization [19].

Some studies attempted to apply the Artificial Neural Networks (ANN) model for safe marine navigation. These models learn the navigation officer's actions and predict maneuver actions [20]. Also, the [21] tried to predict the waves' effect on the vessel yaw motion.

However, the big issue for ANN is to predict the vessel collision. Haris and Amdahl [22] proposed the simplified analytical method. The authors used the interaction between the deformation on the striking and the struck vessels. This method could be categorized into three methods: experiments, numerical simulations, and simplified analytical

methods. All methods could be applied from the local element to the level of the small scale vessels.

[23] proposed the reasoning system for the vessel turning problems. This method was applied for ocean navigation and collision. The application of automated navigational aids reduces the risk of collision. The intelligent decision-making facilities require an anomaly or dangerous situation detection to increase marine navigation safety [24]. Also, [25] proposed the radar filtration algorithm for the Vessel Traffic Centre (VTC) to raise awareness. [26] used the fuzzy inference system in the collision avoidance system.

The [27] calculated vessels' optimal turning angles to avoid the collision. [28] studied a method to analyze the results of trainee maneuvering to avoid collisions on the simulators using the target's obstacle zone. The ANN was applied in [29–32] works for the marine traffic prediction. The ANN was used to train the machine learning model to predict marine traffic instead of developing the mathematical model. The [30] used machine learning methods combined with DBSCAN based algorithm to find the nautical traffic patterns for ocean-going vessels. Kim et al. [29, 31] proposed the method to predict vessel trajectory in the harbors. He trained the regression model combined with the dead reckoning model.

In recent years, Deep learning (DL) found attention in marine traffic patterns or anomalies, too [33]. This technology seems to have wide applications in further analysis of the marine geospatial data.

3 The Problem of Navigation in the Marine Traffic

The navigation in marine traffic is a complex problem. This study aims to apply the RL agent to act as it would be in marine traffic. At first, the RL agent must avoid collision with an obstacle. The second aim is to reach the position of the destination. So, the agent is trained to act in some areas of marine traffic. This area was created as a fixed size grid structure with an active obstacle. The task is different than it is on video games. The difficulty comes from unpredictable scenarios. Due to this complication, the agent tries to learn and act in an environment with one obstacle. Such limitation to one obstacle simplifies the task. However, it is usual in everyday marine navigation because decisions are made according to one most risky obstacle.

Thus, like the navigator on the vessel, the Q-learning algorithm allows the agent to learn about the current state in the environment and to decide about the action. So, the agent or vessel should avoid obstacles and make safe maneuvering depending on the situation. As it was mentioned, the agent must reach the point of destination. This point of destination represents, e.g., the turning waypoint, where the route consists of many turning points. Usually, turning waypoints are preplanned and fixed on the chart. The agent must start from one turning point and safely reach the next turning point.

4 Methodology

In this paper, the possibilities to apply a Deep Q-learning algorithm to control the vessel are investigated. The Q-learning algorithm is model-free and belongs to DRL. It was proposed in [34]. The algorithm is an active reinforcement method, and its agents policy

or Q-table are generated on the fly. This table uses state-action pairs to index a Q-value. This value is described as an expected discounted future reward of taking action and in a particular state. Thus, every state has an assigned value. This is expressed by Q-update equations. The Q-learning algorithm exploits off-policy control to maximize the expected value. It allows isolating the deferral policy from the learning policy. The action is selected by the Bellman optimal equations and the e-greed policy [34]. The Q-learning algorithm implicates the agent in the environment (Fig. 1). The agent has a set of states and actions per state. The agent acts in the environment and gets the rewards.

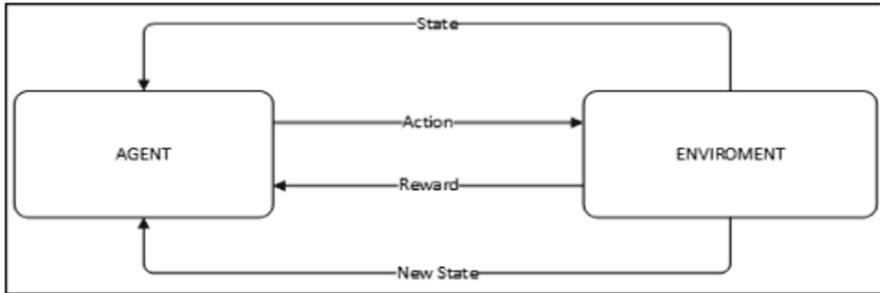


Fig. 1. Q-learning algorithm training

AI in recent years has made significant progress because of the development of machine learning, especially deep learning (DL) tools. For example, the DeepMind team has recently successfully trained an agent using deep Q-learning to play the game Go, eventually defeating a human professional player. Deep Q-learning belongs to deep reinforcement learning (DRL), which is widely used to control autonomous robots. The DRL is an important research field in machine learning and consists of two parts: deep learning (DL) and reinforcement learning (RL). The RL was proposed by Sutton in 1984 [35], used a reinforcement signal to critique the action, gain experience from the environment through several trial and error to improve the strategies to adapt to the environment and achieve good control quality. The neural network tries to predict the action based on the environment. Thus, the neural network is getting states as input. The expected action is the output of the neural network.

In this paper, the Deep Q-learning algorithm was challenged to navigate the vessel through other vessels or obstacles. According to the restrictions mentioned above, the Q-learning algorithm simulates vessel control. Thus, the non-human expert can be used to train to maneuver in the vicinity of other vessels and obstacles. The reward system was used in the environment of learning. The agent is rewarded when the vessel reaches the point of destination. In this environment, the agent (in our case – the vessel) tries to learn to maneuver through obstacles. The agent has two primary purposes: to achieve the destination position and do not appear in the same position with obstacles at the same time. In our experiment, the obstacle moves like other usual vessels in marine traffic. Thus, the agent should plan the moves in a dynamic environment.

5 Results

The proposed experiment was conducted on the 5×5 grid. The agent has space for four possible actions. The experiment consisted of episodes. In these episodes, the agent tried to learn about the environment and solve the maneuvering/ optimization problem. Therefore, the episode consisted of one agent, one obstacle, and a point of destination. However, the main goal of the agent is to reach the destination point. The environment is dynamic. Thus, the obstacle could move independently along the grid. Moreover, the position of the obstacle is chosen at random in every episode. If the agent reaches the point of the destination, it yields a reward of 1. If the agent makes a collision with an obstacle, it causes a penalty. The agent should learn to navigate through various obstacles or other moving vessels. Furthermore, the agent should predict the next position of the moving obstacle to avoid the collision. The examples of situations for q-learning are presented in Fig. 2: a – initial state, b – some intermediate state. A hexagon denotes the agent. Square denotes the destination. Rhombus denotes the obstacle. Arrows denote the direction of motion. The motion of the agent starts from the left upper corner and tries to reach a square, as shown in Fig. 2.

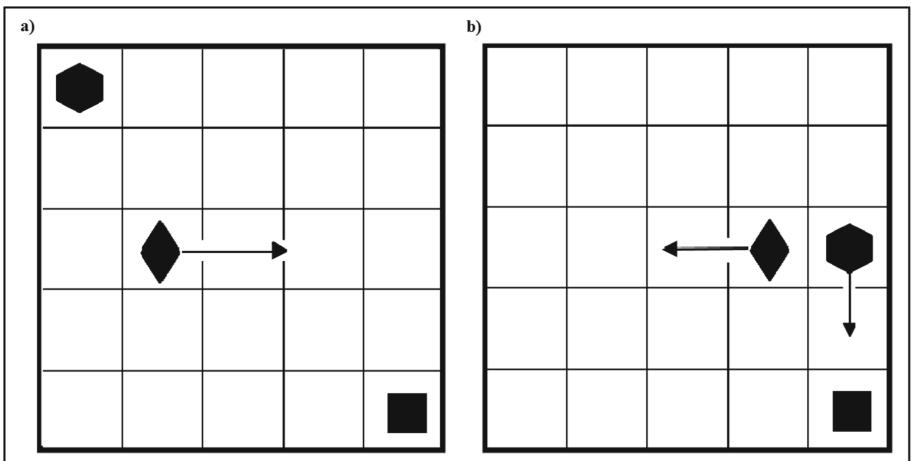


Fig. 2. Examples of the episodes for Q-learning agent training: a) – initial state, b) intermediate state.

The agent and obstacle move non-stop. The agent is self-training to avoid obstacles and other vessels. If an agent successfully reaches the destination, it gets 1 point of reward. If the agent collides with an obstacle, the penalty of 1 point is subtracted. If the agent reaches the point of destination with collision, it fails to learn the episode. The agent should reach the destination point in 25 steps. Otherwise, it fails – returns 1 point of reward. The Q-learning made a lookup table (Q-table) of maximum expected rewards at each state. The Q-table guides agent for the best action at each state. The environment initially generates an obstacle in a random place in every episode. The randomly generated obstacle makes learning tasks as close as real situations.

After every episode, the agent collects information about the environment. This information is used to find an optimal strategy to achieve the goals. This exploration is called the decaying Epsilon exploration method. The decay of Epsilon expresses the progress of the optimal strategy. The Epsilon decay method tries to decrease the percentage of exploration for an optimal strategy. So, the decaying process is representing this. As shown in Fig. 3, the agent improves the strategy in every episode.

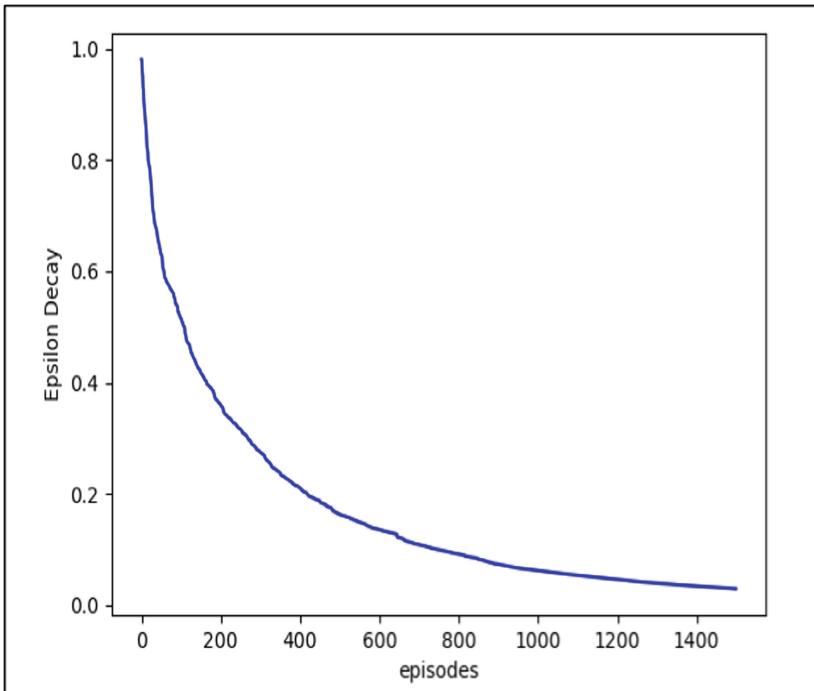


Fig. 3. Epsilon decay progress depending on episodes

Another essential parameter that allows us to evaluate the training progress is the number of successful episodes. The results are given in Fig. 4. During the first steps, the agent does not achieve good performance. However, the performance achieves better results with every later step.

The agent starts learning for possible actions in the environment (Fig. 5). The primary episodes endure quite long because the agent tries to find the point of destination. This is the reason for insufficient learning progress. After 50 episodes, the agent achieves 41,18 % precision. Respectively, after 100 episodes, achieve 49,5% precision. The agent is learning continuously about the environment and available action with a growing number of episodes. The further, the better the results agent performs (see Fig. 6). The agent reaches about 90% after 1500 episodes.

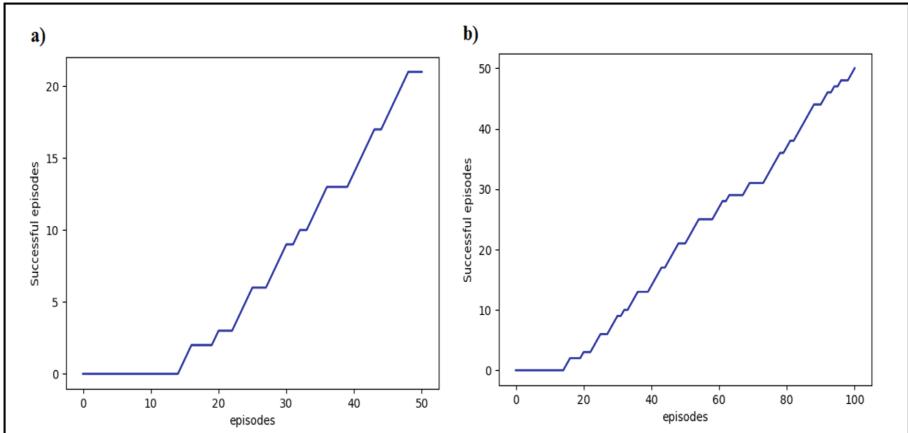


Fig. 4. Agent learning progress. a) after 50 episodes, b) after 100 episodes

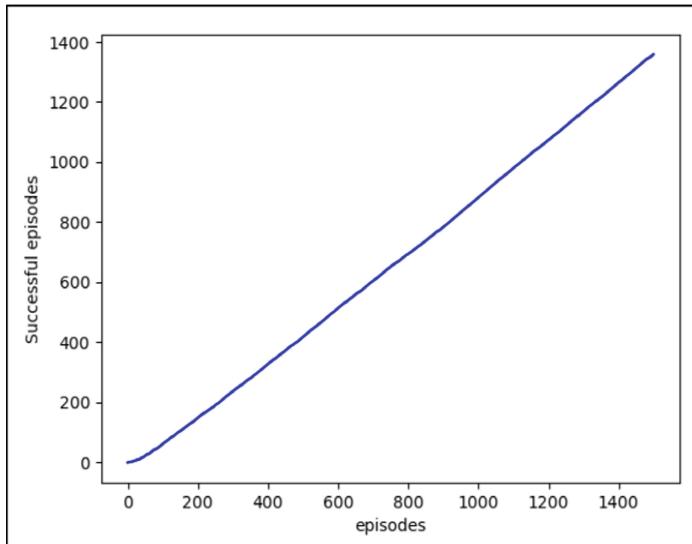


Fig. 5. Agent learning progress

6 Discussions and Conclusion

For the sake of clarity, the experiment was done in a simplified environment as compared to the real marine traffic. The main problem is the distance/time complexity when a Deep Q-learning algorithm is applied to navigate in a real environment. Due to learning limitations, the Deep Q-learning algorithm cannot operate in a wide area. Moreover, in marine traffic, the decision on further action is stretched in time. Deep Q-learning algorithm tries to find a decision in each step despite on necessity of the action. Therefore, the best is to switch on the Deep Q-learning decision algorithm when it is necessary

– maneuvering starts or a dangerous situation appears. After detecting a dangerous situation at an appropriate distance, the Deep Q-learning algorithm could successfully be applied to make a decision to solve a safe maneuvering problem in such a situation.

The goal of this paper was to disclose the capability of a Deep Q-learning algorithm to navigate through marine traffic. The DSS was realized to make the decision and control the vessel through the agent. It is shown that a marine navigation control based on the deep reinforcement learning algorithm is capable of avoiding obstacles or, accordingly, the vessel in marine traffic. The proposed network was trained by the reward system. Thus, this reward system could be adopted for more complex learning. The agent is trained by a randomly generated obstacle. It allows the agent to reach about 90% accuracy. The demonstrated agent could act as a decision support system in real marine traffic. Moreover, the Deep Q-learning algorithm could be improved to work in any changing navigation environment, for example, navigating in severe weather conditions.

References

1. Statheros, T., Howells, G., McDonald-Maier, K.: Autonomous ship collision avoidance navigation concepts, technologies and techniques. *J. Navig.* **61**, 129–142 (2008). <https://doi.org/10.1017/S037346330700447X>
2. Kobyliński, L.: Smart ships – autonomous or remote controlled? *Zesz. Nauk. Akad. Morskiej w Szczecinie*. **53**, 28–34 (2018). <https://doi.org/10.17402/262>
3. Karlis, T.: Maritime law issues related to the operation of unmanned autonomous cargo ships. *WMU J. Marit. Aff.* **17**, 119–128 (2018). <https://doi.org/10.1007/s13437-018-0135-6>
4. Skredderberget, A.: Yara Birkeland; The first zero emission, autonomous ship; Yara International. <https://www.yara.com/knowledge-grows/game-changer-for-the-environment/>
5. MUNIN. <https://www.unmanned-ship.org/munin/about/>
6. Haikkola, P.: One Sea Roadmap towards commercial autonomous shipping in 2025 (2017). https://www.oneseaecosystem.net/wp-content/uploads/sites/2/2017/08/onesea_roadmaps-august-2017_paivi-haikkola_rev.pdf
7. Tam, C.K., Bucknall, R., Greig, A.: Review of collision avoidance and path planning methods for ships in close range encounters. *J. Navig.* **62**, 455–476 (2009). <https://doi.org/10.1017/S0373463308005134>
8. Campbell, S., Naeem, W., Irwin, G.W.: A review on improving the autonomy of unmanned surface vehicles through intelligent collision avoidance manoeuvres. *Annu. Rev. Control.* **36**, 267–283 (2012). <https://doi.org/10.1016/j.arcontrol.2012.09.008>
9. Lisowski, J.: Comparison of dynamic games in application to safe ship control. *Polish Marit. Res.* **21**, 3–12 (2014). <https://doi.org/10.2478/pomr-2014-0024>
10. Lisowski, J.: Computational intelligence methods of a safe ship control. *Proc. Comput. Sci.* **35**, 634–643 (2014). <https://doi.org/10.1016/j.procs.2014.08.145>
11. Szlapczynski, R.: A new method of ship routing on raster grids, with turn penalties and collision avoidance. *J. Navig.* **59**, 27–42 (2006). <https://doi.org/10.1017/S0373463305003528>
12. Liu, Y., Bucknall, R.: Path planning algorithm for unmanned surface vehicle formations in a practical maritime environment. *Ocean Eng.* **97**, 126–144 (2015). <https://doi.org/10.1016/j.oceaneng.2015.01.008>
13. Kuczkowski, Ł., Śmierzchalski, R.: Comparison of single and multi-population evolutionary algorithm for path planning in navigation situation. *Solid State Phenom.* **210**, 166–177 (2014). <https://doi.org/10.4028/www.scientific.net/SSP.210.166>

14. Szlapczynski, R., Szlapczynska, J.: Customized crossover in evolutionary sets of safe ship trajectories. *Int. J. Appl. Math. Comput. Sci.* **22**, 999–1009 (2012). <https://doi.org/10.2478/v10006-012-0074-x>
15. Tam, C., Bucknall, R.: Cooperative path planning algorithm for marine surface vessels. *Ocean Eng.* **57**, 25–33 (2013). <https://doi.org/10.1016/j.oceaneng.2012.09.003>
16. Xue, Y., Clelland, D., Lee, B.S., Han, D.: Automatic simulation of ship navigation. *Ocean Eng.* **38**, 2290–2305 (2011). <https://doi.org/10.1016/j.oceaneng.2011.10.011>
17. Perera, L.P., Carvalho, J.P., Soares, C.G.: Bayesian network based sequential collision avoidance action execution for an ocean navigational system. *IFAC Proc.* **43**, 266–271 (2010). <https://doi.org/10.3182/20100915-3-DE-3008.00046>
18. Tam, C., Bucknall, R.: Path-planning algorithm for ships in close-range encounters. *J. Mar. Sci. Technol.* **15**, 395–407 (2010). <https://doi.org/10.1007/s00773-010-0094-x>
19. Lazarowska, A.: Ship's trajectory planning for collision avoidance at sea based on ant colony optimisation. *J. Navig.* **68**, 291–307 (2015). <https://doi.org/10.1017/S0373463314000708>
20. Westrenen, F.V.: Towards a decision making model of river pilots. *IFAC Proc.* **28**, 217–222 (1995). [https://doi.org/10.1016/S1474-6670\(17\)46728-0](https://doi.org/10.1016/S1474-6670(17)46728-0)
21. Nicolau, V., Aiordachioae, D., Popa, R.: Neural network prediction of the wave influence on the yaw motion of a ship. In: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, pp. 2801–2806. IEEE (2004)
22. Haris, S., Amdahl, J.: Analysis of ship–ship collision damage accounting for bow and side deformation interaction. *Mar. Struct.* **32**, 18–48 (2013). <https://doi.org/10.1016/j.marstruc.2013.02.002>
23. Dixena, D., Chakraborty, B., Debnath, N.: Application of case-based reasoning for ship turning emergency to prevent collision. *IEEE International Conference on Industrial Informatics*, pp. 654–659 (2011). <https://doi.org/10.1109/INDIN.2011.6034956>
24. Daranda, A., Dzemyda, G.: Navigation decision support: Discover of vessel traffic anomaly according to the historic marine data. *Int. J. Comput. Commun. Control.* **15**(3), 3864 (2020). <https://doi.org/10.15837/IJCCC.2020.3.3864>
25. Bukhari, A.C., Tusseyeva, I., Lee, B.G., Kim, Y.G.: An intelligent real-time multi-vessel collision risk assessment system from VTS view point based on fuzzy inference system. *Expert Syst. Appl.* **40**, 1220–1230 (2013). <https://doi.org/10.1016/j.eswa.2012.08.016>
26. Ahn, J.H., Rhee, K.P., You, Y.J.: A study on the collision avoidance of a ship using neural networks and fuzzy logic. *Appl. Ocean Res.* **37**, 162–173 (2012). <https://doi.org/10.1016/j.apor.2012.05.008>
27. Xiong, W., Hu, H., Xiong, N., Yang, L.T., Peng, W.C., Wang, X., Qu, Y.: Anomaly secure detection methods by analyzing dynamic characteristics of the network traffic in cloud communications. *Inf. Sci.* **258**, 403–415 (2014). <https://doi.org/10.1016/j.ins.2013.04.009>
28. Szlapczynska, J., Szlapczynski, R.: Heuristic method of safe manoeuvre selection based on collision threat parameters areas. *TransNav Int. J. Mar. Navig. Saf. Sea Transp.* **11**, 591–596 (2017). <https://doi.org/10.12716/1001.11.04.03>
29. Kim, J.S.: Vessel target prediction method and dead reckoning position based on SVR seaway model. *Int. J. Fuzzy Log. Intell. Syst.* **17**, 279–288 (2017). <https://doi.org/10.5391/IJFIS.2017.17.4.279>
30. Xiao, Z., Ponnambalam, L., Fu, X., Zhang, W.: Maritime traffic probabilistic forecasting based on vessels' waterway patterns and motion behaviors. *IEEE Trans. Intell. Transp. Syst.* **18**, 3122–3134 (2017). <https://doi.org/10.1109/TITS.2017.2681810>
31. Kim, J.S., Jeong, J.S.: Extraction of reference seaway through machine learning of ship navigational data and trajectory. *Int. J. Fuzzy Log. Intell. Syst.* **17**, 82–90 (2017). <https://doi.org/10.5391/IJFIS.2017.17.2.82>

32. Zhang, H., Xiao, Y., Bai, X., Yang, X., Chen, L.: GA-support vector regression based ship traffic flow prediction. *Int. J. Control Autom.* **9**, 219–228 (2016). <https://doi.org/10.14257/ijca.2016.9.2.21>
33. Venskus, J., Treigys, P., Bernatavičienė, J., Tamulevičius, G., Medvedev, V.: Real-time maritime traffic anomaly detection based on sensors and history data embedding. *Sensors*. **19**, 3782 (2019). <https://doi.org/10.3390/s19173782>
34. Watkins, C.J.C.H., Dayan, P.: Q-learning. *Mach. Learn.* **8**, 279–292 (1992). <https://doi.org/10.1007/bf00992698>
35. Sutton, R.S.: Temporal credit assignment in reinforcement learning. Dr. Diss. Available from Proquest (1984)



Improved Multi-scale Fusion of Attention Network for Hyperspectral Image Classification

Fengqi Zhang^{1,2,3}, Lina Yang^{1(✉)}, Hailong Su², and Patrick Shen-Pei Wang³

¹ School of Computer, Electronics and Information, Guangxi University,
Nanning 530004, People's Republic of China

lnyang@gxu.edu.cn

² School of Electronics and Information Engineering, Tongji University,
Shanghai, China

hailongsu@aliyun.com

³ Computer and Information Science, Northeastern University,
Boston, MA 02115, USA

patwang@ieee.org

Abstract. With the development of remote sensing technology, hyperspectral images that carry both spectral information and spatial information have attracted much attention. As a significant task, hyperspectral image (HSI) classification needs to fully extract spectral and spatial features to better determine the category. However, when extracting spatial features, the difference in object scale often affects the classification effect. It is difficult to distinguish both larger and smaller objects at the same time. In this paper, we propose an improved spatial multi-scale fusion scheme for the spatial extraction network, combining spatial patches sampled at different scales to achieve the effect of accurately classifying objects of different sizes. And the entire network is based on the attention mechanism, following the attention setting of SSAN (*Spectral-Spatial Attention Networks for Hyperspectral Image Classification*), which makes the network pay more attention to pixels in key bands and key spaces. The experimental results on the data set Pavia Center prove that our method has achieved a greater accuracy improvement, reaching expected performance.

Keywords: Hyperspectral image classification · Multi-scale fusion · Spatial patch · Attention mechanism

1 Introduction

Hyperspectral images are different from general images in that they have more spectral channels. It contains rich spectral and spatial information, and can better characterize the internal structure details of objects. This item is widely used in agriculture, environment, urban monitoring and so on. It is worth mentioning

that pixel-level hyperspectral image classification is the basis of these applications. Each pixel in the image is a high-dimensional vector. Such abundant spectral information provides many meaningful and distinguishable features for classification, but the dimensional curse [1] (Hughes phenomenon [2]) and the finiteness of the data size also bring difficulties to judge the class. Therefore, feature selection [3] and feature extraction [4] were proposed to alleviate this problem by dimensionality reduction. Among them, PCA (Principal Component Analysis) [5] is an effective feature extraction method.

In addition, because hyperspectral images are greatly influenced by conditions such as environment, atmosphere, and illumination, the same object may have different spectral features, and objects that are not identical may have similar spectral information [6]. This spectral heterogeneity makes Classification cannot only consider spectral features. Moreover, more and more fine spatial resolution also makes it necessary to study a small part of the pixel patch [7]. Therefore, the paper by M. Fauvel et al. took space into consideration [8], but the handcrafted features represented by EMPs [9] in the early days were too dependent on empirical knowledge, and the extracted features were relatively primitive, so deep learning methods that are more in line with end-to-end characteristics and have deeper features replaced handcrafted method. The robust features extracted by deep learning methods can more effectively deal with spatial complexity and intra-class variability. However, early deep learnings ways (such as SAE [10] and DBN [11]) have a large number of fully connections, numerous parameters, which destroy the original spatial structure. Therefore, the CNN with fewer parameters and more suitable for space is used to extract spatial features. On the other hand, for the reason that the adjacent channels of the spectral vector of a pixel are highly dependent and similar, it is suitable to use RNN to extract spectral features, and the value of a channel is related to the previous and subsequent channels, so the Bi-RNN become the final choice.

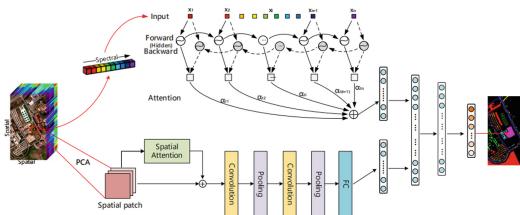


Fig. 1. SSAN architecture diagram.

SSAN [12] uses the attention mechanism in terms of the reason that different spectral channels have different contributions to category information. Similarly, different pixels in the space around the central pixel also have discrepant effects on the classification of the center point, so it's a natural thought of the attention mechanism. As in [13], this mechanism can increase the weight of the parts that are helpful for classification, and reduce the weight of irrelevant parts, just like we

need to focus when taking pictures, so focusing on the main area will remarkably increase classification efficiency and accuracy. SSAN makes full use of attention, but does not consider that objects of different scales in the image will bias the judgement result. Some objects are larger, and a larger sampling patch is needed to capture more structural features related to it, so as to better judge the class of the pixel at the center; some objects will be tiny, but large patches will be sampled to some redundant, irrelevant information. This information doped into useful information, which is not conducive to classification, so relatively smaller pieces is the goal at this state. Therefore, when extracting spatial features, the selection of patches and the grasp of object scales matters.

Inspired by SSAN, this paper fully considers the influence of patch. The original patch only uses a fixed scale. Obviously, it cannot take care of objects that are too large or too small. Therefore, we use three types of patches for sampling, which are suitable for large, small, and medium-sized objects. According to the classification accuracy of the final test, the probability values of these three patches are fused. It can also be said that a simple attention mechanism is used here. For a particular category, the classification value of which patch is better to distinguish this category has greater weight, that is, to make it more effective when encountering objects in this class.

Innovation of this paper:

1. On the basis of SSAN, multi-scale fusion of sampling information of different spatial patches is added, and this fusion itself also has the idea of attention, which makes the classification network more robust to objects of various scales.
2. The idea of combining scale mechanism and attention mechanism is proposed, which will continuously optimize network performance.

The remaining part of this paper is structured as follows: Sect. 2 mainly introduces the SSAN algorithm that is the source of the improvement of the article and explains the principle of the multi-scale mechanism. Section 3 begins to focus on describing our new method of combining scale fusion. And the display and analysis of experimental results will appear in Sect. 4. Section 5 summarizes the paper again.

2 Related Work

This chapter briefly introduces the SSAN algorithm and multi-scale mechanism.

2.1 SSAN Algorithm

As a method to be improved, here is a general description of the algorithm. When extracting spatial features, due to the large number of hyperspectral image channels, PCA is used for dimensionality reduction first, and several channels that can cover most of the information are taken. Finally, the original image

is transformed into a picture with only four channels and used to input into the subsequent spatial sub-network. Similar to SSUN [14], this algorithm also integrates the spectral feature extraction network and the spatial feature extraction network into one network, and the final feature is the combined spectral-spatial feature, which is advantageous to the joint optimization of the network. In addition, different from integrated networks (such as 3D-CNN [15]) and pre-processing-based networks (such as [16]), it is a post-processing-based network, which is to generate high-level networks through two sub-networks. Then the two features are fused in a fully connected layer (stacking the two feature vectors) to generate a combined spectral-spatial feature, and finally they are classified. See Fig. 1 for details.

This algorithm adds the attention mechanism to both sub-networks. The attention here is the same as the mechanism in the paper [13], which multiplies different weights on elements with different contributions. The Bi-RNN model with attention is used in the network for extracting spectral features. This model includes a forward GRU and a backward GRU layer, which can fully take into account the relationship between a channel and its upper and lower channels. The attention weight is obtained through a layer of neural network, and then the attention weight is multiplied by the corresponding output, and finally passed to the fully connected layer.

In the branch of space, the attention mechanism is also used. Different from the spectral branch, the spatial branch uses the CNN network, and an attention layer is added before the patch is fed to the CNN to generate the weight matrix. This matrix indicates where the area that provides useful information is. The larger the weight, the more similar it is to the central pixel (that is, the pixel to be classified), the closer the relationship, and the greater the probability of being from the same class. These places and the intermediate pixels form the internal details of a certain type of target, making the features easier to distinguish.

Finally, the features of the two sub-networks are stacked and fed into the fully connected layer to learn joint features and effectively achieve classification.

2.2 Multi-scale Mechanism

In the image field, whether it is object detection, instance segmentation, or hyperspectral image classification, there will be scale problems. This problem is unavoidable in this general direction, because our goal is nothing more than to determine the class of pixels or objects, but different categories have wide range of changes in measurement, which will undoubtedly increase the difficulty of our classification. Many studies have given their own feasible paths for scale problems. For example, in the field of object detection, for the situation that the range of the objects scale in the picture is too large, there have been image pyramids, feature map pyramids, feature map multi-scale fusion and other solutions.

The hyperspectral image classification is different from other problems in that it will have a scale problem when sampling the pixel patches around the pixels. It is particularly necessary to select a suitable patch for objects of different scales. Because only patches with enough coverage of the corresponding object of the

pixel and few other types of noise can accurately grasp the information and not miss informative places, such classification accuracy will have a big leap.

3 Proposed Method

This chapter focuses on the method proposed in this paper with motivation.

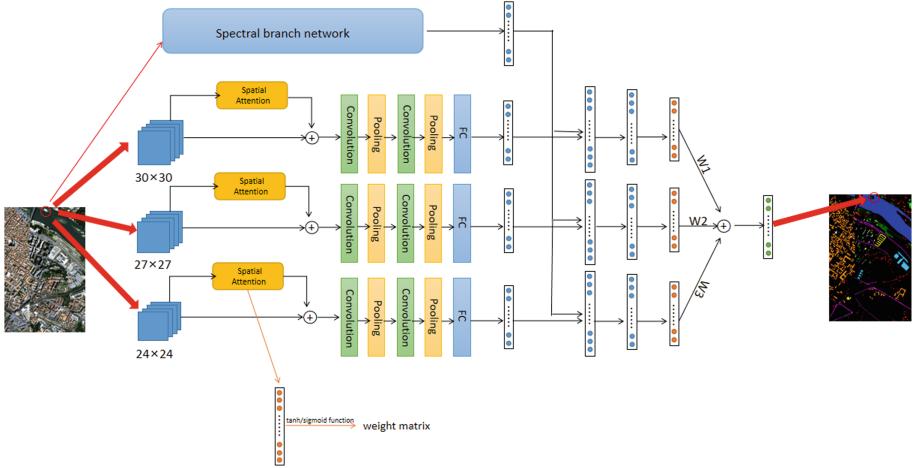


Fig. 2. The framework of the fusion model is shown in the picture above. Among them, the spectral branch network follows the structure and settings in SSAN, and the focus is on improving the space part. The Spatial Attention part uses a layer of neural network to calculate the weight of different pixels in the patch, which represents the information density of different pixels. W_1 , W_2 , and W_3 in the figure are the weights of the predicted probabilities for the three patch sizes respectively.

3.1 Switch Different Sizes of Patches

The patch used in SSAN is 27×27 . If this specification is compared with windows of other sizes, the OA, AA and Kappa values corresponding to the 27×27 model in Table 2 are all higher than other models after the experiment. So it is obviously the best, which shows that it is close to the patch size of as many pixels to be classified as possible. It is suitable for situations where the size of the target class is not much different and the variance is small. However, there are some extreme cases, such as a certain class is too large or too small, that is, it is too far from the average of the sizes of many classes. In this situation, the 27×27 window will not have better classification accuracy for these classes. In order to see if there is a big difference in size among the nine categories in Pavia Center, we took three windows with specifications of 24×24 , 27×27 ,

and 30×30 . Experiment with them separately, and save the accuracy of their screening of each class and their probability of predicting each class during the test (that is, the probability after the softmax layer).

3.2 Multi-scale Fusion

As shown in Fig. 2, three sampling patches of different sizes represent the three models. Assuming that the model is represented by M_i (where $i = 1, 2, 3$), the accuracy of the i -th model for class j (where $j = 1, \dots, 9$) is a_{ij} , the probability of predicting different categories for a certain pixel target is p_{ij} , where the probability is the judgment of the network model for the pixel category, and the category with the highest probability is the prediction class. In order to fully combine the classification advantages of the three specifications for objects of different sizes, we further processed the final predicted probability, which is the fusion part.

The combination of three scale advantages is achieved by multiplying the weight by the predicted probability. Specifically, for the case of (i, j) , the corresponding weight is calculated in Eq. (1), that is, the accuracy of the current model is divided by the accumulation of the accuracy of all models for this class. In order to obtain the final probability P predicted by the fusion model, Eq. (2) is used to achieve weighted accumulation. The above formulas are here (Where $p_i = \{p_{i,1}, p_{i,2}, \dots, p_{i,9}\}$):

$$W_{i,j} = \frac{a_{ij}}{\sum_{s=1}^4 a_{s,j}} \quad (1)$$

$$P = \sum_{i=1}^4 W_i \cdot p_i \quad (2)$$

Figure 2 completely shows the network structure of the whole idea. In this way, the model corresponding to the patch size with higher classification accuracy for a certain class has a larger weight, and then it has a larger proportion of the probability of predicting the class, so that the class with the highest probability is more likely to be the correct class. In fact, when you think about it carefully, this fusion method also implies the idea of attention. This fusion mechanism focuses on highlighting the effect of better size models for different classes, and strives to avoid the negative effects of models with poor usefulness, and finally makes the detection accuracy has been boosted.

4 Experiments and Analysis

4.1 Experimental Setup

The data set used in this article is the Pavia Center data set. The data set is made by ROSIS sensors. It has 115 bands. After 13 noise bands are deleted, 102 useful bands remain. The original resolution of the picture is 1096×1096 , which becomes 1096×715 after removing the 381-pixel black band. There are

nine objects in this data set: lawn, water, brick, tree, tile, shadow, bitumen, and asphalt. In the setting of the number of training set, validation set and test set, we continue to inherit SSAN, as shown in Table 1.

Table 1. Pavia center dataset's experimental data settings

Label	Class name	Labeled (training)	Validation (training)	Testing
1	Waters	100	100	65,771
2	Trees	100	100	7398
3	Asphalt	100	100	2890
4	Self-blocking bricks	100	100	2485
5	Bitumen	100	100	6384
6	Tiles	100	100	9048
7	Shadows	100	100	7087
8	Meadows	100	100	42,626
9	Bare Soil	100	100	2663
	Total	900	900	146,352

In terms of parameter setting, many of the parameters in the 27×27 model have been used in SSAN. A total of 10,000 epochs have been trained. The size of a batch is 128, and the number of blocks in Bi-RNN is still set to 512. But here the learning rate is reduced to $1e-4$, and the initialization weight of the subsequent convolutional network in the spatial branch was changed from 1 to 0.1, and also reduced the regularization coefficient of the spatial attention layer weight to $8e-6$. These changes make the model more accurate although the convergence speed is reduced, 10,000 epochs can guarantee the final convergence of the network. Moreover, the initial values of the filter and bias of the spatial network are obviously approached to zero, and the regularization coefficient is also reduced. For the characteristics of the convolutional network and the single-layer Attention network, it is obviously more helpful to avoid the appearance of convolutional network overfitting and the occurrence of insufficient fitting with Attention layer.

4.2 Analysis of Experimental Results

The experimental results are presented in Table 2. First observe the test results of the 27×27 model experiment. We will find that the accuracy of the model for the second, fifth and seventh categories is 94.81%, 96.17% and 94.71% respectively, which is obviously not as good as other categories. But what is surprising is that the 30×30 model has an accuracy of 96.22% for the second category, and the 24×24 model has an accuracy of 99.19% and 98.19% for the fifth and eighth categories, which shows that different patch model has obvious diversities in the classification effects of different classes, which also proves our previous

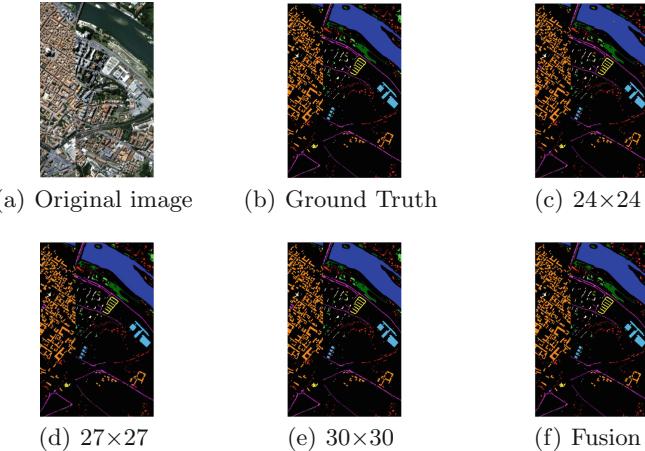


Fig. 3. Figure a is the original remote sensing image of the Pavia dataset. And figure b is the color map after completely successful classification (that is, the ground truth). Figure c to Figure e are the color maps of three single-size models, and figure f is the color map after our multi-scale algorithm classification. It can be seen that its classification effect is significantly better than the previous model.

Table 2. Classification performance of different models on the Pavia center data set

Label	Class name	“24 × 24” model	“27 × 27” model	“30 × 30” model	Fusion model
1	Waters	99.67	99.55	98.94	99.76
2	Trees	95.14	94.81	96.22	97.23
3	Asphalt	94.92	98.72	94.99	98.37
4	Self-blocking bricks	99.88	99.92	99.92	99.92
5	Bitumen	99.19	96.17	96.21	98.00
6	Tiles	93.75	98.26	98.02	98.66
7	Shadows	98.19	94.71	95.80	98.14
8	Meadows	98.48	98.71	98.67	99.02
9	Bare Soil	100.0	99.62	100.0	99.92
<hr/>					
	OA	98.55	98.59	98.35	99.17
	AA	97.69	97.83	97.64	98.78
	Kappa	97.94	98.00	97.66	98.82

statement, that is, the classification effects of fixed-size patches on objects of different sizes are distinctly divergent.

After training the three models, start testing them. At the same time, we saved their prediction probabilities for these classes. After the test, we also saved their accuracy for subsequent calculation of weights. Then perform weighted fusion, and finally calculate the accuracy of the fusion. It can be seen that the accuracy of most classes is improved compared with the previous single-size model. In particular, the previous three categories with poor 27×27 classification have been greatly advanced in our scale fusion model. In addition, OA (overall accuracy), AA (average accuracy), and Kappa (Kappa coefficient) are

Table 3. The comparison with the previous method on the Pavia center data set

Class name	KNN	Spec-SVM	EMPs-SVM	CNN	CNN-PPF	SSAN	Our fusion method
Waters	99.15	99.34	99.65	99.52	99.77	99.55	99.76
Trees	88.76	94.49	95.83	94.32	95.04	94.81	97.23
Asphalt	76.32	96.15	97.80	96.88	97.44	98.72	98.37
Self-blocking bricks	81.92	96.57	99.07	98.26	99.11	99.92	99.92
Bitumen	86.39	96.67	95.59	95.73	96.75	96.17	98.00
Tiles	91.44	96.02	96.56	96.52	98.82	98.26	98.66
Shadows	81.31	91.37	93.79	93.56	93.69	94.71	98.14
Meadows	93.67	97.58	98.18	98.11	98.72	98.71	99.02
Bare Soil	97.18	99.88	99.62	99.62	99.92	99.62	99.92
OA	92.66	97.34	98.02	98.23	98.45	98.59	99.17
AA	88.22	96.69	97.16	97.29	98.28	97.83	98.78
Kappa	90.13	97.72	97.47	97.54	97.88	98.00	98.82

used here to measure the classification effect. The experimental results show that the three measurement standards of the fusion model are significantly higher than the three single-size patches. In Fig. 3, you can see the intuitive effect of the comparison of several models. Obviously, our fusion model is better than other situations.

In addition, this paper compares some previous experimental methods with ours. The results are shown in Table 3, it is obvious that our results are better.

5 Conclusion

The method proposed in this study is to combine multi-scale fusion and attention mechanism, because both methods can obviously improve the classification performance of hyperspectral images. Therefore, based on the SSAN algorithm that has done a good job in attention, this article adds a scale fusion mechanism. According to the spatial feature extraction network, the size of the image patch sampled around the center pixel (pixel to be classified) in the network turns the original model into three networks of different sizes. And according to the test performance of these networks, the weights for different classes are calculated. Finally, the multi-scale fusion mechanism is achieved by weighting the predicted probabilities of different networks, and the target effect is further optimized on the basis of the attention mechanism.

Subsequent research can continue to innovate the attention method thoroughly. In addition, it is possible to consider a simpler scale method from the perspective of speed, so that more innovative attention mechanisms can be combined with different multi-scale methods to obtain more efficient results. And we will consider different fusion rules (such as Choquet and Choquet-based integrals), and use it as future directions. This is our next step to explore ideas.

Acknowledgment. This work is financially supported by the Nature Science Foundation with No. 61862005, the Guangxi Nature Science Foundation with No. 2017GXNSFBA198226, the Scientific Research Foundation of Guangxi University with

No. XGZ160483, the Higher Education Undergraduate Teaching Reform Project of Guangxi with No. 2017JGB108, and the project with No. DD3070051008.

References

1. Lu, X., Zhang, W., Li, X.: A hybrid sparsity and distance-based discrimination detector for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **56**(3), 1704–1717 (2017)
2. Hughes, G.: On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **14**(1), 55–63 (1968)
3. Chang, C.I., Wang, S.: Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **44**(6), 1575–1585 (2006)
4. Bruce, L.M., Koger, C.H., Li, J.: Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Trans. Geosci. Remote Sens.* **40**(10), 2331–2338 (2002)
5. Licciardi, G., Marpu, P.R., Chanussot, J., Benediktsson, J.A.: Linear versus non-linear PCA for the classification of hyperspectral data based on the extended morphological profiles. *IEEE Geosci. Remote Sens. Lett.* **9**(3), 447–451 (2011)
6. Liu, B., Yu, X., Zhang, P., Yu, A., Fu, Q., Wei, X.: Supervised deep feature extraction for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **56**(4), 1909–1921 (2017)
7. Benediktsson, J.A., Ghamisi, P.: Spectral-Spatial Classification of Hyperspectral Remote Sensing Images. Artech House (2015)
8. Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R.: Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **46**(11), 3804–3814 (2008)
9. Benediktsson, J.A., Palmason, J.A., Sveinsson, J.R.: Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **43**(3), 480–491 (2005)
10. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. *Adv. Neural. Inf. Process. Syst.* **19**, 153–160 (2006)
11. Chen, Y., Zhao, X., Jia, X.: Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **8**(6), 2381–2392 (2015)
12. Mei, X., et al.: Spectral-spatial attention networks for hyperspectral image classification. *Remote Sens.* **11**(8), 963 (2019)
13. Chen, L.C., Yang, Y., Wang, J., Xu, W., Yuille, A.L.: Attention to scale: scale-aware semantic image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3640–3649 (2016)
14. Xu, Y., Zhang, L., Du, B., Zhang, F.: Spectral-spatial unified networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **56**(10), 5893–5909 (2018)
15. Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **54**(10), 6232–6251 (2016)
16. Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y.: Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **7**(6), 2094–2107 (2014)



Predictive Models in the Assessment of Tax Fraud Evidences

Fabíola Cristina Venturini^(✉) and Ricardo Mattos Chaim

University of Brasília, Brasília, Federal District, Brazil

fabiola.cristina@aluno.unb.br, ricardoc@unb.br

<http://ppca.unb.br/>

Abstract. The aim of the work is to verify the possibility of improving the selection of taxpayers to be inspected through projections of the results of future audits, based on the results of the inspections already carried out. The analysis of information about the process, obtained from the auditors involved in the selection of taxpayers and in the inspection of companies, allowed the selection of the variables used in the models, and the literature review allowed to define the techniques and tools necessary for their creation and training. The research generated predictive models of logistic regression and neural networks, whose forecasts identified sets of companies that correspond to approximately half of the audited companies and account for more than 80% of the credit constituted (89% in the case of the model neural network), of so that these models have the potential to optimize the application of available resources and maximize results, assisting in the selection of indications of irregularities and fraud with greater potential for the constitution of the due credit.

Keywords: Data mining · Predictive models · Logistic regression · Neural networks · Tax frauds

1 Introduction

The relationship between the State and citizens has changed throughout human history, but some aspects go back to antiquity, such as the obligation to pay taxes, the existence of government structures for collecting them [2] and the perception that the tax burden must be compatible with the current reality. There are records of tax reforms resulting from an excessive tax burden where these 3 aspects can be identified in the 23rd century BC [15], in Sumeria.

On the other hand, the purpose of state taxes and obligations has undergone major changes over time. The Enlightenment ideas, concepts of resource redistribution, promotion of fair competition, valuation of clear rules and right of appeal were incorporated into the tax relationship [2].

However, neither the age of the topic nor the importance of taxation for public finances and for the community itself guarantees the payment of taxes due. The evasion is a complex phenomenon related to economic, ethical and

cultural aspects. In Brazil, tax defaults are related to the continuous increase in the tax burden, the complexity of the tax system, the feeling that public spending is not efficient and the perception that it is possible for defaulters to escape punishment [17].

To combat tax evasion, the legislation imposes on the taxpayer a series of ancillary obligations that involve the delivery of information in digital media to the tax authorities. In the 1990s, taxpayers delivered only one file per month, with summarized tax information, several changes occurred over the years and from 2018 the use of non-electronic invoices was prohibited. Currently, invoices and books of all commercial operations are sent to the Tax Authorities in structured layouts and this information is the main input for the detection of fraud. Data is processed to detect signs of irregularities using data mining techniques, and improving the techniques used is part of an ongoing process.

As part of this process, this work aimed to verify if the use of predictive models, based on machine learning, could be used to improve the step of selecting the signs of irregularities to be inspected.

This article is organized in 5 parts: Sects. 2 presents the theoretical reference; the research method is covered in the Sect. 3; Sect. 4 presents the results achieved in a case study; and Sect. 5 brings the conclusions.

2 Data Mining Process

Data mining allows you to explore a large amount of data in search of patterns, such as association rules or time sequences, and allows you to detect the existence of systematic, implicit, previously unknown and potentially useful relationships, serving both to understand the data and to do predictions [13, 22]. The use of algorithms to extract data patterns is a specific step in the process of discovering useful knowledge from the data, Knowledge Database Discovery (KDD) [5].

In this work, the supervised learning strategy was adopted, which aims to learn a mapping of inputs and outputs of previously labeled pairs, the inputs are related to the characteristics of the previous inspections and the outputs correspond to their results [12, 13, 19].

2.1 Modeling

The choice of the algorithm to solve a specific problem is almost an art, given the variety of existing algorithms and considering that there is no algorithm that solves all problems, since “each technique usually addresses some problems better than others” [5]. The option to use logistic models and neural networks, based on machine learning, in this paper was driven both by the volume and complexity of the data involved and by the fact that the methodology was used for similar purposes, with works on financial data processing [4], tax fraud [3, 7], classification [9], and credit risk classification [12].

Logistic Regression. Regression analysis is a statistical technique that allows to infer and model the relationship of a response dependent variable with explanatory variables. It is widely used and often misused by developing statistics between variables that are not completely related in the sense of cause and effect [14]. The design of experiments is pointed as a way to determine cause and effect relationships and, thus, avoid this problem [14].

Logistic regressions are used when the response variable is categorical, presenting a quality, such as “success” and “failure”, which can be associated with values 0 and 1 and can be treated as the result of a Bernoulli test. For categorical independent variables, it is necessary to create dummy variables for each category (binary conversion) [14, 22], and the values obtained for the dependent variable will always be between 0 and 1, results above 0.50 are classified as an event, by convention.

Neural Networks. Artificial neural networks are computational models inspired by the human brain, whose inputs are analyzed by processing units connected together to transmit information, such as brain neurons [8, 13]. Perceptrons are arranged in layers, each layer can have several perceptrons and each network several layers, however there are limitations in practical terms related to the increase in the size of the networks, both due to the complexity and the processing time [6, 13, 22]. Despite the wide use and the known advantages, dealing with neural networks involves some aspects that deserve to be highlighted:

- Theoretical understanding: the underlying theory of deep learning methods is not well understood, there are doubts about which architectures would perform better, how many layers or how many nodes per layer are suitable for a given task [8].
- Time complexity: the larger the network, the greater the computational resources needed the longer the network takes to provide an output [8].
- Overfitting: when training models some specific noise for the training data set and not just the general patterns in the data, the model is ineffective to predict new results. This problem usually occurs when using a limited set of data, and possible solutions include cross-validation and other strategies [5].
- Convergence to local minimums: in some cases the network does not achieve an efficient adjustment of the weights because the weights converge to the local minimums. However, recent theoretical and empirical results strongly suggest that local minimums are not a serious problem in general [11].

To combine use of neural networks with other resources is as a way to deal with the uncertainties associated with the use of neural networks [8, 10, 21]. In this article, the use of neural networks and logistic regression was adopted to deal with such uncertainties.

2.2 Results Assessment

The confusion matrix is the cross tabulation between the true classification and that predicted by the trained model. The correct classifications are true positives

(TP) and true negatives (TN), and the incorrect are false positives (FP) and false negatives(FN). The model's quality measures (accuracy, recall, precision and F1 Score) consider both the correction per class of response and the general correction [22].

3 Methodology

This applied and exploratory research involved a literature review and a case study. The R software was used to generate the models.

3.1 Literature Review

Articles related to machine learning [1, 9, 12] and the use of data mining models, such as logistic regression [3, 18, 20] and neural networks [16] were analyzed, for different applications, including in the areas of financial [4, 9] and tax fraud detection [7, 23]. Although none of the articles has exactly the same scope as this research, which focuses on incorporating predictive data mining models in the process of selecting signs of tax fraud to be inspected, they do contribute to the choice of the methodology and techniques used.

3.2 Information and Data Analyzed

Information on the selection of taxpayers to be inspected was obtained in meetings with the auditors responsible for the activity. To assess the possibility of obtaining comparison standards between types of inspections based on the opinion of the auditors, a questionnaire was applied on the characteristics of 20 types of inspections, the result of which is shown in the case study.

Table 1 presents the variables created from the characteristics of taxpayers, tax inspections and the value of the infraction notices(IN).

Table 1. Dummy variables for main economic activity

Group of variables		Qty.	Dummy variable
Input	Main economic activity	6	VCme(Trade); VCmu(Communication); VInd(Industry); VTran(Transport); VServ(Services); Vatacado(Wholesale)
Input	Location	10	VIBra(Brasília); VIBrz(Brazilândia); VCeí(Ceilândia); VIBan(Bandeirante); VPla(Planaltina); VIsob(Sobradinho); VIGam(Gama); VITag(Taguatinga); VISia(Sia); VlOut(Other)
Input	Inspection type	4	VDil(Specific investigation); VAud(Audit) VAec(Concentrated audit); VMon(Monitoring)
Input	Algorithm	17	VAfi; VAli; VAnt; VAut; VCar; VCer; VCre; VWal; VReg VOut; VCst; VDel; VImo; VLos; VMis; Vecf; VSub
Output	Value of the IN	6	VN1; VFinedQQ (categorical); VN3; VFinedN3(categorical)

The analyzed data are stored in the corporate systems. Dummy variables were created to replace categorical variables, output variables VN1 and VfinnedQQ are used for IN of any value, and VN3 and VfinnedN3 for IN above R\$ 1 million.

4 Data Analysis - Case Study

Table 2 presents the annual collection of taxes on transactions with goods (ICMS) and services (ISS), the amounts charged in the infraction notices drawn up, including taxes and fines, and the comparison in percentage terms.

Table 2. ICMS and ISS: Tax revenue collected x Infraction notices (IN) (R\$ thousand)

Year	Tax revenue collected	IN Values	%	Year	Tax revenue collected	IN Values	%
2009	4,892,566	922,166	19%	2015	8,281,246	1,505,878	18%
2010	5,543,231	1,089,976	20%	2016	9,226,484	1,002,480	11%
2011	6,171,451	1,175,278	19%	2017	9,550,407	1,750,887	18%
2012	6,821,347	534,151	8%	2018	10,041,904	1,140,136	11%
2013	7,502,109	1,953,460	26%	2019	10,189,433	1,821,393	18%
2014	8,228,595	2,528,881	31%	Total	86,448,773	15,424,686	18%

Although the amount of the tax assessment notices issued in the period is significant when compared to the amount collected, many tax lawsuits are closed without the assessment notices being assessed (55%), as shown in Table 3.

Table 3. Dummy variables for type of inspection

Type of inspection	Performed	IN drawn up	IN Values	VAec	VAud	VMon
Concentrated audit	4,057	2,864	6,335,670,423.03	1	0	0
Audit	841	620	6,049,841,529.42	0	1	0
Monitoring	3,402	238	262,600,632.84	0	0	1
Total	8,300	3,722	12,648,112,585.29			

These occurrences, combined with the fact that many signs of fraud are not inspected due to lack of human and material resources, drive the constant search for improvement in the choice of taxpayers targeted by the inspection.

Thus, predictive models were created to compare the results of previous inspections and the characteristics of the companies inspected, in order to look for possible hidden patterns that allow improving the future selection of taxpayers to be inspected, as shown in Fig. 1.

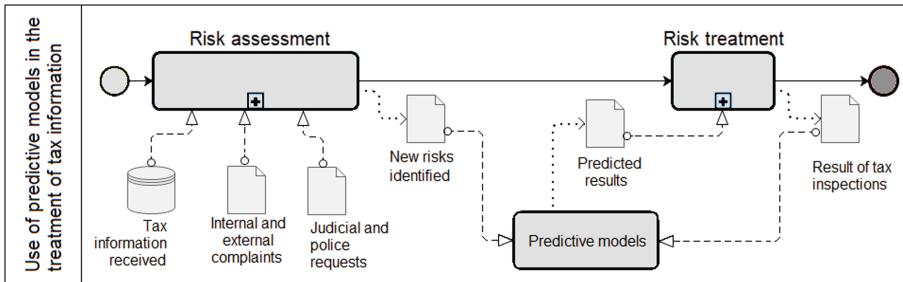


Fig. 1. Predictive models in the tax information treatment process

4.1 Data Understanding

A graphical analysis was performed comparing IN values and quantities by economic activity, location and type of tax action, but such analysis did not allow to find a pattern between these characteristics and the results of the inspections. The frequency distribution of the IN value takes the form of the chi-square curve with two degrees of freedom.

The questionnaire was answered by 68% of the auditors, however only 2 types of tax actions had one of the characteristics evaluated by more than 50% of the auditors. Thus, it was not possible to establish comparison standards for the selected characteristics based on the auditors' perception, and these data were not incorporated into the models developed.

4.2 Logistic Regression Models

Logistic regression was used to search for relationships between data entries (taxpayer characteristics and evidence of fraud) and outputs (invoices) to make predictions of the output category for new entries.

The following models were analysed:

- LRM1: dependent variable presents 1 for infraction Notices of any value and 0 for "Not Fined", all records are used (8.300 observations);
- LRM2: dependent variable presents 1 for infraction Notices above R\$ 1 million and all records are used;
- LRM3: dependent variable presents 1 infraction Notices of any value and only records related to concentrated audits are used (4,057 observations);
- LRM4: Infraction Notices above R\$ 1 million, only concentrated audits.

The functions `glm`, `train`, `predict` and `create.partition` of R Software were used and the cross-validation technique was implemented (3/4 for training and 1/4 for validation). The confusion matrix was constructed and the forecasts made compared to the real value of the infraction notices drawn up in order to observe which portion of notices would be kept and which would be lost if the predictions had guided the decision to inspect the company or not.

Results of Logistic Regression Models. Table 4 shows the comparison between the predictions made and the actual result of the tax actions.

Table 4. Value of the infraction notices drawn up x predicted infraction notices

Model	VFinedQQ		Value of infrac. notices	Qty. taxpayers	Total taxpayers	% Value	% Qty
	Predicted	Performed					
LRM1	Fined	Fined	2,189,971,868.17	719	878	83.93	42.33
		Not Fined	0	159			
	Not fined	Fined	419,106,935.39	211	1,196	16.06	57.66
		Not fined	0	985			
LRM2	Fined	Fined	1,022,390,825.27	100	145	32.40	6.99
		Not fined	(*) 0	45			
	Not fined	Fined	2,132,983,197.4	214	1,929	67.60	93.00
		Not fined	(*) 0	1,715			
LRM3	Fined	Fined	1,135,119,762.29	656	785	90.10	77.42
		Not fined	0	129			
	Not fined	Fined	124,850,505.5	60	229	9.90	22.58
		Not fined	0	169			
LRM4	Fined	Fined	668,422,469.22	109	162	48.85	15.99
		Not fined	(*) 0	53			
	Not fined	Fined	592,020,681.15	128	851	43.26	84.00
		Not fined	(*) 0	723			

(*) individual values less than R\$ 1 million were considered equal to zero.

The predictions of the LRM1 model imply maintaining 83.9% of the value of the infraction notices drawn up and avoiding 1,196 inspections carried out (57.66%), while the LRM3 model reaches 90.10% of the value and avoiding 226 inspections.

Table 5 shows the result of the confusion matrix for each model and, in the last column, the result of the analysis made in Table 4 (percentage of the values of the infraction notices maintained).

Table 5. Logistic regression models

Model			Accuracy	Recall	Precision	F1 Score	% Value
Name	Type of inspection	Value IN					
LRM1	Every type	All	81.43%	93.76%	72.72%	81.91%	83.93%
LRM2	Every type	1 million up	87.51%	31.84%	68.96%	43.57%	32.40%
LRM3	Concentrated audit	All	81.36%	91.62%	83.56%	87.40%	90.10%
LRM4	Concentrated audit	1 million up	82.13%	45.99%	67.28%	54.63%	48.85%

The percentage of value maintained is low (below 50%) when considering only the tax assessment notices above R\$ 1 million (LRM2 and LRM4).

4.3 Neural Network Models

The h2o.deeplearning and h2o.predict functions were used to create four models of neural networks that have the same dependent variables and characteristics of the evaluated logistic regression models. All models use 2 hidden layers and 1000 training epochs, because they showed less variation between tests than 1 hidden layer models and tests using 10,000 epochs did not show significant variations.

Results of Neural Network Models. Table 6 presents the comparison between the predictions made and the actual result of the tax actions.

Table 6. Value of the infraction notices drawn up x predicted infraction notices

Model	VFinedQQ		Value of infrac. notices	Qty. taxpayers	Total taxpayers	% Value	% Qty
	Predicted	Performed					
NN1	Fined(1)	Fined(1)	2,683,902,509.29	829	1,040	89.57	50.14
		Not fined(0)	0	211			
	Not fined(0)	Fined(1)	312,322,526.79	101	1,034	10.42	49.85
		Not fined(0)	0	933			
NN2	Fined(1)	Fined(1)	1,404,118,707.31	140	226	47.89	10.89
		Not fined(0)	(*) 0	86			
	Not fined(0)	Fined(1)	1,527,726,055.73	174	1,849	52.1	89.1
		Not fined(0)	(*) 0	1,675			
NN3	Fined(1)	Fined(1)	1,516,644,057.63	627	775	86.52	76.42
		Not fined(0)	0	148			
	Not fined(0)	Fined(1)	236,219,363.58	89	239	13.47	23.57
		Not fined(0)	0	150			
NN4	Fined(1)	Fined(1)	662,326,142.33	100	171	54.24	16.84
		Not fined(0)	(*) 0	71			
	Not fined(0)	Fined(1)	558,647,142.04	138	844	45.75	83.15
		Not fined(0)	(*) 0	706			

(*) individual values lower than R\$ 1 million were considered equal to zero.

Table 7 summarizes the evaluation of neural network models.

Table 7. Neural network models

Model			Accuracy	Recall	Precision	F1 Score	% Value
Name	Type of inspection	Value IN					
NN1	Every type	All	85.08%	89.00%	79.96%	84.24%	89.57%
NN2	Every type	1 million up	87.71%	42.35%	64.25%	51.05%	47.89%
NN3	Concentrated Audit	All	78.20%	95.39%	78.41%	86.07%	86.52%
NN4	Concentrated Audit	1 million up	80.00%	41.59%	60.73%	49.37%	54.24%

All models of neural networks are highly accurate, however, Table 6 shows that predictions of the NN1 model imply maintaining 89.57% of the value charged

avoiding 1,034 inspections (49.85% of the total), while the NN3 model reaches 86.52% of the value and avoids 239 inspections.

The results indicate that the use of the LRM1 or NN1 models would have allowed the maintenance of more than 80% of the amount charged, with about half of the inspections.

5 Conclusion

The amount of digital information received by the Tax Authorities, due to the increasing computerization of procedures related to compliance with tax obligations, allows a better understanding of the reality of taxpayers and the complexity of their relationships. Converting the data received into useful information to guide inspection is a constant challenge, much has been done in this regard and much remains to be done.

Much evidence of fraud is associated with amounts and is traditionally inspected in decreasing order of these amounts, with priority over other factors, so that many inspections do not result in collection. As resources for enforcement are limited, many indications of fraud are not monitored. But it is not possible to discard the verification of evidence of fraud without a strong reason, at the risk of weakening the selection process of the companies to be audited.

The search for a scientific method that allows a probability of confirmation to be associated with the value of the fraud indication to select the most likely situation to be confirmed aims to optimize the work of the Tax Administration and maintain an impersonal and technical choice.

According to the results of the study, the application of the resources available for carrying out the inspection could have been optimized with the use of predictive models based on the result of previous tax inspections, with promising results being obtained only for these predictive models whose dependents the variable received the result ‘success’ when an infraction notice of any value was drawn up.

Models that use infraction notices of any value have the greatest potential to optimize the use of available resources, since the actions that would not be carried out correspond to approximately half of the actions carried out and the amount of the credit constituted was greater than 80%, on both models. The linear regression and neural network models that use all inspections (NN1 and LRM1) have a better result than models that use only the concentrated audit (NN3 and LRN3), because they present a greater amount of inspections that could be avoided, and consequently, higher percentage of resources released for the development of other activities.

The contribution of this research is to open up the possibility of using predictive models in the process of selecting signs of tax fraud to be inspected.

References

1. Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H.: State-of-the-art in artificial neural network applications: a survey. *Heliyon* **4**(11) (2018)
2. Alink, M., Van Kommer, V.: Translated by Vinícius Pimentel de Freitas: Handbook on Tax Administration, IBFD - International Bureau of Fiscal Documentation (2011)
3. Babu, S.K., Vasavi, S.: Predictive analytics as a service on tax evasion using gaussian regression process. *Helix* **7**(5), 1988–1993 (2017)
4. Choi, D., Lee, K.: An artificial intelligence approach to financial fraud detection under iot environment: a survey and implementation. *Secur. Commun. Netw.* (2018)
5. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: Knowledge discovery and data mining: towards a unifying framework (1996)
6. Ferneda, E: Redes neurais e sua aplicação em sistemas de recuperação de informação. *Ci. Inf.*, Brasília, 35 jan./abr(1), 25–30 (2006)
7. González, P.C., Velásquez, J.D.: Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Syst. Appl.* (40), 1427–1436 (2012)
8. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: a review. *Neurocomputing* **187**, 27–48 (2016)
9. Hajek, P., Henriques, R.: Mining corporate annual reports for intelligent detection of financial statement fraud –a comparative study of machine learning methods. *Knowl.-Based Syst.* (128), 139–152 (2017)
10. Huang, W., Lai, K.K., Nakamori, Y., Wang, S., Yu, L.: Neural networks in finance and economics forecasting. *Int. J. Inf. Technol. Decis. Making* **6**(1), 113–140 (2007)
11. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
12. Li, Y., Jiang, W., Yang, L., Wu, T.: On neural networks and learning systems for business computing. *Neurocomputing* (275), 1150–1159 (2018)
13. Mishra, P.: R Data Mining Blueprints. Packt Publishing Ltd., Birmingham (2016). ISBN 978-1-78398-968-38
14. Montgomery, D.C., Runger, G.C.: Estatística Aplicada e Probabilidade para Engenheiros, 5^aed. Publisher, Livros Técnicos e Científicos Ltda (2012)
15. Morales, A.M.C., Pineda, C.M.R. and Monsalve, O.O.V.: La primerareforma tributaria en la historia de la humanidad. *Entramado* **15**(1), 152–163 (2019)
16. Prieto, A., Prieto, B., Ortigosa, E.M., Ros, E., Pelayo, F., Ortega, J., Rojas, I.: Neural networks: an overview of early research, current frameworks and new challenges. *Knowl.-Based Syst.* (128), 139–152 (2017)
17. Siqueira, M.L., Ramos, F.S.: A economia da sonegação. *Revista Econ. Contemporânea*. RJ (2005)
18. Silva, A.A., Cerqueira, A.F.: *Fraudes Contábeis Repercussões Tributária Enfoque no ICMS*. Juruá Editora (2018)
19. Soltoggio, A., Stanley, K.O., Risi, S.: Born to learn: the inspiration,progress, and future of evolved plastic artificial neural networks. *Neural Netw.* **108**, 48–67 (2018)
20. Stulp, F., Sigaud, O.: Many regression algorithms, one unified model: a review. *Neural Netw.* **69**, 60–79 (2015)
21. Tkáč, M., Verner, R.: Artificial neural networks in business: two decades of research. *Appl. Soft Comput. J.* **38**, 788–804 (2016)

22. Witten, I.H., Frank, E., Hall, M.A.: Data Mining Practical Machine Learning Tools and Techniques, 3th edn. Morgan Kaufmann Publishers - Elsevier, Burlington (2011)
23. Wu, R.S., Ou, C.S., Lin, H.Y., Chang, S.I., Yen, D.C.: Using data mining technique to enhance tax evasion detection performance. *Expert Syst. Appl.* (39), 8769–8777 (2012)



Mobile Manipulator Robot Control Through Virtual Hardware in the Loop

Byron S. Jorque^(✉), Jéssica D. Mollocana^(✉), Jessica S. Ortiz^(✉),
and Víctor H. Andaluz^(✉)

Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
`{bsjorque, jamollocana, jsortiz4, vandaluz1}@espe.edu.ec`

Abstract. This article focuses on the implementation of the Hardware-in-the-Loop technique to evaluate advanced control algorithms to a mobile manipulator robot that comprised of 3DOF anthropomorphic type robotic arm mounted on a unicycle type platform. The implementation of HIL includes the use of Unity 3D graphic engine for the development of a virtual environment that allows to visualize the execution of the movements of the robot through the implemented control algorithm. In addition, it is considered the kinematic model and dynamic model of the robot that represent the characteristics and restrictions of movement of the mobile manipulator robot. Finally, experimental results achieved through the implementation of the HIL technique are presented, in which the behavior of the robotic system and the evolution of control errors when executing locomotion and object manipulation tasks can be verified.

Keywords: HIL · Mobile Manipulator · Algorithm controller · Kinematic · Dynamic

1 Introduction

Robotics have evolved in the last decades to the point that it is essential in the industry for the automation of production lines. Autonomous robots execute repetitive tasks with great velocity and precision [1, 2]. However, the current challenge in robotics is to transcend from industrial robotics to service robotics, in which robots are specifically designed for the service of mankind [3]. Service robotics is a field that focuses on assisting humans outside the industrial environment. For instance, they can perform domestic, security, surveillance, and transportation tasks. Service robotics is additionally an active field of research due to its numerous personal and professional applications. For example, they can be used as servants, company robots, and nursing aid robots [6]. Other uses of service robots may use multiple actuators; for instance, the cooperative control of mobile manipulators [4], or human - robot collaboration [5]. Finally, there are robots specifically designed for a task. In this way, robots can be aerial, aquatic, and terrestrial, and they can move using wheels, legs, fins, and propellers [7].

There are complex tasks that require both locomotion and manipulation. That is why new kinds of robots are being developed, which combine both functionalities; they are

called mobile manipulators [10]. In other words, a mobile manipulator is a mechanic structure comprised of a robotic arm coupled with a locomotion system [8]. This kind of robots, therefore, combine the dynamic aspect of the platform with the ability of object manipulation of the arm [8]. This in turn makes the robot more versatile since it has more work range and flexibility [9]. Nowadays, mobile manipulators are used in several fields, providing services like domestic cleaning, personal assistance [11], and working in the construction and mining industries [12].

Between the techniques used to control a mobile manipulator robot is the execution of the complete simulation, construction of the robot and the implementation of the controller in hardware and that developed in recent years; Hardware in the Loop. HIL is a new alternative for the control of complex physical systems [13], which includes the development of a real time simulation environment to test processes. For the development of a HIL environment it is essential to incorporate into software the mathematical models that represent a physical system and physical hardware devices that act as controllers [14]. In this way, simulation platform is obtained that emulates a real process and at the same time provides the control unit, electrical signals similar to those obtained from a real process.

There are several control techniques for mobile manipulators, such as complete simulation, mechanical construction, and Hardware in the Loop (HIL), which is the most recent one. The later one is a new alternative for the development of simulation environments, and the control of complex physical systems [13]. This technique is not limited to the software representation of the system or the complete implementation, which can be expensive, but rather it combines the advantages of both techniques. HIL incorporates the mathematical models that represent a physical system with the hardware that control the behavior of the robot [14]. HIL also allows including additional physical devices that interact with the simulated environment [15]. In this way, a robust simulation platform is obtained, which emulates a real process and provides control units with electrical signals that are similar to those of a real process.

In this paper we present the implementation of the technique of Hardware in the Loop for the autonomous control of a mobile manipulator robot that comprised of a robotic arm mounted on a mobile platform type unicycle. For the implementation of the Hardware in the Loop technique, the development of a virtual environment in the Unity 3D graphic engine is considered [16, 17] that allows to visualize the behavior of the robot when executing autonomous tasks. The proposed control scheme considers a cascade system consisting of 2 subsystems. The first subsystem considers a kinematic controller, while the second subsystem comprises a dynamic compensation to decrease the velocity and tracking error. Also, it is analyzed the stability and robustness of the control algorithm proposed through Lyapunov's theory considering as perturbations errors in the maneuvering velocities. Finally, several results obtained by implementing the HIL technique are presented, evaluating the behavior of the control algorithms and the evolution of control errors, which converge asymptotically to zero when there are no disturbances and is limited when considering disturbances in the maneuverability commands.

This article is presented in 6 sections, including the Introduction. In the second section, the structure of the HIL environment can be appreciated, as well as the communication channels. Section 3 shows the kinematic and dynamic modeling of the mobile manipulator, considering its characteristics and movement restrictions. Section 4 details the design of the control algorithm, as well as the stability and robustness analysis for the validation of the controller. Experimental results are shown in Sect. 5, and finally Sect. 6 shows the conclusions.

2 System Structure

In this section, the structure of the implemented HIL environment is detailed. Figure 1 shows the HIL scheme, which is comprised of three blocks: real time simulation, communication channel and end hardware.

HIL environment is carried out in the Unity-3D graphic engine, as well as external resources such as the CAD software, and additional input and output devices. The CAD software contains the configurations of the real robot and the 3D models of the environment that resembles a factory where the robot is tested. The virtual environment then allows visualizing the locomotion and object manipulation actions taken by the robot.

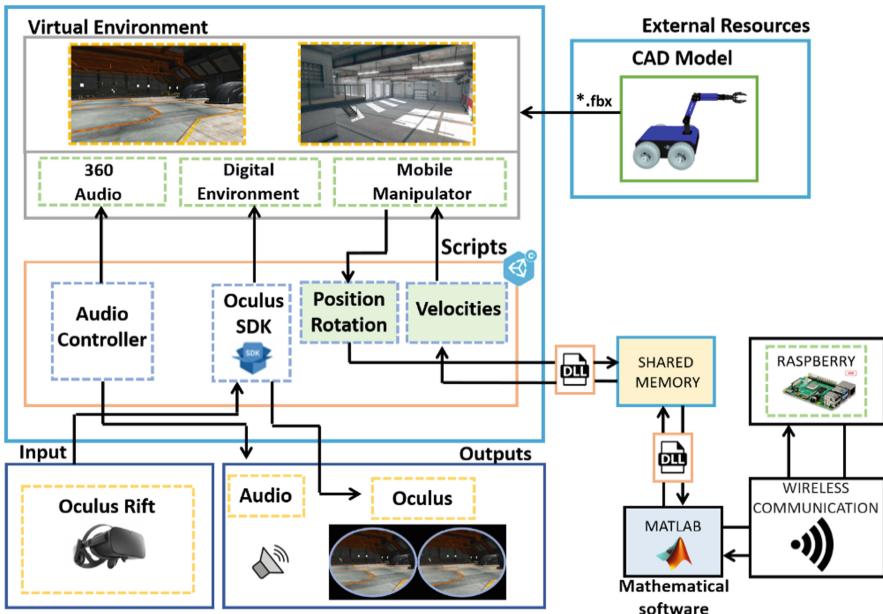


Fig. 1. Hardware in the loop.

Hill technique also allows connecting the virtual environment with additional I/O devices using a set of scripts. The purpose of the I/O devices is to have visual and auditory feedback of the proper functionality of the robot. The control algorithm is embedded in

a low-cost Raspberry-Pi board, which executes and sends the control actions wirelessly using the ZigBee protocol to Matlab.

The link with Matlab is carried out using a Dynamic Link Library (DLL), which allows different software to exchange information through shared memory. This memory enables introducing the control actions of the mobile manipulator inside the simulation environment. In this way, the controller sends the proper control actions at a given time, and receives feedback from the simulation to compute the corresponding error.

3 Mobile Manipulator Robot

This section shows the kinematic and dynamic modeling of a mobile manipulator robot. The robot is comprised of a mobile platform and a robotic arm, which are considered as a single robotic system. Those elements also determine restriction and motion characteristics of the robot.

The kinematics of a robot is given by its position and orientation with respect to its working environment, the geometrical relationship between its components, as well as motion restrictions [18]. Figure 2 illustrates the configuration of the mobile manipulator robot that was used in this work. The point p expresses the position of the robot, \mathbf{h} is the interest point, which represents the position and orientation of the manipulator's end effector with respect to the reference frame $\{R\}$.

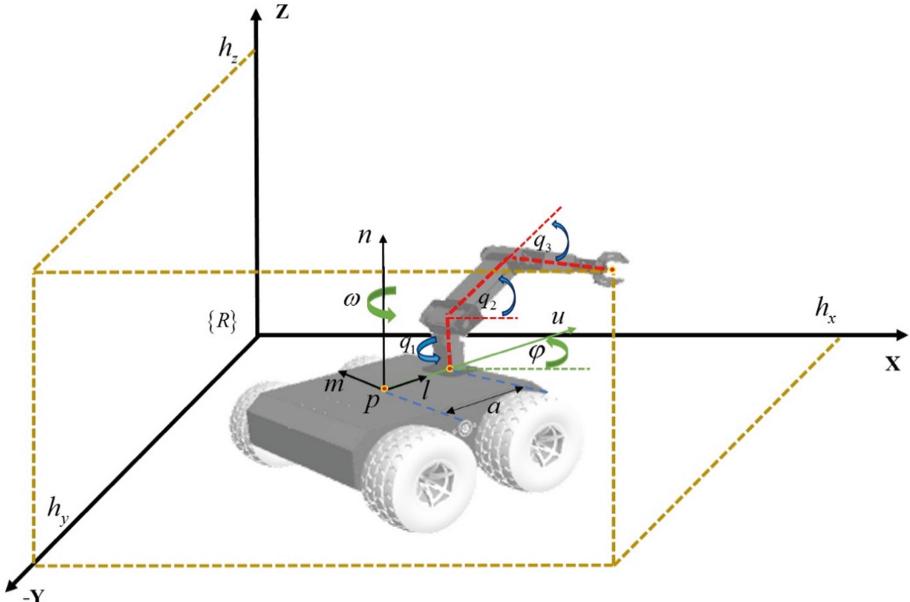


Fig. 2. Configuration of the mobile manipulator robot.

The kinematic model of the robot is obtained by applying the derivative to the end effector. The position and orientation of the end effector are expressed as a function of

the position of the mobile platform and the configuration of the robotic arm.

$$\dot{\mathbf{h}}(t) = \mathbf{J}(\mathbf{q})\mathbf{v}(t), \quad (1)$$

where $\mathbf{v} = [\dot{h}_x \dot{h}_y \dot{h}_z]^T$ is the end effector's velocity vector, $\mathbf{v} = [\mu \omega \dot{q}_1 \dot{q}_2 \dot{q}_3]^T$ is the robot's velocity vector, $\mathbf{q} = [x \ y \ \varphi \ q_1 \ q_2 \ q_3]$ is the position vector, and $\mathbf{J}(\mathbf{q})$ is the Jacobian matrix that establishes a linear relationship between velocities, and is based on the rotation angles of the joints in the manipulator [18].

Robot characteristics and motion restrictions are represented by the Jacobian matrix. For instance, motion restrictions are given by:

$$\dot{x} \sin(\varphi) - \dot{y} \cos(\varphi) + a\omega = 0, \quad (2)$$

where \dot{x} and \dot{y} are the velocities of the mobile platform over the x and y axis, a is a constant that defines the distance between the base of the robotic arm and the mobile platform point p of coordinates (x, y) , and ω represents the angular velocity of the mobile platform with respect to the z axis. All of this in relationship with the reference frame $\{R\}$.

On the other hand, in order to determine the motion characteristics of the mobile manipulator at executing tasks, it is important to define the dynamics of the robotic system. This is carried out using the Euler-Lagrange proposal, given by:

$$L = K - P, \quad (3)$$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{q}}} \right) - \frac{\partial L}{\partial \mathbf{q}} = \boldsymbol{\tau}, \quad (4)$$

where L is the Lagrange function, which defines the balance between kinetic and potential energies of the different elements of the robot, and $\boldsymbol{\tau}$ is the torque vector applied to the robot, i.e. the torque due to the translation and rotation of the robot. By associating (3) and (4), the matrix model (5) is obtained. It contains all the forces generated by the robot:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{v}}(t) + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\mathbf{v}(t) + \mathbf{g}(\mathbf{q}) = f(t), \quad (5)$$

using (5), the motion equations of the actuators are included, as well as the robot's configuration. In this way, the expression for the dynamical model is derived:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{v}}(t) + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\mathbf{v}(t) + \mathbf{g}(\mathbf{q}) = \mathbf{v}_{ref}(t), \quad (6)$$

where $\mathbf{M}(\mathbf{q})$ is the Inertial matrix; $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$ is the matrix of centripetal force and Coriolis, $\mathbf{g}(\mathbf{q})$ is the gravitational vector, which represents the effects of gravity on the robot's components, and $\mathbf{v}_{ref}(t)$ is the vector of control velocity.

4 Control Scheme

This section presents the design of the control algorithm, which is comprised of a kinematic control and a dynamic compensator. The former is based on the robot's kinematics while the later has the objective to reduce the velocity error by compensating its dynamics. Additionally, the stability and robustness of the proposed controller is analyzed in this section. Figure 3 illustrates the control scheme for the mobile manipulator robot implemented using HIL.

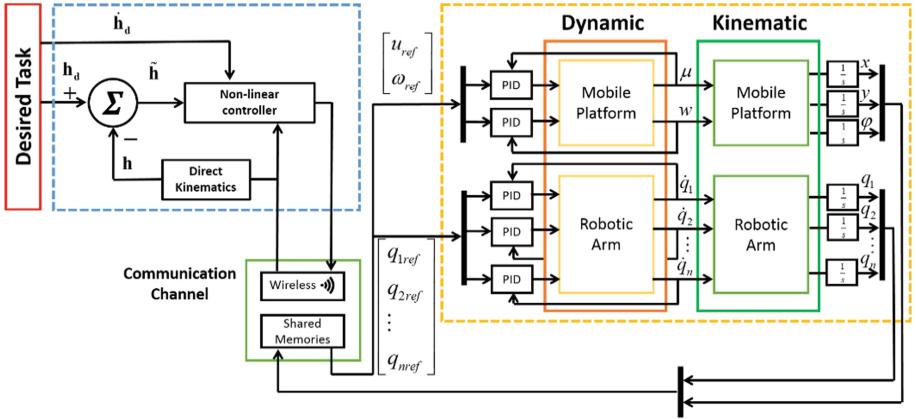


Fig. 3. Autonomous control scheme.

4.1 Kinematic Control

The kinematic model of the robotic mechanism (1) contributes to the development of the kinematic controller. The velocity vector of the robot can be expressed as the velocity vector of the end effector using the pseudo-inverse of the Jacobian matrix.

$$\mathbf{v}(t) = \mathbf{J}^\#(\mathbf{q})\dot{\mathbf{h}}(t) \quad (7)$$

where, $\mathbf{J}^\#(\mathbf{q}) = \mathbf{W}^{-1}\mathbf{J}^T(\mathbf{J}\mathbf{W}^{-1}\mathbf{J}^T)^{-1}$ and \mathbf{W} is a positive symmetric matrix that exerts the control actions. Thus, the velocity vector is given by:

$$\mathbf{v}(t) = \mathbf{W}^{-1}\mathbf{J}^T(\mathbf{J}\mathbf{W}^{-1}\mathbf{J}^T)^{-1}\dot{\mathbf{h}}(t) \quad (8)$$

Therefore, the control law that commands the velocities of the manipulator, as well the desired trajectory of the robot, is formulated:

$$\mathbf{v}_c = \mathbf{J}^\#(\dot{\mathbf{h}}_d + \mathbf{K} \tanh(\tilde{\mathbf{h}})) + (\mathbf{I} - \mathbf{J}^\# \mathbf{J}) \mathbf{D} \tanh(\eta) \quad (9)$$

where, \mathbf{h}_d is the vector of desired positions, $\dot{\mathbf{h}}_d$ is the vector of desired velocities at operative extreme, $\tilde{\mathbf{h}}$ represents the control error given by $\tilde{\mathbf{h}} = \mathbf{h}_d - \mathbf{h}$; \mathbf{K} and \mathbf{D} are defined positive gain matrices, and finally η defines the position error vector of the robotic arm, its function is to obtain maximum manipulability [19].

Once the control law based on the kinematic model has been established, it is important to grant its stability. For this purpose, the behavior of error control $\tilde{\mathbf{h}}$ is carried out, considering a perfect velocity tracker $\mathbf{v} = \mathbf{v}_c$. By replacing (9) in (1), the following expression is obtained:

$$\dot{\tilde{\mathbf{h}}} + \mathbf{K} \tanh(\tilde{\mathbf{h}}) = 0 \quad (10)$$

For the stability analysis, the following function is considered a candidate Lyapunov $V(\tilde{\mathbf{h}}) = \frac{1}{2}\tilde{\mathbf{h}}^T\tilde{\mathbf{h}}$; where its first time derivate is defined as $\dot{V}(\tilde{\mathbf{h}}) = \tilde{\mathbf{h}}^T\dot{\tilde{\mathbf{h}}}$. Now replacing (10) in the Lyapunov candidate function is obtained:

$$\dot{V}(\tilde{\mathbf{h}}) = \tilde{\mathbf{h}}^T \mathbf{K} \tanh(\tilde{\mathbf{h}}) \quad (11)$$

this equation implies that the closed loop control system is asymptotically stable; such that $\tilde{\mathbf{h}}(t) \rightarrow 0$ con $t \rightarrow \infty$.

4.2 Dynamic Compensation

The dynamic compensation block has as objective reducing the velocity tracking error by compensating the dynamics of the system. The desired velocities \mathbf{v}_c are fed to the controller for the dynamic compensation, and they originate the reference velocity for the mobile manipulator robot \mathbf{v}_{ref} . Thus the dynamic compensation is given by:

$$\mathbf{v}_c = \mathbf{J}^\# (\dot{\mathbf{h}}_d + \mathbf{K} \tanh(\tilde{\mathbf{h}})) + (\mathbf{I} - \mathbf{J}^\# \mathbf{J}) \mathbf{D} \tanh(\eta) \quad (12)$$

where, $\mathbf{v}_{ref} = [\mu_{ref} \omega_{ref} \dot{q}_{1ref} \dot{q}_{2ref} \dot{q}_{3ref}]^T$ represent the control action; $\tilde{\mathbf{v}}$ defines the velocity error given by $\tilde{\mathbf{v}} = \mathbf{v}_c - \mathbf{v}$, and finally, $\dot{\mathbf{v}}_c$ is the vector accelerations.

Following the procedure, previously established in the stability analysis of the kinematic controller, and considering a Lyapunov candidate function as the quadratic error, it is possible to determine that $\dot{\tilde{\mathbf{v}}}(t) \rightarrow 0$ asymptotically, when $t \rightarrow \infty$. This grants the stability of the proposed control law.

Similarly, performing a robustness analysis is a fundamental part of the implemented controller. This allows determining the validity of the control errors. Thus, we have that:

$$\dot{V}(\tilde{\mathbf{h}}) = \tilde{\mathbf{h}}^T \delta_{\dot{\mathbf{h}}} - \tilde{\mathbf{h}}^T \mathbf{K} \tanh(\tilde{\mathbf{h}}) \quad (13)$$

where, $\delta_{\dot{\mathbf{h}}}$ is the variation difference between the desired velocity and the real ones. This is under the condition of perfect tracking defined as $\delta_{\dot{\mathbf{h}}} = \dot{\mathbf{h}}_d - \dot{\mathbf{h}}$.

In order for $\dot{V}(\tilde{\mathbf{h}})$ to be negative, the following expression must be true:

$$|\tilde{\mathbf{h}}^T \mathbf{K} \tanh(\tilde{\mathbf{h}})| > |\tilde{\mathbf{h}}^T \delta_{\dot{\mathbf{h}}}|, \quad (14)$$

for $\tilde{\mathbf{h}}$ with high values, $\mathbf{K} \tanh(\tilde{\mathbf{h}}) \approx \mathbf{K}$; Therefore $\dot{V}(\tilde{\mathbf{h}})$ will be negative only if $\|\mathbf{K}\| > \|\delta_{\dot{\mathbf{h}}}\|$. In this way the error $\tilde{\mathbf{h}}$ diminishes. On the other hand, $\tilde{\mathbf{h}}$ with low values, $\mathbf{K} \tanh(\tilde{\mathbf{h}}) \approx \mathbf{K}\tilde{\mathbf{h}}$. In this case, (13) is written as:

$$\|\tilde{\mathbf{h}}\| > \|\delta_{\dot{\mathbf{h}}}\| / \lambda_{\min}(\mathbf{K}) \quad (15)$$

implying that the error $\tilde{\mathbf{h}}$ is limited by,

$$\|\tilde{\mathbf{h}}\| \leq \frac{\|\delta_{\tilde{\mathbf{h}}}\|}{\lambda_{\min}(\mathbf{K})} \quad (16)$$

And, if $\delta_{\tilde{\mathbf{h}}} \neq 0$, $\tilde{\mathbf{h}}(t)$ ultimately limited by (16).

5 Experimental Results

This section presents the most relevant experiments and results. They serve to evaluate the behavior of the mobile manipulator robot in the virtual environment. For the experimental tests a laptop containing the mathematical model that simulates in behavior of the robot inside a virtual scenario created in the Unity 3D graphic engine was considered. The computer has 16 GB of RAM memory with a GPU of 6 GB. The controller was implemented in a Raspberry Pi-4 model B of 4 GB of RAM. It represents the hardware component of HIL. Finally, to establish a wireless communication channel, Xbee S2C devices were used. Figure 4 shows the physical implementation of the HIL system.



Fig. 4. Physical implementation of the Hardware in the Loop environment.

The experiments carried out recreate the simulation environment of the mobile manipulating robot with the law of control implanted. The mechanical design of the robot in the Unity-3D graphic engine is shown in Fig. 5, as well as the main variables that allow to visualize the displacement of the robot considering an infinite type path in order to excite the dynamics of the robot and to check the evolution of the control algorithm. To achieve the desired trajectory, the following parametric functions are used $h_{xd} = 2.45 \cos(0.05t)$, $h_{yd} = 1.35 \sin(0.1t)$ and $h_{zd} = 0.6 + 0.2 \cos(0.1t)$.

In order to verify the implementation of the HIL technique, the Fig. 6 shows the real-time error variation for the position in the $X - Y - Z$ axis, as well as angular and linear velocity variation of every link in the mobile manipulator robot. Its observed how the

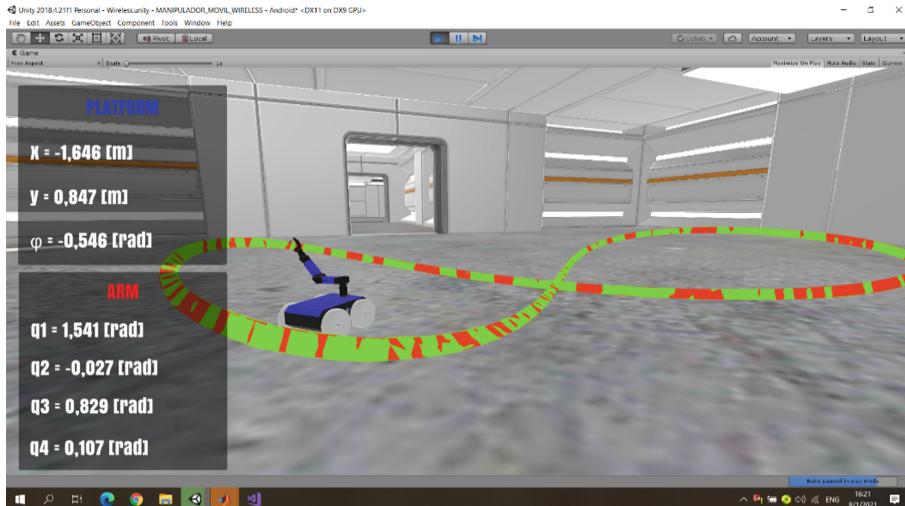
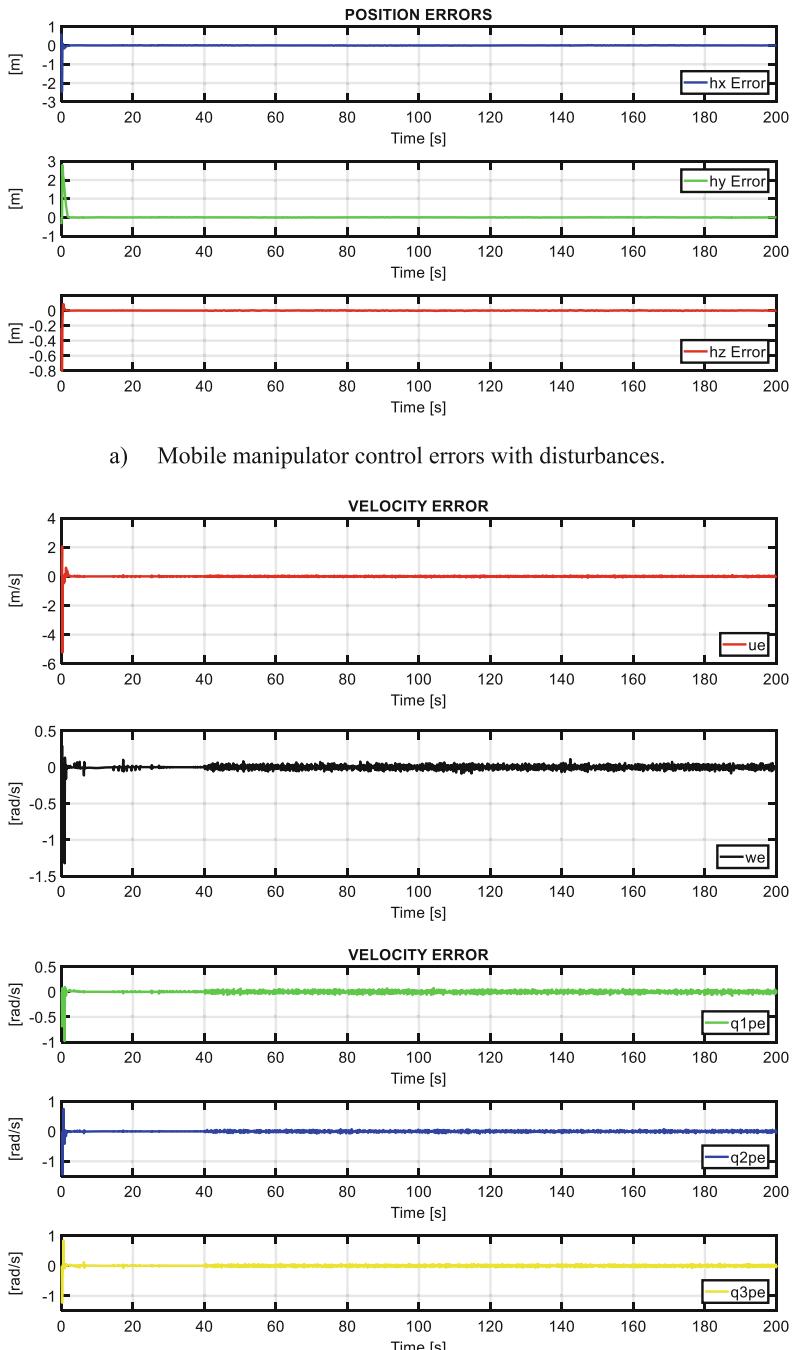


Fig. 5. Mobile manipulator movement.

control actions make adjustments in the robot's joints so that the end effector can follow the desired path, even when the robot suffers perturbations. The Fig. 6a shows control errors with perturbations while Fig. 6b shows velocities errors with perturbations. The errors associated with the dynamic function, do not tend to cero, but rather oscillate in a value near cero. However, the system still follows the desired trajectory, which indicates that the controller is indeed robust, as defined in (16). Therefore, by means of the control law, the errors tend to cero in function of the gain matrix \mathbf{K} .

In this way, the performed tests illustrate how the implementation of a HIL environment constitutes an effective technique to control mobile manipulator robots. This technique allows implementing control algorithms embedded in virtual environments. Unlike alternative techniques, HIL permits incorporating input/output devices in the system. Those devices produce additional visual and auditive feedback of the robot's actions inside the virtual environment, making the system more user friendly. In contrast, physical construction is more expensive and time consuming, and total simulation does not enable incorporating physical components that interact with the mechanism.



b) Mobile manipulator velocity error with disturbances.

Fig. 6. Control and velocity errors with disturbances.

6 Conclusions

Hardware in the Loop is a technique that assist in the development of complex robotic test environments, without the need of building a mechanical system. The virtual environment developed in the graphic engine Unity 3D enables visualizing the robot's motion, which is commanded by the control law implemented in the end hardware. This control unit, through wireless information exchange, performs important parameter control, such as the motion of the joints in the robotic arm and the position of the mobile platform. Obtained results, regarding the robot performing autonomous tasks, validate the efficacy of the implemented technique, as well as the performance of the control law. Steady state errors in these tests tended to zero, as designed.

Acknowledgements. The authors would like to thank the Coorporación Ecuatoriana para el Desarrollo de la Investigación y Academia- CEDIA for their contribution in innovation, through the CEPRA projects, especially the project CEPRA-XIV-2020-08-RVA “Tecnologías Inmersivas Multi-Usuario Orientadas a Sistemas Sinérgicos de Enseñanza-Aprendizaje”; also the Universidad de las Fuerzas Armadas ESPE and the Research Group ARSI, for the support for the development of this work.

References

1. Andaluz V.H., Guamán S., Sánchez J.S. Autonomous march control for humanoid robot animation in a virtual reality environment. *computational kinematics: mechanisms and machine science*, vol. 20, pp. 70–78. Springer (2018)
2. Aliev K., Antonelli, D.: Analysis of cooperative industrial task execution by mobile and manipulator robots, pp. 248–260. Springer (2019)
3. Soto, D., Ramírez, J., Gazeau J.: Towards an autonomous airborne robotic agent, pp. 62–69. Springer (2017)
4. Andaluz V., et al.: Coordinated cooperative control of mobile manipulators. In: *IEEE International Conference on Industrial Technology*, Auburn, AL, pp. 300–305 (2011)
5. Su, H., Yang, C., Ferrigno, G., De Momi, E.: Improved human-robot collaborative control of redundant robot for teleoperated minimally invasive surgery. *IEEE Robot. Autom. Lett.* **2**(4), 1447–1453 (2019)
6. Baumgarten, S., Jacobs, T., Graf, B.: The robotic service assistant - relieving the nursing staff of workload. In: *ISR 2018; 50th International Symposium on Robotics*, Munich, Germany, pp. 1–4 (2018)
7. Ortiz, J.S., et al.: Modeling and kinematic nonlinear control of aerial mobile manipulators, pp:87–95. Springer (2017)
8. Ortiz J., et al.: Coordinated control of a omnidirectional double mobile manipulator, pp. 278–286. Springer (2018)
9. Andaluz V., et al.: Numerical methods for cooperative control of double mobile manipulators, pp. 889–898. Springer (2017)
10. Varela, J., et al.: User centred design of rehabilitation robots, pp. 97–109. Springer (2017)
11. Meng, Z., et al.: Modelling and control of a 2-link mobile manipulator with virtual prototyping. In: *13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, Xi'an, pp. 363–368 (2016)
12. Amaya, E., et al.: Diseño y construcción de una plataforma robótica para reconstrucción 3D mediante un sistema de procesamiento embebido. *IEEE Lat. Am. Trans.* **1**(16), 5–19 (2018)

13. Nguyen, K., Ha, C.: Development of hardware-in-the-loop simulation based on gazebo and pixhawk for unmanned aerial vehicles. *JASS* **19**, 238–249 (2018)
14. Kumar, A., et al.: Hardware in the loop based simulation of a robotic system with real time control and animation of working model. In: International Conference on Inventive Systems and Control (ICISC), Coimbatore, pp. 1–5 (2017)
15. Carvajal, C., et al.: Autonomous and tele-operated navigation of aerial manipulator robots in digitalized virtual environments, pp. 496–515. Springer (2018)
16. Andaluz, V.H., et al.: Unity3D-MatLab simulator in real time for robotics applications, pp. 246–263. Springer (2016)
17. Molina, M., Ortiz, J.: Coordinated and cooperative control of heterogeneous mobile manipulators, pp. 483–492. Springer (2018)
18. Varela-Aldás, J., Andaluz, V.H., Chicaiza, F.A.: Modelling and control of a mobile manipulator for trajectory tracking. In: 2018 International Conference on Information Systems and Computer Science (INCISCOS), Quito (2018)
19. Andaluz, V., Flavio, R., Toíbero, J.M., Carel, R.: Adaptive unified motion control of mobile manipulators. *Control Eng. Pract.* **20**(12), 1337–1352 (2012)



Optimizing Regularized Multiple Linear Regression Using Hyperparameter Tuning for Crime Rate Performance Prediction

Alexandra Vultureanu-Albișă^(✉) and Costin Bădică

University of Craiova, Craiova, Romania

alexandra.vultureanu@edu.ucv.ro, cbadica@software.ucv.ro

Abstract. Multiple Linear Regression is a well-known technique used to experimentally investigate the relationship between one dependent variable and multiple independent variables. However, fitting this model has problems, for example when the sample size is large. Consequently, the results of traditional methods to estimate the model can be misleading. So, there have been proposed regularization or shrinkage techniques to estimate the model in this case. In this work, we have proposed a methodology to build a crime rate performance prediction model using multiple linear regression methods with regularization. Our methodology consists of three major steps: i) analyzing and preprocessing the dataset; ii) optimizing the model using k -fold cross-validation and hyperparameter tuning; iii) comparing the performance of different models using accuracy metrics. The obtained results show that the model built using lasso regression, outperforms the other constructed models.

Keywords: Multiple linear regression · Overfitting · Regularization · Crime prediction

1 Introduction

The simplest case of a single scalar predictive variable x and a single scalar response variable y is known as *simple linear regression*. Extension to multiple predictive variables \mathbf{X} is known as *multiple linear regression*. Multiple regression is a statistical machine learning technique that uses several explanatory variables, also called independent variables to predict the outcome of a response variable, also called dependent variable.

Machine learning tends to have a difficult time discriminating strong correlations from false ones. Consequently, there is a risk of making incorrect inferences about possible correlations between attributes. One problem emerging from this difficult situation is *overfitting*. Understanding it is possible by distinguishing between prediction errors caused by bias versus variance in the model. Bias is when models “consistently learn the same wrong thing” and variance is “the tendency to learn random things irrespective of the real signal” [1]. While bias produces underfitting, variance causes overfitting.

The preference for simplicity in machine learning modelling is closely associated with the concerns about models fitting on noise. Simplicity (i.e. Ockham's razor) works as a safeguard against relying on relationships that might be false correlations discovered by an overfitted and inaccurate model. So, overfitting is a violation of Ockham's razor that states that “entities should not be multiplied beyond necessity” [1]. In machine learning, this is often accepted due to the fact that, given two models with the same training error, the simpler of the two will probably have the smaller test error.

Regularization is one of the techniques to offset or mitigate the risk of overfitting. *Regularization* is a component of a machine learning model that discourages selecting a more complex model with the goal to learn the data patterns and filter out the noise in the data set by introducing additional information to a problem to choose the right solution. This is based on using an additional term for penalizing the error measure function that is thought to bring a higher predictive and generalization powers. According to Occam's razor, regularization attempts to find the simplest model that explains the data.

The problem of finding the optimal set of hyperparameters is identified as hyperparameter optimization. A hyperparameter is a predefined value that controls the performance of our model. It is not derived by the training process, as the usual model parameters, but it is rather preset by the user. Different values of hyperparameters will determine different performances of the model, so hyperparameters can be subject to a numerical optimization associated to the model definition process. For example, in our case, a hyperparameter can control how much we want to regularize the regression model. To select its best value, we must do hyperparameter tuning. Models can have multiple hyperparameters and finding the most appropriate combination can be approached using specialized optimization algorithms. A standard strategy is Grid Search that searches for the best hyperparameter combination from a grid of values.

The aim of this paper is to evaluate performance of existing methods of regularization using cross-validation, by identifying those best suited to prediction and decide when their performance is optimal. We experimentally evaluate the performance of four regularized linear regression models to predict violent crimes per capita. The motivation behind the choice of this problem is based on facts from official statistics. According to [2], crime, violence and vandalism in European Union reached a rather significant percentage. Even though these figures have decreased from 2010 to 2019, safety continues to be an issue, being undoubtedly one of important human rights.

The linear models that were consistently compared are ridge regression, lasso regression and elastic net regression, while the novelty of our approach refers to the addition of support vector regression to this list and we applied all of them on our dataset with the crime rate. This paper introduces the regression methods with regularization, as well as our obtained experimental results on real benchmark data set with an implementation that involves a general-purpose library, thus contributing to a better understanding of the practice behind these techniques.

2 Related Works

There are various and rather well-known methods for regularization. For example, comparisons recently made between ridge regression, lasso regression and elastic net regression for correct detection of measurement bias followed by experimentation on data sets are presented in [3,4].

Since lasso regression has an influence on the extraction of characteristics, many authors have used it to create prediction models. For instance, one recently studied presented in [5] proposed the use of lasso regression to estimate the patient-specific dental arch in cone beam computed tomography (CBCT) images.

Ridge regression is an important machine learning technique that introduces a regularization parameter for the analysis of data suffering from multicollinearity. An improved quantum algorithm for this regularization method was proposed in [6]. The algorithm was then applied on an exponentially large data set. The technique of parallel Hamiltonian simulation to simulate a number of Hermitian matrices in parallel is proposed and used to develop a quantum version of the k -fold cross-validation approach.

Support vector regression (SVR hereafter) is a widely used regression technique that was inspired from support vector classification (SVC hereafter). Both SVR and SVC are based on the kernel trick, which maps data to a higher dimensional space using a (possibly nonlinear) kernel function. A new method for parameter selection for linear SVR was proposed in [7]. An air quality prediction model constructed on the multiple piecewise linear model, using a standard support vector regression (SVR) with a quasi-linear kernel, is proposed in paper [8].

An interesting approach to show the usefulness of the elastic net was proposed in paper [9]. To allow the Brain-Computer Interfaces user to adapt more easily to decoding weights, the features that are only moderately useful was removed to control an additional degree of freedom.

An important comprehensive review of regularization was realized in [10]. The authors of the paper introduced a general conceptual approach to regularization.

Optimization problems are characterized by the presence of one or more objective maximizing or minimizing functions and various restrictions that must be met so that the solution is valid. Over the last decade, optimization was a concern for many authors, who have tried to optimize the performance of their proposed models.

An interesting approach was proposed in [11], where the core focus of the paper involves to create an efficient optimization framework for electrical drive design. The authors are exploiting well known and widely applied genetic algorithms, for the purpose of optimizing the design of electrical drives. For improving the computational time of a multi-objective evolutionary algorithm they approximated the actual function through the surrogate models. Authors presents the most popular methods used for constructing surrogate models, Artificial Neural Networks, because they offer parameterization options that allow for an adequate degree of control over the complexity of the resulting model.

In this paper [12], a new optimization algorithm called Conflict Monitoring Optimization (CMO) is proposed. The algorithm attempts to use models of fear processing and modulation of brain conflicts as inspiration for prioritizing meta-heuristics while adjusting each of the hyper-parameters and condition settings associated.

When linear programming or Lagrange multipliers are not feasible in optimization problems, neural networks such as heuristics are used in these cases. In [14] is presented the solution for these problems, a multilayer perceptron applied to approximate the objective functions.

Reference [13] describes a comparative study providing the purpose to examine the dependence of specific force coefficients. These coefficients are used in mechanistic cutting force models utilizing two methods for determining them: linear regression method and nonlinear optimization method. This paper emphasizes the need for optimization.

A recent approach to optimization is represented by deep neural networks architectures utilizing particle swarm optimization for image classification presented in paper [15]. In this work, authors propose a novel algorithm named psoCNN, capable to automatically search for meaningful deep convolutional neural networks architectures for image classification tasks.

3 System Design

3.1 Problem Description

Prediction using machine learning models allows us to make highly accurate guesses of an unknown output variable based on the values of input variables. Linear regression assumes that there is an independent scalar variable and a dependent variable (actually a vector of scalar variables in the general case of multiple linear regression). The vector of the independent variable represents the factors that are used to compute the dependent variable or outcome.

In this paper we aim to optimize the models that use linear multiple regression with the help of hyperparameter adjustment. This means that we are interested to determine the values of the hyperparameters such that to obtain the best predictions using the data sets from our experiments.

Let us suppose that we are looking for a statistical relationship between a scalar response (or dependent) variable (y) and one or more scalar explanatory (or independent) variables x_1, x_2, \dots, x_p . Dependent variables are also called features in machine learning. A *multiple linear regression model* is written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \quad (1)$$

Let us suppose that we have a data set comprising n observations or examples $(y_i, [x_{i1}, x_{i2}, \dots, x_{ip}])$ for $i = 1, 2, \dots, n$. Then we have:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i \quad (2)$$

for each observation $i = 1, 2, \dots, n$. Here:

- p is the number of features.
- n is the sample size, i.e. the number of examples.
- x_{ij} are the values of independent variables $j = 1, 2, \dots, p$ for each example $i = 1, 2, \dots, n$. They are sometimes called attributes or features.
- $X_i = [1, x_{i1}, x_{i2}, \dots, x_{ip}] \in 1 \times p + 1$ is the row vector of feature values of the i -th example of the sample.

$$- \mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \in n \times p + 1 \text{ is the matrix representing the whole data set.}$$

- $Y = [y_1, y_2, \dots, y_n]^T \in n \times 1$ is the column vector of values of the response variable.
- β_0 is the value of the response when each independent variable is 0. It is also called an intercept. β_j for $j = 1, 2, \dots, p$ are the regression coefficients.
- $\beta = [\beta_0, \beta_1, \dots, \beta_p]^T \in p + 1 \times 1$ is the column vector containing all the coefficients of the regression.
- ϵ_i is the i^{th} random error term that represents the difference between the linear model and a particular observed value. The column vector of error values for each example is written as $\epsilon = [\epsilon_1, \epsilon_2, \dots, \epsilon_n]^T \in n \times 1$.

Note that using the vector notation, linear regression model can be simplified to:

$$Y = \mathbf{X}\beta + \epsilon \quad (3)$$

3.2 Data Set Description and Analysis

We have selected the *Communities and Crime Data Set* [16] for our problem domain from the *The UCI Machine Learning Repository*. This dataset includes 128 attributes and 1994 records and gathers information from different communities in the United States about several factors that can highly influence some common crimes such as murder, rape, robbery, and assault.

All numeric data were normalized into the decimal range 0.00–1.00. Because of some controversy in some states about the number of rapes that led to a lack of values for rape crime, the number of violent crimes per capita might be incorrect. We first removed non-predictive features, aiming to improve the results of the prediction: *state*, *county*, *community*, *fold*.

The second step in analyzing and preprocessing the data set was to check the missing values. Out of 123 predictive characteristics, 23 contained missing values. Because the *OtherPerCap* attribute had only one missing value, we kept it by filling it with the average value. All other features each had 1675 missing values, which were removed from the dataset.

3.3 Methods

Ridge regression was proposed as an improvement of least square method, by tolerating any correlations between the independent variables. The least square

method can achieve its goals only by eliminating redundant data, while ridge regression is a stronger method, being able to tolerate unreasonable data, as well as by shrinkage of samples. Specifically, it is able to perform well with many correlated predictors, as they are poorly determined and have high variance. Ridge regression decreases the correlation coefficients close to zero, but it does not discard them. Ridge regression can be defined as follows:

$$\hat{\beta}(\text{Ridge}) = \arg \min_{\beta} \frac{1}{n} \|Y - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_2^2 \quad (4)$$

where recall that we have n observations and $\|Y - \mathbf{X}\beta\|_2^2 = \sum_{i=1}^n (y_i - X_i\beta)^2$ is the l_2 -norm (quadratic) loss function, $\|\beta\|_2^2 = \sum_{j=0}^p \beta_j^2$ is the l_2 - norm penalty on β and $\lambda \geq 0$ is called the tuning/ penalty/ regularization coefficient.

Lasso regression uses an l_1 , rather than l_2 norm, for the regularization term. Lasso improves both the prediction accuracy and the model interpretability by combining the good qualities of the ridge regression and subset selection. If there is a large correlation in the group of predictors then Lasso regression chooses only one of them and it restricts the others to zero. Consequently, it reduces the variability of estimates by decreasing some of the coefficients exactly at zero, thus producing easy-to-interpret models. Lasso is a feature selection approach based on a linear regression model with l_1 regularization:

$$\hat{\beta}(\text{Lasso}) = \arg \min_{\beta} \frac{1}{n} \|Y - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \quad (5)$$

where $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$ is the l_1 - norm penalty on β and $\lambda \geq 0$ is the regularization coefficient.

Elastic Net regression was designed to avoid the imbalance of the Lasso or the Ridge solution paths when predictors are highly correlated. This method is meant to enjoy the computational advantages of both Ridge and Lasso regression. The Elastic net uses a mixture of the l_1 – Lasso and l_2 – Ridge penalties and it can be defined as follows:

$$\hat{\beta}(\text{ElasticNet}) = \arg \min_{\beta} \frac{1}{2n} \|Y - \mathbf{X}\beta\|_2^2 + \alpha\rho\|\beta\|_1 + \frac{\alpha(1-\rho)}{2} \|\beta\|_2^2 \quad (6)$$

where parameter α controls the penalty factor and parameter $\rho \in [0, 1]$ controls the convex combination of l_1 and l_2 . Part of l_1 norm of the Elastic net performs self-regulating predictor selection, while part l_2 encourages group selection and balances solution paths in terms of random sampling, thus improving prediction Note that:

- for $\rho = 0$ the penalty function becomes an l_2 penalty as in Ridge regression.
- for $\rho = 1$ the penalty function becomes an l_1 penalty as in Lasso regression.
- for $0 < \rho < 1$ the penalty function is a convex combination of l_1 and l_2 norms.

Some texts define Elastic net as follows:

$$\hat{\beta}(\text{ElasticNet}) = \arg \min_{\beta} \|Y - \mathbf{X}\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \quad (7)$$

and we clearly have the correspondence: $\lambda_1 = 2n\alpha\rho$ and $\lambda_2 = n\alpha(1 - \rho)$.

Linear SVR. To avoid overfitting of the training data in linear support vector regression (SVR), regularization parameter (often referred to as C) and error sensitivity parameter (often referred to as ϵ) are used. Proper selection of parameters is very essential to get a good model, but the search process can be complicated and time consuming.

$$\hat{\beta}(\text{LinearSVR}) = \arg \min_{\beta} f(\beta, C, \epsilon) \quad (8)$$

where:

$$f(\beta, C, \epsilon) \equiv \frac{1}{2} \|\beta\|^2 + C \cdot L(\beta, \epsilon) \quad (9)$$

According to (9), $\|\beta\|^2/2$ is the penalty term and $L(\beta, \epsilon)$ is the sum of training losses defined as:

$$L(\beta, \epsilon) = \begin{cases} \sum_{i=1}^n \max(|X_i\beta - y_i| - \epsilon, 0) & \text{Case of } l_1 \text{ loss,} \\ \sum_{i=1}^n \max(|X_i\beta - y_i| - \epsilon, 0)^2 & \text{Case of } l_2 \text{ loss.} \end{cases} \quad (10)$$

where $C > 0$ is the regularization parameter.

Note that coefficients λ , α and C play somehow similar roles in Eqs. (4), (5), (6) and (9). Their only constraint is to be positive real values. On the other hand, ρ controls the involvement of the l_1 term and l_2 term Eq. (6) and it is constrained to be in interval $[0, 1]$.

4 Results and Discussions

4.1 Description of Environment

We have developed our experiments in Python v3.7.4 language using at the core the *scikit-learn* package. Nevertheless, more libraries have been used to perform all the experimental tasks: *numpy*, *pandas*, *scikit-learn*, and *matplotlib*. The types of regression were developed using the *Ridge()*, *Lasso()*, *ElasticNet()* and *linearSVC()* classes representing linear models in *scikit-learn* and using methods presented Sect. 3.

The Python package *pandas* have been used to automate the process of data inspection. It was also used to read the data from *CSV* files into a specialized data structure called *data frame*. The data set split ratio is an important factor used by cross-validation that helps to execute the predictive model with the best use of the data set. Partitioning is done by using the *train_test_split* method of *model_selection* library of *scikit-learn*. The data set splitting was done with 70% for training and 30% for testing.

All regression methods that we implemented used regularization to prevent overfitting. Cross-validation was used to accurately estimate how well do our trained models perform on test data. Then we used the measures computed using cross-validation to choose the best value of the hyperparameters. We have used 10-fold cross-validation in our experiments. For the implementation we have used the *KFold* method of *model_selection* library of *scikit-learn*. Basically, for each configuration of the hyperparameters, we trained the model on the training set, then we tested the learnt model on the test set, and this validation procedure was performed on every fold such that the resulted validation metric is aggregated across each validation step. While designing the model, we have evaluated the impact of multiple values of hyperparameters of the models (λ , α , ρ and C) for controlling the regularization fit on our training set. We have used *GridSearchCV*, a member function of *scikit-learn*'s *model_selection* package. For the evaluation, we have used the *R-squared metric*, fitting the model and getting the best value of the regularization parameters of each model.

4.2 Testing the Performance and Discussion of Results

Model training was performed using the proposed linear regression methods. The parameters controlling the regularization play a vital role. The considered values of the (λ , α , C) parameters, playing similar roles in all models, were $\{0.001, 0.01, 0.1, 1.0, 10.0\}$. The considered values of ρ were $\{0.1, 0.2, \dots, 0.9\}$.

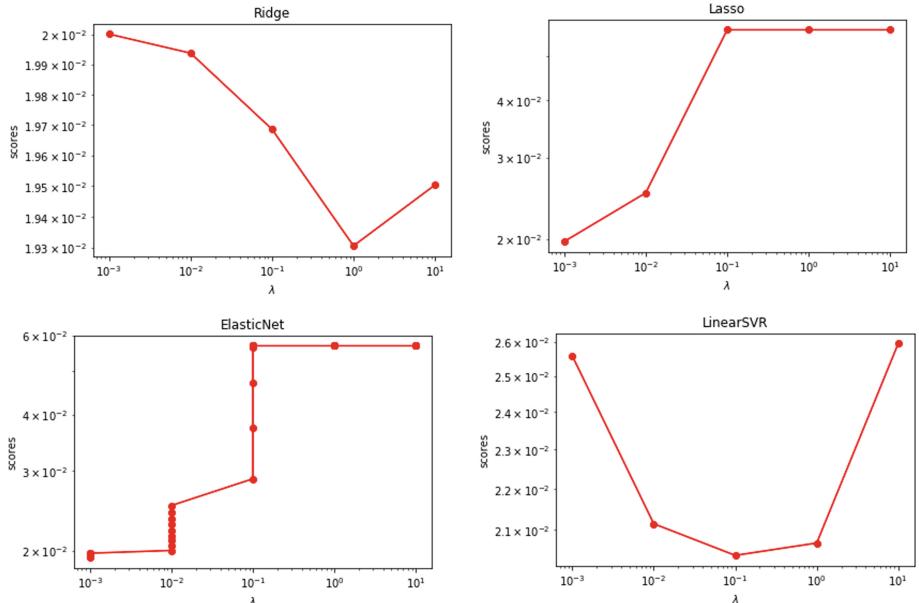


Fig. 1. The mean cross-validation error at different λ

The best values obtained for the controlling parameters are as follows: i) Ridge: $\lambda = 1$; ii) Lasso: $\lambda = 0.001$; iii) Elastic net: $\alpha = 0.001$ and $\rho = 0.1$; and iv) Linear SVR: $C = 0.1$.

After completing the grid search, we plotted different mean errors obtained for 10-fold cross-validation, for different values of λ (see Fig. 1).

Furthermore, we present the effect of regularization by plotting regression coefficients vs the regularization penalty. The output of these plots using our chosen values for the regularization parameters is presented in Fig. 2.

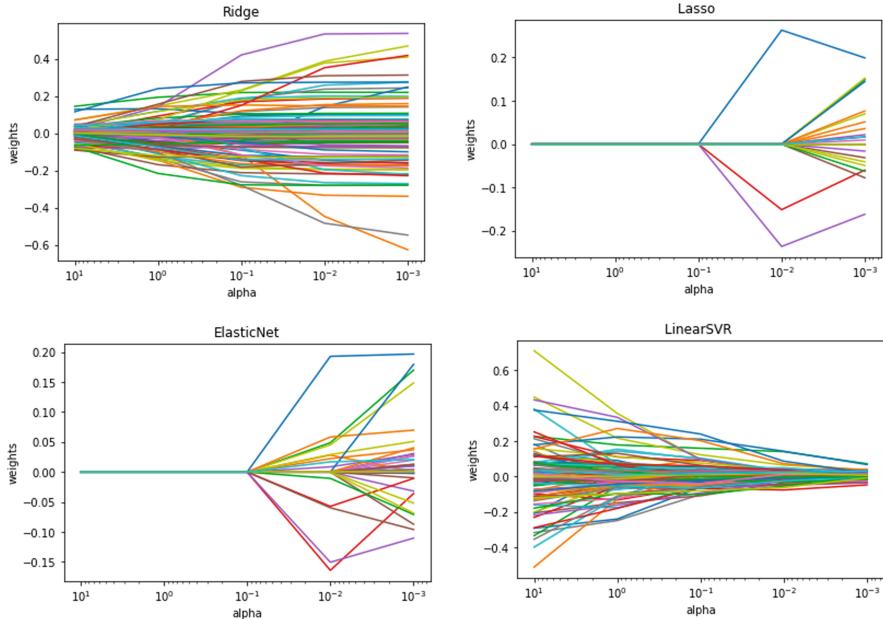


Fig. 2. Coefficients weights as a function of regularization parameters

The final testing of the regularized linear regression models was done on the basis of validation and making new predictions based on the data set divided by 30% (data set testing) using the chosen values of the regularization parameter for each model. The *sklearn.metrics module* was used to evaluate the performance measures. The metrics for evaluating the overall quality of regression models are:

1. *R-squared* (R^2). Measures the squared correlation between the actual values and the values predicted by the model. The larger the adjusted R^2 , the better is the model.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

where \hat{y}_i is the predicted outcome, y_i is the observed outcome and \bar{y} is the average of observed outcomes

2. *Root Mean Squared Error (RMSE)*. Measures the average magnitude error made by the model while predicting the result of an observation. It is the square root of the average square residue. Residues are the difference between actual and predicted values. The smaller the RMSE, the better is the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{N}} \quad (12)$$

3. *Mean Absolute Error (MAE)*. It measures average absolute differences between observed and predicted outcomes. The lower the MAE, the better is the model.

$$MAE = \frac{1}{N} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

After testing the performance of our models, we have obtained for each model the values of the metrics RMSE, R-square and MAE. These values are presented in Table 1.

Table 1. Metrics performance matrix

	Training dataset			Testing dataset		
	R^2	RMSE	MAE	R^2	RMSE	MAE
Ridge	0.697	0.131	0.092	0.633	0.132	0.095
Lasso	0.664	0.138	0.095	0.645	0.130	0.092
ElasticNet	0.670	0.136	0.095	0.641	0.131	0.093
LinearSVR	0.666	0.137	0.088	0.648	0.129	0.088

Ridge and LinearSVR regression does not eliminate any feature. It is interesting to note that the values of the model coefficients have been greatly reduced for insignificant features. Lasso has picked 23 features and Elastic net regression has picked 37 features and eliminated the other features for the model development. From these experiments it is quite clear that the problem of multicollinearity has been removed by introducing regularization techniques. Moreover, the Elastic net does not make zero as many coefficients as Lasso does. Therefore, the insight of the key features in our data set was best highlighted by Lasso that made the model smart enough to consider approx. 25% of the features for predicting per capita violent crimes with the highest accuracy.

5 Conclusions and Future Works

The aim of our work was to experimentally evaluate and compare the results obtained by optimizing regularized multiple linear regression using hyperparameter tuning for predicting per capita violent crimes. Our methodology consisted

in the application of different methods for selecting the most important variables and then using the selected variables to build different multiple linear regression models. Regression techniques that we evaluated were: ridge, lasso and elastic net to which we added support vector regression (SVR).

According to Ockham's razor, the simpler model between two models will probably have the smaller test error. After comparing the performances of the built models, we have found that the most performing was Lasso regression. It is interesting to note that Lasso reduces the coefficient of the less important feature to zero, thus eliminating some features altogether. So this works well for feature selection because we have a large number of features, being consistent with Ockham's razor principle.

As future work we would like to expand the experiment of multivariate regression involving multiple predictors and to compare the performance of the regression methods evaluated in this paper with other regression methods.

References

1. Domingos, P.: A few useful things to know about machine learning. *Commun. ACM* 78–87 (2012). <https://doi.org/10.1145/2347736.2347755>
2. Eurostat - Crime, violence or vandalism in the area - EU-SILC survey, Website https://ec.europa.eu/eurostat/en/web/products-datasets/-/ILC_MDDW03
3. Harimurti, R., Yamasari, Y., Ekohariadi, M., Asto, B.I.G.P.: Predicting student's psychomotor domain on the vocational senior high school using linear regression. In: International Conference on Information and Communications Technology (ICOIACT), pp. 448–453. Yogyakarta (2018). <https://doi.org/10.1109/ICOIACT.2018.8350768>
4. Liang, X., Jacobucci, R.: Regularized structural equation modeling to detect measurement bias: evaluation of lasso, adaptive lasso, and elastic net. *Struct. Eq. Modeling: Multidisc. J.* **27**(5), 722—734 (2019). <https://doi.org/10.1080/10705511.2019.1693273>
5. Ghozatlou, O., Zoroofi, R.A.: Patient-specific dental arch estimation via LASSO regression analysis in CBCT images. In: 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME), pp. 124–128. Tehran, Iran (2019). <https://doi.org/10.1109/ICBME49163.2019.9030426>
6. Yu, C., Gao, F., Wen, Q.: An improved quantum algorithm for ridge regression. *IEEE Trans. Knowl. Data Eng.* <https://doi.org/10.1109/TKDE.2019.2937491>
7. Hsia, J.-Y., Lin, C.-J.: Parameter selection for linear support vector regression. *IEEE Trans. Neural Netw. Learn. Syst.* <https://doi.org/10.1109/TNNLS.2020.2967637>
8. Zhu, H., Hu, J.: Air quality forecasting using SVR with quasi-linear kernel. In: 2019 International Conference on Computer Information and Telecommunication Systems (CITS), pp. 1–5, Beijing, China (2019). <https://doi.org/10.1109/CITS.2019.8862114>
9. Kelly, J.W., Degenhart, A.D., Siewiorek, D.P., Smailagic, A., Wang, W.: Sparse linear regression with elastic net regularization for brain-computer interfaces. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4275–4278, San Diego, CA (2012). <https://doi.org/10.1109/EMBC.2012.6346911>

10. Bickel, P.J., Li, B.: Regularization in statistics. In: *Test*, vol. 15, no. 2, pp. 271–344 (2006). <https://doi.org/10.1007/BF02607055>
11. Zăvoianu, A.-C., Bramerdorfer, G., Lughofer, E., Silber, S., Amrhein, W., Klement, E.P.: Hybridization of multi-objective evolutionary algorithms and artificial neural networks for optimizing the performance of electrical drives. *Eng. Appl. Artif. Intell.* **26**, 1781–1794 (2013). <https://doi.org/10.1016/j.engappai.2013.06.002>
12. Moattari, M., Moradi, M.H.: Conflict monitoring optimization heuristic inspired by brain fear and conflict systems. *Int. J. Artif. Intell.* **18**(1), 45–62 (2020)
13. Rubeo, M.A., Schmitz, T.L.: Mechanistic force model coefficients: a comparison of linear regression and nonlinear optimization. *Precis. Eng.* **45** (2016). <https://doi.org/10.1016/j.precisioneng.2016.03.008>
14. Villarrubia, G., De Paz, J.F., Chamoso, P., De la Prieta, F.: Artificial neural networks used in optimization problems. *Neurocomputing* **272**, 10–16 (2018)
15. Junior, F.E.F., Yen, G.G.: Particle swarm optimization of deep neural networks architectures for image classification. *Swarm Evol. Comput.* **49**, 62–74 (2019). <https://doi.org/10.1016/j.swevo.2019.05.010>
16. UCI Machine Learning Repository. <https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime>



Modelling a Deep Learning Framework for Recognition of Human Actions on Video

Flávio Santos¹ , Dalila Durães¹ (✉) , Francisco Marcondes¹ , Marco Gomes¹ , Filipe Gonçalves² , Joaquim Fonseca² , Jochen Wingbermuehle², José Machado¹ , and Paulo Novais¹

¹ Centre Algoritmi, University of Minho, 4710-057 Braga, Portugal
{flavio.santos,dalila.duraes,
francisco.marcondes}@algoritmi.uminho.pt, {marcogomes,jmac,
pjon}@di.uminho.pt
² Bosch Car Multimedia, 4705-820 Braga, Portugal
{filipe.goncalves,joaquim.fonseca2,
jochen.wingbermuehle}@pt.bosch.com

Abstract. In Human action recognition, the identification of actions is a system that can detect human activities. The types of human activity are classified into four different categories, depending on the complexity of the steps and the number of body parts involved in the action, namely gestures, actions, interactions, and activities [1]. It is challenging for video Human action recognition to capture useful and discriminative features because of the human body's variations. To obtain Intelligent Solutions for action recognition, it is necessary to training models to recognize which action is performed by a person. This paper conducted an experience on Human action recognition compare several deep learning models with a small dataset. The main goal is to obtain the same or better results than the literature, which apply a bigger dataset with the necessity of high-performance hardware. Our analysis provides a roadmap to reach the training, classification, and validation of each model.

Keywords: Action recognition · Deep learning models · Video intelligent solutions

1 Introduction

Intelligent solutions of action recognition have been studied, with different perspectives, for several disciplines, including psychology, biomechanics, and computer vision [1, 2]. However, in recent years there has been a rapid growth in production and consumption of a wide variety of video data due to the popularization of high quality and relatively low-price video devices [3]. Smartphones and digital cameras contributed a lot to this factor. Simultaneously, on YouTube, there are about 300 h of video data updates every minute [4]. New technologies such as video captioning, answering video surveys, and video-based activity/event detection are emerging every day along with the growing

production of video data [5, 6]. From the video input data, human activity detection indicates which activity is contained in the video and locates the regions in the video where the action occurs [7]. Also, from the computer vision community point of view, we can use visual tracking for the process of locating, identifying, and determining the dynamic configuration of one or many moving (possibly deformable), objects (or parts) in each frame of one or several cameras [8].

This paper conducted an experience of action recognition, comparing several deep learning models and obtaining better results. Our analysis provides a roadmap to reach the training, classification, and validation of each model with a dataset with a fewer class. The organization of this paper was: firstly, Sect. 2 introduces the concepts with state of the art, namely models and dataset; then, Sect. 3 presents materials and methods, with training data and validation data; next, Sect. 4 presents result and discussion; and finally, Sect. 5 concludes by performing a global conclusion and some future work.

2 State of Art

Human action recognition used several deep learning models [3, 4, 9]. However, our goal is to develop models that cover a multisensory integration process. In this stage, we will focus on optimizing the video signal learning process and afterwards expanding the architectures for efficient human action recognition by applying audiovisual information. The reason for choosing this path is twofold the different “learning dynamics” between the visual and audio information – audio generally train much faster than visual ones, which can lead to generalization issues during joint audiovisual training [9]. There are several architectures for human action recognition [3, 4]. However, the most used are C2D-Resnet 50, SlowFast, and I3D [8]. Furthermore, these architectures allow them to be combined with audio so, we will focus on these architectures.

2.1 C2D – Resnet 50

C2D is a standard 2D convolution network. A convolution network is a neural network that uses convolution in place of a fully connected matrix multiplication in at least one layer [10].

The Residual Network (ResNet) was conceived to explore a neural network depth [11, 12]. It aims to handle the vanishing/exploding gradient problem that worsens according to the number of the layers raises because of a network difficulty on learning identity functions [13]. The numeral 50 denotes the network depth, i.e., the number of layers.

In short, the ResNet aims to handle the gradient descent problem caused by identity function by skipping the layers expected to compute these functions [11, 14]. Notice that ResNet cannot be directly applied to C3D. This is because the search for temporal data significantly increases the resources consumption.

2.2 SlowFast Network

The generic architecture of a SlowFast network can be described as a single stream architecture that operates at two different temporal rates (Slow pathway and Fast pathway), which are fused by lateral connections. The underlying idea is to model two tracks

separately, working at low and high temporal resolutions. One is designed to capture fast-changing motion but fewer spatial details (fast pathway) and the other as a lightweight version more focused on the spatial domain and semantics (slow path) [15].

As presented in [15], the fast pathway data is fed into the slow pathway via lateral connections throughout the network, allowing the slow path to become aware of the fast pathway results. To do it, it requires a match to the sizes of features before fusing. At the end of each pathway, SlowFast performs global average pooling, a standard operation intended to reduce dimensionality. It then concatenates the results of the two tracks. It inserts the concatenated result into a fully connected classification layer, which uses Softmax to classify which action is taking place in the image [15].

2.3 Inflated 3D ConvNet (I3D)

By adding one dimension into a C2D (e.g. $k \times k$) it becomes a C3D (e.g. $t \times k \times k$) [16]. Inflating is not a plain C3D but a C2D, often pre-trained, whose kernels are extended into a 3D shape. Growing is as simple as including an additional, usually temporal, dimension [12]. The I3D stands for two-stream inflated 3D convolution network [16]. Therefore, I3D is a composition of an inflated C2D with optical flow information [12, 16].

2.4 Dataset

Kinetics 700 Dataset

The Kinetics dataset is a project that provides a large scale of video clips for human action classification, covering a varied range of human actions. This dataset contains real-world applications with video clips having a duration of around 10 s. The dataset's primary goal is to represent a diverse range of human actions, which can be used for human action classification and temporal localization. Another characteristic is that clips also contain audio so that the dataset can be used for multi-modal analysis. The fourth version, created in 2019, was the Kinetics-700 dataset with 700 classes, each with 700 video clips [17].

This dataset is essentially focused on human actions, where the list of action classes includes three types of actions: person actions, person-person actions, and person-object actions. The person-actions are a singular human action and include drawing, drinking, laughing, and pumping fist. The person-person actions cover human actions like kissing, hugging, and shaking hands. Finally, the person-object actions contain actions like opening a present and washing dishes. Furthermore, some actions required more emphasis on the object to be distinguished, such as playing different wind instruments. Other actions required temporal reasoning to distinguish, for example, different types of swimming [17].

AVA Kinetics Dataset

The AVA Kinetics dataset creates a crossover of the two datasets. The AVA-Kinetics dataset builds upon the AVA and Kinetics-700 datasets by providing AVA-style human action and localization annotations for many of the Kinetics videos.

The AVA-Kinetics dataset extends the Kinetics dataset with AVA style bounding boxes and atomic actions. A single frame is annotated for each Kinetics video, using a frame selection procedure described below. The AVA annotation process is applied to a subset of the training data and all video clips in the validation and testing sets from the Kinetics-700 dataset. The procedure to annotate bounding boxes for each Kinetics video clip was as follows: person detection, key-frame selection, missing box annotation, human action annotation, and human action verification [18].

3 Materials and Methods

As mentioned in Sect. 1, the idea was to classify activities in video. The first step was to download the AVA-Kinetics datasets and cross between AVA Actions and Kinetics datasets. Downloading files from YouTube was relatively slow since the program itself blocks excess downloads. During the download IP some problems have occurred like “this video is no longer available because the YouTube account associated with this video has been terminated”, the owner of this video has granted you access, please sign in: “This video is private”, and this video is no longer available because the uploader has closed their YouTube account. On the second step, we evaluate the top-60 most frequent classes, and our dataset has 283 videos of the 430 videos from AVA v2.2 and 100 classes from Kinetics-700 datasets, where each class has between 650 and 1000 videos.

The annotation format presented was the `video_id`, `middle_frame_timestamp`, `person_box`, `action_id`, and `score`. The `video_id` is a YouTube identifier. The `middle_frame_timestamp` is measured in seconds from the start of the video. The `person_box` is normalized at upper left (x_1, y_1) and lower right (x_2, y_2) about the frame size, where (0.0, 0.0) corresponds to the upper left corner and (1.0, 1.0) corresponds the bottom right corner. The `action_id` is a whole identifier of an action class, from `ava_action_list_v2.2_for_activitynet_2019.pbtxt`. Moreover, finally, the `score` is a float indicating the score for that labelled box.

3.1 Architectures Networks

C2D – Resnet 50

Initially, we began training with the C2D-ResNet 50 architecture. All characteristic of this architecture is presented in Table 1.

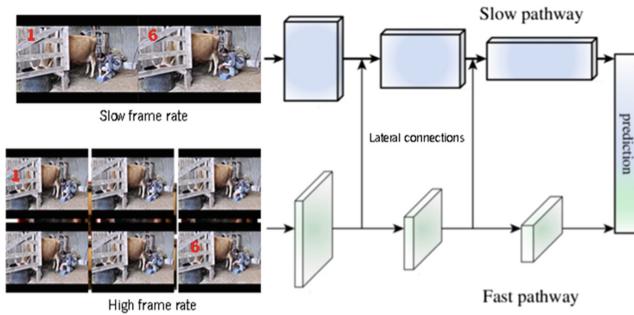
The batch size was 12, LR: 0.1, the optimizer: SGD with 85 epochs and Cross-Entropy loss for this architecture.

Table 1. Global average pool of the architecture C2D-ResNet 50.

	layer	output size
conv1	$7 \times 7, 64, \text{stride } 2, 2, 2$	$16 \times 112 \times 112$
pool1	$3 \times 3 \times 3 \text{ max, stride } 2, 2, 2$	$8 \times 56 \times 56$
res2	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$8 \times 56 \times 56$
pool2	$3 \times 1 \times 1 \text{ max, stride } 2, 1, 1$	$4 \times 56 \times 56$
res3	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$4 \times 28 \times 28$
res4	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$4 \times 14 \times 14$
res5	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$4 \times 7 \times 7$
	global average pool, fc	$1 \times 1 \times 1$

SlowFast

Figure 1 presents a slowfast network’s generic architecture, which can be described as a single stream architecture that operates at two different temporal rates (Slow pathway and Fast pathway), which are fused by lateral connections.

**Fig. 1.** A slowfast network.

Inflated 3D ConvNet (I3D)

Figure 2 presents the approach for I3D architecture. This approach begins with a 2D architecture and inflates all the filters and pooling kernels, adding a dimension layer (time).

3.2 Training Data

Hence, we began the training only with the classes we had a download. However, for these 100 classes, we did not have the 650–1000 videos for each class. Because some videos are no longer available, or the owner has changed the video to private, or the video is no longer available on YouTube. Thus, the following data visualization shows the



Fig. 2. Non-local action recognition example.

difference between the downloaded videos and full dataset. Table 2 compares the total of videos for the complete training dataset and the video download training dataset.

Table 2. Comparison between the videos of the complete training dataset and the download training dataset.

	Completed	Download
q1	510.5	454
q3	888.5	812.5
Max	997	972
Min	393	127
Median	683	602

3.3 Validation Data

Regarding the validation of the videos, the script had a full validation of 50 videos. Table 3 present a comparison for the total of videos of complete validation dataset and the videos download validation dataset.

Table 3. Comparison between the videos for the complete validation dataset and the download validation dataset.

	Completed	Download
q1	48	46
q3	50	48
Max	50	50
Min	44	40
Median	49	47

Remember that the accuracy is obtained with the number of correct predictions, based on the total number of predictions.

4 Results and Discussion

This section presents the results and discuss the data presented in Sect. 3, based on state-of-art, illustrated in Sect. 2.

As we can see, Fig. 3 show the training data loss for epoch in the three different architectures. In this case, we can observe that SlowFast and I3D present the worst results, and the best results were obtained for the C2D-Resnet50 architecture.

Figure 4 and Fig. 5 present the training data evaluation Top1 and Top5, respectively. The best accuracy for epoch was obtained for C2D-Resnet50 architecture in Top1 and Top5.

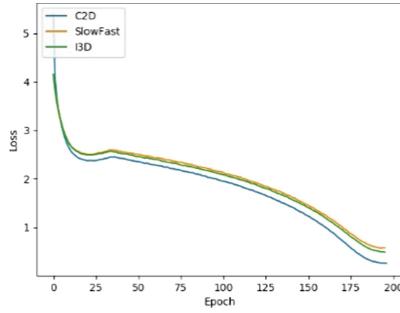


Fig. 3. Training data Loss for epoch.

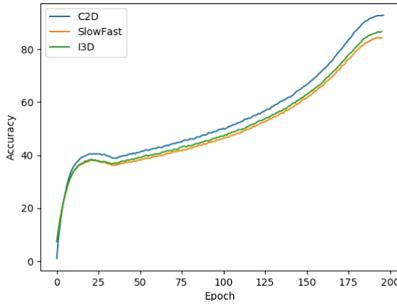


Fig. 4. Training data evaluation Top1.

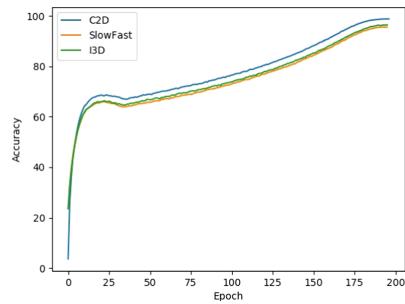
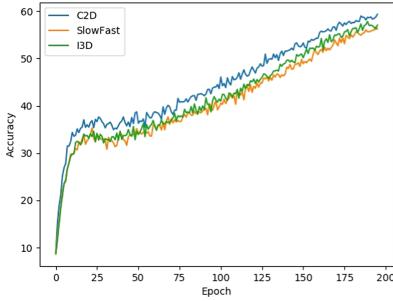
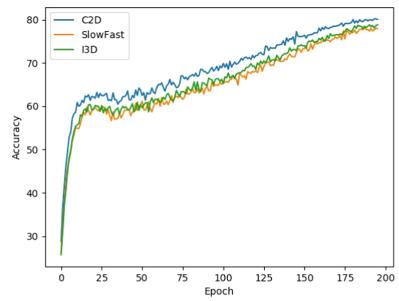


Fig. 5. Training data evaluation Top5.

For the validation data Top 1 and Top5, the results are presented in Figs. 6 and 7, respectively. Also, the best accuracy for epoch was obtained for C2D-Resnet50 architecture in Top1 and Top5.

Table 4 presents a comparison of results for top1 and top 5 training and validation. It is possible to observe that C2D-Resnet50 architecture has better results compared with the other architecture.

Table 4 shows that the training Top1 for C2D - Resnet 50 architecture was the best accuracy of 92.79 versus 86.70 for I3D versus 84.33 for SlowFast. In the case of training Top5, C2D-Resnet50 architecture was also the best accuracy of 98.82 compared with

**Fig. 6.** Validation data – evaluation Top1.**Fig. 7.** Validation data – evaluation Top5.**Table 4.** Comparison of results for top 1 and top 5 training and validation.

Architecture	Train – Top 1	Train – Top 5	Val – Top 1	Val Top 5
C2D-ResNet 50	92.79	98.82	59.36	80.19
SlowFast	84.33	95.64	56.69	78.15
I3D	86.70	96.47	57.83	78.84

94.47 for I3D, and 95.64 for SlowFast. Furthermore, the validation Top1 the C2D – Resnet 50 architecture was the best accuracy with 59.36 compared with 57.83 of I3D, and 56.69 of SlowFast. Finally, validation Top5 the C2D-Resnet 50 architecture was also the best accuracy 80.19, compared with 78.84 of I3D, and 78.15 of SlowFast.

5 Conclusions and Future Work

This paper conducted an experience on video Human action recognition, which is to compare several deep learning models and obtain better results with a fewer class dataset. We began to compare three architectures C2D-Resnet 50, SlowFast, and I3D with the same baseline parameters after downloading the dataset. Comparing the results of training and validation, we can observe that C2D-Resnet 50 obtained better accuracy results for the three architectures. Our experiment results are consistent with the present in the literature, and we used a small dataset.

We intend to extend these architectures to work with the synchronized audio information to achieve better results in the next steps. Moreover, we intend to introduce Attention Models to learn which frames are most important in the classification process. Another future intends it is to apply the late fusion for the audio and video models.

Acknowledgement. This work is supported by: European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project n° 039334; Funding Reference: POCI-01-0247-FEDER-039334].

References

1. Ko, T.: A survey on behavior analysis in video surveillance for homeland security applications. In: 37th IEEE Applied Imagery Pattern Recognition Workshop, pp. 1–8. IEEE (2008)
2. Analide, C., Novais, P., Machado, J., Neves, J.: Quality of knowledge in virtual entities. In: Encyclopedia of Communities of Practice in Information and Knowledge Management, pp. 436–442. IGI Global (2006)
3. Durães, D., Marcondes, F.S., Gonçalves, F., Fonseca, J., Machado, J., Novais, P.: Detection violent behaviors: a survey. In: International Symposium on Ambient Intelligence, pp. 106–116. Springer, Cham (2020)
4. Marcondes, F.S., Durães, D., Gonçalves, F., Fonseca, J., Machado, J., Novais, P.: In-vehicle violence detection in carpooling: a brief survey towards a general surveillance system. In: International Symposium on Distributed Computing and Artificial Intelligence, pp. 211–220. Springer, Cham (2020)
5. Durães, D., Carneiro, D., Jiménez, A., Novais, P.: Characterizing attentive behavior in intelligent environments. Neurocomputing **272**, 46–54 (2018)
6. Costa, R., Neves, J., Novais, P., Machado, J., Lima, L., Alberto, C.: Intelligent mixed reality for the creation of ambient assisted living. In: Portuguese Conference on Artificial Intelligence, pp. 323–331. Springer, Heidelberg (2007)
7. Zhu, Y., Zhao, X., Fu, Y., Liu, Y.: Sparse coding on local spatial-temporal volumes for human action recognition. In: Asian Conference on Computer Vision, pp. 660–671. Springer, Heidelberg (2010)
8. Jesus, T., Duarte, J., Ferreira, D., Durães, D., Marcondes, F., Santos, F., Machado, J.: Review of trends in automatic human activity recognition using synthetic audio-visual data. In: International Conference on Intelligent Data Engineering and Automated Learning, pp. 549–560. Springer, Cham (2020)
9. Shokri, M., Harati, A., Taba, K.: Salient object detection in video using deep non-local neural networks. J. Vis. Commun. Image Represent. **68**, 102769 (2020)
10. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceeding of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016).
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
13. Hochreiter, S., Bengio, Y., Fransconi, P., Schmidhuber, J.: Gradient flow in recurrent nets: the difficulty of learning long-terms dependencies (2001)
14. Huang, G., Yu, S., Zhuang, L., Daniel, S., Killian, Q.W.: Deep networks with stochastic depth. In: European Conference on Computer Vision, pp. 646–661. Springer Cham (2016)
15. Feichtenhofer, C., Fan, H., Malik, J., He, K.: Slowfast networks for video recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 6202–6211 (2019)
16. Carreira, J., Andrew, Z.: Quo vadis, action recognition? A new model and the kinetics dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6299–6308 (2017)
17. Carreira, J., Noland, E., Hillier, C., Zisserman, A.: A short note on the Kinetics-700 human action dataset. arXiv, vol. preprint, no. 1907.06987 (2019)
18. Li, A., Thotakuri, M., Ross, D.A., Carreira, J., Vostrikov, A., Zisserman, A.: The AVA-kinetics localized human actions video dataset. arXiv preprint 2005.00214 (2020)



Torque Control of a Robotic Manipulator Joint Using LQG and LMI-Based Strategies with LTR

José N. N. Júnior^(✉), Gabriel F. Machado, Darielson A. Souza,
Josias G. Batista, Ismael S. Bezerra, Antônio B. S. Júnior,
Fabrício G. Nogueira, and Bismark C. Torrico

Department of Electrical Engineering, Federal University of Ceará,
Fortaleza, Ceará, Brazil
{juniornogueira,gabrielfreitas}@alu.ufc.br
{darielson,josiasgb,barbosa,fnogueira,bismark}@dee.ufc.br

Abstract. This paper presents two control methodologies to obtain a robust performance of a robot manipulator. A dynamic model of the manipulator driven by three-phase induction motors is formulated. A torque control of one of the joints is presented. Torque control is very important, because you can determine the critical load that can be carried by the manipulator. Furthermore, using the inverse dynamics model it is possible to determine the positions and speeds of the manipulator joint. In this work, robust control techniques were implemented, such as Linear Quadratic Gaussian (LQG) and Linear Matrix Inequalities (LMI), and these two approaches are compared in performance. In addition, a Loop Transfer Recovery (LTR) procedure is used to achieve robustness to the uncertainties in the state estimation.

Keywords: Robotic manipulator · Linear Quadratic Gaussian · Linear matrix inequalities · Loop transfer recovery

1 Introduction

In the last decades, the necessity of optimisation and robustness in complex manufacture systems contributed for the advance of different control strategies. Various alternatives for the classical control techniques such as PI/PID (Proportional-Integral-Derivative) were developed in order to achieve better performance and guarantee robust stability of the closed-loop control systems. These modern control strategies are commonly applied in the state-space and in the frequency domain, and they are not limited to mono-variable systems [1, 2].

Industrial robots designed to manipulate manufactured products sometimes are limited in performance because of not-programmed situations. In general, it is always desirable precision and efficiency in a production line, but with inevitable environment limitations and modelling errors sometimes a robust performance

is acceptable for applications. Thus, in large-scale multi-variable or in complex dynamics mono-variable plants, to provide the production requirements, optimal and robust control design techniques can be implemented [3,4].

A classical solution in the field of optimal control theory is the Linear Quadratic Gaussian (LQG). It is a combination of the well-known Linear Quadratic Regulator (LQR) and the Kalman Filter (KF) estimator and can be applied in linear time-invariant or linear time-varying systems. However, because of its limitations due to uncertainties in the state estimation it is necessary to guarantee the robustness of the closed-loop system by a Loop Transfer Recovery (LTR) procedure [5,6,8].

Other methods of control design are based in Linear Matrix Inequalities (LMIs), where the performance and robustness objectives can be treated in different matrix inequalities and the control parameters are found minimising a criterion. Also in this case, when considering a state estimator, it is necessary to provide robustness to uncertainties so, as in the case of the LQG control, it is possible to use the LTR strategy [7,9].

This paper aims to design a torque control of a rotating joint of a robotic manipulator driven by a three-phase induction motor using LQG with LTR and LMI with LTR control strategies. It is presented the characteristics of the manipulator robot as the design procedure of the controllers. Results are compared and the performance evaluation criterion of each controller used is presented. This research has as main contribution the implementation of LQG with LTR and LMI with LTR controllers applied to a joint of a robotic manipulator.

2 Robotic Manipulator

A cylindrical robotic manipulator that is driven by three phase induction motors has been studied in this work. As can be seen in Fig. 1 the first joint moves around the main axis of the structure (rotational motion), the second and third joints have linear (prismatic) movements, which defines the manipulator as a RPP (Rotational-Prismatic-Prismatic).

2.1 Forward Kinematics

The kinematics exposes the relative motion of the reference systems, as the structure moves by relating reference systems to the various portions of the structure [10,11].

Any position of the end-effector can be found in the Cartesian space from the coordinates in the joint space, as noted in Eq. (1):

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} -\sin(\theta_1)(d_3 + 0.35) \\ \cos(\theta_1)(d_3 + 0.35) \\ 0.245 + d_2 \end{bmatrix}. \quad (1)$$

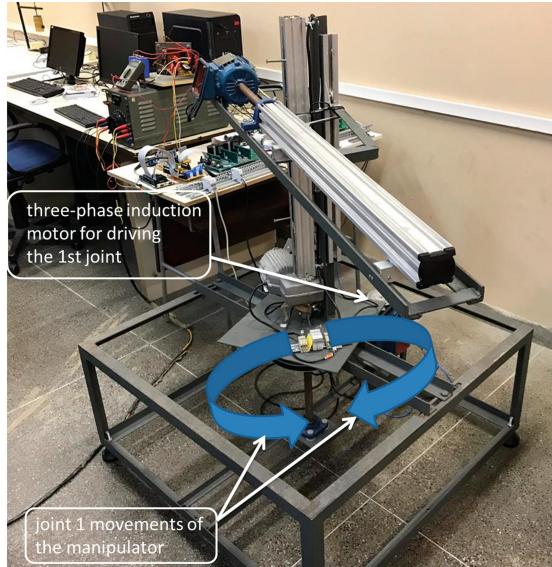


Fig. 1. Setup of the studied cylindrical manipulator.

2.2 Dynamic Modelling

The dynamics of the manipulator displays the position-speed-acceleration-torque relationship of the joints. Therefore, the dynamic modelling of an industrial robot aims to know the relationship between the movement of the robot and the forces applied to it [10, 12, 13].

Thus, considering the kinetic energy of the manipulator its dynamic equation can be written in a simplified formulation:

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = \tau, \quad (2)$$

where, $q, \dot{q}, \ddot{q} \in \mathbb{R}^n$ indicate the joint's positions, speeds and accelerations, respectively; $M(q)$ is the inertial matrix; $C \in \mathbb{R}^n$ is the matrix that describes the centripetal and Coriolis forces and $G \frac{\partial g}{\partial q} \in \mathbb{R}^n$ is the gravity matrix.

Applying the Lagrange formulation we can obtain the torque equation of joint 1 of the manipulator [11]. This equation of the motion describing the torque of joint 1 is [15]:

$$\begin{aligned} \tau_1 = & -[(4m_1 \sin \theta_1 - 4m_2 \cos \theta_1)d_3 + I_3]\ddot{\theta}_1 + [(m_1 + m_2)(\sin \theta_1 \cos \theta_1)d_3]\ddot{d}_3 \\ & + [(m_1 \sin \theta_1 - m_2 \cos \theta_1)d_3]\dot{\theta}_1^2 - [m_1 \cos \theta_1 + m_2 \sin \theta_1]\dot{d}_3^2. \quad (3) \\ & - [(m_1 + m_2)(\sin \theta_1 \cos \theta_1)d_3]\dot{\theta}_1 \dot{d}_3 \end{aligned}$$

The terms $\ddot{\theta}_1, \ddot{d}_3$ in the torque equations are related to the angular accelerations of the joint, the terms $\dot{\theta}_1^2, \dot{d}_3^2$ are the centripetal accelerations, and the term $\dot{\theta}_1 \dot{d}_3$ is the Coriolis acceleration [12].

3 Controllers Design

In this section the design of LQG with LTR and LMI with LTR controllers are presented. The control scheme of these strategies for this work application is presented in Fig. 2.

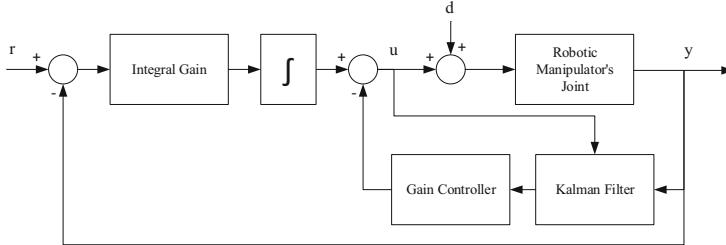


Fig. 2. Control scheme of the state-feedback control strategies with KF estimator and integral action.

3.1 Linear Quadratic Gaussian with Loop Transfer Recovery (LQG with LTR)

In the LQG control strategy, due to the separation principle, basically two conceptions can be merged: an optimal regulator and an optimal state estimator [16].

To minimise a specific criterion or cost function a linear state-space feedback control is designed, and in the case of LQG this cost function is quadratic, it is the LQR method. Also, the KF is chosen for a linear-quadratic state estimation solution. Therefore, to compose a linear and easy to implement control law, the LQR and the KF can be combined.

The LQR objective is to find the feedback gain K that minimises the cost function

$$J = \int_0^{\infty} (x^T Q x + u^T R u). \quad (4)$$

The gain is found by the algebraic equation:

$$K = R^{-1} B^T P_o, \quad (5)$$

where P_o satisfies the Riccati equation:

$$A^T P_o + P_o A - P_o B R^{-1} B^T P_o + Q = 0. \quad (6)$$

For the optimal estimator, the following state-space system is considered:

$$\begin{cases} \dot{\hat{x}}(t) = Ax(t) + Bu(t) + Bw(t), \\ y(t) = Cx(t) + v(t), \end{cases} \quad (7)$$

where $v(t)$ and $w(t)$ are the measurement and the input noise, respectively.

For the LQG purposes, the output signal measured presents Gaussian noise and the initial state variables are expected to be components of a random Gaussian vector.

Optimally, it is desirable to minimise the cost $J = E\{e(t)e^T(t)\}$, where $e(t) = x(t) - \hat{x}(t)$, obtaining the Riccati equation:

$$\dot{P}_o = AP_o + P_oA^T - P_oC^T R_v^{-1} C P_o + BR_w B^T, \quad (8)$$

where

$$\begin{aligned} R_v &= E\{v(t)v^T(t)\}, \\ R_w &= E\{w(t)w^T(t)\}. \end{aligned} \quad (9)$$

The KF gain is then computed as

$$K_o = P_o C^T R_v^{-1}. \quad (10)$$

3.1.1 Integral Action in the State-Feedback Control

Assuming that it is desired to follow a reference input, it is necessary to increase another state variable to the state-space feedback control design: the error between the reference and the measured output.

This suggests an expansion of the matrices in (7) obtained from Eq. (3) with the matrices written as in (11) and the regulation of this new state variable implies in an integral action that guarantees reference tracking.

$$A_a = \begin{bmatrix} A & 0 \\ -C & 0 \end{bmatrix}; \quad B_a = \begin{bmatrix} B \\ 0 \end{bmatrix}; \quad C_a = [C \ 0]; \quad D_a = \begin{bmatrix} D \\ 0 \end{bmatrix}. \quad (11)$$

3.1.2 Loop Transfer Recovery (LTR)

Considering the control strategies presented in this work, the measurement noise affects the corresponding estimators. Then, the optimal performance of the LQG is not achieved and its robustness is affected. Thus, to work around this problem it is necessary to use the LTR method [17].

The procedure is a simple adjustment to define the LQG stability margins as the LQR margins with direct state-feedback [19]. Thus, a real parameter must be included in the Riccati equation of the KF. Considering $q \in \Re$ a scalar used as a project parameter, the covariance matrix is described as $q^2 BR_w B^T$, modifying the Riccati equation and leading to a new gain for the KF:

$$K_o \rightarrow qB(R_w R_w^{-1})^{1/2}. \quad (12)$$

As a result, when $q \rightarrow \infty$ the LQG asymptotically get close to the LQR robustness characteristics [20].

3.2 Linear Matrix Inequalities with Loop Transfer Recovery (LMI with LTR)

The strategy consists in a description by Linear Matrix Inequalities (LMIs) of the performance and robustness restrictions. In the controller design procedure, LMIs can be used to ensure stability by Lyapunov, defining a minimum level of disturbance rejection capability and a region for the pole placement. Thus, the synthesis is completed when numerical values of static gains are found, satisfying the imposed restrictions [21].

3.2.1 Stability Based on the Lyapunov's Quadratic Function

The Lyapunov's Theorem states that to guarantee the stability of a A matrix system

$$\frac{d}{dx}x(t) = \dot{x} = Ax(t), \quad (13)$$

a P , positive defined ($P > 0$) and symmetric matrix must exist and the inequality

$$A^T P + PA < 0 \quad (14)$$

has a solution.

These conditions classifies the system as asymptotically stable. Where A is the state matrix of the system and P is the Lyapunov matrix. Thus, Eq. (14) is an LMI that which solution guarantees system stability.

3.2.2 H_∞ Norm

The H_∞ norm is used to achieve the stability and robustness of the system due to external disturbances. A system with an external input disturbance can be written generically in state-space as

$$\begin{cases} \dot{x}(t) = Ax(t) + B_1 w(t), \\ z(t) = C_z x(t) + D_z w(t). \end{cases} \quad (15)$$

where $x \in R^n$ is the state vector, $w \in R^{n_w}$ is the input disturbance and $z \in R^{n_z}$ is the controlled variable vector.

The transfer function that relates the disturbance to the controlled variable is

$$T_{wz} = C_z(sI - A)^{-1}B_1 + D_{zw}. \quad (16)$$

The strategy most used to define the Eq. (16) is the H_∞ norm. This norm is related to the highest gain that can be obtained between an input and an output. By definition, it is the maximum value between the output and input signal energy [16],

$$\begin{aligned} \|T_{wz}\|_\infty &= \sup \frac{\|z\|_2}{\|w\|_2} < \gamma \\ \|w\|_2 &\neq 0 \end{aligned} \quad (17)$$

where γ is a scalar greater than zero. Therefore, the system's H_∞ norm is the lowest value of γ .

This norm can also be defined through the *Bounded Real Lemma* [22], which states that the A matrix is asymptotically stable if there is a Lyapunov matrix X_∞ that is positive and it is based on the minimisation of γ :

$$\min \gamma, X_\infty = X_\infty^T > 0, \begin{pmatrix} A_{cl}X_\infty + X_\infty A_{cl}^T & B_{cl} & X_\infty C_{cl\infty}^T \\ B_{cl}^T & -\gamma I & D_{cl\infty}^T \\ X_\infty C_{cl\infty} & D_{cl\infty} & -\gamma I \end{pmatrix} < 0. \quad (18)$$

4 Results

In this section the simulations results of the two control strategies are presented and compared initially by two performance criteria: the rise time (T_r) and the overshoot (OS).

Figure 3 shows simulation results for a comparative analysis of torque control. Where a $1.0 \text{ N} \cdot \text{m}$ step reference is applied at time 1 second and at time 13 s the step reference is changed to $1.5 \text{ N} \cdot \text{m}$. An input step disturbance of value -10.0 is applied in time 20 s.

One can observe that the LMI-based control has better performance than the LQG as the speeds of the reference tracking and the disturbance rejection responses are superior.

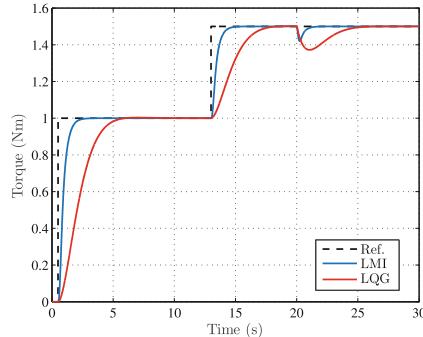


Fig. 3. Tracking and step disturbance rejection response of the LMI-based and LQG control systems.

In Fig. 4 the control signals are shown for both controllers used. One can evaluate that the cost for the better tracking performance of the LMI-based control is a bigger control signal by means of magnitude.

Furthermore, considering the disturbance rejection response, it is evident that the cost for a much better performance of the LMI-based approach is similar to the LQG by means of control effort. Thus, clearly the LMI-based control in disturbance rejection response offers a better trade-off between performance and less control effort.

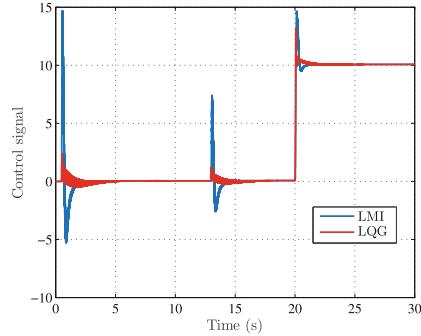


Fig. 4. Control signal of the LMI-based and LQG control systems.

4.1 Discussions

For a quantitative comparison and a more detailed analysis, some performance indices based in the error signal can be used. To quantify the performance of the two control strategies the IAE (Integral Absolute Error) index is used to simulate a disturbance rejection (IAE_q) and a reference tracking (IAE_r) situation.

Table 1 presents the indices (T_r , OS , IAE_q and IAE_r) that quantify the performance of the controllers considering a reference step input of $1.0 \text{ N} \cdot \text{m}$ and a step disturbance of value -10.0 .

According to Table 1, the LMI-based controller has a better performance compared to the LQG, as its IAE indices are considerably small.

Table 1. Performance indices for a reference step input of $1.0 \text{ N} \cdot \text{m}$ and a step disturbance of value -10.0 .

Indices	LQG	LMI
T_r [s]	5.03	2.60
OS [%]	0.03	0.0
IAE_q	0.2938	0.0407
IAE_r	1.7569	0.4703

Other simulation results in Fig. 5 show the performance of the two control strategies considering measurement noise, arbitrary reference inputs and step disturbances. The two controllers present a good performance in reference tracking and disturbance rejection.

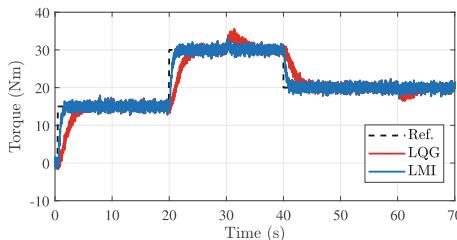


Fig. 5. Reference tracking and step disturbance rejection response of the LMI-based and LQG control systems with measurement noise.

5 Conclusions

This work presented the modelling and torque control of a joint of a robotic manipulator driven by a three-phase induction motor. The two control strategies presented are designed to achieve a robust performance and can be used not only in SISO (single-input and single-output) systems.

The performance quantified by means of indices such as the *IAE* was analysed for the corresponding application. It was seen that for the joint torque control of the studied manipulator, the LMI-based control strategy offers better results in both reference tracking and disturbance rejection capability.

Furthermore, analysing the control effort, one can see a similarity in the LMI and the LQG strategies in disturbance rejection. Thus, for this application the choice of the LMI-based control method is suitable and satisfactory. In the results, the presence of the measurement noise does not prevent the robust performance of the two control strategies. This reaffirms the suitability of these methods in applications involving industrial robot manipulators.

As a future work, the authors intend to control all the manipulator joints in parallel, that is, a MIMO system application.

Acknowledgements. The authors thank the Coordination of Superior Level Staff Improvement (CAPES) and the Brazilian National Council for Scientific and Technological Development (CNPq) for the financial support to this research.

References

1. Batista, J., et al.: Performance comparison between the PID and LQR controllers applied to a robotic manipulator joint. In: IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, vol. 1. IEEE (2019)
2. Batista, J.G., et al.: Modelagem Dinâmica e Simulação de Um Controlador PID e LQR para um Manipulador Cilíndrico. Em: Anais do 14º Simpósio Brasileiro de Automação Inteligente. Campinas: GALOÁ (2020)
3. Ambrose, R.O.: IEEE/IFR Innovation & Entrepreneurship Award. Evaluation 2020, 06-17 (2020)
4. Kelly, R., Davila, V.S., Loría Perez, J.A.: Control of Robot Manipulators in Joint Space. Springer, Heidelberg (2006)

5. Murcia, H.F., Gonzlez, A.E.: Performance comparison between PID and LQR control on a 2-wheel inverted pendulum robot. In: 2016 IEEE Colombian Conference on Robotics and Automation (CCRA), pp. 1–6. IEEE, September 2016
6. Wang, L., Ni, H., Zhou, W., Pardalos, P.M., Fang, J., Fei, M.: MBPOA-based LQR controller and its application to the double-parallel inverted pendulum system. *Eng. Appl. Artif. Intell.* **36**, 262–268 (2014)
7. Stoustrup, J., et al.: An LMI Approach to Fixed Order LTR Controller (1996)
8. Yong, W.A.N.G., Zhigui, X.U., Zhang, H.: A novel control method for turboshaft engine with variable rotor speed based on the Ngdot estimator through LQG/LTR and rotor predicted torque feedforward. *Chinese J. Aeronautics* **33**(7), 1867–1876 (2020)
9. Yahia, R., et al.: Robust control of a robotic manipulator using LMI-based high-gain state and disturbance observers. In: 15th International Multi-Conference on Systems, Signals & Devices (SSD). IEEE (2018)
10. Sanz, P.: Robotics: modeling, planning, and control (siciliano, b. et al.: 2009)[on the shelf]. *IEEE Robot. Autom. Mag.* **16**(4), 101–101 (2009)
11. Spong, M.W., et al.: Robot Modeling and Control. Wiley, New York (2008)
12. Siciliano, B., Sciavicco, L., Villani, L., Oriolo, G.: Robotics: Modelling, Planning and Control. Springer, Heidelberg (2010)
13. Potkonjak, V.: Dynamics of Manipulation Robots: Theory and Application. Springer, Berlin (1982)
14. Kozlowski, K.R.: Modelling and Identification in Robotics. Springer, Heidelberg (2012)
15. Batista, J., et al.: Dynamic model and inverse kinematic identification of a 3-DOF manipulator using RLSPSO. *Sensors* **20**(2), 416 (2020)
16. Skogestad, S., Postlethwaite, I.: Multivariable Feedback Control: Analysis and Design, vol. 2. Wiley, New York (2007)
17. Lewis, F.L., Syrmos, V.L.: Optimal Control. Wiley, Hoboken (1995)
18. Doyle, J.C.: Guaranteed margins for LQG regulators. *IEEE Trans. Automatic Control* **23**(4), 756–757 (1978)
19. Doyle, J., Stein, G.: Robustness with observers. *IEEE Trans. Automatic Control* **24**(4), 607–611 (1979)
20. Maciejowski, J.M.: Multivariable Feedback Design. Electronic Systems Engineering Series. Addison-Wesley, Wokingham (1989)
21. Boyd, S., et al.: Linear Matrix Inequalities in System and Control Theory, vol. 15. Siam (1994)
22. Chilali, M.: H/sub/spl infin//design with pole placement constraints: an LMI approach. *IEEE Trans. Automatic Control* **41**(3), 358–367 (1996)
23. Duan, G.-R., Yu, H.-H.: LMIs in Control Systems: Analysis, Design and Applications. CRC Press (2013)



Forecasting the Retirement Age: A Bayesian Model Ensemble Approach

Jorge M. Bravo¹ and Mercedes Ayuso²

¹ NOVA IMS - Universidade Nova de Lisboa, Université Paris-Dauphine PSL, MagIC, CEFAGE-UE, Lisbon, Portugal
jbravo@novaims.unl.pt

² Department of Econometrics, Statistics and Applied Economy, Riskcenter-UB, Faculty of Economics and Business, University of Barcelona, Barcelona, Spain
mayuso@ub.edu

Abstract. In recent decades, most countries have responded to continuous longevity improvements and population ageing with pension reforms. Increasing early and normal retirement ages in an automatic or scheduled way as life expectancy at old age progresses has been one of the most common policy responses of public and private pension schemes. This paper provides comparable cross-country forecasts of the retirement age for public pension schemes for selected countries that introduced automatic indexation of pension ages to life expectancy pursuing alternative retirement age policies and goals. We use a Bayesian Model Ensemble of heterogeneous parametric models, principal component methods, and smoothing approaches involving both the selection of the model confidence set and the determination of optimal weights based on model's forecasting accuracy. Model-averaged Bayesian credible prediction intervals are derived accounting for both stochastic process, model, and parameter risks. Our results show that statutory retirement ages are forecasted to increase substantially in the next decades, particularly in countries that have opted to target a constant period in retirement. The use of cohort and not period life expectancy measures in pension age indexation formulas would raise retirement ages even further. These results have important micro and macroeconomic implications for the design of pension schemes and individual lifecycle planning.

Keywords: Retirement age · Bayesian Model Ensemble · Mortality forecasting · Life expectancy gap · Pension design and policy · Stochastic methods

1 Introduction

In recent decades, most countries have responded to continuous longevity improvements, low fertility, population ageing, and declining market returns with systemic and/or gradual parametric pension reforms to restore solvency and to alleviate public finance pressure [1]. Parametric reforms include modifying the pension system rules and parameters (e.g., retirement ages, contribution rates, benefit formula, indexation

rules, pension accruals, qualifying conditions, pension decrements/increments). Several countries adopted more systemic reforms by profoundly changing the nature of their public pension schemes (e.g., the Notional Defined Contribution (NDC) system's adoption in Sweden, Poland, Italy, Norway and Latvia) or by strongly supporting the introduction of new (mandatory or voluntary) pillars. Many countries have also acted on the revenue side of the system, for instance, by ear-marking tax revenue for the public pension system, by reducing tax reliefs and allowances on pension benefits¹. For public national pension schemes, a common element of most reforms has been to introduce an automatic link between future pensions and life expectancy developments. The link has been strengthened in at least seven different ways [2]: (i) by linking life expectancy and/or other demographic markers (e.g., sustainability factors) to initial pensions (e.g., Finland, Portugal); (ii) by indexing the full and early retirement ages to life expectancy (e.g., Denmark, the Netherlands, Portugal, Slovakia, Italy, Finland, Cyprus, UK); (iii) by linking the eligibility requirements to longevity developments (e.g., France); (iv) by indexing pension decrements (increments) for early (late) retirement to longevity markers (e.g., Portugal); (iv) by replacing traditional Notionally Defined Benefit (NDB) public PAYG schemes with NDC schemes; (vi) by conditioning pension indexation (e.g., the Netherlands). (vii) by phasing in national FDC plans (e.g., Chile).

Increasing early and normal retirement ages in an automatic or mechanical way as life expectancy at old age progresses and closing routes into premature retirement has been one of the most common policy responses of public and private pension schemes to population aging [3]. To this end, countries have been pursuing different retirement policy strategies [10]: (a) implementing fixed schedules (e.g., Germany, Spain, United States), (b) automatically indexing retirement age to life expectancy, (c) targeting a constant expected number of years in retirement, (d) targeting a constant balance (ratio) between time spent in work (contributing) and in retirement, (e) targeting a constant ratio of adult life (or total lifespan) spent in retirement, (f) targeting a stable old-age dependency ratio, (g) following simple ad-hoc rules to share the longevity risk burden between workers and pensioners. The way these policies have been introduced suffers, however, from several flaws. First, unisex life expectancy measures computed from official period life tables have been used to index retirement ages to longevity developments, neglecting the sizable systematic difference between period and cohort life expectancy (life expectancy gap), generating unintended and sizable ex-ante tax/subsidies from future to current generations, giving a false signal and an unfair actuarial link between the contribution effort and pension entitlements, distorting labour supply decisions leading to macroeconomic inefficiency, incorrectly signalling solvency prospects and, as a result, delaying or slowing down pension reforms [1, 2]. Second, they adopt uniform rules neglecting longevity heterogeneity between socioeconomic groups and high lifespan inequality at retirement [12, 13]. Predicting state pension ages is important for many reasons including assessing the long-term sustainability of pensions and other social benefits, planning for retirement, macroeconomic forecasting, healthcare, qualification and taxation policies, intergenerational fairness valuations.

¹ In private individual or employer-sponsored pension plans, insurance and non-insurance longevity risk-sharing devices have been proposed and/or implemented, the benefit structure switched from DB to DC and conservative ALM strategies have been adopted [4–9, 15].

This paper provides comparable cross-country forecasts of the normal retirement age for public national pension schemes of selected countries (The Netherlands, Denmark, Portugal, Slovakia) that introduced automatic indexation of pension ages to life expectancy. The four countries pursued alternative retirement age policies and are thus a good sample for policy analysis and discussion. We evaluate to what extent the use of a period and not a (more adequate) cohort approach to life expectancy computation affects the retirement age path. To forecast retirement ages by age, sex and calendar year, life expectancy measures must be estimated from stochastic mortality models. The traditional approach to age-specific mortality rate forecasting is to use a single deemed to be «best» model for each population selected from a set of candidate models using some method or criteria, often neglecting model risk for statistical inference purposes. To this end, a significant number of single and multi-population discrete-time and continuous-time stochastic mortality models have been proposed in the actuarial and demographic literature². To tackle both the model risk problem and the need to generate comparable cross country and subpopulation estimates, we follow [1, 10, 14] and use an adaptative Bayesian Model Ensemble of heterogeneous models comprising Generalised Age-Period-Cohort (GAPC) stochastic mortality models, principal component methods, and smoothing approaches. The novel strategy is motivated by the model confidence set procedure developed by [16] and involves both the selection of the subset of superior models using a fixed-rule trimming scheme and considering the model's out-of-sample forecasting performance in the validation period, and the determination of optimal weights³. To derive BME prediction intervals for the quantities of interest, we use the Model-Averaged Tail Area construction proposed by [19] accounting for both stochastic process and model and parameter risk. Model combination has a long tradition in the statistical and forecasting literature but has received little attention in the actuarial and demographic arena. Ensemble learning methods have proven to improve traditional and machine learning forecasting results [20]. The empirical results show that: (i) normal retirement ages are forecasted to increase substantially in the next decades, particularly in countries targeting a constant period in retirement; (ii) the use of cohort instead of period life expectancy measures in the indexation formula would raise retirement ages even further. The results have important micro and macroeconomic implications for the design of pension schemes and for individual lifecycle labour market, consumption and saving decisions. The structure of this article is as follows. Section 2 outlines the key concepts and research methods used in the paper. Section 3 reports summary results for the forecasted pension age together with the reference period (and cohort) life expectancy measures. Section 4 critically discusses the results and concludes.

2 Materials and Methods

2.1 Life Expectancy Measures

Let $\tau p_x(t)$ denote the τ -year survival rate of a reference population cohort aged x at time t , defined as $\tau p_x(t) := \exp(-\int_0^\tau \mu_{x+s}(s)ds)$, where $\mu_x(t)$ is a stochastic force of mortality

² See, e.g., [21–28] and references therein.

³ This contrasts with previous approaches focusing either on the selection of optimal combination schemes and weights [17] or assigning equal weights to the set of superior models [18].

process. For the discretized stochastic process, we assume that $\mu_{x+\xi}(t + \varepsilon) = \mu_x(t)$ for any $0 \leq \xi, \varepsilon < 1$, from which $\mu_x(t)$ is approximated by the central death rate $m_x(t)$ and $p_x(t) = \exp(-m_x(t))$. The complete cohort life expectancy for an x -year old individual in year t is computed as follows

$$\dot{e}_{x,g}^C(t) := \frac{1}{2} + \sum_{k=1}^{\omega-x} \exp\left(-\sum_{j=0}^{k-1} m_{x+j,g}(t+j)\right), \quad (1)$$

whereas the corresponding period life expectancy is given by

$$\dot{e}_{x,g}^P(t) := \frac{1}{2} + \sum_{k=1}^{\omega-x} \exp\left(-\sum_{j=0}^{k-1} m_{x+j,g}(t)\right), \quad (2)$$

with ω denoting the highest attainable age. The concept of life expectancy gap [2] at age x in year t , $\dot{e}_{x,g}^{Gap}(t)$, is given by $\dot{e}_{x,g}^{Gap}(t) := \dot{e}_{x,g}^C(t) - \dot{e}_{x,g}^P(t)$.

2.2 Retirement Age Policies

This section briefly resumes the retirement age policies adopted in the mandatory part of public PAYG pension schemes of selected countries (The Netherlands, Denmark, Portugal, and Slovakia) to automatically index the full pension age to life expectancy. In The Netherlands, the scheme comprises the universal state pension (Dutch: *Algemene Ouderdomswet*, AOW) (first pillar), occupational pension schemes (second pillar), disability benefits and survivor benefits. Before the 2012 reform, it provided all residents a flat-rate pension benefit as from the age of 65. In 2012 the government passed a reform increasing the eligibility age for the public pension and the creation of incentives for a similar movement in 2nd and 3rd pillar pensions. The plan was to raise the eligibility age by one month per year between 2013–2015, three months per year between 2016–2018 and four months per year in 2019–2021, reaching the age of 67.2 by 2021 [29]. From that year on, the pension age will be linked to period life expectancy computed at age 65 as projected by Statistics Netherlands, in the following way⁴:

$$\Delta x_{R,t} = [\dot{e}_{65}^P(t) - 18.26] - (x_{R,t-1} - 65), \quad (3)$$

where $\Delta x_{R,t}$ is the increase of the eligibility age (in years), $x_{R,t-1}$ is the previous year eligibility age, and the remaining variables keep their previous meaning. The law requires the government to announce the automatic increases at least 5 years before implementation. In case $\Delta x_{R,t}$ is negative or less than 0.25 years, the value of $\Delta x_{R,t}$ will be set at zero (pension age decreases are ruled out by law). The increases are not continuous but set at 3-month steps. The policy goal underlying the indexation rule (3) is that expected

⁴ On July 2, 2019, the Dutch parliament passed a law that slows the rate of scheduled increases in the retirement age for public pensions, under which the retirement age will remain at the 2019 through 2021 and will rise gradually to age 67 from 2022 to 2024. Starting in 2025, the retirement age will automatically rise based on increases in life expectancy at age 65.

years in retirement should be constant and equal to 18.26. The precise way in which it has been designed implies, however, that the expected remaining lifetime at retirement exceeds that target.

The Danish multi-pillar system comprises state organized pensions, privately and collectively organized occupational pensions and private and individual pension savings. In pillar one, all citizens above the state pension age (in 2020, 66 years of age for women and 67 for men) are entitled to a universal tax financed, flat rate pension. The first pillar includes a mandatory and fully funded supplementary benefit (ATP) covering wage earners, unemployed and disability pensioners [29]. Following the 2006 and 2011 reforms, the retirement will gradually be raised from 65 to 67 in 2022. In addition, the state pension age was linked to period life expectancy, using 1995 as baseline. The expected period in retirement is targeted at 14.5 years (17.5 including VERP⁵), based on period life-expectancy at age 60 for the total population. The indexation mechanism of the retirement age is as follows:

$$x_{R,t} = 60 + \dot{e}_{60}^P(t - 15) - 14.5. \quad (4)$$

Changes in old-age pension age are decided 15 years before they occur (12 years for VERP), with the first increase due to decision in 2015-indexation to be in 2030 (2027 for VERP). The maximum increase in the retirement age is restricted to 1 year every 5 years, with increases rounded to the nearest half year.

The Portuguese pension system is based on three pillars of differing importance: the dominant earnings-related old-age state pension system (first pillar), the occupational pension provision (second pillar), and the personal pension provision (third pillar). The first pillar combines an earnings-related, defined benefit (DB), mandatory public PAYG scheme, comprising two separate but convergent schemes: (i) a private-sector workers scheme and (ii) a civil service pension scheme covering public servants enrolled before December 2005. Occupational pension schemes and accident insurance form the second pillar. The third pillar, personal pension provision, is voluntary and consists of various private personal funded schemes [30, 31]. There is a common time-dependent normal retirement age for both men and women which, from 2015 onwards, is automatically linked to period life expectancy computed at age 65 as follows

$$x_{R,t} = 66 + \frac{m_t}{12}, \text{ with } m_t = \frac{2}{3} \left[\sum_{j=2015}^t 12 \times [\dot{e}_{65}^P(j-2) - \dot{e}_{65}^P(j-3)] \right] \quad (5)$$

where m_t denotes the number of months to be added to the statutory retirement age (rounded to the nearest integer). As of 2020, the normal retirement age is 66 years and 5 months for both men and women. The policy goal underlying (5) is to extend the working life by two thirds of period life expectancy gains observed at age 65.

Finally, the pension system in Slovakia consists of an earning related universal pension system (PAYG, mandatory, DB points system) covering almost all pensioners in the country, the armed forces pension scheme and voluntary fully funded 2nd and 3rd pillar DC schemes. Until 2003, the normal retirement age was 60 years for men and 53–57 years for women (depending on the number of children raised). Since 2004, that

⁵ Voluntary early retirement pension (VERP).

age has been gradually converging to 62 for both men and women. The 2012 pension reform, effective as from 2017, linked the retirement age to the Y-O-Y difference (in days) of 5-year moving average of the unisex period life expectancy as follows:

$$x_{R,t} = x_{R,t-1} + [\dot{\bar{e}}_x^P(t-7:t-3) - \dot{\bar{e}}_x^P(t-8:t-4)], \quad (6)$$

where $\dot{\bar{e}}_x^P(t-7:t-3)$ is the 5-year moving average observed between years $t-7$ and $t-3$ at the age of round down $x_{R,t-1}$. This indexation rule (6) allocates the burden of longevity improvements to future pensioners only. As of 2020, the normal retirement age is 62 years and 6 months for women and 62 years and 8 months for men. In March 2019 there was a reform reversal, with the retirement age now capped age 64 after which there will be no further increases.

2.3 Bayesian Model Ensemble Approach to Mortality Forecasting

This section summarizes the Bayesian Model Ensemble (BME) approach for mortality modelling and forecasting proposed in [1] and adopted here. Let each candidate model be denoted by M_l , $l = 1, \dots, K$ representing a set of probability distributions comprehending the likelihood function $L(y|\theta_l, M_l)$ of the observed data y in terms of model specific parameters θ_l and a set of prior probability densities for said parameters $p(\theta_l|M_l)$. Consider a quantity of interest Δ present in all models, such as the future observation of y . The marginal posterior distribution across all models is

$$p(\Delta|y) = \sum_{k=1}^K p(\Delta|y, M_k) p(M_k|y), \quad (7)$$

where $p(\Delta|y, M_k)$ denotes the forecast PDF based on model M_k alone, and $p(M_k|y)$ is the posterior probability of model M_k given the observed data. The posterior probability for model M_k is denoted by $p(M_k|y)$ with $\sum_{k=1}^K p(M_k|y) = 1$. To identify the model confidence set and compute model weights, for each subpopulation we first rank the models according to their out-of-sample predictive accuracy. We conducted a backtesting exercise considering a 5-year forecasting horizon for all models and populations and use the symmetric mean absolute percentage error (SMAPE) to measure forecasting accuracy. To compute $p(M_k|y)$, the normalized exponential function is adopted

$$p(M_k|y) = \frac{\exp(-|\xi_k|)}{\sum_{l=1}^K \exp(-|\xi_l|)}, \quad k = 1, \dots, K, \quad (8)$$

with $\xi_k = S_k / \max\{S_k\}_{k=1, \dots, K}$ and S_k is SMAPE for model k and population g . The normalized exponential function assigns larger weights to models with smaller forecasting error, with the weights decaying exponentially. The sampling distribution of the BME estimate of the quantity of interest Δ is a mixture of the individual model sampling distributions. We derive model-averaged Bayesian credible intervals using the Model-Averaged Tail Area (MATA) construction [19]. Let $\phi = g(\Delta)$ be a transformation of the variable of interest with sampling distribution $\hat{\phi}_k = g(\hat{\Delta}_k)$

approximately normal given that M_k is true. The $(1 - 2\alpha)100\%$ MATA-Wald confidence interval for Δ is given by the values Δ_L and Δ_U which satisfy the pair of equations: (i) $\sum_{l=1}^K w_k(1 - \Phi_{L,k}) = \alpha$ and (ii) $\sum_{l=1}^K w_k(\Phi_{U,k}) = \alpha$, where $z_{L,k} = (\widehat{\Delta}_k - \Delta_L)/se(\widehat{\Delta}_k)$, $z_{U,k} = (\widehat{\Delta}_k - \Delta_U)/se(\widehat{\Delta}_k)$, $\Phi_L = g(\Delta_L)$, $\Phi_U = g(\Delta_U)$ and $\Phi(\cdot)$ is c.d.f. of the standard normal distribution.

2.4 Candidate Stochastic Mortality Models

The set of candidate stochastic mortality models considered in the implementation of the adaptive BME comprises six widely used single population GPCM models, one single-population univariate functional demographic time-series model (weighted Hyndman-Ullah method), one bivariate functional data model (Regularized SVD model) and the two-dimensional smooth constrained P-splines model. Table 1 summarizes the analytical structure of the nine candidate models considered in this study.⁶

Table 1. Analytical structure of the stochastic mortality models

Model	Analytical structure	Reference
LC	$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)}$	[21]
APC	$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}$	[23]
RH	$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(0)} \gamma_{t-x}$	[22]
CBD	$\eta_{x,t} = k_t^{(1)} + (x - \bar{x})k_t^{(2)}$	[24]
M7	$\eta_{x,t} = k_t^{(1)} + (x - \bar{x})k_t^{(1)} + ((x - \bar{x})^2 - \sigma)k_t^{(1)} + \gamma_{t-x}$	[32]
Plat	$\eta_{x,t} = \alpha_x + k_t^{(1)} + (x - \bar{x})k_t^{(2)} + (x - \bar{x})k_t^{(3)} + \gamma_{t-x}$	[26]
HUw	$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, i = 1, \dots, p \quad t = 1, \dots, n$	[33]
CPspl	$\eta = B\alpha$	[34]
RSVD	$m(x, t) = d_1 U_1(t) V_1(x) + \dots + d_q U_q(t) V_q(x) + \varepsilon(x, t)$	[35]

Source: Author's preparation

The set includes: [LC] the standard age-period Lee-Carter model under a Poisson setting for the number of deaths; [APC] the age-period-cohort model; [RH] an extension of the Lee-Carter model to include cohort effects in the linear predictor $\eta_{x,t}$, particular substructure obtained by setting $\beta_x^{(0)} = 1$ and additional approximate identifiability constraint; [CBD] the Cairns-Blake-Dowd model considering a predictor structure with substructure $\beta_x^{(1)} = 1$ and $\beta_x^2 = (x - \bar{x})$, with \bar{x} the average age in the data; [M7] an extension of the original CBD model with cohort effects and a quadratic age effect; [Plat] the Plat model [26] with $\kappa_t^{(3)} = 0$; [HUw] the weighted Hyndman-Ullah Functional Demographic Model (FDM) considering geometrically decaying weights; [CPspl]

⁶ See [1] and references there in for technical details.

the two-dimensional P-splines model with demographic constraints; [RSVD] the Regularized Singular Value Decomposition (RSVD) model. Some of the GAPC models described in Table 1 are nested within one of the others. In these cases, trimming models and determining a model confidence set leads to better estimates of each model's weight in the combined forecast. To implement the adaptative BME procedure, we use a fixed-rule trimming scheme in which the number of GAPC models to be discarded is fixed exogenously (three out of six) and determine the statistically superior set of best models based on the model's out-of-sample forecasting performance in the validation period. To forecast age-specific mortality rates, we first calibrate the models using each country population data from 1960 to the most recent year available and for ages in the range 60–95. We derive prediction intervals for mortality rates accounting for both stochastic process and parameter risk using a bootstrap approach [36]. For each model and population, we consider 5000 bootstrap samples. The model fitting, forecasting and simulation procedures have been implemented using a R software routine. The datasets used in this study consist of observed death counts, $D_{x,t}$, and exposure-to-risk, $E_{x,t}$, classified by age at death $x \in [0, 110+]$, year of death $t \in [1960, 2018]$ and sex, obtained from the Human Mortality Database [11] and national pension age data.

3 Results

Figure 1 plots, for the total population of each country, the model confidence set (vertical axis) and the BME model weights (horizontal axis). We can observe that the model confidence set varies between countries and their predictive accuracy is population specific. We find that no single model dominates based on the predictive accuracy criteria. Figure 2 plots the BME forecast of the unisex period and cohort life expectancy measures at the reference indexation age from 2020 to 2100 (2050 for cohort longevity measures) and corresponding 95% MATA confidence intervals, together with the point forecast of the pension age using the actual legislated formulas (3)–(6). In addition, it plots also forecasts of the pension age that would emerge if cohort and not period life expectancy measures were used. In The Netherlands, the period (cohort) life expectancy at age 65 is forecasted to increase from 20.42 (21.82) years in 2020 to 23.32 (24.81) years in 2050. The life expectancy gap growths from 1.40 in 2020 to 1.49 in 2050, which represents a constant implicit tax of 7% from future to current pensioners.

The Dutch pension age is forecasted to increase from the current 66.3(3) years in 2020 to 70.06 years in 2050 and to 74.21 years by 2100, considering the current indexation rule based on $\dot{e}_{x,g}^P(t)$. If, instead, the cohort life expectancy had been considered, the results show that the normal pension age would need to raise nearly 1.5 years to 71.55 years in 2050 to be consistent with the target of delivering 18.26 expected years in retirement. In Denmark, the period (cohort) life expectancy at age 60 is forecasted to increase from 23.88 (25.87) years in 2020 to 27.15 (29.00) years in 2050, with a significant positive life expectancy gap of 1.85 years in 2050. The Danish pension age is forecasted to increase from 67 years in 2020 to 71 years in 2050 and to 76 years by the end of the century. If, instead, the cohort life expectancy had been considered, the results show that the normal pension age would need to raise one extra year to 72 years in 2050 to be consistent with the target of delivering 14.5 expected years in retirement. Compared to

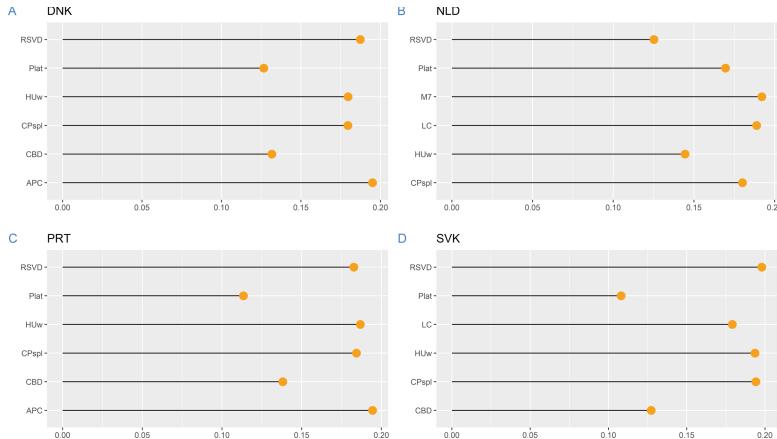


Fig. 1. Model confidence set and BME model weights per country, total population. Notes: DNK = Denmark, NLD = The Netherlands; PRT = Portugal; SVK = Slovakia.

the Dutch case, the pension age increases in Denmark are mitigated by the constraint that imposes a cap on retirement age increments (one extra year every 5 years). The BME results for Portugal show that the unisex period (cohort) life expectancy at age 65 is forecasted to increase from 20.24 (21.59) years in 2020 to 23.30 (25.04) years in 2050 and to 27.54 years (period) in 2100. As a result, the Portuguese normal pension age is forecasted to increase from the current 66.41(6) years for both men and women in 2020 to 68.41(6) years in 2050 and to 71.3 years by the end of the century. If, instead, the cohort life expectancy had been considered, we forecast that the normal pension age would need to raise to 68.6(6) years in 2050. Compared to other countries, we note that the retirement age indexation mechanism adopted in Portugal is less sensitive to the choice of the life expectancy measure, with deviations emerging only to the extent that $\dot{e}_{x,g}^P(t)$ and $\dot{e}_{x,g}^C(t)$ exhibit different trends.

We note also that the retirement age policy adopted, splitting the burden of longevity increments between active life and retirement periods, results in smaller pension age increments when compared to the Dutch and Danish cases which target a constant period in retirement. Finally, the results for the Slovak Republic show that the period (cohort) life expectancy at age 63 is forecasted to increase from 18.80 (19.83) years in 2020 to 21.07 (22.09) years in 2050 and to 24.04 in 2100 (period approach). As a result, the normal pension age resulting from the indexation rule (6) was forecasted to increase from the current 62.6(6) years for both men and women in 2020 to 65.04 years in 2050 and to 67.97 years by the end of the century. Once again, if the cohort life expectancy had been considered, the results show that the statutory pension age would need to raise to 65.05 years in 2050, almost the same as that obtained using the period approach.

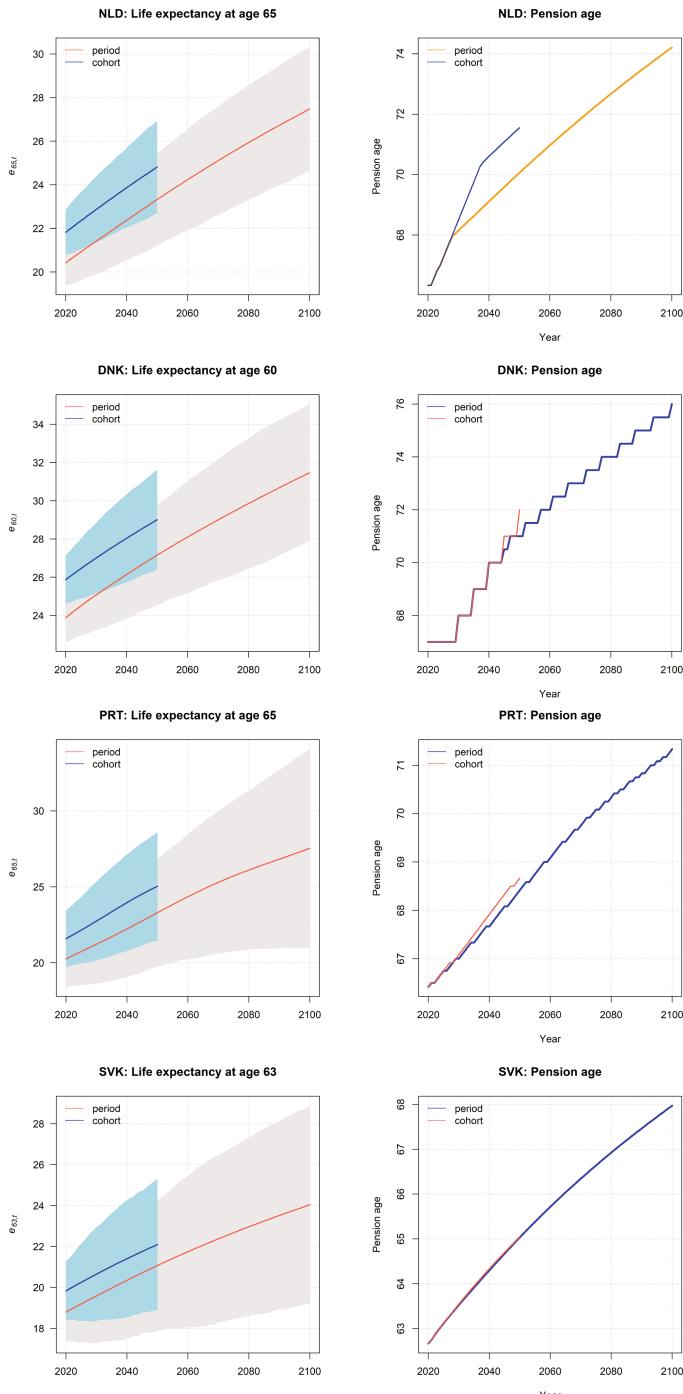


Fig. 2. BME Forecast of unisex period and cohort life expectancy and pension age with 95% CIs.

4 Conclusion

The purpose of linking retirement ages to life expectancy developments is chiefly to minimize the impact of demographic and economic shocks on the financing of pension schemes, but they also carry an implicit objective of introducing economic/actuarial rationality for justifying changes, avoiding the thorns of regular political negotiations to adopt the required adjustments enhancing the credibility of the system, preventing unexpected public finance burdens in the future. The way they have been introduced in pension schemes suffers, however, from several flaws, including the use of inappropriate longevity measures, lack of actuarial fairness across generations, longevity heterogeneity between socioeconomic groups and high lifespan inequality. This paper provides forecasts of statutory retirement ages for selected countries using a novel adaptative BME approach and briefly discusses its implications for retirement planning. Retirement age increases have significant micro and macroeconomic implications and raise new challenges such as reducing inequality, guaranteeing fairness across generations, adapting labour markets to older workers, adopting uniform or differentiated rules by sex and socioeconomic group, raising healthy life expectancy in tandem with life expectancy, reforming health care systems, investing in lifetime qualification policies for enhanced productivity at old ages, or tackling the case of women with children.

References

1. Bravo, J.M., Ayuso, M., Holzmann, R., Palmer, E.: Addressing the Life Expectancy Gap in Pension Policy. *Insurance: Mathematics and Economics* (2021, accepted/in press)
2. Ayuso, M., Bravo, J.M., Holzmann, R.: Getting life expectancy estimates right for pension policy: period versus cohort approach. *J. Pension Econ. Finan.* **20**(2), 212–231 (2021). <https://doi.org/10.1017/S1474747220000050>
3. European Commission: Pension Reforms in the EU since the Early 2000's: Achievements and Challenges Ahead. Brussels: European Economy Discussion Paper 42 (2016)
4. Bravo, J.M., Pereira da Silva, C.M.: Immunization using a stochastic process independent multifactor model: the Portuguese experience. *J. Bank. Finan.* **30**(1), 133–156 (2006)
5. Milevsky, M., Salisbury, T.: Optimal retirement income tontines. *Insur.: Math. Econ.* **64**, 91–105 (2015)
6. Bravo, J., El Mekkaoui de Freitas, N.: Valuation of longevity-linked life annuities. *Insur. Math. Econ.* **78**, 212–229 (2018)
7. Bravo, J.M.: Funding for longer lives: retirement wallet and risk-sharing annuities. *Ekonomiaz* **96**(2), 268–291 (2019)
8. Bravo, J.M.: Longevity-linked life annuities: a Bayesian model ensemble pricing approach. In: CAPSI 2020 Proceedings. Atas da 20ª Conferência da Associação Portuguesa de Sistemas de Informação 2020, p. 29 (2020). <https://aisel.aisnet.org/capsi2020/29>
9. Bravo, J.M., Coelho, E.: Forecasting subnational demographic data using seasonal time series methods. *Atas da Conferência da Associação Portuguesa de Sistemas de Informação* (2019)
10. Bravo, J.M., Ayuso, M., Holzmann, R., Palmer, E.: Intergenerational actuarial fairness when longevity increases: amending the retirement age to cope with life expectancy developments. *Scand. Actuar. J.* (2021, submit for publication)
11. Human Mortality Database: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) (2020)

12. Ayuso, M., Bravo, J.M., Holzmann, R.: On the heterogeneity in longevity among socioeconomic groups: scope, trends, and implications for earnings-related pension schemes. *Glob. J. Human Soc. Sci.-Econ.* **17**(1), 31–57 (2017)
13. Holzmann, R., Ayuso, M., Bravo, J.M., Alaminos, E., Palmer, E.: Reframing lifecycle saving and dissaving by low-, middle-, and high-income groups: initial hypotheses, literature review, and ideas for empirical testing (2021, submitted for publication)
14. Bravo, J.M., Ayuso, M.: Previsões de mortalidade e de esperança de vida mediante combinação Bayesiana de modelos: Uma aplicação à população portuguesa. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, E40, 128–144 (Dec 2020). <https://doi.org/10.17013/risti.40.128-145>
15. Bravo, J.M.: Pricing participating longevity-linked life annuities: a Bayesian model ensemble approach. *Eur. Act. J.* (2021, revised and resubmitted)
16. Hansen, P., Lunde, A., Nason, J.: The model confidence set. *Econometrica* **79**, 453–497 (2011)
17. Andrawis, R., Atiya, A., El-Shishiny, H.: Forecast combinations of computational intelligence and linear models for the NNS time series forecasting competition. *Int. J. Forecast.* **27**(3), 672–688 (2011)
18. Samuels, J.D., Sekkel, R.M.: Model confidence sets and forecast combination. *Int. J. Forecast.* **33**(1), 48–60 (2017)
19. Turek, D., Fletcher, D.: Model-averaged wald confidence intervals. *Comput. Stat. Data Anal.* **56**(9), 2809–2815 (2012)
20. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: Statistical and machine learning forecasting methods: concerns and ways forward. *PLoS ONE* **13**(3), e0194889 (2018)
21. Brouhns, N., Denuit, M., Vermunt, J.: A Poisson log-bilinear regression approach to the construction of projected life tables. *Insur. Math. Econ.* **31**, 373–393 (2002)
22. Renshaw, A.E., Haberman, S.: A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insur.: Math. Econ.* **38**(3), 556–570 (2006)
23. Currie, I.: Smoothing and forecasting mortality rates with P-Splines. Heriot Watt Un (2006)
24. Cairns, A., Blake, D., Dowd, K.: A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *J. Risk Insur.* **73**, 687–718 (2006)
25. Hyndman, R., Ullah, S.: Robust forecasting of mortality and fertility rates: a functional data approach. *Comput. Stat. Data Anal.* **51**, 4942–4956 (2007)
26. Plat, R.: On stochastic mortality modeling. *Insur. Math. Econ.* **45**(3), 393–404 (2009)
27. Hunt, A., Blake, D.: On the structure and classification of mortality models. *North Am. Actuar. J.* (2020). <https://doi.org/10.1080/10920277.2019.1649156>
28. Bravo, J.M., Nunes, J.P.V.: Pricing longevity derivatives via fourier transforms. *Insur. Math. Econ.* **96**, 81–97 (2021)
29. European Commission: The 2018 ageing report: economic and budgetary projections for the EU Member States (2016–2070), European Economy, Institutional Paper 079 (2018)
30. Bravo, J.M., Herce, J.A.: Career breaks, broken pensions? Long-run effects of early and late-career unemployment spells on pension entitlements. *J. Pension Econ. Finan.* 1–27 (2020). <https://doi.org/10.1017/S1474747220000189>
31. Bravo, J.M.: Taxation of pensions in Portugal: a semi-dual income tax system. CESifo DICE Rep. – J. Inst. Comp. **14**(1), 14–23 (2016)
32. Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Ong, A., Balevich, I.: A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North Am. Actuar. J.* **13**(1), 1–35 (2009)
33. Shang, H.L., Booth, H., Hyndman, R.J.: Point and interval forecasts of mortality rates and life expectancy: a comparison of ten principal component methods. *Demogr. Res.* **25**, 173–214 (2011)
34. Camarda, C.G.: Smooth constrained mortality forecasting. *Demogr. Res.* **41**(38), 1091–1130 (2019)

35. Huang, J.Z., Shen, H., Buja, A.: The analysis of two-way functional data using two-way regularized singular value decompositions. *J. Am. Stat. Assoc.* **104**(488), 1609–1620 (2009)
36. Brouhns, N., Denuit, M., Van Keilegom, I.: Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scand. Actuar. J.* **3**, 212–224 (2005)



Using Bayesian Dialysis and Tetrads to Detect the Persistent Characteristics of Fraud

The Case of Vat and Corporate Tax in Spain

Ignacio González García[✉] and Alfonso Mateos

Decision Analysis and Statistic Group, Departamento de Inteligencia Artificial de la Universidad Politécnica de Madrid, Madrid, Spain
amat eos@fi. upm. es

Abstract. In this paper, we propose a methodology combining Bayesian and big data tools designed to optimize the investigation of fraud. This methodology is called Bayesian dialysis. We address three issues: a) Is it possible to capitalize on the evidence provided by data indicating fraud without a parametric model and using an interpretable approach? b) If so, would it be the best solution in any case? c) What is the effect size of all unobservable, even unknown, variables? We prove the viability of a new method using as an exemplary case the selection for VAT control in the Spanish Tax Agency (Agencia Estatal de Administración Tributaria—AEAT). The new method improves fraudster detection precision by 12.29%, which is increased from an average of 82.28% to 94.36%. We also use 2018–2019 corporate tax data to test the scope of this approach. Finally, based on the concept of tetrads, we propose a method to quantify the effect of unknown latent variables on models analysis.

Keywords: Bayes · Fraud · VAT · Beta

1 Introduction

Of the many manifestations of fraud [1], tax fraud is particularly important. It has been estimated to amount annually to € 825,000 million in the EU [2] and € 25,648 million in Spain [3].

Efforts to combat fraud include identifying new models [4], using tested statistical methods [5] and creating new methods [6]. A review of the state of the art in the tax field [7] shows that all kinds of techniques are used, including data mining [8, 9]. These techniques are not able to capitalize on experience, except by recalculating model parameters. They are insufficient in problems such as tax fraud, where two taxpayers may or may not defraud motivated by their personal convictions or background, which are constructs characterized by unobservable variables. This paper presents a method to capitalize on the evidence from checks previously performed by the experts. Section 2 provides the problem setting and describes variables. Section 3 shows the research applied to data on VAT inspections carried out by the AEAT in the period 2009–2018 and corporate tax in the year 2018–2019 in response to the above three questions.

2 Methodology

2.1 Notation

Let x_i , for $i = (1, 2, \dots, N)$, elements of a set \mathbf{U}_t , used to investigate possible cases of fraud. For each period t , a subset $I_t \subset \mathbf{U}_t$ is selected and inspected. Experts try to maximize the fraud detected with a M_t model that evolves with experience. Evidence, E_t , includes cases of fraud, F_t , and compliance, C_t , with $I_t = F_t \cup C_t$. Thus, we can identify potential explanatory variables of fraud.

2.2 Problem Setting

Problem 1. *Is it possible to capitalize on the evidence provided by data indicating fraud without a parametric model and using an interpretable approach?*

Problem 2. *If so, would it be the best solution in any case?*

Problem 3. *What is the effect size of all unobservable, even unknown, variables?*

2.3 Valuation of the Evidence

Step 1. Gather Evidence. Every year experts extract subsets I_t from \mathbf{U}_t , taxpayer census, for inspection, using a M_t model built using experience-based rules or statistical tools. Using the data from past inspections, E_{t-1} , it is possible to determine the proportion \bar{p} of fraud detected by each combination of variables (i.e., value added tax (VAT) fraud by automobile repair shops). It is also possible to determine the confidence interval [CI] of \bar{p} using a beta distribution. For two fictitious groups, $I_1 = \{F_1 = 400, C_1 = 100\}$ and $I_2 = \{F_2 = 9, C_2 = 1\}$, $\bar{p}_1 = 400/500 = 0.8$ and $\bar{p}_2 = 0.9$, respectively. The respective confidence intervals for $\alpha = 5\%$ would be (0.762, 0.832) and (0.587, 0.977). The variability of the latter interval is higher because it includes a smaller number of elements.

Step 2. Select Variables and Propose a Model. We could select a set of potentially explanatory variables of fraud (V) using either the variables taken into account by the experts or statistical techniques to predict fraud (Y) from potential predictors (X). Experts use models to select variables. However, they do not know the real causes of fraud and the magnitude of the relations between variables, and they may overlook key latent variables and misinterpret confounding bias.

In our research, we accounted for four variables: $V = \{v_1 = \text{VAT paid quota}, v_2 = \text{volume of sales according to Art. 121 of the Law}, v_3 = \text{difference between declared sales and income attributed by third parties}, v_4 = \text{percentage year-over-year sales increase}\}$.

These variables were used to build directed acyclic graphs (DAGs), which are an accepted economic form of knowledge representation [10, 11]. As we have the evidence, I_{t-1} , we can determine the annual joint probability distribution and, if we have a model of fraud, build a Bayesian network (BN).

The number of causal graphs that can be built with N variables is $4^{\frac{N!}{2(N-2)!}}$. If $N = 4$, we have $4^6 = 4,196$, and, if $N = 6$, this number increases to 1,073 million. Even confined to DAGs, this is an unmanageable number [12]. Therefore, we have used a model proposed by experts.

Step 3. Measure Evidence. A subset of \mathbf{U} structured according to variables is referred to as a cube (actually a hypercube if the number of variables is greater than 3). For each continuous variable, we determine the percentiles (D_i), which are quartiles $i \in \{1, 2, 3, 4\}$ in some cases and deciles $i \in \{1, \dots, 10\}$ in others. If we use quartiles in a problem with three variables, each data cube will be divided into $4^m = 4^3$ basic cubes or, if we use deciles, into 1000. The number is potentially much greater for categorical variables. For the analysis of a customs fraud problem, the number of Harmonized System subheading codes is 5,212. For tax checks, such as VAT in Spain, the number of economic activity statistical codes of is 1,219.

The value of \bar{p} is different in each basic cube and is greater in some cubes than in others. If the variables are $V = \{v_1, \dots, v_N\}$, the hypercube I_t contains a number of elementary hypercubes given by:

$$\text{Number of components} = \prod_{i=1}^N \prod_j^{M_i} v_{ij}, \quad (1)$$

where M_i is the number of components of the variable v_i .

A number of these cubes, ID , contain data, evidence, and we define the coverage coefficient φ as:

$$\varphi = \frac{\#ID}{\#I}. \quad (2)$$

For example, if we analyse the fraud in Spain's import flows in the year 2019 using the Harmonized System and country of origin codes as variables, the result would be: $\varphi_{HC} = \frac{25,607}{407,620} = 6.23\%$, because fraud was detected in 25,607 out of the 407,620 possible combinations of commodity and country codes. If we include a third dimension, importer code, the value of the coverage coefficient will be smaller because this variable is subject to the curse of dimensionality: $\varphi_{HCI} = \frac{118,193}{12,849,448} = 0.91\%$.

Step 4. Model quality metrics. Decision theory provides many tools to deal with the uncertainty associated with fraud prediction. One classical tool is the confusion matrix (Table 1).

It is common to use ratios like $Inspection\ Rate = IR = \frac{TP+FP}{TP+TN+FP+FN}$, $Accuracy = A = \frac{TP+TN}{TP+FP+TN+FN}$, $Recall = R = \frac{TP}{TP+FN}$ and $F1 = \frac{2RP}{R+P}$.

In our case, evidence is confined to the first column, those cases where fraud has been predicted and companies have been inspected. This differs from clinical trials, where there is information about the cases treated with placebo. We can use:

Table 1. Confusion matrix

Observed	Prediction of the model	
	Fraud	Compliance
Check		
Fraud	True positives (TP)	False negatives (FN)
Compliance	False positives (fp)	True negatives (TN)

$$Precision = P = \frac{TP}{TP + FP} = \frac{TP}{I} = \bar{p} \text{ and} \quad (3)$$

$$Failure = F = \frac{FP}{TP + FP} = \frac{TP}{I} = \bar{c}, \quad (4)$$

which represent the percentage of cases of fraud and the percentage of cases of compliance found in I .

We also use confidence intervals, calculated from a beta distribution of P and F , output after controlling the top $n\%$ of elements of a subset ordered by level of risk. 3.

3 Results

3.1 Question 1. The Case of VAT Control

The 2018 AEAT taxpayer census, U_{2018} , contains 69.9 million taxpayers.

The AEAT performs checks, triggered by omissions and rules of compliance, and intensive inspections. In 2018, AEAT conducted 58,819 intensive inspections of cases, of which 19,738 focused on VAT, detecting 16,635 cases of fraud, see Table 2, where $\bar{p} = 84.28\%$ with CI (83.79, 84.80). In the case of import flow, the raud detected (2019) was $\bar{p} = 2.32\%$. We find that the precision of the checks depends on the problem characteristics. Precision is much higher whenever companies are selected for an inspection after a scrutiny of their record resulting in serious grounds for suspicion than for decisions based on systems of general rules applied in seconds by the system. The existence of fraud is identified by the “amount instructed” variable. If this magnitude is greater than cero, it implies existence of fraud.

Table 2 illustrates the taxpayer census (1), the number of taxpayers selected for VAT inspection (2), cases of compliance (3), cases of fraud (4), percentage of fraud.

Table 2. Fraud detected by taxpayer type

Census	Results (I_{2018})			
	I_{2018}	C_{2018}	F_{2018}	F/I
Total	nº	nº	nº	%
69,918,752	19,738	3,103	16,635	84.28

Table 3 shows the data from an estimation of the fraud rate and CIs by crossing data on company size and type. A business, like a bar, owned by a natural person may have several other natural persons on its payroll.

Table 3. Number of cases, percentages and confidence interval.

Natural persons		Legal entities					
7,177		90.76	12.561		81.06	19,738	85.07
	90.37			81			80,89
89.38		6,466	79.67		10,169	84.07	
16,695							

\bar{p} = % fraud

b = Interval 0.975

c = Interval 0.025

The correlation between interdecilic numbers and variables is: -0.14 with v_1 = VAT paid quota, -0.3 with v_3 and -0.4 with v_4 . For example, the correlation between (1, ..., 10) and the average fraud values found in the deciles of v_2 is -0.07 . Therefore, there is no correlation between the proportion of fraud detected in the interval and the magnitude of the quota.

Figure 1 (left) depicts the number of cases inspected from 2009 to 2018 (dotted line), with 294,321 inspections, $\bar{p} = 85.39\%$ and $\sigma = 0.2$. Figure 1 (right) shows the deciles for three variables $\{v_2, v_3, v_4\}$ and each combination of interdecilic intervals representing detected fraud. There are 343 N.A. elements ($\phi = 0.2$), and there is no continuous gradient. Therefore, it would not be a good strategy to choose the elements in the last deciles, because the highest fraud is not there. The cubes with the densest concentration, like the red spheres, should be identified and filtered.

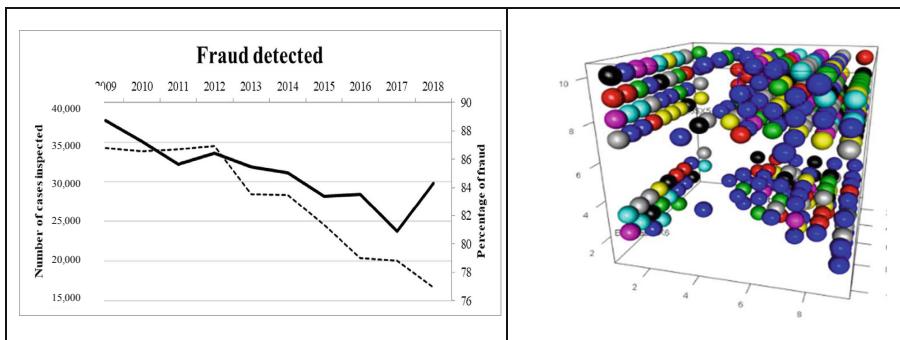


Fig. 1. Granularity and homogeneity (Own elaboration)

Once we have identified the elementary hypercubes (EH) (spheres in Fig. 1) in which most fraud is concentrated, the number of selected EH can be represented, as in Fig. 2, on the abscissa and the percentage of fraud for which they account on the ordinate for each set of analysed variables $\{v_1, v_2, v_3\}$, $\{v_1, v_3, v_4\}$, $\{v_1, v_2, v_3, v_4\}$. For the first combination, there are 345 basic elements with information ($\phi = 0.35$) including an average of 58 cases of fraud and, for the second, $\phi = 0.696$, with an

average of 28 fraud cases. The first combination $\{v_1, v_2, v_3\}$ is more parsimonious (see Fig. 2), as it provides more information with fewer elements.



Fig. 2. Accumulated distribution of fraud (Own elaboration)

We capitalize on the evidence from inspections carried out in the year $t-1$ (i.e., 2017), choosing the elementary hypercubes with information. We create what we define as N-dimensional *Bayesian filter* \mathcal{BF}_{2017}^N : filter because some of the inspections are selected, and *Bayesian* because the confidence interval of a beta distribution is then applied to the selection. Dialysis is run on I_{2018} , using the three-dimensional filter based on the selected tree variables created with data from the previous year.

$$\mathcal{DI}_{2017-2018} = \mathcal{BF}_{2017}^3(\mathcal{I}_{2018}).$$

If the \bar{p} of $\mathcal{DI}_{2017-2018}$ is statistically greater than 82.28% (the real result output according to expert criteria), dialysis can be considered to capitalize on experience.

I_{2017} contains 21,055 elements with 17,499 cases of fraud ($\bar{p} = 83.33\%$), see Table 4. For each of the two sets of analysed candidate variables, $\{v_1, v_2, v_3\}$ and $\{v_1, v_3, v_4\}$, we determined the deciles and selected the 3D hypercubes with evidence of fraud.

We selected two levels of n%, with $\bar{p} > 85\%$ (Level 1) or with $\bar{p} > 90\%$ (Level 2) to create a $\mathcal{DI}_{2017-2018}$ using variables $\{v_1, v_2, v_3\}$ and $\{v_1, v_3, v_4\}$ with 154 and 240 elements, respectively, see Table 4. We filtered I_{2018} using both levels for cubes with evidence and found 183 and 371 using $\{v_1, v_2, v_3\}$ and $\{v_1, v_3, v_4\}$, respectively. Level 1 returned 12,211 cases of fraud (67.73% of the total) with a \bar{p} of 89.97%. For Level 2 ($\bar{p} > 90\%$), $\varphi = 0.39$ in $\{v_1, v_2, v_3\}$ and $\varphi = 0.269$ in $\{v_1, v_3, v_4\}$.

The data are shown in Table 4. The \bar{p} values detected in both dialyses (89.04 and 96.30 for variables $\{v_1, v_2, v_3\}$ and 89.97 and 94.36 for $\{v_1, v_3, v_4\}$) are better than the real values detected using the expert model. This is self-evident for 2017 data, because we select only the top performers, but, importantly, also for 2018. This goes to show that we can capitalize on evidence, that is, detect the effect of persistent and unknown causes.

To select the best option, we apply a Bayesian approach using CI criteria instead of \bar{p} .

The $CI_{0.85}\{v_1, v_2, v_3\}$ interval is calculated using $F = 9,349$ and $C = 1,151$ (10,500 inspections) and is (88.42, 89.62). Repeating the process, $CI_{0.90}\{v_1, v_2, v_3\}$ is (93.82, 96.09). Similarly, $CI_{0.85}\{v_1, v_3, v_4\}$ is (89.43, 90.49) and $CI_{0.9}\{v_1, v_3, v_4\}$ is (95.51, 96.60). We could choose:

Table 4. Dialysis results

	Level	{v ₁ ,v ₂ ,v ₃ }				{v ₁ ,v ₃ ,v ₄ }			
		Ft	No.	%Ft	#HE	Ft	No.	%F _t	#HE
Filter built in 2017	85	9,483	10,826	87.59	154	9,120	10,234	88.93	240
	90	1,353	1,449	93.37	112	3,301	3,514	93.94	175
	All	17,499	21,055	83.11	256	17,499	21,055	83.11	422
Dial sis 2018	85	9,349	10,500	89.04	183	10,987	12,211	89.97	371
	90	1,335	1,405	96.30	139	4,705	4,896	94.36	269
	All	3,013	19,738	84.28	693	3,013	19,738	84.28	677

- a) $DI_{2017-2018}$ for $\{v_1, v_2, v_3\}$ with $\bar{p} = 96.30$ and a lower bound of the $CI_{0.90}$ equal to 93.82,
- b) $DI_{2017-2018}$ for $\{v_1, v_2, v_3\}$ with $\bar{p} = 94.36$ and a lower bound of the $CI_{0.90}$ equal to 95.51.

If we take costs into account and want to reduce the number of false positives, the second option (b) offers a (1–0.025), that is, 97.5%, possibility of detecting fraud, greater than the 95.51% offered by option a).

This method. a) provides excellent results; b) is an alternative to “black box” approaches, that can’t be used when the solution must be interpretable, as in the case in the fiscal field; and c) let consider the variability associated to the evidence.

3.2 Question 2. The Case of Corporate Tax Checks

We repeat the process now with corporate tax data for the period 2018–2019, using three categorical variables: v_1 = economic activity code at the item level (with 615 values), v_2 = tax residency code (with 17 values), v_3 = foreign trade operator (with 2 values). In this case, the number of EH is $615 \times 17 \times 2 = 20,910$, 3,046 of which provide evidence. Therefore,

$$\varphi = \frac{3,046}{20,910} = 0.15$$

We use a one-dimensional filter $BF^1_{2018}(I_{2019})$ in this case, code of activity. We detect #HE 197 and 217 at the 85% level (2018) and at the 90% level, respectively, with $\bar{p} = 98.13\%$ and $\bar{p} = 95.35\%$.

The real inspection strategy is to change the inspected business types every year, covering different sectors each year. There are more than 69 million taxpayers and only about 25,000 can be inspected intensively (a further half million extensive checks are conducted).

Using $BF^1_{2018}(I_{2019})$ for the 5-digit activity code variable, we determined 43 and 57 one-dimensional hypercubes at the 90% level, with $\bar{p} = 96.53\%$, and at the 85% level, with $\bar{p} = 95.36\%$, respectively. The results of applying expert criteria are not as good: $\bar{p}_{2019} = 76.30\%$ and $\bar{p}_{2018} = 76.05\%$. This follows since they cannot capitalize

upon evidence because their strategic decision to inspect different types of taxpayers every year to stop impunity.

The response to Question 2 is that we can only capitalize upon the evidence in relation to variables that are used in the model to determine the subset for control.

3.3 Question 3. Aggregation of Evidence Under Latent Causes

The combination of simple learners is a well-known machine learning strategy [13]. We now try to combine the two potential explanations and consider the possibility of the best explanation being provided by $\{v_1, v_2, v_3, v_4\}$. It is evident that, although these variables do correlate with fraud, they are not the real causes and that there is a latent factor PF “propensity to commit fraud” that has an influence on the declared data, as illustrated in Fig. 3.

Spearman [14] derived a set of equations, called vanishing tetrads, that are necessarily true in a causal process of this type. He argued that if these equations held, there was evidence of a *latent common cause*. We use this idea to estimate the importance of F . For notational simplicity, let's denote $A = v_1$, $B = v_2$, $C = v_3$ and $D = v_4$.

$$e_1 = \rho_{AB}\rho_{CD} - \rho_{AD}\rho_{BC} \quad (5)$$

$$e_2 = \rho_{AC}\rho_{BD} - \rho_{AD}\rho_{BC} \quad (6)$$

$$e_3 = \rho_{AB}\rho_{BD} - \rho_{AB}\rho_{CD} \quad (7)$$

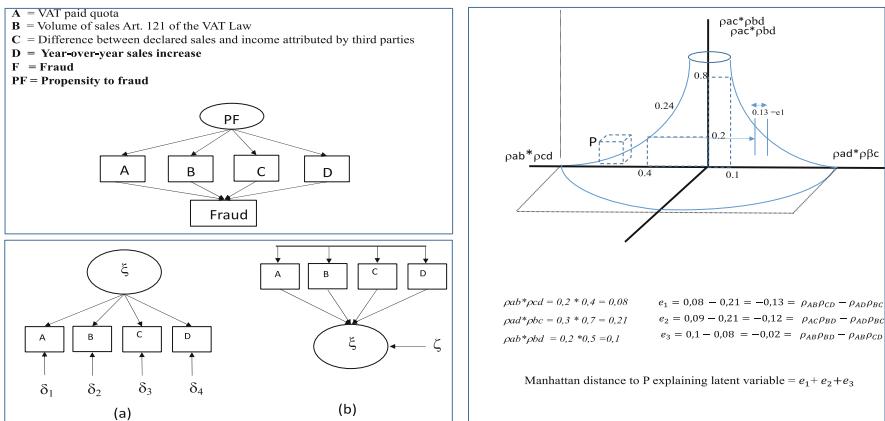


Fig. 3. Scheme used to apply tetrads

These equations must hold regardless of the (non-zero) values of the path coefficients. Figure 3 shows a possible value of ρ_{AB} , ρ_{CD} (i.e., 0.4 and 0.2). If (5) holds, the value of $\rho_{AD} * \rho_{BC}$ must be the same, and there are as many valid combinations as points in the

hyperbole (i.e., 0.1 and 0.8). *The geometric position of points for which the vanishing tetrad condition holds is the depicted hyperboloid of revolution.* (when every $\rho_{i,j} > 0$ or $\rho_{i,j} < 0$; If it is not the case the solution is the same and can be represented by symmetry). If the vanishing condition does not hold, the Manhattan distance between point P , real value of correlations, and the hyperboloid is a measure of the distance to the case where there is one and only one common cause. For VAT variables, $\rho_{AB} = 0.11$; $\rho_{Ac} = 0.29$; $\rho_{AD} = 0.08$; $\rho_{BC} = -0.06$; $\rho_{BD} = 0.34$; $\rho_{CD} = 0.04$. Therefore, $e_1 = 0.11*0.04 + 0.08 * 0.06 = 0.0092$, $e_2 = 0.29*0.34 + 0.08*0.06 = 0.1034$ and $e_3 = 0.11*0.34 - 0.11*0.04 = 0.033$, and the sum is 0.1456, that is, 4.85% of its maximum value of 3.

Spearman's reasoning has been used [15] to distinguish causes from effects in plausible models, as those depicted in the left inferior box of Fig. 3 where, if there is a common cause Fig. 3 (a), condition of vanishing tetrads holds. We apply the idea to the relation between PF and v_i and we conclude that: a) the combination of the four variables provides a much better explanation for fraud than the model of experts; and b) we can sustain that this model, with a common cause, the propensity to frau, is a plausible causal explanation.

Acknowledgements. This project has been supported by the Ministry of Economy and Competitiveness. Project MTM2017–86875-C3–3-R.

References

1. Onwubiko, C.: Fraud matrix: a morphological and analysis-based classification and taxonomy of fraud. *Comput. Secur.* **96**, 101900 (2020)
2. CASE: Study and Reports on the VAT Gap in the EU-28 Member States: 2018 Final Report TAXUD/2015/CC/131 (2018). https://ec.europa.eu/taxation_customs/sites/taxation/files/2018_vat_gap_report_en.pdf
3. REAF- REGAF Asesores fiscales Consejo de Economistas: Reflexiones sobre el fraude fiscal y el problema de las estimaciones:20 propuestas para reducirlo (2017). <https://www.reaf-regaf.economistas.es>
4. Stankevicius, E., Leonas, L.: Hybrid approach model for prevention of tax evasion and fraud. *Procedia – Soc. Behav. Sci.* **213**, 383–389 (2015)
5. Baesen, B., Van Vlasselaer, V., Verbecke, W.: Fraud Analytics Using Predictive, and Social Network Techniques. A Guide to Data Science for Fraud Detection. Wiley (2015)
6. Matos, T., Macedo, J., Lettich, F., Monteiro, J., Renso, C., Perego, R., Nardini, F.: Leveraging feature selection to detect potential tax fraudsters. *Expert Syst. Appl.* **145**, 113128 (2020)
7. González, P., Velásquez, J.: Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Syst. Appl.* **40**(5), 1427–1436 (2013)
8. Martikainen, J.: Data mining in tax administration- using analytics to enhance tax compliance. Aalto University School Business (2012). <https://epub.lib.aalto.fi/en/ethesis/id/13054>
9. González, I.: Analytics and big data: the case of AEAT. *Tax Administration Review*, no. 44, October 2018
10. González, I., Mateos, A.: Social network analysis tools in the fight against fiscal fraud and money laundering. In: Proceedings of the 15TH International Conference on Modelling Decisions for Artificial Intelligence (MDAI 2018) (2018)

11. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kauffman, San Francisco (1988)
12. Glymour, G., Schines, R., Spirtes, R., Kelly, K.: Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modelling. Academic Press, Orlando (1987)
13. Fratello, M., Tagliaferri, R.: Decision trees and random forests. In: Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C. (eds.) Encyclopedia of Bioinformatics and Computational Biology. Academic Press, Orlando (2019)
14. Spearman, C.: General intelligence determined and measured. *Am. J. Psychol.* **15**(201), 93 (1904)
15. Bollen, K.A., Ting, K.-F.: A tetrad test for causal indicators. *Psychol. Methods* **5**(1), 3–22 (2000)



Benchmark of Encoders of Nominal Features for Regression

Diogo Seca¹(✉) and João Mendes-Moreira^{1,2}

¹ LIAAD - INESC TEC, Porto, Portugal

jose.d.seca@inesctec.pt

² FEUP, University of Porto, Porto, Portugal

Abstract. Mixed-type data is common in the real world. However, supervised learning algorithms such as support vector machines or neural networks can only process numerical features. One may choose to drop qualitative features, at the expense of possible loss of information. A better alternative is to encode them as new numerical features. Under the constraints of time, budget, and computational resources, we were motivated to search for a general-purpose encoder but found the existing benchmarks to be limited. We review these limitations and present an alternative. Our benchmark tests 16 encoding methods, on 15 regression datasets, using 7 distinct predictive models. The top general-purpose encoders were found to be Catboost, LeaveOneOut, and Target.

Keywords: Nominal encoders · Categorical encoders · Feature engineering · Mixed-type data · Regression · Supervised machine learning

1 Introduction

In real-world datasets, it is common to find both qualitative and quantitative features. In particular, nominal features pose a problem for many learning algorithms such as neural networks, which are only able to deal with quantitative features.

Prior to modeling, a nominal feature with high cardinality on a mixed-type dataset must be either be dropped or encoded as quantitative values. Dropping the feature could result in the loss of important information, and is, therefore, an uncommon practice. On the other hand, transforming the nominal feature using commonly used methods like One Hot Encoding could lead to growth in the dimensionality of the data, and increase the risk of effects of the “curse of dimensionality” during modeling.

Some novel techniques have been proposed to face this problem and transform all features, both qualitative and quantitative, into new representations. One

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020.

such example is VAEM, a generative auto-encoding neural network [8]. Despite being promising, we couldn't find any other experiments with VAEM other than the work of the authors. Moreover, due to VAEM's computational requirements and long training time, we will not be considering this method within the scope of this study.

Approaches to encoding qualitative features have been studied and benchmarked before. We find the prior benchmarks lacking due to not using a diverse and reasonable amount of datasets for regression [1, 9, 14]. The only prior benchmark that used diverse regression datasets was done by Pargett et al. [11] but lacks some important methods such as the Target encoder, which resulted in top performance in Kaggle competitions [1]. Moreover, we could not find any comprehensive benchmark that was open-sourced.

We aim to correct these issues by providing an open-sourced benchmark that is both comprehensive and supports including new encoding methods as well as new datasets.

1.1 Encoders

Qualitative features, also known as categorical features, include both nominal and ordinal features. Unlike ordinal features, nominal features do not follow a particular order. Our focus is on nominal encoders.

Encoders can also be described as Unsupervised or Supervised. Supervised encoding methods require that the specification of which variables should be used as features, and which variable or variables should be used as target variables.

In particular, supervised techniques may cause target leakage. Target leakage is the result of introducing bias into the modeling process by including information about the target feature, that would not be present during out-of-sample tests. This phenomenon may increase the risk of overfitting, especially for small datasets. In order to prevent overfitting and validate our results, we use two-fold shuffled cross-validation.

In this study, we experiment and evaluate supervised and unsupervised encoders from the scikit-learn category_encoders package [9]. The unsupervised encoders tested include Backward Difference Contrast, BaseN, Binary, Count, Hashing, Helmert Contrast, Ordinal, One-Hot, Polynomial Contrast, and Sum Contrast. The supervised encoders tested include Target Encoding, LeaveOneOut, CatBoost, M-estimator, Generalized Linear Mixed Model, and James-Stein Estimator. Moreover, we also used a Drop Encoder, which consists of simply dropping the nominal features. The Drop Encoder is introduced in the benchmark as the baseline encoder.

Some authors distinguish between determined, algorithmic, and automatic encoders [7]. Determined encodings are characterized as having low running time complexity. Algorithmic techniques are more sophisticated than determined techniques and require intensive computation, which is undesirable for big data projects. Automatic techniques are machine learning algorithms, most frequently neural networks, that encode qualitative data based on a previously learned representation. The most common automatic techniques are also known as entity embedding techniques, e.g.: word2vec, Gamma-Poisson [4, 5],

Table 1. Characterization of the datasets

Dataset	Target	Instances	Features	Numerical	Nominal	Cardinality
codling	dead	299	10	9	1	[7]
nassCDS	injSeverity	26063	14	5	9	[5, 2, 2, 2, 2, 2, 3, 2, 2]
races2000	time	77	4	3	1	[5]
terrorism	nkill.us	45	13	11	2	[4, 4]
midwest	percollege	437	27	23	4	[320, 5, 2, 16]
mpg	cty	234	11	5	6	[15, 38, 10, 3, 5, 7]
msleep	sleep_total	29	9	6	3	[29, 4, 12]
txhousing	sales	7126	9	8	1	[46]
gtcars	mpg.c	46	14	8	6	[18, 25, 4, 2, 9, 5]
water	mortality	61	3	2	1	[2]
ca2006	Bush2004	25	10	8	2	[2, 2]
UKHouse	y1	519	12	10	2	[519, 80]
myeloid	futime	136	7	5	2	[2, 2]
us_rent	estimate	103	4	2	2	[52, 2]
Baseball	years	263	23	17	6	[2, 2, 24, 23, 2, 24]

and GEL [6]. Entity embedding techniques, or simply embedding techniques, are based on the extraction of morphological information. The text that composes the initial features are encoded to a high-dimensional real-valued (sometimes complex-valued) vector space. For this reason, these encoding techniques are sometimes also referred to as vector representations. Similar words or expressions are expected to be at a shorter distance within the vector space than dissimilar words or expressions.

Our study focuses on determined techniques. Determined techniques are more suitable for encoding quantitative data in large datasets because they are less computationally expensive than automatic or algorithmic techniques [7].

We paired several learning algorithms with the qualitative encoders presented by [9], including both encoders for nominal features and for ordinal features. Our suspicion is that ordinal encoders should result in poorer performance relative to nominal encoders.

In the following section, we discuss important considerations in case the user wishes to test a novel encoder, in the context of regression. The study concludes with practical advice on which encoders to choose from.

2 Materials and Methods

In this section, we will go over the most important experimental details. We tested several combinations of datasets, learning algorithms, and encoders, in order to evaluate how the performance of encoders changes when coupled with different learning algorithms and datasets.

2.1 Datasets Used

The datasets described in Table 1 were gathered from a collection of R datasets maintained by Vincent Arel-Bundock [2]. Table 1 notes the datasets after

dropping the instances with missing values and dropping the features with unique identifiers or nominal features with extremely high cardinality, relative to the number of instances. The datasets are specified of number of instances, number of features, number of numerical features, nominal features, and cardinality of nominal features. Among the nominal features, we also included binary nominal (aka dichotomous) features.

2.2 Encoders

The implementation of the presented encoders can be found in Scikit-learn's python package category_encoders (version 2.11), except for DropEncoder, which was custom programmed. Moreover, the only encoder from category_encoders not found in our benchmark is the Weight of Evidence Encoder, as it only applies to binary classification problems [3].

2.3 Learning Algorithms

The following consists of the supervised learning algorithms experimented: Feed-forward Neural Network with 1 layer of neurons (NN-1L), Feed-forward Neural Network with 2 layers of neurons (NN-2L), Random Forest (RF), Gradient Boosted Trees (GBT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Elastic Net (EN). Both neural networks have 100 neurons per layer.

We used the scikit-learn implementations for these learning algorithms [12], due to its maturity and ease of use.

2.4 Performance Metrics

In order to measure the performance of the learned encoders and predictive models upon different regression tasks, we first compute the Root Mean Squared Error (*RMSE*) for each validation set. We then calculate $\bar{\epsilon}$, i.e. the mean *RMSE* for all f validations folds, given: an encoder e , a learning algorithm a , and a dataset d .

$$\bar{\epsilon}(e, a, d) = \frac{1}{F} \sum_f^{F} RMSE(e, a, d, f)$$

We aimed to study how different combinations of encoders and learning algorithms compare to the best performing combination for a given dataset d . We transformed our relative scale of error into an absolute scale, by subtracting the minimum error achieved for each prediction task. After this transformation, error 0 can be viewed as the lowest error achieved for a particular prediction task. We scale these results by dividing them by the interquartile range (*IQR*), in order to allow the comparison of results between datasets.

2.5 Experimental Procedure

The experimental procedure is summarized in the following steps:

1. aggregation of mixed-type datasets, i.e. containing both qualitative and quantitative features;

2. pre-processing of datasets, including removal of features, missing values, and definition of the nominal features to encode;
3. shuffled-k-fold, with folds parameter $k = 2$ and shuffled 15 times, resulting in 30 pairs of {training sample, validation sample};
4. for each pair of samples, standardize the quantitative features of both samples using the means and standard deviations estimated from the training sample;
5. for each pair {training sample, validation sample}, learn several encoders and predictive models from the training sample and measure the quality of the predictions on the validation sample.

$$Score(e, a, d) = \frac{\bar{\epsilon}(e, a, d) - \min_{e,a} \bar{\epsilon}(e, a, d)}{IQR \bar{\epsilon}(e, a, d)}$$

2.6 Resource Monitoring

Two metrics were extracted during the initialization, training, and application of encoders: *Max RAM*, which measures the maximum RAM in megabytes; *CPU Time*, which measures the CPU time in seconds that elapsed.

2.7 Extending the Benchmark

Should the reader be interested in evaluating the performance of a novel encoder, he is able to fork our [our Github repository diogoseca/encoders-benchmark](#) and extend this benchmark with few lines of code.

3 Results and Discussion

We experimented with all combinations of learning algorithms and nominal encoders across 15 datasets, 16 encoders, and 7 learned models. Each dataset resulted in 30 pairs of {training sample, validation sample}.

3.1 Encoders Performance

The scores for each combination of encoders and learning algorithms are described in Table 2. The *MeanScore* is calculated as the mean of the scores for a particular encoder.

In general, the encoders CatBoost, LeaveOneOut, and Target are the top-performing. In the case where the learning algorithm are Neural Networks, be it NN-1L or NN-2L, the GLMM encoder also achieved a decrease in error, comparable to the error when using CatBoost, LeaveOneOut or Target encoders. For tree methods such as GBT or RF, the OneHot encoder should be preferred. Determined encoders like Count and Ordinal result in some of the poorest performance, which is expected, given that they assume the features to be ordinal features, not nominal. The results suggest that generally, the supervised encoding of nominal features results in better performance than to simply drop them.

Table 2. Scores of encoders by learning algorithm, sorted by the MeanScore

Encoder	EN	GBT	KNN	NN-1L	NN-2L	RF	SVM	MeanScore
CatBoost	2.909	0.469	0.859	0.420	0.430	0.467	0.758	0.902
LeaveOneOut	2.902	0.470	0.882	0.419	0.434	0.465	0.775	0.907
Target	2.902	0.473	0.868	0.458	0.447	0.468	0.765	0.911
GLMM	2.909	0.563	0.881	0.372	0.422	0.586	0.784	0.931
MEstimate	2.909	0.581	0.875	0.425	0.467	0.622	0.825	0.958
Polynomial	2.982	0.485	0.895	0.537	0.525	0.521	0.776	0.960
OneHot	2.998	0.427	0.895	0.608	0.568	0.432	0.818	0.964
Binary	2.998	0.477	0.957	0.532	0.559	0.450	0.799	0.967
BaseN	2.998	0.472	0.957	0.556	0.558	0.452	0.799	0.970
Sum	2.914	0.428	1.005	0.582	0.602	0.443	0.841	0.973
JamesStein	2.983	0.559	0.912	0.529	0.504	0.607	0.962	1.008
Hashing	2.999	0.520	1.017	0.632	0.585	0.491	0.845	1.013
Drop	2.999	0.629	0.997	0.501	0.533	0.598	0.904	1.023
BackwardDiff	2.998	0.477	1.254	0.704	0.700	0.466	0.865	1.066
Helmert	3.030	0.436	2.159	2.047	1.772	0.455	2.067	1.709
Ordinal	3.065	0.486	2.086	2.380	1.909	0.467	1.828	1.746
Count	2.802	0.490	1.596	14.147	11.321	0.459	2.529	4.764

The results differ significantly between learning algorithms. Neural networks achieved overall lower error. Elastic Nets achieved overall higher error.

According to the results from Table 3, the encoding of binary features in datasets such as water, ca2006, and myeloid does not show a clear benefit.

While we see a improvement in performance by using CatBoost, LeaveOneOut, or Target encoders when applied to the UKHouse dataset, on the dataset midwest, it would be preferable to simply drop the nominal features. Therefore, our results do not allow us to argue on the benefits of using supervised encoders for high cardinality nominal features.

3.2 Resource Monitoring

Hashing encoder is one of the fastest encoders, as observed in Table 5. However, the HashingEncoder is not recommended due to its high usage of available memory, as observed in Table 4.

Although the GLMM encoder achieved a decrease in error for Neural Networks, we do not recommend using the GLMM encoder, as it is significantly slower than the other encoders in the benchmark, as is observed in Table 5.

3.3 Comparing the Results

A previous benchmark by Vorotyntsev has found the top encoders to be: JamesStein, CatBoost, Target, and LeaveOneOut encoder [14]. These results are consistent with our study, except for the JamesStein encoder, which instead of

ranking as 1, ranks as 11 in our benchmark. This could be due to different implementations of the JamesStein encoder or simply due to the choice of datasets. In accordance with the No free lunch theorem [15], each learning algorithm has a subset of problems to which it is most fitting. The difference between classification tasks and regression tasks between our benchmark and that of Vorotyntsev could explain this discrepancy in results.

Pargent F. points out that the supervised encoders performed best when having regularization [11]. This is consistent with our results, as CatBoost, LeaveOneOut, and Target encoders are all variants of the former, and have regularization techniques so as to avoid overfitting. However, contrary to the findings of Pargent F., the GLMM Encoder did not achieve overall top performance, but instead ranks as fourth, according to the MeanScore. One notable finding is that GLMM achieves top performance only when used with NN-1L and NN-2L, suggesting synergy when it is used together with neural networks methods. Moreover, the GLMM is not efficient in terms of CPU time requirements, as seen in Table 5.

Potdat et al. argue that the BackwardDifference and Sum encoders offer better results than other non-target encoding methods, namely OneHot, Ordinal, Helmert, Polynomial, and Binary encoders. This does not match with the results found in our experiences [13]. The results from Potdat et al. are based on a single small dataset and could be therefore specific to the dataset used.

3.4 On Fairness

By encoding features such as age, sex, or race, we introduce algorithmic bias into the transformed numerical features. Our produced benchmark does not compensate for this bias. If the practitioner seeks fairness, we advise conditioning the results on original features, and not the ones produced by an encoder. For more information on the topic, see the survey by Mehrabi et al. [10].

4 Conclusions

In general, we recommend using the following supervised encoders: CatBoost, LeaveOneOut, and Target. In the case of neural networks, experimenting with GLMM may also be useful, if there are enough computational power and available time. In the case of decision tree learning methods GBT and RF, we advise also experimenting with OneHot encoder, as these algorithms seem to be not as prone to “curse of dimensionality” effects that OneHot encoder typically introduces.

The current benchmark can be improved by adding more regression datasets. Moreover, we suggest that future extensions also include automatic encoders that use natural language processing techniques to quantify morphological information from text. Another interesting extension would be to combine techniques that both quantify morphological information and quantify context information.

Appendix: Tables of Encoder Scores and Resource Monitoring

Table 3. Scores of encoders by datasets, sorted by mean score of each row.

Encoder	Baseball	UKHouse	ca2006	colding	gtcars	midwest	mpg	msleep	myeloid	nassCDS	racess2000	terrorism	txhousing	us_rent	water	MeanScore
CatBoost	0.698	1.007	0.664	1.058	0.902	1.092	0.751	0.912	0.746	0.698	0.868	0.509	2.412	0.579	0.628	0.902
LeaveOneOut	0.721	1.026	0.669	1.071	0.937	1.004	0.793	0.926	0.758	0.697	0.856	0.530	2.409	0.572	0.633	0.907
Target	0.707	1.018	0.678	1.067	1.023	1.081	0.764	0.926	0.728	0.698	0.868	0.500	2.415	0.571	0.627	0.911
GLMM	0.698	1.015	0.668	1.070	0.924	1.279	0.782	1.099	0.738	0.695	0.871	0.457	2.422	0.620	0.627	0.931
MEstimate	0.701	1.030	0.669	1.051	0.953	1.513	0.756	1.059	0.734	0.695	0.872	0.508	2.412	0.777	0.635	0.958
Polynomial	0.849	1.144	0.653	1.073	0.761	0.906	0.751	1.268	0.803	0.830	0.903	0.502	2.459	0.878	0.628	0.960
OneHot	0.784	1.245	0.644	1.072	0.719	1.058	0.690	1.188	0.826	0.832	0.893	0.517	2.492	0.874	0.620	0.964
Binary	0.789	1.212	0.664	1.046	0.902	1.024	0.708	1.168	0.818	0.827	0.886	0.506	2.498	0.828	0.636	0.967
BaseN	0.776	1.214	0.652	1.046	0.947	1.115	0.719	1.125	0.818	0.825	0.878	0.501	2.498	0.813	0.626	0.970
Sum	0.896	1.231	0.646	1.125	0.840	1.002	0.830	1.233	0.850	0.830	0.891	0.489	2.462	0.705	0.572	0.973
JamesStein	0.699	1.037	0.673	1.033	1.081	1.712	0.762	1.169	0.742	0.693	0.885	0.516	2.498	0.983	0.629	1.008
Hashing	0.830	1.090	0.610	1.058	1.218	1.099	0.788	1.177	0.752	1.018	0.890	0.511	2.528	0.983	0.637	1.013
Drop	0.657	0.963	0.709	1.015	0.993	1.032	0.680	0.826	0.719	1.728	0.850	0.515	2.558	0.836	1.266	1.023
BackwardDiff	0.920	1.584	0.664	1.048	1.063	1.258	0.824	1.198	0.785	0.887	0.505	2.472	1.311	0.649	1.066	
Helmert	2.656	4.486	0.665	1.344	1.532	2.682	1.477	1.764	0.849	0.828	0.926	0.498	2.655	2.705	0.575	1.709
Ordinal	1.318	6.556	0.647	1.018	1.942	2.385	1.503	1.689	0.715	0.773	0.896	0.497	2.694	2.922	0.637	1.746
Count	3.445	1.155	0.856	1.136	2.088	1.639	3.437	1.115	2.032	46.937	1.119	0.520	3.714	1.546	0.713	4.764

Table 4. Maximum RAM (megabytes) used by the encoders for each dataset, sorted by the mean of each row.

Encoder	Baseball	UKHouse	ca2006	colding	gtcars	midwest	mpg	msleep	myloid	nassCDS	races2000	terrorism	txhousing	us-rent	water
BackwardDiff	166.6	166.5	165.9	101.0	165.9	165.6	165.7	165.8	166.5	166.2	165.1	165.8	166.7	165.9	
BaseN	164.6	164.4	164.4	100.7	164.3	164.0	164.1	164.1	164.4	165.6	164.0	163.8	164.1	164.5	164.3
Binary	164.7	164.6	164.5	101.1	164.2	164.2	164.4	164.4	164.6	165.6	164.0	164.0	164.1	164.4	164.4
CatBoost	161.2	161.1	161.1	101.3	161.2	161.2	161.3	161.3	161.3	161.5	161.0	161.0	161.3	161.2	161.2
Count	161.1	161.0	160.8	100.9	160.8	160.5	160.7	160.6	161.0	161.4	160.6	160.4	160.7	161.0	160.8
Drop	145.6	145.5	145.4	97.2	145.4	145.3	145.3	145.3	145.5	145.2	145.2	145.2	145.4	145.5	145.4
GLMM	130.2	130.2	129.7	102.0	129.7	118.3	118.6	118.4	130.2	120.7	118.2	118.3	129.7	130.2	129.7
Hashing	1526.0	1525.5	1524.6	939.0	1524.4	1522.5	1521.6	1523.8	1523.5	1621.0	1534.8	1521.8	1531.4	1525.8	1524.2
Helmut	163.6	163.3	162.6	101.0	162.7	162.4	162.4	162.6	162.5	162.7	161.7	161.8	162.6	163.2	162.7
JamesStein	160.2	160.0	159.9	101.1	159.8	159.3	159.4	159.5	160.0	160.1	159.0	159.1	159.7	160.1	159.9
LeaveOneOut	159.8	159.7	159.7	101.3	159.4	159.3	159.4	159.4	159.6	164.7	159.2	159.1	159.3	159.8	159.6
MEstimate	160.4	160.2	160.2	101.3	160.1	160.1	159.6	159.7	159.8	160.3	165.0	159.2	159.4	160.3	160.1
OneHot	168.4	168.4	168.1	100.7	167.9	167.7	167.7	167.9	167.9	168.1	166.5	168.0	168.4	168.1	168.1
Ordinal	161.5	161.3	161.3	100.7	161.3	160.8	160.8	160.9	161.4	161.3	160.7	160.7	161.0	161.4	161.2
Polynomial	164.7	164.6	164.0	101.5	163.9	163.8	163.8	163.8	164.6	163.0	162.1	162.0	163.9	164.7	164.0
Sum	162.9	162.8	162.1	100.8	162.1	161.8	161.9	162.0	162.7	162.4	161.2	161.3	162.0	162.8	162.1
Target	160.1	160.0	159.8	101.0	159.8	159.5	159.6	159.6	160.0	165.0	159.2	159.2	159.7	160.1	159.8

Table 5. CPU time (seconds) used by the encoders for each dataset, sorted by the mean of each row.

Encoder	Baseball	UKHouse	ca2006	colding	gtcars	midwest	mpg	msleep	myloid	nassCDS	races2000	terrorism	txhousing	us-rent	water
BackwardDiff	0.844	0.690	0.682	0.655	0.857	0.679	0.997	0.767	0.678	1.525	0.740	0.621	0.652	0.807	0.705
BaseN	0.858	0.690	0.714	0.664	0.918	0.750	1.025	0.775	0.689	1.123	0.927	0.644	0.680	0.799	0.894
Binary	0.822	0.614	0.655	0.610	0.816	0.610	0.947	0.739	0.677	1.051	0.910	0.618	0.626	0.677	0.853
CatBoost	0.852	0.673	0.702	0.662	0.911	0.708	0.942	0.772	0.683	1.184	0.681	0.628	0.647	0.734	0.689
Count	0.550	0.704	0.736	0.668	0.629	0.679	0.645	0.778	0.736	0.708	0.720	0.687	0.697	0.779	0.734
Drop	0.583	0.640	0.651	0.656	0.652	0.560	0.673	0.679	0.679	0.648	0.693	0.633	0.658	0.708	0.703
GLMM	9.783	2.106	0.915	1.279	2.447	3.263	3.413	1.130	2.098	14.583	1.119	1.568	1.138	4.340	0.754
Hashing	0.638	0.710	0.704	0.679	0.737	0.560	0.763	0.756	0.739	0.718	0.792	0.689	0.729	0.770	0.779
Helmut	0.802	0.686	0.706	0.671	0.879	0.660	0.880	0.758	0.712	0.943	0.713	0.660	0.654	0.742	0.723
JamesStein	0.694	0.587	0.633	0.590	0.761	0.573	0.773	0.683	0.617	0.830	0.638	0.587	0.605	0.647	0.639
LeaveOneOut	0.858	0.703	0.731	0.664	0.873	0.714	0.927	0.755	0.718	0.989	0.714	0.667	0.681	0.737	0.731
MEstimate	0.666	0.638	0.638	0.774	0.748	0.547	0.750	0.659	0.680	0.840	0.817	0.637	0.655	0.834	
OneHot	0.864	0.672	0.719	0.670	0.866	0.649	0.865	0.734	0.718	0.913	0.730	0.646	0.697	0.773	0.754
Ordinal	0.628	0.806	0.836	0.728	0.699	0.556	0.736	0.845	0.818	0.801	0.777	0.742	0.747	0.865	0.799
Polynomial	0.774	1.579	0.661	0.655	0.829	1.845	0.940	0.753	0.653	1.079	0.708	0.621	0.644	0.696	0.670
Sum	0.836	0.691	0.715	0.667	0.892	0.780	0.928	0.767	0.717	1.027	0.734	0.676	0.651	0.758	0.726
Target	0.824	0.658	0.699	0.656	0.814	0.639	0.966	0.775	0.699	0.951	0.795	0.633	0.717	0.757	0.714

References

1. An, S.: 11 categorical encoders and benchmark, August 2020. <https://kaggle.com/subinium/11-categorical-encoders-and-benchmark>
2. Arel-Bundock, V.: A collection of datasets originally distributed in various R packages, May 2020. <https://vincentarelbundock.github.io/Rdatasets/index.html>
3. Bhalla, D.: Weight of evidence (WOE) and information value (IV) explained, March 2015. <https://www.listendata.com/2015/03/weight-of-evidence-woe-and-information.html>
4. Cerda, P., Varoquaux, G.: Encoding high-cardinality string categorical variables. *IEEE Trans. Knowl. Data Eng.* **1** (2020)
5. Cerda, P.R.: Statistical learning with high-cardinality string categorical variables. Ph.D. thesis, Université Paris-Saclay (2019)
6. Golinko, E., Zhu, X.: Generalized feature embedding for supervised, unsupervised, and online learning tasks. *Inf. Syst. Front.* **21**(1), 125–142 (2019)
7. Hancock, J.T., Khoshgoftaar, T.M.: Survey on categorical data for neural networks. *J. Big Data* **7**(1), 28 (2020)
8. Ma, C., Tschiatschek, S., Hernández-Lobato, J.M., Turner, R., Zhang, C.: VAEM: a deep generative model for heterogeneous mixed type data. [arXiv:2006.11941](https://arxiv.org/abs/2006.11941), June 2020
9. McGinnis, W.D., Siu, C., Andre, S., Huang, H.: Category encoders: a scikit-learn-contrib package of transformers for encoding categorical data. *J. Open Source Softw.* **3**(21), 501 (2018)
10. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. [arXiv:1908.09635](https://arxiv.org/abs/1908.09635), August 2019
11. Pargent, F., Bischl, B., Thomas, J.: A benchmark experiment on how to encode categorical features in predictive modeling (2019)
12. Pedregosa, F., et al.: Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
13. Potdar, K., Pardawala, T.S., Pai, C.D.: A comparative study of categorical variable encoding techniques for neural network classifiers. *IJCAI* **175**(4), 7–9 (2017)
14. Vorotyntsev, D.: Benchmarking categorical encoders - towards data science, July 2019. <https://towardsdatascience.com/benchmarking-categorical-encoders-9c322bd77ee8>
15. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**(1), 67–82 (1997)



Optimizing Model Training in Interactive Learning Scenarios

Davide Carneiro^{1,2(✉)}, Miguel Guimarães¹, Mariana Carvalho¹,
and Paulo Novais²

¹ CIICESI, ESTG, Politécnico do Porto, Porto, Portugal
`{dcarneiro,8150520,mrc}@estg.ipp.pt`

² Algoritmi Center/Department of Informatics, University of Minho, Braga, Portugal
`pjon@di.uminho.pt`

Abstract. In the last years, developments in data collection, storing, processing and analysis technologies resulted in an unprecedented use of data by organizations. The volume and variety of data, combined with the velocity at which decisions must now be taken and the dynamism of business environments, pose new challenges to Machine Learning. Namely, algorithms must now deal with streaming data, concept drift, distributed datasets, among others. One common task nowadays is to update or re-train models when data changes, as opposed to traditional one-shot batch systems, in which the model is trained only once. This paper addresses the issue of *when* to update or re-train a model, by proposing an approach to predict the performance metrics of the model if it were trained at a given moment, with a specific set of data. We validate the proposed approach in an interactive Machine Learning system in the domain of fraud detection.

Keywords: Interactive Machine Learning · Meta-learning · Error prediction

1 Introduction

Machine Learning is a scientific field that is constantly growing and evolving, and its contributions in different domain areas are obviously significant. The development of new and improved algorithms, frameworks and tools in machine learning, alongside with computational evolution, led to the rise of new successful applications.

Interactive Machine Learning (iML) [4] is one research area that has received a lot of attention in recent years [8] in the most distinct domain areas. For instance, in [7] the authors discussed the effectiveness of the IML-“human-in-the-loop” previously developed by the authors in [5] using the Ant Colony Optimization algorithm along with the Traveling Salesman Problem. In healthcare, [5] shows how one can take advantage of using interactive machine learning in small sets of data (with insufficient training samples), instead of using traditional

machine learning algorithms. Specific applications exist, including in the annotation of medical imaging [2], in biomedical images [9], in knowledge discovery in bioinformatics [6], or in music [13].

In iML, unlike in traditional Machine Learning, learning is a continuous process that relies on the contribution of Human experts [1]. Humans keep up with the whole process, analyze and evaluate the returned predictions of a specific trained model, providing feedback that would be incorporated into the new versions of that same model. The human expert feedback could consist of new variables to be added to the model, or removing irrelevant ones. This kind of interactive approach usually leads to better model performance indicators.

One key characteristic of iML systems is thus that the data, the statistical properties of the variables or even the structure of the dataset changes over time. This is, however, not exclusive to iML systems and happens in other applications of ML such as in learning from streaming data [10] or from data with concept drift [14]. The main consequence of this change in the data is that existing ML models quickly become outdated and no longer represent all the relevant patterns in the data.

Two challenges arise in these situations: *how* to update a model and *when* to update a model. The former may include completely retraining a model or only parts of it. This was addressed by the research team in previous work [11, 12]. The latter is the main focus of this paper. This paper therefore proposes a method that can, in the future, be fully automated, to determine objectively if it is worth it to update or retrain a model at a given moment in time, given the statistical properties of the dataset and the expected computational complexity of doing it. The key idea is to rely on meta-learning to predict the key performance indicators of the model (e.g. RMSE, MAE, F1-score) or the training time, as a measure of computational complexity. This work is, therefore, not targeted at specific Human users. Instead, the main goal is for it to be integrated into automated iML systems, so that models can be updated when appropriate instead of on a regular and pre-determined basis. We expect that this can result on a more efficient use of computational resources, especially on big data scenarios.

We validate the proposed approach in an existing iML system previously developed by the research team in the domain of financial fraud detection, that is also briefly described in this paper.

2 Background

The approach proposed in this paper, described in detail in Sects. 3 and 4, was devised to solve one challenge that emerged during the development and implementation of an interactive Machine Learning system for Fraud Detection: to determine when is the correct moment to update a model. This section describes this system in a from a high-level point of view (Fig. 1) since it is not the focus of this paper.

The system has as main data input SAF-T (Standard Audit Files for Tax) files. The SAF-T file is an XML file describing accounting data from organizations, that is sent to national tax authorities and/or external auditors for audit

and compliance purposes. These files go through a feature extraction process that creates some relevant features for audit, and through a Rule-based System whose main goal is to further enrich the data based on a set of rules. At the end of this process data is stored in the so-called *unlabeled dataset*.

From this point on resides the interactive nature of the system. The auditor picks instances from the unlabeled data, analyzes them, provides her/his own structured feedback, and saves the changes. Auditor feedback may include changing the values in the database and/or providing details and a justification about the rationale followed. Cases that have been processed by the auditor are regarded as having been validated by an expert and are then added to the *labeled dataset*. This dataset constitutes the input for the training of models that can predict, for instance, the likeliness of fraud of a given instance.

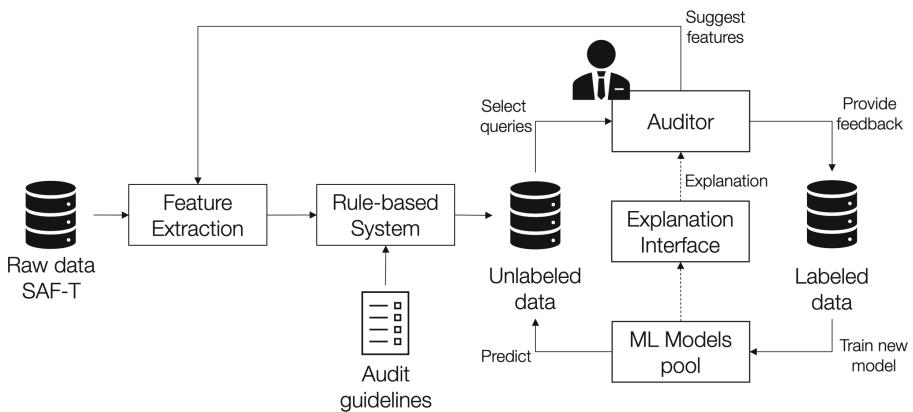


Fig. 1. Overview of the interactive learning system for fraud detection.

The auditor can also suggest new features and new values for existing features (when they are enumerations) which will, if approved, be added to the feature extraction process. These user-defined features have however a particularity: they may not result from the feature extraction phase (because they do not depend directly on the input data, for instance) they can be “guessed” by specifically trained models. That is, once there is enough labeled data, models can be trained to predict the value of each of these user-defined variables. These predictions are later validated by the auditors. This means that in any given time there are multiple models in the so-called model-pool: one for predicting the main dependent variable (likeness of fraud) and others for predicting the values of specific user-defined variables. The system also includes an explanatory interface, for creating human-friendly justifications for the predictions, developed in previous work [3].

One key aspect about this system is thus that the labeled dataset changes over time, as auditors provide their feedback. This has as main implications that

models run the risk of becoming outdated. One solution to this problem would be to frequently re-train these models, in order to keep them up-to-date. However, given the potential number of models and the increasing amount of data, this operation becomes computationally very expensive with time. Moreover, most of the times it can happen that the new data is not significantly different from the existing one, and the new models are also similar to the previous ones in terms of performance. Training new models in this situation is thus a waste of resources.

The approach followed for addressing this issue consists in trying to predict the key performance indicators of a new model before its training, so that it is only trained if the statistical properties of the new data promise to result in a potentially better model. This approach is described in the following two sections.

3 Methodology

Interactive Machine Learning systems, such as the one described in Sect. 2 must deal with additional issues when compared to traditional one-shot batch learning systems. Some of these challenges result from the dynamic nature of the data and the models. That is, models need to be updated with some frequency to adapt to changes in the data. One of the issues is thus *when* to update models, whether it is by training a new model or by updating parts of it.

A small period between model updates may result in a more dynamic system, one that adapts faster to changes in the data, but also one more sensitive to noise. It will also be a more costly system in the sense that it will require more computational resources and machine time for the frequent training of the models. On the other hand, a large period between updates may result in a system that, while requiring less computational resources, is also slower to respond to changes in the data. There is thus a trade-off involving computational resources, sensitivity to noise and the ability of the system to adapt to changes.

In this paper we argue that the most appropriate criterion for deciding if and when to update a model is a prediction of the error measure of the projected new model. The rationale is thus that resources should only be employed in the training of the model if that is expected to result in a model with significantly smaller error.

Indeed, the existence of new data is not, by itself, a guarantee that a new model will be better than the previous one. For instance, the new data may be very similar to existing data (so it contributes with no new patterns), or the new data may have problems that actually decrease the overall quality, such as biases or missing data.

In order to predict the error of a new model beforehand, we propose an approach based on meta-learning, as follows. The key idea is to use a so-called meta-model that predicts the expected error of training a given model on a dataset with given statistical properties. Statistical properties are called meta-features and detail aspects related to the shape of the dataset (e.g. number of

rows, number of columns, ratio between these), statistical properties of variables (e.g. mean kurtosis, mean skewness), information-theoretic (e.g. class entropy, mean entropy of attributes, noise signal ration), among others.

Meta-features constitute the independent variables of the so-called meta-dataset and are extracted from the training of a model with a specific dataset. That is, they describe the characteristics of the dataset on the moment of training as well as the configuration of the algorithm used. The dependent variables are the performance metrics of the resulting model (e.g. RMSE, MAE, F1-score) as well as the training time (as a measure of complexity). Each instance of the meta-dataset thus describes the conditions in which a model was trained and how well that model behaved during validation.

Once the meta-dataset is large enough, a meta-model can be trained to predict any of the performance metrics and/or the expected training time of the model. This information can then be used by the system, automatically or from the perspective of Human decision-support, to determine if a model should be updated at a given time. If the new model is not expected to be significantly better than the current one and/or the expected training time is too large, then maybe it is not worth to train it at all.

Figure 2 depicts, in more detail, the components of the system that are involved in the training and use of the meta-model. The process is run with a given periodicity, which can be given in terms of time (e.g. daily, weekly) or in terms of experience (e.g. at every new 1000 instances in the labeled dataset). When the process starts it attempts to train a new model using the labeled data.

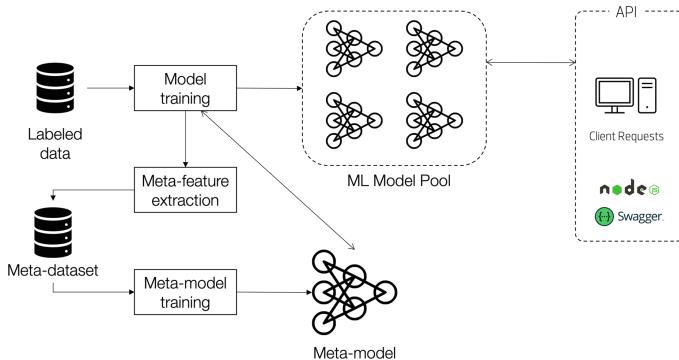


Fig. 2. Detail of the components in the system involved in the training and use of the meta-model.

Before doing so, the system checks if there is a meta-model trained. If no meta-model exists, the system trains a new model and adds it to the pool of models, as would usually happen. However, if there is a meta-model, the system asks for a prediction of the expected measure of error and/or the expected time of training. To this end, the system provides as input the intended algorithm

to train the model as well as the meta-features that currently characterize the dataset and the intended metric. The meta-model responds with the expected metric of error/complexity. If multiple metrics are required (e.g. RMSE, MAE, training time) multiple meta-models must be trained, one for each metric. However, the meta-dataset is the same, only the selection of the dependent variable changes.

Based on this response, the system will decide whether or not to train a new model. Whenever a new model is trained, its meta-features are extracted and added to the meta-dataset, thus enriching the system's ability to make predictions regarding the quality of future models. The models available in the pool can be used by external and/or internal entities, to make predictions given the appropriate input data.

The following section describes the validation of the proposed approach in a specific case study in the domain of fraud detection, and the results obtained.

4 Validation and Results

This section describes how the proposed system was validated in an iML application in the domain of fraud detection. To this end, a real proprietary fraud detection dataset of 20.000+ validated instances is used. The dataset contains 28 variables, of which 4 were proposed by Human users (named here f1 to f4) and 1 is the variable that denotes the likeliness of fraud. The experiments were performed on a system with a 2.7 GHz quad-core processor and 16 GB 2133 MHz RAM.

It is important to note that in this work, variables f1 to f4 of the dataset are, in some cases, the independent variable, and in other the dependent variable. The system trains one classification model to predict each of the Human-defined variables (as the other variables are extracted from the raw data), and a regression model to predict the likeliness of fraud. Thus, multiple models may exist simultaneously in the pool. The training of each of these models contributes with data for the meta-dataset.

To validate this approach, the following method was followed. Multiple models were initially trained with a subset of the labeled data, and the previously described meta-features extracted. This generated a meta-dataset that was used to train a meta-model to predict RMSE. A Random Forest algorithm was used to train the meta-model, composed of 20 trees. This meta-model was then used throughout time to predict the error before the training of the model. In this case, nonetheless, models were always trained, in order to obtain the real value of RMSE for validation purposes.

In order to facilitate the validation of the system, the 20.000+ instances of labeled data were streamed to simulate its arrival over time. In order to assess the previously mentioned trade-off between the ability to adapt, use of computational resources and sensitivity to noise, models were trained with five different periodicities (measured in number instances): $T = 200$, $T = 500$, $T = 1000$, $T = 2500$ and $T = 5000$.

Figure 3 details how error varies in each case over time. In this case, the error is measured as the difference between the observed and predicted values, so it includes both positive and negative values. The Figure shows how there is an initial significant variation of error at the beginning of each simulation, especially in the three smaller periodicities. The error then tends to stabilize around zero as more data arrives. This shows how the system is possibly too sensitive to changes in the data when there is few data, pointing out that higher periods for model update should perhaps be used. It must also be noted that in Fig. 3 the scales are different in each plot, which may partially hide the significance of the differences. Original scales were however maintained to improve the visualization of the evolution of error over time.

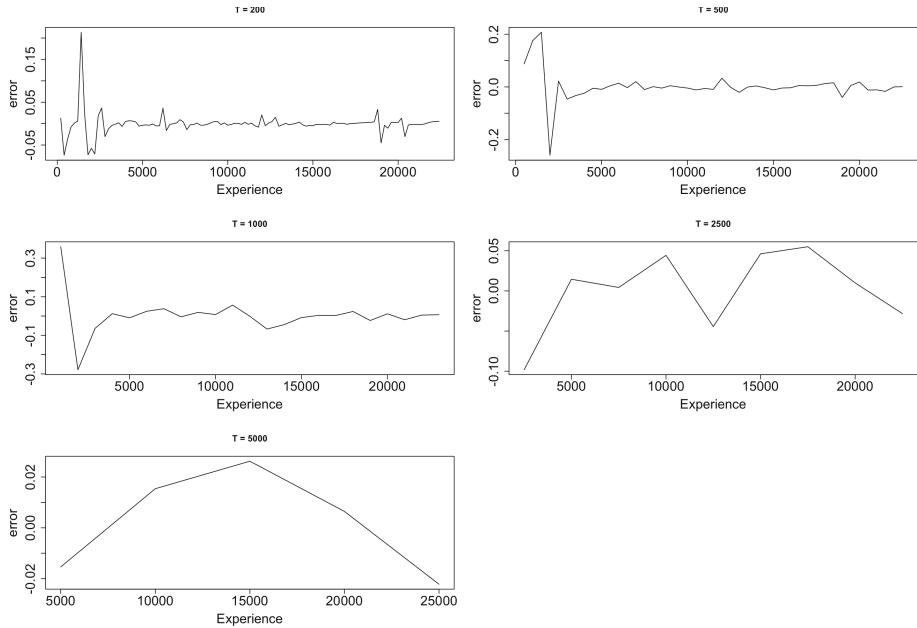


Fig. 3. Evolution of prediction error over time, with different periodicities.

Figure 4, on the other hand shows, for each case, a plot of the observed values of RMSE against the predicted ones, together with the line of perfect correlation in each case. The Figure shows that, in general, predictions are very good. Table 1 details some metrics for these experiments, namely the RMSE, the MAE and the standard deviation of the observed error. The table shows that lower prediction errors are obtained with a $T = 200$ and a $T = 5000$. However, it also shows that the variation of error is much higher for $T = 200$ than for $T = 5000$. This points out that it is preferable to train models at a slower rate, which decreases sensibility to noise and minimizes use of computational resources. This is valid,

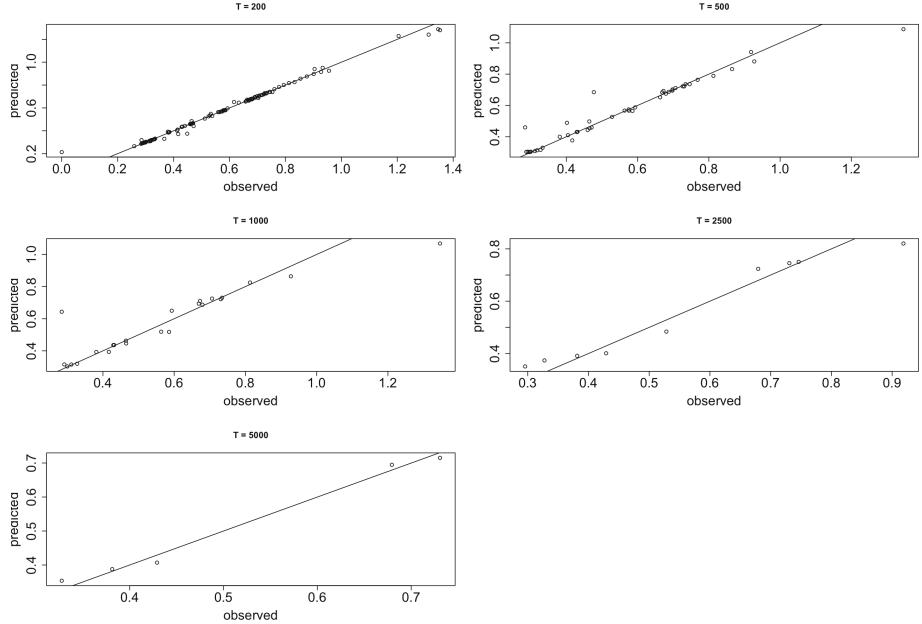


Fig. 4. Observed vs. predicted RMSE values, with different periodicities.

Table 1. RMSE, MAE and error deviation for the five different periodicities considered (in number of instances).

T	RMSE	MAE	σ_{error}
200	0.026	0.010	0.024
500	0.059	0.027	0.053
1000	0.099	0.047	0.089
2500	0.047	0.038	0.029
5000	0.018	0.017	0.008

however, for this scenario in which there is no concept drift in the data. If it were, results could be different and would potentially favor smaller periodicities.

5 Conclusions

In this paper we addressed a challenge that is ever more frequent in today's ML applications: to decide when is it a good moment to update a given ML model. Researchers and ML engineers must face this challenge especially in domains in which data, their statistical properties, or even its structure change over time. This is a fairly common issue nowadays, in which learning is increasingly done from streaming data rather than from batch data. Moreover, data is increasingly

more dynamic, resembling the changes that occur ever faster in all aspects of society. Examples in which this decision problem exists are systems based on interactive Machine Learning, systems that learn from streaming data, or system in which the data contains concept drift.

The approach proposed in this paper is based on the use of meta-learning. The meta-model, based on data extracted from the training of multiple ML models, allows to predict the performance metrics of a potential model as well as its training time. As detailed in Sect. 4, which focused on RMSE, error metrics can be predicted with a fairly good accuracy. The approach was validated with a real proprietary dataset, in an iML system developed in the domain of fraud detection.

In future work we will apply the system in a scenario with live streaming data with concept drift, as this is the best application scenario for this work. This was not done in this paper as the goal was to validate the proposed approach in different scenarios (e.g. periodicity of model updates). To this end, we simulated the streaming of the data in order to simulate its arrival to the system and to assess the performance of the meta-model throughout time. We will also take into consideration the possibility of dropping older data instead of using the whole dataset, and assess how that impacts the performance of the model and of the meta model, as a way to react appropriately to changes in the data over time.

Acknowledgments. This work was supported by the Northern Regional Operational Program, Portugal 2020 and European Union, through European Regional Development Fund (ERDF) in the scope of project number 39900 - 31/SI/2017, and by FCT – Fundação para a Ciência e Tecnologia within projects UIDB/04728/2020 and UID/CEC/00319/2019.

References

1. Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T.: Power to the people: the role of humans in interactive machine learning. *AI Mag.* **35**(4), 105–120 (2014)
2. Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., et al.: Ilastik: interactive machine learning for (bio) image analysis. *Nat. Methods* **16**, 1–7 (2019)
3. Carneiro, D., Silva, F., Guimarães, M., Sousa, D., Novais, P.: Explainable intelligent environments. In: International Symposium on Ambient Intelligence, pp. 34–43. Springer (2020)
4. Fails, J.A., Olsen Jr., D.R.: Interactive machine learning. In: Proceedings of the 8th International Conference on Intelligent User Interfaces, pp. 39–45 (2003)
5. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inform.* **3**(2), 119–131 (2016)
6. Holzinger, A., Jurisica, I.: Knowledge discovery and data mining in biomedical informatics: the future is in integrative, interactive machine learning solutions. In: Interactive Knowledge Discovery and Data Mining in Biomedical Informatics, pp. 1–18. Springer (2014)

7. Holzinger, A., Plass, M., Kickmeier-Rust, M., Holzinger, K., Crișan, G.C., Pintea, C.M., Palade, V.: Interactive machine learning: experimental evidence for the human in the algorithmic loop. *Appl. Intell.* **49**(7), 2401–2414 (2019)
8. Jiang, L., Liu, S., Chen, C.: Recent research advances on interactive machine learning. *J. Vis.* **22**(2), 401–417 (2019)
9. Khan, N.M., Abraham, N., Hon, M., Guan, L.: Machine learning on biomedical images: Interactive learning, transfer learning, class imbalance, and beyond. In: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 85–90. IEEE (2019)
10. Krawczyk, B., Minku, L.L., Gama, J., Stefanowski, J., Woźniak, M.: Ensemble learning for data stream analysis: a survey. *Inf. Fusion* **37**, 132–156 (2017)
11. Ramos, D., Carneiro, D., Novais, P.: evoRF: an evolutionary approach to random forests. In: International Symposium on Intelligent and Distributed Computing, pp. 102–107. Springer (2019)
12. Ramos, D., Carneiro, D., Novais, P.: Using a genetic algorithm to optimize a stacking ensemble in data streaming scenarios. *AI Commun.* (Preprint) **1–14** (2020)
13. Visi, F.G., Tanaka, A.: Interactive machine learning of musical gesture. arXiv preprint [arXiv:2011.13487](https://arxiv.org/abs/2011.13487) (2020)
14. Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. *Mach. Learn.* **23**(1), 69–101 (1996)



Early Prediction of student's Performance in Higher Education: A Case Study

Mónica V. Martins¹ , Daniel Toledo¹, Jorge Machado¹ ,
Luís M. T. Baptista¹ , and Valentim Realinho^{1,2}

¹ Polytechnic Institute of Portalegre (IPP), Portalegre, Portugal
mvmartins@ipportalegre.pt

² VALORIZA - Research Center for Endogenous Resource Valorization, Portalegre, Portugal

Abstract. This work aims to contribute to the reduction of academic failure at higher education, by using machine learning techniques to identify students at risk of failure at an early stage of their academic path, so that strategies to support them can be put into place. A dataset from a higher education institution is used to build classification models to predict academic performance of students. The dataset includes information known at the time of student's enrollment – academic path, demographics and social-economic factors. The problem is formulated as a three category classification task, in which there's a strong imbalance towards one of the classes. Algorithms to promote class balancing with synthetic oversampling are tested, and classification models are trained and evaluated, both with standard machine learning algorithms and state of the art boosting algorithms. Our results show that boosting algorithms respond better to the specific classification task than standard methods. However, even these state of the art algorithms fall short in correctly identifying the majority of cases in one of the minority classes. Future directions of this study include the addition of information regarding student's first year performance, such as academic grades from the first academic semesters.

Keywords: Academic performance · Machine learning in education · Imbalanced classes · Multi-class classification · Boosting methods

1 Introduction

One of the challenges that higher education institutions around the globe face is to effectively deal with the different student's learning styles and different academic performances, as a means to promote student's learning experience and institution's formative efficiency. The ability to predict and anticipates student's potential difficulties is of interest for the institutions that aim to build strategies to provide support and guidance to students that might be at risk of academic failure or dropout. At the same time, large amount of data is collected each year by the institutions, including information regarding the academic path of the student, as well as demographics and socio-economic factors. The two combined factors make this a fertile ground for the contribution of machine learning approaches to predict student's performance.

In this study, data from Polytechnic Institute of Portalegre (IPP), Portugal, is used to build machine learning classification models to predict students that might be at risk of failing to succeed in finishing their degrees in due time. The main goal is to provide a system that allows to identify, at a very initial stage, students with potential difficulties in their academic path, so that strategies to support the students can be put into place.

Some of the aspects of the work here presented that differ from a number of similar works found in the literature are as follows: (i) it doesn't focus on any specific field of study, because the focus is to build a system that generalizes to any course of IPP. Therefore, the dataset includes information from students enrolled in the several courses of the four different schools belonging to IPP; (ii) it only relies on information available at the moment of enrollment, because the focus is to develop a system that helps to segment students as soon as possible from the beginning of their path at higher education. This means that no information regarding academic performance after enrolment is used; (iii) differently from the usual approach of restricting the model categories to failure / success, it's also used a third intermediate class (relative success), because the kind of interventions for academic support and guidance might be quite different for students who are in moderate risk from those who are at high risk of being unsuccessful. As a result, methods to deal with the unbalanced nature of the resulting classes must be considered; (iv) besides using standard classification models, it also uses state of the art boosting algorithms to build the classification models.

The remainder of this paper is structured as follows: Sect. 2 presents a brief review of the literature; Sect. 3 describes the methodology, including a description of the data, the methods used to deal with the unbalanced dataset and the procedures for training and evaluation of the classification models; Sect. 4 presents and analyses the results and Sect. 5 indicates some directions for future work.

2 Related Work

The task of predicting academic success in higher education is not a new one, and many researchers have tried different approaches, using different models and different information. Mostly, researchers and higher education institutions are interested in being able to predict if a student is at risk of not completing his program, or of dropping out, because this information might be valuable for putting into place strategies to help those students move forward. In reviews such as [1–4] one can find extensive information regarding the different approaches that have been used to study this issue. Here, we review a few very recent works (published in the last 4 years) that find similarities with the work presented in this paper.

Beaulac and Roosenthal [5] analyze a large data set (38 842 students) from a large university in Canada to predict academic success using Random Forests (RF). The authors use the first few courses attempted and grades obtained by students in order to predict whether the student will complete their program; and if yes, which major they will complete. For the prediction of program completion, an overall 79% accuracy is obtained, with 91% for the class of students who completed their program, and 53% for the students who didn't. Regarding the prediction of the major, 47% accuracy was obtained.

Hoffait and Schyns [6] use a dataset of 6845 students and standard classification methods (RF, Logistic Regression (LR) and Artificial Neural Networks(ANN)) to identify freshmen's profiles likely to face major difficulties to complete their first academic year. The obtained accuracy for the majority class is about 70%, and for the minority class less than 60%, regardless of the algorithm used. They then use RF to develop a strategy to improve the accuracy of the prediction for some classes of major interest. The developed approach does not always lead to an increase in the identification of the number of students at risk.

Miguéis et al. [7] use the information available at the end of the first year of students' academic path to predict their overall academic performance, inferring academic success both from the average grade achieved and the time taken to conclude the degree. Their prediction models use information regarding demographics and social factors as well as academic measures, including assessments from first year courses. They use a dataset of 2459 students from a European Engineering School to build several models using Support Vector Machines, Naïve Bayes, Decision Trees (DT), RF, Bagging Decision Trees and Adaptive Boosting Decision Trees, obtaining the higher scores (overall accuracy of 96%) with Random Forests and Adaptive Boosting Decision Trees.

One of the common problems in classification student's success or dropout is class imbalance. Class imbalance happens when one or more of the classification categories have significant lower number of records than a majority class. This might result in a high prediction accuracy driven by the majority class at the expense of very poor performance on the minority classes. In the case of student's performance, this is a common problem because only a minority of students will underperform or drop out. Nevertheless, these are the classes of students that researchers aim to best identify.

In [8] Thammasiri et al. conduct a study using different class balancing strategies and several standard classification methods to predict dropout in a dataset with 21654 students. They compare class balancing techniques based on random under sampling, random oversampling, or synthetic oversampling. Best results are obtained with this last approach, named synthetic minority over-sampling technique (SMOTE) [9]. This approach has indeed shown to successfully tackle the imbalanced classification issue in different domains [10] and will be further explained in Sect. 3.2.

3 Methodology

In this section we present the dataset, the methods used to deal with the imbalanced nature of the data, and the methodology used to build and evaluate the classification models.

3.1 Data

In this study we use institutional data (acquired from several disjoint databases) related to students enrolled in undergraduate courses of Polytechnic Institute of Portalegre, Portugal. The data refers to records of students enrolled between academic years 2008/09–2018/2019 and from different undergraduate degrees, such as agronomy, design, education, nursing, journalism, management, social service and technologies.

We performed a rigorous data preprocessing to handle data from anomalies, unexplainable outliers and missing values, and dropped records that couldn't be classified as explained below (last 3 or 4 academic years, depending on the course duration). The final dataset consisted of 3623 records and 25 independent variables.

The data contains variables related to demographic factors (age at enrollment, gender, marital status, nationality, address code, special needs) socio-economic factors (student-worker, parent's habituations, parent's professions, parent's employment situation, student grant, student's debt) and student's academic path (admission grade, retention years at high school, order of choice for enrolled course, type of course at high school). We limit the academic information to factors observable prior to registration, excluding internal assessments after enrollment.

Each record was classified as Success, Relative Success and Failure, depending on the time that the student took to obtain her degree. Success means that the student obtained the degree in due time; Relative Success means that the student took until three extra years to obtain the degree; Failure means that the student took more than three extra years to obtain the degree or doesn't obtain the degree at all. This somehow corresponds to three levels of risk: 'low-risk' students with high probability of succeeding; 'medium-risk' students, for whom the measures taken by the institution might contribute to success; and the 'high-risk' students, who have a high probability of failing. The distribution of the records among the three categories is shown in Fig. 1.

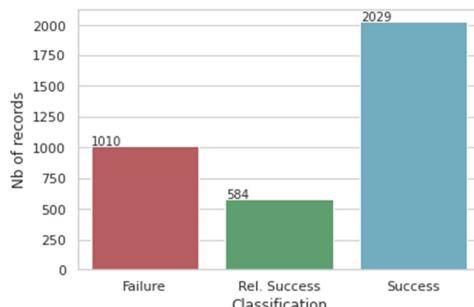


Fig. 1. Distribution of student's records among the three categories considered for academic success.

The distribution of the records among the three categories is imbalanced, with two minority classes, "Failure" and "Relative Success". "Failure" accounts for 28% of total records, and "Relative Success" accounts for 16% of total records, while the majority class, "Success", accounts for 56% of the records. The classes that we most aim to correctly identify are the minority ones, since the students from these classes are the ones that might benefit from planned interventions for academic support and guidance. How we tackled this multi-class imbalanced classification task will be explained in the next sections.

3.2 Data Sampling Techniques

Sampling strategies are often used to overcome the class imbalance problem, either eliminating some data from the majority class (under-sampling) or duplicating data of the minority class (over-sampling) or adding some artificially generated data to the minority class. The under-sampling techniques have the disadvantage of reducing the size of the data set; the over-sampling by data duplication approaches have the disadvantage of adding no new information to the models. The data augmentation approaches by synthesis of new data from the minority class have shown to produce very good results in imbalanced classification.

We used two strategies for data augmentation with SMOTE based sampling methods [9] applied to the two minority classes in our dataset: the plain SMOTE algorithm and the Adaptive Synthetic (ADASYN) [11] algorithm.

The SMOTE algorithm works by finding neighbor examples from the minority class in the feature space and synthetizing a new example in the space between the neighbors. This procedure is used as many times as needed to create a balance between the number of samples in the classes. The ADASYN algorithm is derived from SMOTE, and features one important difference that has to do with how it chooses the points for synthetizing new examples, biasing its choice towards non homogeneous neighborhoods.

We used the implementations available at the *imbalanced-learn* module in *scikit-learn*, in Python [12], which allow to deal with multiple minority classes, such as the present case. For comparison and evaluation purposes, a Logistic Regression model was built with the original dataset. Then SMOTE and ADASYN were applied to the dataset, and a Logistic Regression model was built for each case. Therefore, for this part of the work, three different models were built.

3.3 Classification Models

Regarding the algorithms used for building the classification models, some of the standard algorithms often reported in the literature for student's performance classification were used as a first approach: Logistic Regression (LR) [13], Support Vector Machines (SVM) [14], Decision Trees (DT) [15] and an ensemble method, Random Forests (RF) [16]. To train these models the implementations available at the Scikit-learn library in Python [17] were used.

Then, a second stage went a step further and boosting methods were also exploited. Albeit being underrepresented in the educational context, boosting methods reportedly work well with imbalanced data classification, including multi-classification tasks [10, 18]. Boosting methods are a class of ensemble methods that build a strong model from the sequential training of weak models. There are many boosting schemes available, but all are variations of a general boosting scheme. Starting with an initial prediction model, in each boosting round a weak classifier is produced that aims to reduce the errors of the previous classifier. After a defined number of rounds, these sequentially built weak classifiers give origin to a single strong prediction model that is much more accurate than any of the previous weak learning models. Following some of the most interesting results reported in the literature for multiclass imbalanced classification [10, 18], we used

four general boosting methods that are applicable to multi-class classification: Gradient Boosting [19], Extreme Gradient Boosting [20], CatBoost [21] and LogitBoost [22].

Details on the model training procedure and evaluation are given in the next section.

3.4 Model Training, Evaluation Metrics and Hyperparameter Tuning

Following the usual procedure, data was divided into training set (80%) and test set (20%). Then, for each model, a 5-fold cross validation procedure was used to avoid overfitting. This means that the training data set was divided into 5 blocks, and the training of each model was done with 4 of the blocks, with the remaining one being used for validation purposes. The process was repeated 5 times, once for each block, thus enabling the maximization of the total number of observations used for validation while avoiding overfitting. The best average cross-validation estimator score was elected. This methodology also included a procedure to assure that every class was well represented in every fold. Then, the overall performance of each elected model was assessed with the test set.

Due to the imbalanced nature of our dataset, accuracy isn't the most adequate measure for model performance, since it's an overall metric that might result in high values based on a good performance solely for the majority class. For imbalanced data, single class metrics are more adequate [18]. In this work we use the F1 measure, which accounts for the trade-off between precision and recall. F1 scores were computed for each class, and the average F1 score for the three classes was also computed. This was the metric used for hyperparameter tuning, as will be explained next. For the optimized model, accuracy was also computed as an overall metric.

All the models went through a process of hyperparameter optimization. One way of tuning the hyperparameters is to perform a grid search, a very exhaustive way of testing many configuration and picking the one that performs better with cross-validation. This approach was used for LG, SVM and DT, using the Grid Search method available at Scikit-learn, using F1-score as the metric to be maximized. For the remaining models, the parameter space was much wider, and a Randomized Grid Search was used instead, where a set of parameter values and combinations is randomly chosen, allowing the control of the number of parameter combinations that are attempted.

4 Results

In this section we present the performance of the proposed models.

4.1 Data Sampling Techniques

Table 1 shows the performance metrics obtained with the test set for the logistic regression models build without correction for the minority classes, or used together either with SMOTE or with ADASYN.

These results show that the LR model without data sampling results in very low F1-score for the “Relative Success” class, the one with fewer samples, albeit resulting in the highest accuracy. This expresses the fact that, in imbalanced classes tasks, accuracy

Table 1. Classification performance without and with data sampling

	Logistic Regression	SMOTE + Logistic Regression	ADASYN + Logistic Regression
F1-score Failure	0.61	0.64	0.63
F1-score Rel.Success	0.06	0.41	0.38
F1-score Success	0.77	0.69	0.69
Average F1-score	0.49	0.58	0.56
Accuracy	0.68	0.61	0.60

alone is not a good performance metric. LR with SMOTE or with ADASYN result in better discrimination for the minority class, although still a low value for F1. SMOTE and ADASYN produce similar results for this dataset. The use of SMOTE leads to the highest F1-scores, either overall and for the individual classes. Therefore, for the remainder of this work, SMOTE was applied to the dataset prior to model training.

4.2 Standard Models

Table 2 presents the metrics obtained for the standard methods used in this work, after hyperparameter optimization. Random Forest lead to the best overall results, which is in line with some of the results reported in the literature [7]. On the other hand, SVM is the worst performer, contrary to what other researchers have obtained [8]. This is not unusual, though, because, depending on the dataset and on the formulation of the problem, any machine learning technique can achieve superior results, prompting an experimental approach to identify the best learner for each task.

Table 2. Classification performance for the standard models

	Logistic Regression	Support Vector Machine	Decision Tree	Random Forest
F1-score Failure	0.63	0.53	0.63	0.66
F1-score Rel.Success	0.41	0.31	0.39	0.37
F1-score Success	0.69	0.71	0.75	0.82
Average F1-score	0.58	0.52	0.59	0.62
Accuracy	0.61	0.60	0.65	0.72

4.3 Boosting Models

Table 3 presents the classification performance for the boosting methods. In general, the models built with the boosting methods outperform the models built with the standard methods, both on the in-class metrics and on the overall metrics. Among these, Extreme Gradient Boosting is the best classifier, although very similar to Gradient Boosting.

In both cases the lowest F1-score among the three classes is obtained for the “Relative Success” class, whereas the highest F1-score is obtained for the majority class.

Table 3. Classification performance for boosting models

	Gradient Boosting	Extreme Gradient Boosting	Logit Boost	CatBoost
F1-score Failure	0.68	0.68	0.69	0.69
F1-score Rel.Success	0.44	0.44	0.41	0.35
F1-score Sucess	0.81	0.83	0.82	0.82
Average F1-score	0.65	0.65	0.64	0.62
Accuracy	0.72	0.73	0.72	0.73

The fact that gradient boosting models outperform most standard machine learning models is in agreement with results obtained in other fields of study. In our case, however, even those models fail in identifying most of the students belonging to the most critical, minority class. This only confirms that the classification of an imbalanced dataset, on a multi class frame is a complicated task [18].

5 Conclusions and Future Work

In this work we use dataset from a higher education institution to build classification models to early predict academic performance of students. The data set includes information known at the time of enrollment – demographics, academic performance prior to enrollment, social-economics - but none information regarding performance after enrollment. We addressed the problem as a three category classification task, in which there's a strong imbalance towards one of the classes. The minority classes are also the target classes for this work. We used algorithms to promote class balancing with synthetic oversampling and built classification models both with standard machine learning algorithms and boosting algorithms. Our results show that boosting algorithms respond better to the specific classification task than standard methods, but even them fail in correctly classifying the minority classes. As a means to improve these results, information regarding the academic performance of students during the first academic semesters will also be included in the dataset for model building.

Acknowledgments. This research is supported by program SATDAP - Capacitação da Administração Pública under grant POCI-05-5762-FSE-000191.

References

1. Romero, C., Ventura, S.: Educational data mining: a review of the state of the art. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **40**, 601–618 (2010). <https://doi.org/10.1109/TSMCC.2010.2053532>
2. Mduma, N., Kalegele, K., Machuve, D.: A survey of machine learning approaches and techniques for student dropout prediction. *Data Sci. J.* **18**, 1–10 (2019). <https://doi.org/10.5334/dsj-2019-014>
3. Shahiri, A.M., Husain, W., Rashid, N.A.: A review on predicting Student's performance using data mining techniques. *Procedia Comput. Sci.* **72**, 414–422. (2015). <https://doi.org/10.1016/j.procs.2015.12.157>
4. Rastrollo-Guerrero, J.L., Gómez-Pulido, J.A., Durán-Domínguez, A.: Analyzing and predicting Students' performance by means of machine learning: a review. *Appl. Sci.* **10**, 1042–1058 (2020). <https://doi.org/10.3390/app10031042>
5. Beaulac, C., Rosenthal, J.S.: Predicting university Students' academic success and major using random forests. *Res. High. Educ.* **60**, 1048–1064 (2019). <https://doi.org/10.1007/s11162-019-09546-y>
6. Hoffait, A.S., Schyns, M.: Early detection of university Students with potential difficulties. *Decis. Support Syst.* **101**, 1–11 (2017). <https://doi.org/10.1016/j.dss.2017.05.003>
7. Miguéis, V.L., Freitas, A., Garcia, P.J.V., Silva, A.: Early segmentation of students according to their academic performance: a predictive modelling approach. *Decis. Support Syst.* **115**, 36–51 (2018). <https://doi.org/10.1016/j.dss.2018.09.001>
8. Thammasiri, D., Delen, D., Meesad, P., Kasap, N.: A critical assessment of imbalanced class distribution problem: the case of predicting freshmen student attrition. *Expert Syst. Appl.* **41**, 321–330 (2014). <https://doi.org/10.1016/j.eswa.2013.07.046>
9. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002). <https://doi.org/10.1613/jair.953>
10. Ali, A., Shamsuddin, S.M., Ralescu, A.L.: Classification with class imbalance problem: a review. *Int. J. Adv. Soft. Comput. Appl.* **7**, 176–204 (2015)
11. He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: Proceedings of the International Joint Conference on Neural Networks, pp. 1322–1328 (2008). <https://doi.org/10.1109/IJCNN.2008.4633969>
12. Lema, G., Nogueira, F., Aridas, C.K.: Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **40**, 1–5 (2015)
13. Hastie, T.J., Pregibon, D.: Generalized linear models. In: Statistical Models in S (2017)
14. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995). <https://doi.org/10.1023/A:1022627411411>
15. Quinlan, J.R.: Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986). <https://doi.org/10.1007/bf00116251>
16. Pavlov, Y.L.: Random forests. *Random Forests* 1–122 (2019). <https://doi.org/10.1201/9780367816377-11>
17. Pedregosa, F., Gaël, V., Gramfort, A., Vincent, M., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **85**, 2825 (2011)
18. Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., Asadpour, M.: Boosting methods for multi-class imbalanced data classification: an experimental review. *J. Big Data.* **7**, 70 (2020). <https://doi.org/10.1186/s40537-020-00349-y>

19. Friedman, J.: Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**(5), 1189–1232 (2001). <https://doi.org/10.1214/aos/1013203451>
20. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016)
21. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: unbiased boosting with categorical features. In: NIPS 2018: Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 6639–6649, December 2018. <https://dl.acm.org/doi/10.5555/3327757.3327770>
22. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. *Ann. Statist.* **28**(2), 337–407 (2000). <https://doi.org/10.1214/aos/1016218223>



Object Detection in Rural Roads Through SSD and YOLO Framework

Luis Barba-Guaman^{1,2(✉)}, Jose Eugenio Naranjo², Anthony Ortiz¹,
and Juan Guillermo Pinzon Gonzalez¹

¹ Artificial Intelligence Laboratory, Universidad Técnica Particular de Loja, Loja, Ecuador

{lrbbarba, ajortiz4, jgpinzon}@utpl.edu.ec

² INSIA, Universidad Politécnica de Madrid, Madrid, Spain

joseeugenio.naranjo@upm.es

<http://www.utpl.edu.ec>, <http://www.upm.es>

Abstract. Object detection is challenging in the computer vision area and is crucial in autonomous driving systems. The largest number of traditional techniques or the use of deep learning are evaluated in the urban area, but in rural areas, there is little research carried out. The goal of this work is object detection in rural roads, this paper presents the use of deep learning frameworks used as You Only Look Once (YOLO) and another that belongs to the same category of one-stage is the well-known Single Shot Multi-Box Detector (SSD), in the state of the literature, produces excellent results in detecting objects in real-time. These models behave differently in network architecture, for this reason, we use images of rural roads with different environments to achieve an optimal balance between precision and precision in the detection of objects. Furthermore, as there is no dataset in these environments, we created our own data set to perform the experiments due to the difficulty of this problem. The result of both detectors has produced acceptable results under certain conditions like lighting conditions, viewing perspectives, partial occlusion of the object.

Keywords: Object detection · Rural roads · Computer vision · Deep learning

1 Introduction

Due to the success Computer Vision (CV) systems have undergone great development in the field of artificial intelligence in recent years. Recently, deep learning (DL) techniques are powerful methods that automatically learn to represent the characteristics of data. Through research and its application, some results have been obtained that have shown progress in most CV challenges [1]. Nowadays, there is a lot of research that has been developed in different areas of CV, one of these with great interest of researchers is object detection. The goal of Object Detection (OD) is to identify some classes of objects that are found in an image or video. Classifying or recognizing means finding the categories present in the scene (all object instances), as well as their respective network confidence values in these detections [2]. Although the OD process seems like a simple task, there are several very important phases that we must consider, these aspects are what make

object detection a real challenge. First, there are several objects that belong to the same class, changes in the perspective of the object, change in ambient lighting, occlusion of objects, all these aspects can generate an important loss of information through the presence of reflection or shadows in the images. Secondly, it is presented in the management of time, memory, and storage that is required to train these models [3]. In addition, there is the possibility that some images contain different types of combinations of the objects they contain (e.g., deformation, small, blur, rotated, motion, and occluded objects). In addition to the accuracy of the detection, another important aspect is how to speed up the detection task. Commonly, the greatest amount of research and development in the area of object detection is solved on urban roads, where driving is easier due to road signs in urban environments or roads in developed countries are good [4]. OD task is extremely difficult in rural scenarios, especially in developing countries, this means that when the roads are well maintained and the road markings or signs are clear and many techniques work successfully, however, these same techniques tend to fail in undeveloped areas or on rural roads [5], in addition to the above is challenging given the observed variability with different times of the day, changing lighting conditions, weather, and variable road conditions [6]. It is important to mention that despite progress in DL based OD and classification techniques, this would not have been possible without improvements in hardware performance. Feng *et al.* [7] comment that advances in the CV are not only based on DL techniques and the use of large image datasets but are also based on powerful hardware architecture that allows improving the time and efficiency in the training of various neuronal networks. In addition, one cannot fail to mention the graphics processing units (GPUs) that have a powerful graphics engine and a computer processor that offers high performance and bandwidth, in order to be able to execute algorithms in parallel and massive [8]. Dhillon *et al.* [9] explain that there are two main types of categories based on Convolutional Neural Networks (CNN): the first is the one-stage method, that is, in a single network it allows the location and classification of objects, the second method is the two-stage, contains two independent networks for each of them to perform a single task. According to the state of literature, within the group of a one-stage have the Single Shot Detector (SSD) [10] a very fast framework used in real-time applications and mobile devices, with a new architecture that makes it fast and innovative, we have You Only Look Once (YOLO) [11]. These frameworks use bounding boxes to detect objects and also allow you to display the class and confidence score. In this second stage, we mention the Region-proposal CNN (R-CNN) [12] with its enhanced versions, these are divided into two neural networks, each of which independently performs a specific task, first analyzes the proposed region, and the other performs the object classification. Other frameworks that belong to this group are Faster Region-based Convolutional Networks (Faster R-CNN) [21] and its follow-up architecture Mask R-CNN [14]. These network models generally produce higher accuracy rates but are slower. The goal of this paper is to an experimental comparative on the application of the SSD and YOLO v3 models for object detection in rural roads, by exploiting the full advantages on the accuracy and processing time of each model analyzed.

1.1 Overview Deep Neural Network (DNN)

There has been steady progress in object feature representations and classifiers for object recognition, but the object detectors can be categorized into two mainstreams: (i) two-stage detectors and (ii) one-stage detectors.

One-Stage Detectors. In general, this process can either make a fixed number of predictions on the grid (one-stage), that is, eliminate the Region of Interest (RoI) extraction stage and directly classify and regress the candidate anchor boxes. Some examples that we can mention are YOLO [15] adopts a unified architecture that extracts feature maps from input images. YOLO v2 [16] YOLO v3 [17] and YOLO v4 [18] are proposed with improved speed and precision. SSD [19] model is another representative one-stage detector, this shortly after the YOLO model, establishing a full convolution network to predict a fixed number of bounding boxes and scores. One-stage detectors demonstrate an optimized trade-off between accuracy and speed only on high-performance desktop GPUs. Figure 1 illustrates the SSD and YOLO level diagrams on object detection.

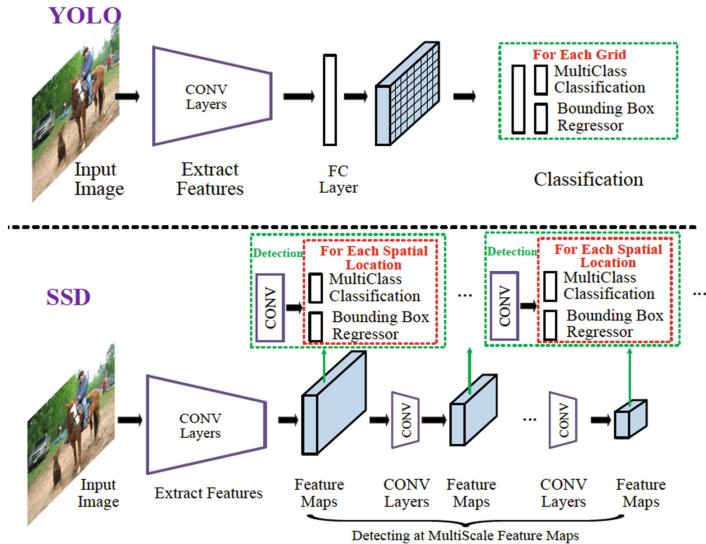


Fig. 1. SSD and YOLO high level diagrams for generic object detection [2]

Two-Stage Detectors. The detection happens in two stages: First, this model has two stages: First, the model proposes a set of regions of interest by select search or regional proposal networks. Next, these regions are sparse as the potential bounding box candidates can be infinite. Second, a classifier only processes the region candidates. A most representative series of two-stage detectors is R-CNN [20] with its extended generations Fast R-CNN [13] and Faster R-CNN [21]. These models are relatively slower inference speed due to the two-stage detection procedure. Figure 2 illustrates the R-CNN, Fast R-CNN, and Faster R-CNN level diagrams on object detection.

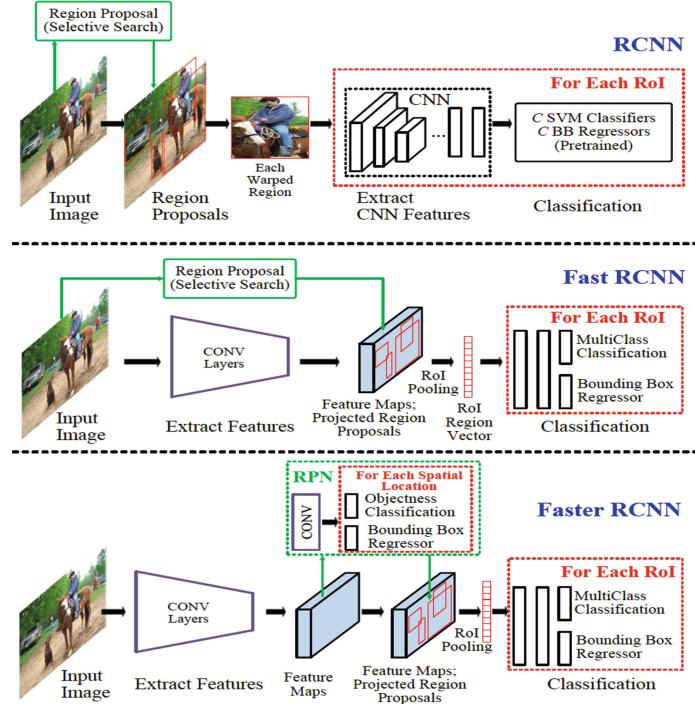


Fig. 2. R-CNN, Fast R-CNN and Faster R-CNN high level diagrams for generic object detection [2].

2 Materials and Methods

Nowadays, there are several investigations and resources that have allowed the development of several object detectors, of all these choices a model in terms of precision and speed, is a very difficult task [10, 15]. In this section we describe SSD MobileNet v2 and YOLO version 3, these are the next generation algorithms used for generic object detection, and these one-stage architectures use bounding boxes to locate and classify an object in the image.

2.1 YOLO Framework

You Only Look Once (YOLO) [15] is a framework involving detection and classification same time. The YOLO framework architecture belongs to single stage detectors. In [15] mentions that using GPU it is possible to obtain 45 FPS in real-time object detection. There is a different version, YOLO v3 [17] is an improvement made over its predecessors, YOLO v2 [16] and YOLO v1 [15]. Figure 3 illustrates the general architecture of YOLO v3.

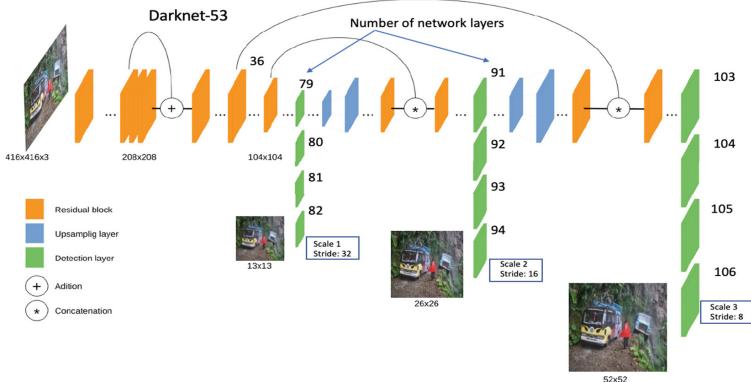


Fig. 3. YOLO v3 general architecture

2.2 SSD Framework

For the efficient detection of objects of different types and sizes, the SSD model uses the hierarchy of pyramidal characteristics of convolutional networks [10]. This model belongs to the group of single deep neural network (DNN) object detector, this framework uses multi-scale feature maps and predetermined boxes for the detection process. It eliminates the bounding box proposals and presents a resampling stage, as a result of this process, the increase in the detection speed will be achieved compared to Faster R-CNN [21] and YOLO v3 [17]. The SSD model implements an algorithm that detects multiple objects, generates good results of confidence when in the image or video there is the presence of any object in each predetermined frame and is valuable in the development of applications on mobile devices and devices or integrated vision cards [2] (Fig. 6).

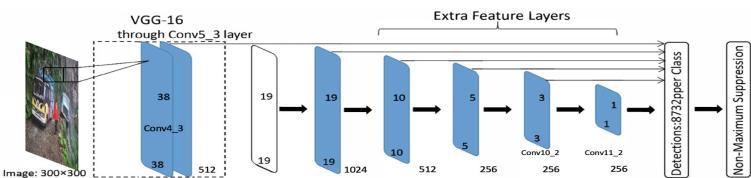


Fig. 4. SSD general architecture

2.3 Dataset

In this research, we had to build our own dataset of images in rural areas, since this type of category is not common in public dataset. A training and test set was used. The training set contains 628 images and 125 in the test set. The instances of objects are cars, bicycles, motorcycles, and people, the dataset base was generated by manually tagging the objects in the images using LabelImg. Several problems were presented with the set of original images. First, a small set of training data, next to the unbalanced of the objects

in the images of rural roads, to solve this problem the data augmentation technique was used (six data augmentation options) to increase our original data set. This technique produced a dataset with 4518 training and validation images. In our study we used 80% and 20% of the set of rural images, respectively, in the training and validation process. It is important to mention that in an image you can find one or more objects that belong to that class. The image resolution with an aspect ratio of 300×300 and 1920×1080 pixels.

2.4 Methodology

Figure 5 represents an overview of the proposed methodology for object detection. It is important to mention that the original data set of the training images was preprocessed and then to increase the training data set the data augmentation technique was used.

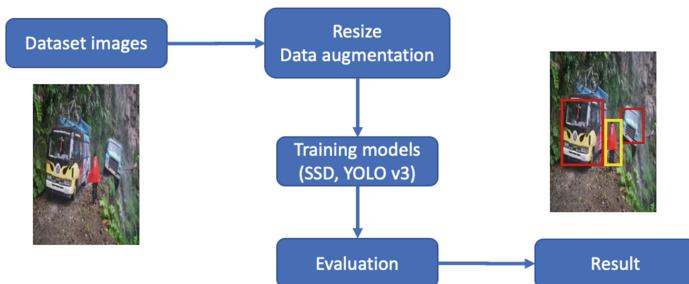


Fig. 5. Main step of the method proposed.

2.5 Metrics

The following metrics were used in the comparison of the models.

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

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

$$\text{F1 score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

Where TP represents the true positives, FN is false negatives, FP represents the false positives, finally TN represents the true negatives. Accuracy and recall are the metrics using in this research. The experiments carried out and the results found are described below.

3 Experimental and Discussions

In this step all models are tested and evaluated, all used the base of the rural road's dataset, their efficiency and development are done via the Google Colab service. This service is provided by Google in the cloud to execute Python code and create Machine Learning models through the Google cloud and with the possibility of making use of its GPUs. The SSD MobileNet v2 model was implemented on TensorFlow v2.0 while YOLO v3 is on Pytorch v1.6. The metrics SSD MobileNet v2 model are shown in Table 1.

Table 1. Object detection result on SSD MobileNet v2 model.

Test	Instances	REC	PRE	F1_score
1	Bike	0.73	0.73	0.68
	Car	0.58	0.58	0.77
	Motorbike	0.71	0.71	0.77
	Persons	0.34	0.34	0.48
2	Bike	0.66	0.95	0.85
	Car	0.46	0.88	0.63
	Motorbike	0.72	0.81	0.76
	Persons	0.24	0.90	0.41

Table 1 shows the REC, PRE, and F1 score variables in the object detection process in different rural road dataset environments. Two tests were carried out. In test one, it is important to mention the detection of the bicycle, car, and motorcycle gives good results, and the values obtained for the precision variable (PRE) are 73%, 58%, and 71% respectively. While in the detection of the person object a low value of precision and F1, the values are 34%. In test two, the configuration was learning rate equal to 0.001, and the momentum variable with a value equal to 0.9. In test two, the values obtained for the precision variable (PRE) are 95%, 88%, 81%, and 90% respectively for bike, car, motorbike, and person detection. High values can be observed in the precision of cars and people, but the recall is very low, its values are 46% and 24% respectively. On the other hand, SSD MobileNet has a higher value for F1-score for bike, car, and motorbike, while for people detection it is low. The values of the configuration carried out in the two tests are, in test one set the learning rate to 0.0045, and the momentum variable was assigned the value 0.9.

Table 2 presents the YOLO v3 model result. It is observed that precision, recall, and F1_score offers high values in the test one and two. The configuration variables in this model were learning rate equal to 0.0001, and the momentum variable with a value equal to 0.9, the number of epochs were 50, 100 and 200 respectively. YOLO v3 with Darknet version handles its own image sizes and performs its conversion according to how it is configured in this case with the width equal to 608 and height equal to 608 pixels. It can be seen from this table that the test two the precision is higher than test three, the values are 93%. Average precision (AP) shows values between 0 and 1, this is calculated between the average precision and recall values, the result obtained in test one is 84%, and in test two it was 85%.

Table 2. Object detection result on YOLO v3 model.

Test	REC	PRE	F1_score	mAP (.50)
1	0.77	0.82	0.78	0.84
2	0.83	0.75	0.77	0.85
3	0.68	0.93	0.78	0.79

Table 3. Object Detection average metrics for SSD MobileNet v2 and YOLO v3 models.

Models	IoU	PRE	REC	F1_score
SSD MobileNet v2	0.55	0.86	0.57	0.66
YOLO v3	0.83	0.83	0.76	0.78

Table 3 contains the values measured for SSD MobileNet v2 and YOLO v3. Note that SSD MobileNet v2 produces a lower IoU (Intersection over Union) result in comparison with YOLO v3 when testing partially rotated and night-illuminated objects. In precision value, both models present a slight difference in their results.

Table 4. Average processing time.

Models	Processing time (ms)
SSD MobileNet v2	0.018
YOLO v3	0.024

Table 4 shows that SSD MobileNet v2 uses less processing time. Both models are fast, SSD has a slight advantage over YOLO. These were the results of the object detection process using different neural network models that support the research.

4 Conclusion

In this paper object detection using the single-stage YOLO v3 and SSD MobileNet v2 frameworks is presented. The objective of this work is the detection of cars, bicycles, motorcycles, and people from images of rural roads.

We began with a theoretical description of both models, briefly presenting the architectural design. In order to analyze the performance, both models were trained and tested with the rural road dataset. The precision, recall, F1_score, and IoU metrics were used for the performance evaluation. Figure 6a illustrates the results using SSD Model. This model performs well in good illumination conditions, it has problems in night scenes.



a) SSD MobileNet v2 result.

b) YOLO v3 result.

Fig. 6. Object detection result with SSD MobileNet v2 and Yolo V3

Figure 6b illustrates the recognition of cars, people, motorbike, and bike in the night environments. The test result shows that YOLO v3 performs better in night conditions. It is important to mention the difficulty of finding tagged in rural areas, in which case, we create our own data set to perform the experiments. Using the SSD and YOLO detectors have produced acceptable results for different sizes and lighting conditions, partial occlusion, and at various night scenes.

Finally, the detection of objects carried out in this work focuses on the search for these in rural roads, several configurations were carried out with the aim of optimizing the models used. This work can serve as a reference in environments in rural roads. As future work is to carry out the implementation in cards dedicated to this type of processing such as Nvidia Jetson Nano and make improvements in the analyzed models.

Acknowledgment. This work is supported by the Artificial Intelligence Laboratory of the Technical University of Loja, Ecuador. University Institute of Automobile Research (INSIA) from Spain.

References

- Rosebrock, A.: Deep Learning for Computer Vision with Python: Starter Bundle. Pyimagesearch (2017)
- Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M.: Deep learning for generic object detection: a survey. *Int. J. Comput. Vis.* **128**, 261–318 (2020)
- Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X.: Object detection with deep learning: a review. *IEEE Trans. NN Lear. Syst.* **30**(11), 3212–3232 (2019)
- Yadav, S., Patra, S., Arora, C., Banerjee, S.: Deep CNN with color lines model for unmarked road segmentation. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 585–589. IEEE (2017)
- Barba-Guaman, L., Eugenio Naranjo, J., Ortiz, A.: Object detection in rural roads using Tensorflow API. In: 2020 International Conference of Digital Transformation and Innovation (2020, in press)
- Barba-Guaman, L., Eugenio Naranjo, J., Ortiz, A.: Deep learning framework for vehicle and pedestrian detection in rural roads on an embedded GPU. *Electronics* **9**(4), 589 (2020)

7. Feng, X., Jiang, Y., Yang, X., Du, M., Li, X.: Computer vision algorithms and hardware implementations: a survey. *Integration* **69**, 309–320 (2019)
8. Ammar, A., Koubaa, A., Ahmed, M., Saad, A.: Aerial images processing for car detection using convolutional neural networks: comparison between faster R-CNN and yolov3. arXiv preprint [arXiv:1910.07234](https://arxiv.org/abs/1910.07234) (2019)
9. Dhillon, A., Verma, G.K.: Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress Artif. Intell.* **9**(2), 85–112 (2020)
10. Yang, F., Chen, H., Li, J., Li, F., Wang, L., Yan, X.: Single shot multibox detector with Kalman filter for online pedestrian detection in video. *IEEE Access* **7**, 15478–15488 (2019)
11. Buric, M., Pobar, M., Ivasic-Kos, M.: Ball detection using YOLO and Mask R-CNN. In: 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 319–323. IEEE (2018)
12. Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., Dietmayer, K.: Deep multi-modal object detection and semantic segmentation for autonomous driving: datasets, methods, and challenges. *IEEE Trans. Intell. Transp. Syst.* **22**(3), 1341–1360 (2021). <https://doi.org/10.1109/TITS.2020.2972974>
13. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
14. Su, H., Wei, S., Yan, M., Wang, C., Shi, J., Zhang, X.: Object detection and instance segmentation in remote sensing imagery based on precise mask R-CNN. In: 2019 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2019, pp. 1454–1457. IEEE (2019)
15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 779–788. (2016)
16. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)
17. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2018)
18. Bochkovskiy, A., Wang, C., Liao, H.: YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934) (2020)
19. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., Berg, A.C.: SSD: single shot multibox detector. In: Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016, pp. 21–37. Springer, Cham (2016)
20. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June, pp. 580–587 (2014)
21. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017)



Study of MRI-Based Biomarkers on Patients with Cerebral Amyloid Angiopathy Using Artificial Intelligence

Fátima Solange Silva¹(✉), Tiago Gil Oliveira^{2,3,4}, and Victor Alves¹

¹ Algoritmi Center, University of Minho, Braga, Portugal
valves@di.uminho.pt

² Life and Health Sciences Research Institute (ICVS), School of Medicine,
University of Minho, Campus Gualtar, 4710-057 Braga, Portugal

³ ICVS/3B's-PT Government Associate Laboratory, Braga/Guimarães, Portugal

⁴ Division of Neuroradiology, Hospital de Braga, Braga, Portugal
tiago@med.uminho.pt

Abstract. Cerebral Amyloid Angiopathy (CAA) is a neurodegenerative disease characterised by the deposition of the amyloid-beta ($A\beta$) protein within the cortical and leptomeningeal blood vessels and capillaries. CAA leads to cognitive impairment, dementia, stroke, and a high risk of intracerebral haemorrhages recurrence. Generally diagnosed by *post-mortem* examination, the diagnosis may also be carried *pre-mortem* in surgical situations, such as evacuation, with observation in a brain biopsy. In this regard, Magnetic Resonance Imaging (MRI) is also a viable a noninvasive alternative for CAA study *in vivo*. This paper proposes a methodological pipeline to apply machine learning approaches to clinical and MRI assessment metrics, supporting the diagnosis of CAA, thus providing tools to enable clinical intervention, and promote access to appropriate and early medical assistance.

Keywords: Machine learning · Artificial intelligence · Cerebral Amyloid Angiopathy · Medical imaging · Biomarkers · MRI

1 Introduction

Cerebral Amyloid Angiopathy (CAA) is a form of cerebral small vessel disease caused by the progressive accumulation of Amyloid-Beta ($A\beta$) in the small and medium leptomeningeal, and cortical vessels and capillaries [1]. The deposition of such peptide weakens the cerebral vessel walls, leading to its rupture and consequently to haemorrhagic events. These damages are typically associated with decline of cognitive capacities, dementia, lobar haemorrhages, microhaemorrhages, ischaemic changes, and white matter lesions [2]. Figure 1 presents lesions associated to hemorrhagic events in patients with possible or probable CAA, visible on T2*-weighted MRI.

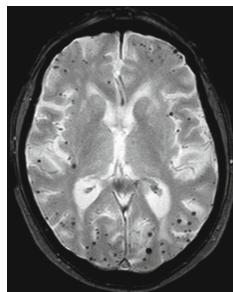


Fig. 1. Patterns of cerebral multiple lobar cerebral microbleeds visible on T2*-weighted MRI, on a patient with possible or probable CAA.

CAA is a common pathology reported in brains of elderly people. It co-exists with several other causes of cognitive decline, being Alzheimer's disease (AD) the most common one, suggesting also a possible role of CAA in the pathogenicity of AD [3]. Moreover, it is important to understand the pathways by which the disease affects patients, and the relationship with other brain pathologies, to facilitate the diagnosis and allow the patients' access to early and targeted medical intervention [4].

The current diagnosis methods of CAA are still at an early stage, depending on the availability of brain tissues for the post-mortem analysis, or during the evacuation of an intracranial hematoma, which gives the possibility to perform a brain biopsy samples pre-mortem, which is not typically performed due to its evasiveness. Therefore, a reliable and non-invasive approach based on the use of existing imaging tools and artificial intelligence technology to reveal clinical and image-related biomarkers becomes necessary, to complement and simplify the diagnosis of CAA [5].

The present study proposes to use artificial intelligence algorithms to identify biomarkers using clinical data and Magnetic Resonance Imaging (MRI) images. To this end, data related to visual rating scales for the assessment of cerebral atrophy and radiomics was generated from patients' MRI scans. The visual rating scales were applied to both hemispheres of six cerebral regions: anterior cingulate, orbito-frontal, anterior temporal, fronto-insula, medial temporal, and posterior.

Moreover, patients diagnosed with CAA and other three neurodegenerative conditions (AD, Parkinson's disease, and mild cognitive impairment) where chosen in order to determine whether automatic methodologies are able to differentiate the pathological conditions, and determine which variables weight the most on the decision.

2 Background

2.1 Clinical Rating Scales

Visual rating scales offer a way to assess brain changes in demented patients. The distinction and diagnosis of neurodegenerative diseases is then crucial for the early access to support and target the treatment for the affected individuals.

The difficulties in the evaluation of tissues for the diagnosis of the disease and the overlapping clinical symptoms for distinct pathologies are obstacles to accurate disease diagnosis and prognostic evaluation. By contrast, the availability and great tissue contrast afforded by MRI approaches allows the assessment of cerebral atrophy with great predictive value [5,6].

Visual rating scales designed to evaluate cerebral changes in patients with cognitive impairment have provided tools to improve sensitivity and reliability of diagnosis-based image interpretation, and findings of value for differential diagnosis of dementia [7,8].

For this study, a visual rating scale proposed by Harper et al. [9], was applied to determine visual biomarkers in patients diagnosed with possible or probable CAA. To conduct the study, MRI from patients selected to the study was analysed according to the guidelines for the assessment of cerebral atrophy in 6 cerebral regions.

2.2 Radiomics

With the advances in the medical field and the increase in the use of digital medical images, there also has been an upgrade of approaches and tools to study this increasing information complexity [10]. Radiomics is a recent area aiming at the extraction of information concealed in biomedical imaging data, which leads to the generation of large amounts of features by means of data-characterisation algorithms and mathematical tools [11]. These features, mostly imperceptible to the human eye, can be classified into different groups: first-order features describing voxels' intensities or spatial distribution, second-order features comparing relationships between adjacent voxels and third-order features exploring relationships between more than two voxels [12]. Figure 2 reveals the stages of the quantitative features.

In this work, we also present a workflow with radiomics analysis to study the application of this methodology in the uncovering of biomarkers to improve decision support.

3 Materials and Methods

3.1 Dataset

The study was conducted in collaboration with the neuroradiology division of the Hospital de Braga, after the study approval by the internal Health Ethics Committee, and all methods were adjusted in accordance with the guidelines of

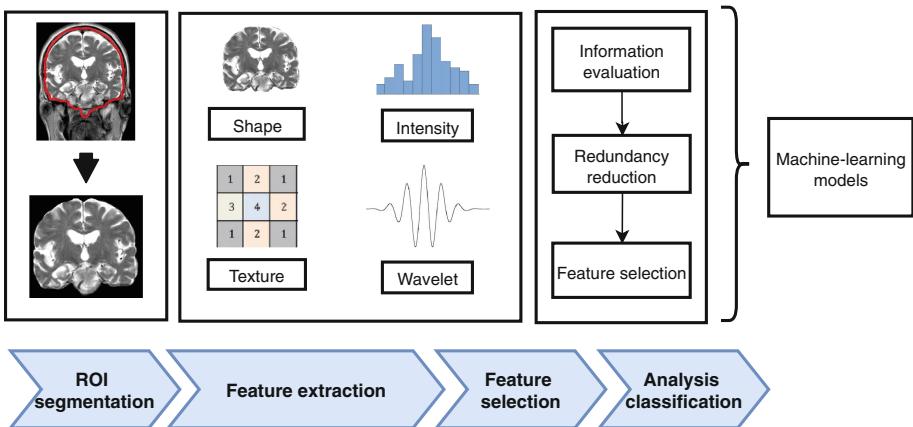


Fig. 2. Pipeline of the radiomic's feature extraction, starting with data standardisation and segmentation of the region of interest (ROI), followed by feature extraction, and culminating in feature evaluation and selection, ready to feed a machine learning model or a statistical pipeline.

the ethics board. The dataset contains data concerning 138 patients admitted to the hospital and diagnosed with possible or probable CAA (according to the Boston criteria [13]), and patients diagnosed with 3 other neurodegenerative diseases: 90 with AD, 19 with Parkinson's disease (PD), and 56 with mild cognitive impairment (MCI), resulting in 303 unfiltered entries.

Information regarding sex and age were collected from the clinical records and stored. Such data is summarised in Table 1. Studies from the patients' brain MRI were also acquired in the DICOM format after anonymisation.

Table 1. Patients' age and sex distribution according to Neurodegenerative disease.

	CAA	AD	MCI	PD
Age	66.09 (\pm 13.9) [31 – 89]	72.48 (\pm 10.13) [33 – 89]	73.08 (\pm 9.30) [51 – 88]	68.29 (\pm 8.50) [58 – 82]
Sex	Male: 63 (61.17%) Female: 40 (38.83%)	Male: 37 (44.05%) Female: 47 (55.95%)	Male: 31 (58.49%) Female: 22 (41.51%)	Male: 7 (41.18%) Female: 10 (58.82%)

3.2 Pre-processing

The raw MRI data was unfiltered, had several MRI sequence acquisitions (T1, T2 and T2* and FLAIR), and the available data per patient was not consistent. Also, some images were distorted, had different contrast, intensity, and image noise. To minimise the effect of these artifacts, and maximise the success of the study, a manual selection of images, sequences, and planes was performed in the 2D images. As the visual assessment of cerebral atrophy was performed on coronal MRI planes, and haemorrhages are best seen in T2 sequences, preference

was given to T2*-coronal acquisitions. Data standardisation, filters were applied to uncover quantitative features.

Additionally, participants with missing data, unsatisfactory scans, or with cerebral lesions, such as tumours or brain damages, were excluded from the study, leaving us with a study population of 252 patients. Moreover, encoding of the categorical information into numeric values, such as the sex, was performed. This way, male gender was set to “0” and female to “1”.

3.3 Visual Rating of Brain Atrophy

The visual evaluation and rating of cerebral atrophy of the patients was performed independently by two neuroradiologists, who reviewed the MRI images, blinded to the patient’s identity. The specialists rated six regions of the brain shown to have potential for differential diagnosis, according to the guidelines proposed by Harper et al. [9]: medial temporal lobe, anterior cingulate, fronto-insula, orbito-frontal, posterior, anterior temporal previously. Each hemisphere was rated separately, and the scores summed up to a global atrophy score, resulting in six atrophy levels per patient. Table 2 represents the sex, age, and examples of the application of the visual rating scales on two patients.

Table 2. Example of the CSV file format for the atrophy level and clinical data.

ID	Sum Anterior Cingulate	Sum Orbito-Frontal	Sum Anterior temporal	Sum Fronto-insula	Sum MTA	Sum Posterior	Age	Sex
I28ZQ8PNN4	4	3	3	3	2	0	61	0
TOQKH2UN27	2	0	2	2	2	3	62	1

Whenever disagreements were significant, the ratings were reviewed, and a consensus was reached. Table 3 sums up the criteria used to evaluate the 6 regions.

Table 3. Description of the criteria used to evaluate the cerebral atrophy for six regions

Regions	0	1	2	3	4
Anterior cingulate	“closed sulcus”	“slight sulcal opening”	“widening of the sulcus”	“acute widening of the sulcus”	–
Fronto-insula	“closed sulcus”	“slight sulcal opening”	“widening of the sulcus along its length”	“acute widening of the sulcus”	–
Orbito-frontal	“closed sulcus”	“slight opening of the sulcus”	“widening of the sulcus”	“severe widening of the sulcus”	–
Anterior temporal	“no visible atrophy”	“slight alteration of the AT sulci”	“widening of the temporal sulci”	“severe atrophy of the gury”	“temporal pole not visible”
Medial temporal	“no visible atrophy”	“slight widening of the choroid fissure”	“visible widening of the choroid fissure, and other sulci”	“severe volume loss of the hippocampus”	“severe atrophy at a terminal stage”
Posterior	“closed sulci”	“opening of the posterior sulci”	“considerable widening of the sulci”	“severe widening of the sulci”	–

3.4 Low Level Feature Extraction with Radiomics

An open-source tool was used, PyRadiomics¹ to proceed to the quantification of radiomic characteristics on regions of interest (ROI) on medical images. The process of extraction of the low-level features from images is as follows:

1. Conversion of the scan DICOM format to NIfTI, removal of non-brain and soft tissues (e.g. skull, fat), definition of the ROI (in this case corresponds to the entire brain), and loading of medical images and correspondent mask of the ROI. The handling and processing (filtering operations, segmentation and registration) of images is done using the SimpleITK library;
2. Application of image filters. The available ones include wavelet and Laplacian of Gaussian filters, implemented using PyWavelets and SimpleITK, as well as several simple ones, including square, square root, logarithm, and exponential filters, implemented with NumPy. Besides the filters, the standardisation option was chosen;
3. Feature generation for the 3D images using five different classes: first-order statistics, shape descriptors, texture classes with grey level co-occurrence matrix, with grey level run length matrix, and grey level size zone matrix.

The resulting features were stored and organised by the anonymised patient ID associated with the original image. Figure 3 presents the workflow to extract the radiomics low level features.

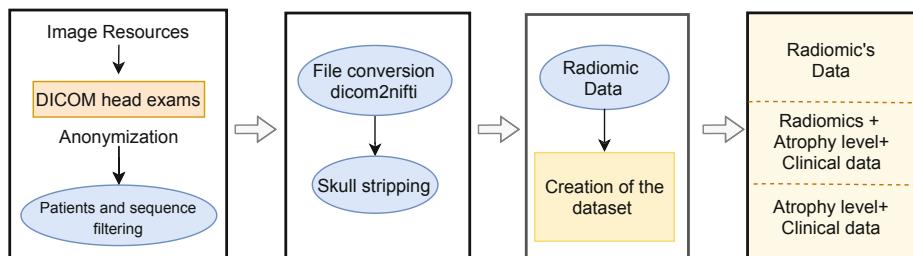


Fig. 3. Overview of the steps to extract low level features from radiomics.

3.5 Machine Learning

Classification models using ML algorithms were created using the generated data. To do so, the study was divided in three analytical approaches: the first using the visual rating scores and clinical data alone; the second one using the radiomics features; and the third consisted of a combination of the two previous types of data, to also study whether the combination of both increments the

¹ PyRadiomics: <https://pyradiomics.readthedocs.io/>.

prognostic value. For each approach, pre-processing was conducted, followed by classification with five ML models. The prepared data for each approach was then split into a subset used for training models (80%) and the subset used to test the models accuracy (20%).

The classification algorithms take the training data and infer a hypothesis, which is used to predict the labels of the test data. In our study, five machine learning classifiers were used: decision trees, random forests, support vector machines, logistic regression, and multi-layer perceptron.

Following the algorithm selection, hyper-parameter tuning was conducted to optimise the hyperparameters, using the Grid-search process to search over the optimal parameter values. Multiple approaches were tested, the results analysed, and adjustments made to optimise the algorithms' parameters and feature selection:

- Performance Evaluation: To evaluate the performance of the different supervised ML algorithms from the three experiments. Precision, recall and the F1-score were estimated to determine the performance of the models. When considering such metrics in imbalanced classifications problems, two distinct measures may be applied depending on the scenario: the “macro-average”, that computes the metric independently of each class weight and the “micro-average”, that considers the contribution of each class to the average metric.
- Feature selection: To further understand the behaviour of the data and the pathologies, feature engineering was applied. For the three approaches, the weight of each feature was estimated. In the first approach, the 8 features were ordered according to their weight in the best algorithm. In the second one, starting with a total of 1722, feature selection aiming at dimensionality reduction was conducted, thus avoiding feature redundancy. Moreover, the features that contributed to about 75% of the accuracy weight were used to compute a second classification step, ending with 20 radiomic features. Finally, in the third approach, the same methodology was repeated, ending with 9 features.

4 Results

To investigate the biomarkers that would allow differentiating between CAA and other 3 neurodegenerative conditions, a total of 1722 radiomic features and 6 visual scores were acquired from the computing and analysis of MRI scans from patients diagnosed with each of the four pathologies. Feature selection and classification training was done using the training cohort, whereas the test cohort was used to estimate the predictive performance.

4.1 Approach 1: Visual Rating Scales Biomarker Exploration

The results of the classification models using the visual rating scores and clinical data is revealed below at Table 4.

Table 4. Results obtained for the first approach.

Algorithm	Accuracy	F1 score	
		Macro-F1	Micro-F1
Logistic Regression	0.41	0.24	0.35
Support vector machine	0.49	0.16	0.29
Random Forest	0.55	0.37	0.51
Multi-layer Perceptron	0.39	0.14	0.22
Decision Tree	0.55	0.29	0.51

After testing the models, the results suggest the algorithm that better classifies the neurodegenerative diseases is the RF.

To determine the features that weighted the most on the classification, feature importance from *sklearn* is estimated. From this, age is revealed as a major feature for the differentiation of CAA from the other pathologies, which was already proven in previous studies, followed by sex, the atrophy in the posterior region, and medial temporal lobe (a common feature in demented patients, frequently found in patients with MCI) [14].

4.2 Approach 2: Classification of Diseases Based on Radiomics Data

Regarding the classification models used to study the radiomic data as possible biomarkers, the results are displayed in Table 5:

Table 5. Results obtained from the classification using radiomic data.

Algorithm	Before feature selection			After feature selection		
	Accuracy	F1 score		Accuracy	F1 score	
		Macro-F1	Micro-F1		Macro-F1	Micro-F1
Logistic Regression	0.48	0.17	0.38	0.44	0.17	0.38
Support vector machine	0.46	0.16	0.29	0.58	0.18	0.42
Random Forest	0.67	0.46	0.63	0.70	0.48	0.66
Multi-layer Perceptron	0.58	0.18	0.42	0.64	0.33	0.52
Decision Tree	0.54	0.45	0.55	0.61	0.59	0.61

Once again, random forest is the classifier with the best accuracy and F1-score. From these results, feature importance was determined, resulting in a list where “*FirstOrderRootMeanSquared*”, measure of the magnitude of the image values, is positioned at the top, followed by “*FirstOrder90Percentile*”, which corresponds to the 90th percentile of the values on the local binary pattern filter, “*GLRLMLongRunEmphasis*”, a measure of the pixels in an image with the same grey level value, with a greater value suggesting a coarse image texture,

and “*GLRLMShortRunLowGrayLevelEmphasis*”, which corresponds to the shorter distance between lower grey-level pixels with the same values.

4.3 Approach 3: Classification of Diseases Based on Combination Data

Finally, for the third approach using the combination of the data used in the previous approaches, the results are displayed in Table 6.

Once again, RF is the algorithm that better classifies the 4 pathologies, with an accuracy of 70% and F1-micro of 66%. This approach also exhibits the best results of all the three. From these results, once again, feature importance was calculated and ordered. “*GLRLMShortRunHighGrayLevelEmphasis*”, a measure of the shorter distance between pixels with the same high grey-level values, “*GLCMJointAverage*”, the mean of the grey level intensities within the matrix, and “*FirstOrderUniformity*”, the sum of the squares of intensity value, a measure of the homogeneity of the image matrix.

Table 6. Results obtained from the classification with the combined data.

Algorithm	Before feature selection			After feature selection		
	Accuracy	F1 score		Accuracy	F1 score	
		Macro-F1	Micro-F1		Macro-F1	Micro-F1
Logistic Regression	0.43	0.24	0.36	0.58	0.18	0.42
Support vector machine	0.43	0.15	0.26	0.58	0.18	0.42
Random Forest	0.70	0.47	0.65	0.70	0.48	0.66
Multi-layer Perceptron	0.43	0.15	0.26	0.49	0.17	0.38
Decision Tree	0.55	0.36	0.52	0.61	0.51	0.60

5 Conclusion

In this study, the joint CAA diagnostic accuracy based on brain imaging data was explored, using radiomics or data derived from visual specialist assessment of MRI of patients diagnosed with probable or possible CAA and other 3 neurodegenerative diseases was explored.

Several ML models were developed for this study. The ML approach was executed applying five algorithms: logistic regression, support vector machines, random forest, multilayer perceptron, and decision tree. The algorithms were run on radiomic data. The results showed that random forest performed better than the four other models, with an accuracy of 0.70, and the weight of the most important features reflect the ability of the computer to pick up texture details from the brain of the patients and use this information to make decisions. This study also highlights the capability of improving the models performance when

combining radiomics and clinical features, corroborating the results published by other researchers [15].

In conclusion, this study highlights the added value of using radiomic features and visual rating scales for the evaluation of cerebral atrophy on acquired MRI of patients with neurodegenerative diseases, in special in patients with CAA. Specifically, our results indicate that age, sex, atrophy in the posterior and medial temporal brain regions, and radiomic texture-related features may help the medical decision making for CAA identification.

Acknowledgments. This work has been supported by FCT - Fundação para a Ciência e a Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

1. Mandybur, T.I.: Cerebral amyloid angiopathy: the vascular pathology and complications. *J. Neuropathol. Exp. Neurol.* **45**(1), 79–90 (1986)
2. Rensink, A.A., Waal, R.M.W., Kremer, B., Verbeek, M.: Pathogenesis of cerebral amyloid angiopathy. *Brain Res. Rev.* **43**, 207–223 (2003)
3. Tetsuka, S., Hashimoto, R.: Slightly symptomatic cerebral amyloid angiopathy-related inflammation with spontaneous remission in four months. *Case Rep. Neurol. Med.* **2019**, 1–5 (2019). <https://doi.org/10.1155/2019/5308208>
4. Pezzini, A., Del Zotto, E., Volonighi, I., Giassi, A., Costa, P., Padovani, A.: Cerebral amyloid angiopathy: a common cause of cerebral hemorrhage. *Curr. Med. Chem.* **16**(20), 2498–2513 (2009)
5. Greenberg, S.M., Charidimou, A.: Diagnosis of cerebral amyloid angiopathy: evolution of the Boston criteria. *Stroke* **49**(2), 491–497 (2018)
6. Pantoni, L.: Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol.* **9**(7), 689–701 (2010)
7. Scheltens, P., Pasquier, F., Weerts, J.G., Barkhof, F., Leys, D.: Qualitative assessment of cerebral atrophy on MRI: inter-and intra-observer reproducibility in dementia and normal aging. *Eur. Neurol.* **37**(2), 95–99 (1997)
8. Tsai, H.H., Tsai, L.K., Chen, Y.F., Tang, S.C., Lee, B.C., Yen, R.F., Jeng, J.S.: Correlation of cerebral microbleed distribution to amyloid burden in patients with primary intracerebral hemorrhage. *Sci. Rep.* **7**, (2017)
9. Harper, L., Fumagalli, G.G., Barkhof, F., Scheltens, P., O'Brien, J.T., Bouwman, F., Burton, E.J., Rohrer, J.D., Fox, N.C., Ridgway, G.R., et al.: MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain* **139**(4), 1211–1225 (2016)
10. Lambin, P., Leijenaar, R.T., Deist, T.M., Peerlings, J., De Jong, E.E., Van Timmeren, J., Sanduleanu, S., Larue, R.T., Even, A.J., Jochems, A., et al.: Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**(12), 749–762 (2017)
11. Yip, S.S.F., Aerts, H.J.W.L.: Applications and limitations of radiomics. *Phys. Med. Biol.* **61**(13), R150–R166 (2016). <https://doi.org/10.1088/0031-9155/61/13/r150>
12. Pyradiomics: Pyradiomics: Radiomic features. <https://pyradiomics.readthedocs.io/en/latest/features.html>
13. Knudsen, K.A., Rosand, J., Karluk, D., Greenberg, S.M.: Clinical diagnosis of cerebral amyloid angiopathy: validation of the Boston criteria. *Neurology* **56**(4), 537–539 (2001)

14. Velickaite, V., Ferreira, D., Cavallin, L., Lind, L., Ahlström, H., Kilander, L., Westman, E., Larsson, E.M.: Medial temporal lobe atrophy ratings in a large 75-year-old population-based cohort: gender-corrected and education-corrected normative data. *Eur. Radiol.* **28**(4), 1739–1747 (2018)
15. Ming, X., Oei, R.W., Zhai, R., Kong, F., Du, C., Hu, C., Hu, W., Zhang, Z., Ying, H., Wang, J.: MRI-based radiomics signature is a quantitative prognostic biomarker for nasopharyngeal carcinoma. *Sci. Rep.* **9**(1), 1–9 (2019)



One-Pixel Attacks Against Medical Imaging: A Conceptual Framework

Tuomo Sipola^(✉) and Tero Kokkonen

Institute of Information Technology, JAMK University of Applied Sciences,
Jyväskylä, Finland
{tuomo.sipola,tero.kokkonen}@jamk.fi

Abstract. This paper explores the applicability of one-pixel attacks against medical imaging. Successful attacks are threats that could cause mistrust towards artificial intelligence solutions and the healthcare system in general. Nowadays it is common to build artificial intelligence models to classify medical imaging modalities as either normal or as having indications of disease. One-pixel attack is made using an adversarial example, in which only one pixel of an image is changed so that it fools the classifying artificial intelligence model. We introduce the general idea of threats against medical systems, describe a conceptual framework that shows the idea of one-pixel attack applied to the medical imaging domain, and discuss the ramifications of this attack with future research topics.

Keywords: Adversarial examples · Artificial intelligence · Cyber security · Machine learning · Model safety · Medical imaging · Healthcare · Security

1 Introduction

Modern networked and digitalized cyber domain is an extremely complex entity that comprises of unpredictable circumstances. As a part of the critical infrastructure, the healthcare sector is one of the major domains of interest from the cyber security perspective. In healthcare, there are numerous networked systems that can be targets for cyber attacks or intrusions. Finland's cyber security strategy indicates healthcare as an area which does not produce cyber security related solutions, services or products, but the activities of which are affected by cyber security, and where possible cyber security incidents will have a significant impact [13].

The state-of-the-art target in the development of the healthcare digitalization is the smart hospital environment. As defined by the European Union Agency for Network and Information Security (ENISA) [17]: “*A smart hospital is a hospital that relies on optimised and automated processes built on an ICT environment of interconnected assets, particularly based on Internet of things (IoT), to improve existing patient care procedures and introduce new capabilities.*” According to

ENISA, one capability of the smart hospital environment are devices that lead to overall smartness. There are numerous systems used in the medical domain with capability of autonomic classification or diagnosis based on machine learning (ML) or deep learning (DL) [1, 10, 15].

Medical imaging technologies such as X-rays, tomography methods and whole-slide imaging digital pathology have become more widespread in the modern medical practice [2, 6]. However, new technologies attract malicious actors who want to profit from the misuse of those technologies or otherwise reach their goals by disrupting normal operations. The medical domain is an especially lucrative target for cyber criminals because of the sensitive nature of the data. For example, in Finland a psychotherapy service and 40,000 of its customers were blackmailed causing public mistrust towards healthcare [7, 8]. This causes long-term side effects from which it might take considerable time to recover. Similar kind of mistrust could be directed to medical imaging systems. Even if such doubts are not known among the public, the experts using imaging systems might lose their trust in AI-based models, and when such models remain in use, their misdiagnoses could cause unneeded overload in the healthcare system.

In this paper, we describe a framework to conduct one-pixel attacks against medical imaging. The remainder of this paper is organized as follows. Section 2 introduces the fooling of AI models using adversarial examples. In Sect. 3, the attack framework is described. Finally, discussion about future research topics is presented in Sect. 4.

2 Adversarial Examples

Fooling AI models using adversarial examples is a known threat. There are many attacks against deep neural networks that analyze images, especially when the goal is image classification. Most of the known attacks are iterative and white-box type, i.e., the inner configuration of neural network models is available to the attacker. However, some defences are available: gradient masking hides the gradient so that attack methods cannot use it, robust optimization uses attacks to re-train the model to be more resistant against attacks and detection methods try to identify attacks again before the input is being passed to the actual AI model [18].

The field of medical imaging is not immune to adversarial attacks. There are examples of crafting images and patches that create unwanted results when using an AI classifier in the medical domain [5, 12, 14]. Ma et al. noted that medical deep neural network models are more vulnerable than those used for natural image detection. However, simple detectors are able to capture the majority of adversarial examples because they contain differing fundamental features [11]. Finlayson et al. demonstrated that the use of projected gradient descent (PGD), natural patches and adversarial patches is effective against fundoscopy, X-ray and dermoscopy imaging [5]. Finlayson et al. have also raised the question of when to intervene regarding these vulnerabilities in medical imaging systems. Acting early could build more resilient systems but also hinder agile development. They

describe the problem of adversarial images similar to the cat-and-mouse game of cyber defence against hacking. As a solution they suggest amending regulatory best practices, for example hash-based fingerprinting of images [4].

One-pixel attacks are a known method of fooling neural network models. Changing just one pixel in the image causes the model to classify an image as being of another class label than the image is in reality. Differential evolution (see, e.g. [3]) can be used to find the optimal perturbation to change the predicted class label of an image. The one-pixel perturbation is encoded with x-y coordinates and RGB values, so that each perturbation is a vector of five elements. This kind of attack applies to different network structures and image sizes but could benefit from more advanced optimization methods [16]. There have been research concerning attack attempts against medical imaging using one-pixel attacks. Although a simplified case of pose estimation of surgical tools, Kügler et al. find adversarial examples near the decision boundary, creating vulnerable regions inside the images [9].

3 Attack Framework

A straightforward way of using an artificial intelligence (AI) solution is to classify medical images. The images are classified either normal or as having indications of disease. This information is accompanied with a score, which indicates how much the image is seen as part of its class. Attacking against medical imaging can be thought as a way of creating mistrust against the healthcare system. The basic principle can be applied in two ways, from normal to indications of disease, and vice versa. Firstly, we have a normal image as a starting point. This image is modified so that the AI model will instead predict the image as having indications of disease. Such a misdiagnosis could create unnecessary use of medical resources. It could also undermine the trust in systems using an AI model because they are producing less accurate results. Secondly, we have an image with indications of disease as a starting point. After appropriately modifying the image, the AI model will classify it as normal. This approach could lengthen the time after which the patient gets treatment. Such misclassifications could be even fatal. These factors could undermine the trust in the healthcare system.

Building and deploying an AI model using machine learning methods is usually broken into two major steps. The first one is the actual training of the model, during which the training images are used to teach the AI model to carry out the classification task as efficiently as possible within the constraints of the training. The second step is the deployment of the AI model so that it predicts or classifies completely unknown images, yielding a result: the classification and the score. If the input images are engineered to intentionally create wrong classification of the said image, we speak of adversarial examples. The image itself could look like healthy tissue; however, the engineered adversarial example could include information that fools the AI model. One such engineering attempt could be a one-pixel attack that changes only one pixel in the image to fool the AI model. This setup is schematically described in Fig. 1, which indicates the training and deployment for predictive/diagnostic use. The one-pixel attack would

be performed by modifying the input images the class label of which is being predicted.

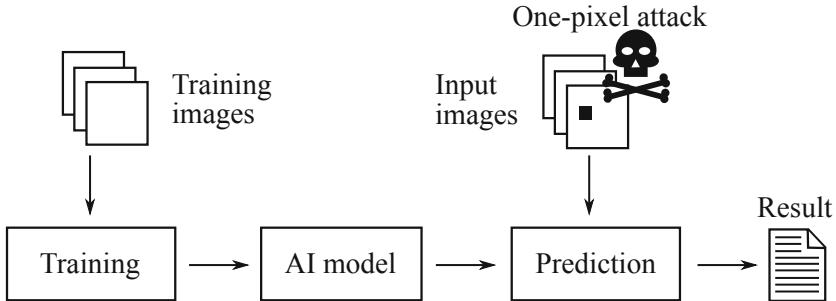


Fig. 1. A schematic presentation of the one-pixel attack against a machine learning model. Adapted from authors' previous paper [14].

Performing the one-pixel attack can be achieved by searching for these images using optimization methods. As seen in the study by Su et al. [16], differential evolution is one suitable candidate for the optimization problem. The problem of finding an adversarial example can be thought as a challenge of finding the necessary change in order to achieve a measurable goal. The goal, measured by a cost function, is to get the AI model produce wrong results. As said, the AI model usually returns a score indicating how confident it is in the classification result. The score is usually expressed in the range of $[0, 1] \in \mathbb{R}$, and it is suitable for acting as the cost function. This attack is a black-box solution because the target AI model is only needed for feeding input and querying the classification result. The inner workings of the AI model are not needed because it is only used as part of the cost function during the optimization. Figure 2 gives the basic idea behind the differential optimization process, where a population of attack images is created. This population is then used as input to the AI model, which predicts the class label and gives a probability score for it. These results are then evaluated, and images that are better at fooling the AI model are retained as the precursors for the future populations. This way of thinking is geared towards the differential evolution method, but it equally applies to many other optimization methods.

Figure 3 showcases the working principle of differential evolution in this scenario. The process is started by giving it an input image and information towards which class label we want the AI model to be fooled. The logic is the same as with the earlier images; however, this is a more detailed view of the differential evolution process when searching for adversarial examples. This process takes one image as a source for its input population, which is initialized based on random or search space spanning one-pixel permutations. In other words, each image in the population will be based on the same source image but have one pixel changed to another by the permutation. The evolutionary process is used

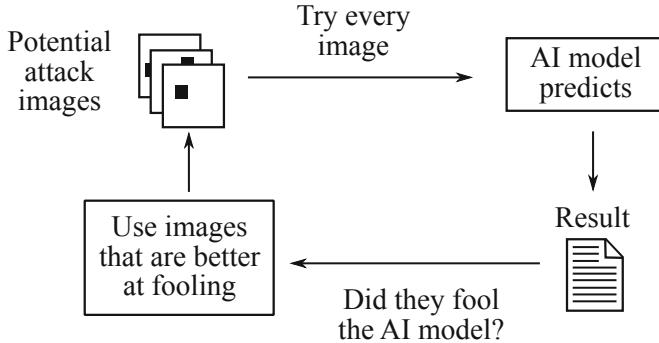


Fig. 2. Basic idea behind the optimization procedure.

to change the images. Then the effectiveness of these attack images is evaluated by using the black-box AI model. This, in turn, makes it possible to select the best images that confuse the AI model. If any image in the population fooled the AI model with acceptable certainty, we can stop and declare that an adversarial image has been found. If no acceptable images can be found, and the optimization does not converge, the search should be stopped.

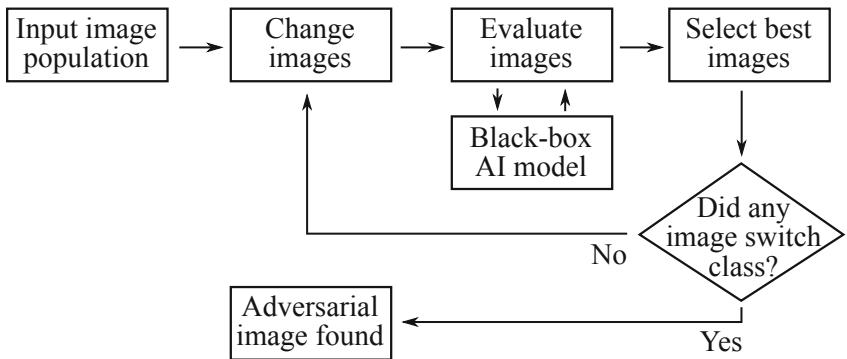


Fig. 3. Block diagram of the procedure of finding an adversarial image using differential evolution.

4 Discussion

The integrity and robustness of medical systems needs to be tested and hardened against known attacks. Furthermore, deeper inspection of robust behavior of machine learning systems will benefit the systems in the medical domain. Such inspection could be directed at least towards two directions. First of them

are theoretical bounds of machine learning systems that warrant more detailed mathematical analysis. Understanding the behavior of AI models and the boundaries of their inaccurate behavior would help create more trustworthy solutions. Secondly, employing robustness strategies during training could harden the AI models against adversarial examples that misuse the theoretical bounds. Bringing these new mitigations using theoretical bounds and defensive robustness strategies into production will be a challenge; however, this ultimately ensures that the professionals and the public trusts in these efficient tools that make the healthcare process faster and more accurate.

One-pixel attack is a decent example of an attack against automatic analysis and diagnosis in medical domain, especially when the pixel is not noticeably prominent. By affecting merely one pixel of an image under analysis, the diagnosis can be incorrect, which can lead to improper treatment. Since the logic of the attack is well understood, it is possible to create uncertainty with a proper attack vector to insert the image into the diagnosis pipeline. The latest real-life attacks have demonstrated that there is a desire to conduct cyber attacks against medical systems, and furthermore, medical systems are seen as valuable targets.

The next step of the continuing research is to research the feasibility and effectiveness of the attack in a real-life scenario with a real dataset and machine learning algorithms. As the concept is quite evident and its targets abundant, studying the feasibility and effectiveness of the attack seems to be a potential way forward.

Acknowledgments. This research is funded by the Regional Council of Central Finland/Council of Tampere Region and European Regional Development Fund as part of the Health Care Cyber Range (HCCR) project of JAMK University of Applied Sciences Institute of Information Technology. The authors would like to thank Ms. Tuula Kotikoski for proofreading the manuscript.

References

1. Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., Greenspan, H.: Chest pathology detection using deep learning with non-medical training. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 294–297 (2015). <https://doi.org/10.1109/ISBI.2015.7163871>
2. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput. Med. Imaging Graph. **31**(4–5), 198–211 (2007). <https://doi.org/10.1016/j.compmedimag.2007.02.002>
3. Feoktistov, V.: Differential Evolution. Springer, Heidelberg (2006). <https://doi.org/10.1007/978-0-387-36896-2>
4. Finlayson, S.G., Bowers, J.D., Ito, J., Zittrain, J.L., Beam, A.L., Kohane, I.S.: Adversarial attacks on medical machine learning. Science **363**(6433), 1287–1289 (2019)
5. Finlayson, S.G., Chung, H.W., Kohane, I.S., Beam, A.L.: Adversarial attacks against medical deep learning systems. arXiv e-print (2019)
6. Ghaznavi, F., Evans, A., Madabhushi, A., Feldman, M.: Digital imaging in pathology: whole-slide imaging and beyond. Annu. Rev. Pathol. **8**, 331–359 (2013)

7. Huhtanen, J.: Potilaiden tietoja vietin psykoterapiakeskuksen tietomurrossa, yritys kertoo joutuneensa kiristykseen uhriksi. Helsingin Sanomat (2020). <https://www.hs.fi/kotimaa/art-2000006676407.html>
8. Kleinman, Z.: Therapy patients blackmailed for cash after clinic data breach. BBC News (2020). <https://www.bbc.com/news/technology-54692120>
9. Kügler, D., Distefano, A., Kuijper, A., Mukhopadhyay, A.: Exploring adversarial examples. In: Understanding and Interpreting Machine Learning in Medical Image Computing Applications, pp. 70–78. Springer (2018)
10. Latif, J., Xiao, C., Imran, A., Tu, S.: Medical imaging using machine learning and deep learning algorithms: a review. In: 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1–5 (2019). <https://doi.org/10.1109/ICOMET.2019.8673502>
11. Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., Lu, F.: Understanding adversarial attacks on deep learning based medical image analysis systems. Pattern Recognit. **110**, 107, 332 (2020). <https://doi.org/10.1016/j.patcog.2020.107332>
12. Paschali, M., Conjeti, S., Navarro, F., Navab, N.: Generalizability vs. robustness: adversarial examples for medical imaging. arXiv e-prints (2018)
13. Secretariat of the Security Committee: Finland's Cyber security Strategy, Government Resolution 3.10.2019 (2019). https://turvallisuuskomitea.fi/wp-content/uploads/2019/10/Kyberturvallisuusstrategia_A4_ENG_WEB_031019.pdf
14. Sipola, T., Puuska, S., Kokkonen, T.: Model fooling attacks against medical imaging: a short survey. Inf. Secur. Int. J. **46**(2), 215–224 (2020). <https://doi.org/10.11610/isij.4615>
15. Soumik, M.F.I., Hossain, M.A.: Brain tumor classification with inception network based deep learning model using transfer learning. In: 2020 IEEE Region 10 Symposium (TENSYMP), pp. 1018–1021 (2020). <https://doi.org/10.1109/TENSYMP50017.2020.9230618>
16. Su, J., Vargas, D.V., Sakurai, K.: One pixel attack for fooling deep neural networks. IEEE Trans. Evol. Comput. **23**(5), 828–841 (2019). <https://doi.org/10.1109/TEVC.2019.2890858>
17. The European Union Agency for Network and Information Security (ENISA): Smart Hospitals, Security and Resilience for Smart Health Service and Infrastructures. Technical report (2016). <https://doi.org/10.2824/28801>
18. Xu, H., Ma, Y., Liu, H.C., Deb, D., Liu, H., Tang, J.L., Jain, A.K.: Adversarial attacks and defenses in images, graphs and text: a review. Int. J. Autom. Comput. **17**(2), 151–178 (2020). <https://doi.org/10.1007/s11633-019-1211-x>

Big Data Analytics and Applications



Implementation of Big Data Analytics Tool in a Higher Education Institution

Tiago Franco¹(✉), P. Alves¹, T. Pedrosa¹, M. J. Varanda Pereira¹,
and J. Canão²

¹ Research Centre in Digitalization and Intelligent Robotics,
Polytechnic Institute of Bragança, Bragança, Portugal

{tiagofranco,palves,pedrosa,mjoao}@ipb.pt

² JCanão, Viana do Castelo, Portugal

jose.canao@jcanao.pt

Abstract. In search of intelligent solutions that could help improve teaching in higher education, we discovered a set of analyzes that had already been discussed and just needed to be implemented. We believe that this reality can be found in several educational institutions, with paper or mini-projects that deal with educational data and can have positive impacts on teaching. Because of this, we designed an architecture that could extract from multiple sources of educational data and support the implementation of some of these projects found. The results show an important tool that can contribute positively to the teaching institution. Effectively, we can highlight that the implementation of a predictive model of students at risk of dropping out will bring a new analytical vision. Also, the system's practicality will save managers a lot of time in creating analyzes of the state of the institutions, respecting privacy concerns of the manipulated data, supported by a secure development methodology.

Keywords: Big data analytics · Web-based tool · Higher education · Data extraction · Machine learning

1 Introduction

The impact of data-based decision making is known as a management revolution. Today's managers benefit from a range of powerful indicators for making crucial decisions, relevant information that was not feasible to obtain years ago [12].

For the field of education, the researchers point to great potential in the use of educational data [3,8,15]. Through the use of new environments that enable learning, such as online communities, discussion forums, chats, Learning Management Systems, among others, a large amount of data is produced inside the educational institutions. This volume is so large that traditional processing techniques cannot be used to process them, forcing educational institutions that want to take advantage of data to explore big data technologies [17].

Concerning the use of educational data, there is a clear difference between institutions that offer fully online training and institutions that have a more traditional education (they usually use online environments, but most of the education remains in person). For online education institutions, data studies are so refined and evolved that these institutions intend to personalize the process working at the individual level, seeking maximum effectiveness by adapting to the difficulties of each student.

For traditional educational institutions, the effective use of data for decision support is still uncommon, generally existing applications are limited to traditional statistical analyzes [6, 16]. This shortcoming is related to several technical challenges that are necessary to deal with the multiple sources of educational data that these institutions store for years [7]. Besides, there is also the difficulty of changing the culture of managers, who are sometimes used to old technologies and are not always convinced that they will produce valuable results by investing in the area [5, 14].

The US Department of Education [2] suggests and other studies agree [1, 5], that implementation should be done progressively, through collaborative projects across departments, and throughout development, it will incorporate other sectors until they get the whole data management ecosystem [14].

Following this line of thought, this article describes a possible software solution that was implemented on the Polytechnic Institute of Bragança (IPB) in Portugal, with the aim of exploring educational data without the need to modify the systems already in operation at the institution.

The purpose of the system is to provide a set of analytical and predictive information about the institution's academic situation for management improvement. As a requirement, the software should have the ability to expand to other educational institutions and has an architecture that aims to guarantee the protection of sensitive data.

Seeking to take advantage of studies already started on improving higher education and the willingness to solve problems already known by IPB, we designed the software in modules. In all, three modules were implemented: the first consists of a machine learning model for predicting students at risk of dropping out; the second consists of a set of graphs and analytical tables that seek to translate the current teaching situation of all courses at the institution; the third module deals with the creation of dynamic reports from the extracted data.

This document is divided into four more chapters that will continue the explanation of the study. Section 2 refers to the developed architecture, detailing its components and connections; The third chapter addresses the reasons why we implemented the 3 modules and their individual characteristics; In Sect. 4, the results of implementing the web tool in the educational institution are presented and discussed; Sect. 5 reports the main conclusions and scientific contributions that we have drawn from this study

2 Architecture

The task of implementing a big data communication structure involving several systems already in operation is not a trivial task. Initially, we seek to deal with the initial phases of the big data analysis process, which are in order of precedence: acquisition, extraction, and integration [3]. For that, we defined that it would be necessary an ETL system (Extraction, Transformation, and Loading) coupled to a server with access to the institution's active systems. Although its name is already very clear about its purpose, an ETL system can also be in charge of other activities that are necessary for the exchange of information between the data sources and the requesting system [18].

The next task in defining the architecture was to create a data collection group component, and its connection with all the other components in order to execute the modules defined. From this, we created the diagram in Fig. 1 to represent the architecture with its components and connections.

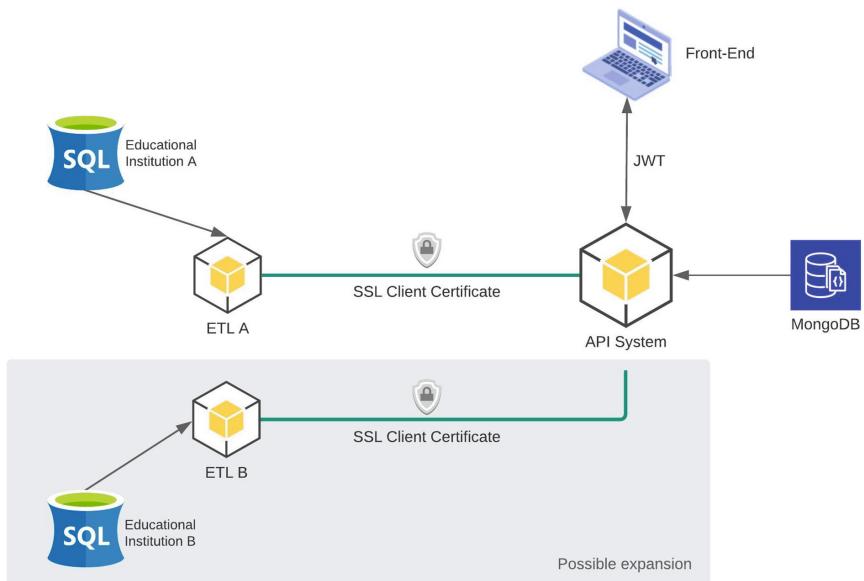


Fig. 1. Structure of the proposed architecture

The API System is a service modeled with the REST architectural style to provide the solution's main feature set. In this component, we manage the users, the triggers for the extraction tasks, process the extracted data, and execute the proposed modules.

This component depends on a secure direct connection to the MongoDB database, where the stored and processed data is located. With this approach,

it is possible to perform complex analyzes to respond to requests from the front end in a few seconds.

The ETL, as already mentioned, is the component that is installed next to the active systems that are intended to extract the data. This component is the only one in the architecture that accesses the educational institution's databases. This strategy makes it possible, when necessary, to encrypt sensitive data, ensuring that they are never stored in MongoDB and protecting them from malicious attacks. The component authentication to the API uses TLS/SSL client certificates to perform secure authentication and prevent dictionary attacks and/or brute force.

During the development of the architecture, we seek to incorporate the current concepts of microservices, big data analytics, and web development. To ensure cybersecurity, we follow the OWASP Secure Coding Practices manual [9]. Also, due to the fact that many users can process personal data, it was decided to consider the mandatory technical requirements defined in the Resolution of the Council of Ministers n° 41/2018 [13].

Thinking about a more commercial aspect of the software, it is possible to expand the application by adding several ETL components in other institutions. As data protection is already guaranteed by ETL on the institution side, the scope of the API component is even greater, as it can group multiple data sources to produce more powerful reports and machine learning models.

3 Implemented Modules

Bearing in mind that the software would not be just a prototype, after the defined architecture, we started to implement the system core. This core included multiple access levels, logging, the establishment of connections between components, the administrative panel, and automatic scheduling of extraction tasks.

To facilitate the audit of extractions, we have created a page to manage them. The main features of this page are: view the ETLs logs sent daily about their operation; check if the extraction tasks were successfully performed and their processing time; manage the frequency of automatic tasks and create an extra sporadic extraction.

With the software core assembled, we compiled a set of analyzes that had passed an initial phase of discussion and could be really implemented, resulting in the following modules.

3.1 School Dropout - Machine Learning Model

One of the main metrics to measure the quality of education in a higher education institution is the student dropout rate. Constantly government entities that deal with education, update their goals on recovery from dropout and seek new solutions to combat it [4].

Likewise, the educational institution that developed this study is also looking for new ways to improve this index. Among the proposed alternatives, the article

[10] proved to be promising for implementation, by validating the capacity of the institution's educational data in predicting possible dropouts through a machine learning model.

For its operation, the model developed takes into account data from three active systems in the institution. The first refers to the student's enrollment information and their status in the course, such as grades, number of approved subjects, academic years. The second system is the attendance record in the classroom. The last refers to the logs generated by the institution's Learning Management System (LMS).

The model requires a weekly snapshot of the student's current situation. With this data, the Random Forest algorithm is used to classify between dropout and non-dropout, using data from the last 4 years of the same week. Afterward, the number of times the student has been classified as dropout during the year is counted as a percentage. This percentage is called the critical rate, indicating students with the highest values as likely dropouts.

In order to implement the visualization of the study results, three pages were implemented. The purpose of the first page is to provide a set of comparisons on the critical rate between schools and courses at the institution. The second page developed, provides a report of the information collected and the results of the model of a specific student.

Seeking to take preventive measures, the third page was designed to support a call system for students at risk of dropping out. In this way, a specialized agent can get in touch with the student and discuss what problems he is facing and possible ways to deal with them. Afterward, the agent can record the reasons for the abandonment (if confirmed) and a report of the conversation.

3.2 Numerical Analysis

Analyzing the difficulties of the educational institution, we observed that an analytical report was produced every year. It's a report used by course coordinators to manage their courses and disciplines. The main analytical components consisted of the sum of the number of students enrolled, evaluated, and approved in each discipline of each course. Our initial idea was to transform this report into dynamic elements on a new page in the application.

In this way, we created a task that could extract all the data related to the material records of a requested year. This extraction resulted in the records being stored in MongoDB, already with the sum of each discipline, that is, it did not bring the students' personal information, only the number of how many students had that record.

With the data available to the front-end, the next step was to develop and increment the elements of the base report. Finally, we personalize the pages by access levels, covering more managers who may use the software.

3.3 Dynamic Reports

Unlike the other modules, the dynamic report was designed in the second stage of implementation. The idea arose from the need to always contact a developer to provide a new analytical element, even if the necessary data has already been extracted.

As the data stored in MongoDB has already been processed and did not contain sensitive data, we developed a tool that allows users to create their own queries to build graphs and tables. In this way, we build a sequence of steps necessary to build one of these elements.

After the user chooses the dataset they want to work with, the first step is to create the filters. Filters make it possible to simplify SQL operators to extract a subset of the original data. The second step is to select the fields that should be returned after the query. The third step is optional, its function is to create fields from calculations in other fields. As an example, it is possible for the teacher to experiment with different formulas to define the composition of the final grade of a discipline.

The fourth step depends on the user's objective, being possible to follow two paths. The first path aims to export a table, having the only functionality of ordering it. The second way is to create a chart. Thus, the fourth stage aims to choose which groupers will compose the chart. Finally, the fifth stage refers to the aesthetic settings of the chart, such as the names of the axes and the title.

With the definition of these elements, we developed a page for the construction of the report effectively. Similar to a text editor, the page has the advantage of being able to import the dynamic elements developed and resize them. With that, different agents of the institution can work with the educational data and create their own reports.

4 Results and Discussion

Following the completion of the software implementation, our work was directed to diagnose the system's impacts on the institution. In this session, a compilation of the impressions found by some managers who used the platform will be presented, comparisons between the results of the machine learning model of the original article [10] and the software developed and particularities found that are worth discussing.

4.1 Analytical Components

A significant improvement was noted in the practicality of creating a numerical analysis of the institution, mainly due to the fact that previously it was necessary to open multiple PDFs to view the same data that is now on a single dashboard. In addition, the amount of information that could be found was expanded, offering a new set of analyzes on teaching.

Similarly, the dynamic reports complement the numerical analyzes, offering the possibility to build an analysis from scratch. In a simplified way, the result

of these components is similar to typical business intelligence software, without the huge set of tools that we usually find, but with the difference of not having the need to import the data into the system, since the data is already available in the software for easy handling.

Figure 2 shows some of the components that can be found on the analysis page. The data displayed is emulated and does not reflect the reality of the institution.



Fig. 2. Screenshot of the analysis page

The first four boxes contain information on the averages of the number of students enrolled, evaluated and approved in the last 4 years of a chosen course. The following two graphs show the decomposition of these averages in the historical form of recent years. At the end of the figure, two graphs are presented referring to enrolled students and dropouts, the first informing the real values and the second in percentage.

4.2 Machine Learning Model

The first tests of the developed module sought to identify whether the extracted and processed data obtain the same result as the article that suggests the model [10]. The results indicate that the behavior of the model was similar to that expected, with a small increase in the number of students selected as dropouts,

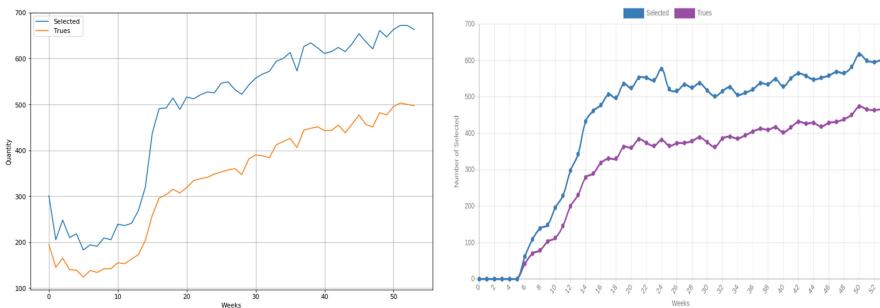


Fig. 3. Performance comparison of machine learning models

but justified by the increase in the number of students enrolled in that year. With that, we can say that the model was successful in its initial purpose.

The two graphs in Fig. 3 illustrate the performance of the predictive model of students at risk of dropping out of the institution in the 2016–2017 academic year. The graphic on the right was taken from the original article and on the left is the product of the implemented software. In both graphs, the upper blue lines represent the number of students that the algorithm selected as dropouts, the bottom lines represent the number of students who actually dropout. Thus, the difference between the lines is the error that the algorithm presents.

Comparing the two graphs, we can see a great divergence between weeks 1 and 20 and a decrease in the distance of the lines in the second graph. The difference in the first few weeks was caused by two factors. The first factor was the adjustments to the semester start dates and the frequency calculation. We found that the old algorithm could incorrectly calculate attendance if the student enrolled late.

The second factor was the addition of two new attributes that improved the model's performance. The first attribute added refers to the number of days it took the student to start the course and the second attribute refers to whether the student works or not during the course. This improvement is reflected throughout the year and can be seen in the decrease in the difference between the two lines of the second graph compared to the first graph.

4.3 Data Extraction

For the predictive model, we initially processed the data for the academic years from 2009 to 2018 in a development environment, leaving us to extract only the data from 2019. After correcting small bugs found, we noticed that the tool took an average of 30 to 40 min to extract all data for a requested week and 2 to 5 min for processing the model.

The differences between the times basically depended on two factors. The first is related to the week in which the data is to be extracted, since at the beginning of the year the amount of information generated by the students was

smaller, resulting in less processing time. The second refers to the time we do the extraction since the server could be overloaded.

In mid-day tests, the most critical moment of use for students, it could take twice as long. With an automatic task scheduler developed, we set up so that the extractions only occur at dawn, a time that had less use of the systems, minimizing the negative impacts of the software on the institution.

As the data required for the analytical module did not require much processing and are not bulky, the extractions took less than 1 min to be successful.

5 Conclusion

In this study, the following steps were described for the implementation of big data analytics tool in a higher education institution. Analyzing the results presented, we can conclude that our system can positively contribute to help the academic managers to improve the success of their students. We can highlight the time savings that the system's practicality has brought to the institution's managers.

Generally, most managers at an educational institution are also teachers, needing to divide their time between multiple tasks. The system presented allows managers to pay more attention only to data analysis since they no longer need to build a series of analytical components. In addition, if necessary, the system allows more complex analyzes to be made, exporting the data, streamlining the process with initial treatment, and covering more external tools.

As described in Sect. 3, two of the three modules implemented were already developed or at least quite discussed. This fact simplified several hours of our work and allowed us to refine the collected studies. This reality can be found in many educational institutions, with papers or mini-projects already developed internally but with several constraints to put it into practice. Once the proposed architecture is developed, it does not require complex refactoring to create a new data extraction in a safe way. Therefore, the inclusion of new modules is encouraged without prejudice to existing ones.

Another positive point observed was in the contact pages of students at risk of dropping out of school. In addition to fulfilling its purpose of enabling preventive contact with students at risk of dropping out, through the record of contacts a history of reasons that led students to dropout was created. In this way, it is possible to analyze the most worrying reasons in order to create a plan of preventive actions by the institution.

A similar strategy that supports the development of the architecture and the whole software can be seen at [11]. The HESA project aims to repeatedly collect a series of data from more than 250 educational institutions in the United Kingdom, creating a government ecosystem of educational data management. Undoubtedly it is an inspiring project for other countries, but this reality depends on a series of facts for its success, besides the monetary, the difficulty of its cooperation of the gigantic team involved. The study presented here, suggests an alternative for institutions that do not intend to wait years for this to happen and already reap the rewards of using data for decision making.

Acknowledgment. This work was supported by FCT - Fundação para a Ciência e a Tecnologia under Project UIDB/05757/2020 and Cognita Project (project number NORTE-01-0247-FEDER-038336), funded by the Norte 2020 - Norte's Regional Operational Programme, Portugal 2020 and the European Union, through the European Regional Development Fund.

References

1. Ali, L., Asadi, M., Gašević, D., Jovanović, J., Hatala, M.: Factors influencing beliefs for adoption of a learning analytics tool: an empirical study. *Comput. Educ.* **62**, 130–148 (2013). <https://doi.org/10.1016/j.compedu.2012.10.023>
2. Bienkowski, M., Feng, M., Means, B.: Enhancing teaching and learning through educational data mining and learning analytics: an issue brief, pp. 1–60 (2014)
3. Bomatpalli, T.: Significance of big data and analytics in higher education. *Int. J. Comput. Appl.* **68**, 21–23 (2013). <https://doi.org/10.5120/11648-7142>
4. European Commission: Education and training monitor 2019 - Portugal (2019)
5. Daniel, B.: Big data and analytics in higher education: opportunities and challenges. *Br. J. Edu. Technol.* **46**(5), 904–920 (2015). <https://doi.org/10.1111/bjet.12230>
6. Daniel, B.: Big data in higher education: the big picture, pp. 19–28 (2017). https://doi.org/10.1007/978-3-319-06520-5_3
7. Daniel, B., Butson, R.: Foundations of big data and analytics in higher education. In: International Conference on Analytics Driven Solutions: ICAS2014, pp. 39–47 (2014)
8. Dutt, A., Ismail, M.A., Herawan, T.: A systematic review on educational data mining. *IEEE Access* **5**, 15991–16005 (2017). <https://doi.org/10.1109/ACCESS.2017.2654247>
9. T.O. Foundation: OWASP secure coding practices quick reference guide (2010)
10. Franco, T., Alves, P.: Model for the identification of students at risk of dropout using big data analytics. In: INTED2019 Proceedings, 13th International Technology, Education and Development Conference, IATED, pp. 4611–4620, 11–13 March 2019. <https://doi.org/10.21125/inted.2019.1140>
11. HESA: About hesa. <https://www.hesa.ac.uk/about>. Accessed 01 Nov 2020
12. McAfee, A., Brynjolfsson, E.: Big data: the management revolution. *Harvard Bus. Rev.* **90**, 60–68 (2012)
13. de Ministros, C.: Resolução do conselho de ministros nº 41/2018. Diário da República nº 62/2018, Série I—28 de março de 2018, pp. 1424 – 1430 (2018). <https://data.dre.pt/eli/resolconsmin/41/2018/03/28/p/dre/pt/html>
14. Murumba, J., Micheni, E.: Big data analytics in higher education: a review. *Int. J. Eng. Sci.* **06**, 14–21 (2017). <https://doi.org/10.9790/1813-0606021421>
15. Romero, C., Ventura, S.: Educational data mining: a survey from 1995 to 2005. *Expert Syst. Appl.* **33**, 135–146 (2007). <https://doi.org/10.1016/j.eswa.2006.04.005>
16. Shacklock, X.: The potential of data and analytics in higher education commission (2016)
17. Sin, K., Muthu, L.: Application of big data in education data mining and learning analytics-a literature review. *ICTACT J. Soft Comput.*: Special Issue Soft Comput. Models Big Data, 4 (2015)
18. Trujillo, J., Luján-Mora, S.: A UML based approach for modeling ETL processes in data warehouses. In: Song, I.Y., Liddle, S.W., Ling, T.W., Scheuermann, P. (eds.) *Conceptual Modeling - ER 2003*, pp. 307–320. Springer, Heidelberg (2003)



Big Data in Policing: Profiling, Patterns, and Out of the Box Thinking

Sónia M. A. Morgado^(✉) and Sérgio Felgueiras

Major Events Laboratory, Research Center (ICPOL), Instituto Superior de Ciências Policiais e Segurança Interna, Lisbon, Portugal
{smmorgado,srfelgueiras}@psp.pt

Abstract. Big Data, being a massive amount of data, which requires technologies, information architecture and systems design, and analytical methods, has a significant impact on modern society and the development of security area strategies. This paper investigates the importance of Big Data in policing and discusses the challenges arising from its appliance to maintain public order. It is presented a theoretical study, based on a literature review in the context of policing that allows the establishment of the construct in which the Portuguese National Public Security developed its own Strategic Information System. For the study, some inclusion and exclusion criteria were used to narrow the gaps for a better understanding of the subject. Even though some questions arise from using the Big Data, such as processing, profiling, parsing algorithms that can conduct excessive normalisations of the data, Big Data is leverage in policing and a tool of predictive policing.

Keywords: Big data · National public security · Predictive policing · Security · Strategic information system

1 Introduction

Digital transformation, data access, technological change are the main elements of modern society and the core for evangelizing infrastructure transformation and cultural paradigms. This metamorphosis implies transformations associated with people, state of mind (trust, predisposition, and change acceptance), time, and leadership.

One of those transversal and longitudinal change, in terms of cultures, organisations, institutions, and countries is technology dependency, which reveals how much enveloped humankind is to the information-process revolution [1].

The concern around the dynamics and challenges of globalisation (economic, cultural, social, and technological) unfolds a set of thoughts that expresses the need to be supported and uphold the decision-making process, and therefore the strategic decision. On the other hand, they are the core of the fundamental elements of police activity and for the use of information from the police database.

Considering the modelling of thoughts, intuitive and analytic, that consider basic structural elements such as search, storage, it is possible to identify the major elements

for a decision. Even though the intuitive system WYSIATI (what you see is all there is), a cognitive bias is formed, from the fact that the level of information and knowledge is scarce as to be acceptable to daily-daily life [2], they induce disruptions such as: i) gaps; ii) negativity; iii) straight forward; iv) fear; v) size; vi) generalisation; vii) destiny; viii) single perspective; ix) guilt; and, x) urgency [3]. The bias conveyed by this approach is not compatible with police decisions because it can have counterproductive effects; however it is important to stress the awareness of the bias process.

The analytic system provides a more robust system capable of highly elaborate and various decisions because they are based on complex models, with predictive variables and modelling restrictions. The possibility of developing behavioural patterns, profiles, and predictive policing enables police to have an essential tool for the decision-making process.

The progress made in the police information data advocates the phenomenon that occurs from the application of knowledge and science to technology. The massive access to data available in the police context is a breeding ground for using advanced analytic tools, a proactive and qualifying way for presenting solutions, modelling behaviour, and reducing or preventing deviant behaviour opportunities. This context converges to the fictional and futuristic universe of “The Minority Report” of Dick [4], that foresee all crime before it occurs, based on the foreknowledge of Precrime. This is a predictive policing policy measure that promotes security and safety with a proactive approach to problems in developed countries [5], putting aside the reactive ones.

Big data utility is as big as a predictive system in policing. Predictive policing requires new organisations, competencies, and structures [6]. According to the author, this is a new way of thinking, a vision of police work, and consequently policing.

As to augment the understanding of the theme (evolution, the conceptual structure of the analysis), a literature revision was made. The computerized literature searches were performed using different platforms, data bases, article repositories, such as science-direct, Scientific Open Access Repository of Portugal (RCAAP), Scientific Electronic Library Online (SciELO), EBSCO, B-on, Proquest, Scopus, Clarivate Analytics - Web of Science e, Institute of Electrical and Electronics Engineers (IEEE).

Inclusion criteria were the following: i) title that contained the terms Big Data, law enforcement, security, national security, predictive policing; ii) published since 1990; iii) complete articles published in journals with peer review; iv) other documents, subjected to evaluation by peers or that had a public presentation (doctoral or master dissertations, book chapters); v) written in English and Portuguese; and, vi) references of consulted articles were considered.

To harmonise the principles of the process and maintaining scientific integrity, the exclusion criteria required the rejection of articles that: i) after a more accurate analysis wasn't related to Big Data; ii) documents not related to law enforcement, national security, policing; and iii) documents nor written in English, or Portuguese, even though the abstract was presented in the idiom of the inclusion criteria.

The paper has a traditional structure for theoretical studies, containing an introduction, state of the art, perspectives, and conclusion.

2 State of the Art

The origins of the taxonomy of Big Data are difficult to measure. The first representation of the term is unveiled by Mashey, identifying future needs from the expansion of technology, such as data storage, processing, and transmission of information [7].

With the ‘90s, Big Data’s concept has become commonplace, and in contemporary society, a technological buzzword. However, the idea incarcerates a multiplicity of dynamics that makes Big Data information a pattern applicable to health, economics, biology, security, and other areas.

Big Data associates with commercial transactions (on-line shopping), cyber research (search engines), access to social media (e.g. Facebook, Instagram), data storage, such as drive (e.g., Cloud, drive), communication networks (e.g., Whatsapp, Houseparty, tik tok). The tracking is managed with algorithms that can predict human behaviour, thus anticipating users’ needs and preferences. Nonetheless, a concept meaning is due. Authors such as Laney [8] convey a definition based on 3 V’s: Volume, Velocity, and Variety. To this model, were added more 2 V’s, for a complete concept: i) Veracity [9]; and, ii) Value [10], transforming it into a 5V’s model [11].

Even though it is a concept in progress, the inclusion of more elements, such as complexity and deconstruction [12, 13], focus on data [14–16], does not deter its importance as the underlying element of the security agenda.

Big Data is also viewed as a distinct set of assumptions that form a paradigm, surmounted by information and the management of the databases, making space for the solution-problem binomial complements [17]. For this reason, information, technology, methods, and impacts are also notions that can endorse the concept of Big Data [18].

In terms of information, the historical moment is digitization. It consisted of transforming the society from analogical signs to digital ones, based on binary code [19]. The change allowed Google to turn available, and consequently universal, access to books [20].

As a natural evolution of the process, “datafication” empowerment results from digitization and presented a more accurate representation of reality [21]. This data processing mechanism allows the structuring of information, pattern extraction, correlation, recognition of linguistic evolution, and tendencies of words and expressions [22]. An example of this application is the repression of the Nazi lexicon [23].

The transition converges to the threefold of data: information–knowledge–wisdom [24]. The definition of Jifa [25] allows us to approach to Maslow hierarchy of needs. In fact, data is the physiological need for Big Data. Information conveys the safety needs – security and safety – and by adding context after the processing and treatment of data, the sense of belongingness. The comprehension provided and the way on how to use information efficiently catapults the esteem, confidence, and accomplishment needs. Ultimately, Maslow’s self-actualisation needs can be perceived in the Big Data by the wisdom of knowing when and why information is used, allowing the growth and challenges provided by society. Collecting data from the variety of disparate sources/silos and transforming it into re-usable, consumable, and executable information converges into being the boost and the underlying fundamentals of Big Data [18].

The growing volume of information available results inexorably in the era of web 3.0, the network society era. The trivialisation of technology (smartphones, tablets,

smartwatches), interaction platforms (zoom, teams, streaming's), and the multitude of other tools convey the biggest challenge for Big Data [26]. The challenge goes alongside how to derive the most value of the data. In this backdrop, Dhankhad [27] considered that 90% of all that were produced in the last two years, even though they didn't go through the triad process of information–knowledge–wisdom.

Concerning the technology approach, the onus is embedded in the capacity of the systems processing the raw and siloed data, deliver either by machine learning or artificial intelligence, into something meaningful is the critical component of Big Data.

Accomplishing the task of generating information requires data transfer, which entails a toolset capable of transferring billions of records and attributes. The tool can arise some questions that are easily overcome using specific techniques of benchmarking [28]. Herein lies the “harsh spot” for data transfer that can be under-evaluated because the velocity of data is overcoming analytical function efficiency [29]. The improvements in the processing are shaping the world of Big Data [30]. The tools (NoSQL, MapReduce, machine learning) that can emphasize bigger and complex analysis are Big Data's mainframe [31].

The methods that allow the blending of datasets, reorganisation, or simple redefinition to drive action, are the propulsors for granting quality an intrinsic value to the information, object of evaluation from the users and consumers [32, 33]. After this first stage, the second one embarks in analytical modes, such as cluster analysis, genetic algorithms, natural language processing, machine learning, neural networks, predictive models, regression and logistics models, social media analysis, feeling and perceptions analysis, signs and data visualization [18, 22, 26, 34]. As such, Big Data, an ongoing evolving process, proposes a fundamental scale of information, which encompasses the most basic expression of a strategic element. Because Big Data it is still undergoing evolution, the decision-making process should be updated and amended regularly in line with upgrades in methods, technology and innovations, analytic, and the results of the evaluations [35].

The integration of the information collected from the Big Data in the decision-making process enables smarter, intelligent, and efficient decisions, the reason which sustains an assertive, effective, and optimized extraction of knowledge [27]. The human intervention for the analysis accordingly with organization's needs, the comprehension of data, is still fundamental [30].

As a source of information, knowledge, and wisdom, the systems present some disruptive components. Those disruptions are information leaks, access, and illegitimate divulgence of personal information, that compromises cybersecurity and the protection and privacy of users and consumers [26]. It is known that the network involved in Big Data catapults information leaks and misuse of information. Applying technological, political, and legal mechanisms can be an assurance of stability and trust. [26]. However, the indiscriminate access to open sources, the tracking of various systems reveals the urgency for regulating access to information, ensuring impartiality, independence, and preventing the information monopoly [36].

The conciliation of the different approaches, can state that Big Data, is an intangible asset, comprised of data that storage volume, velocity, variety, and veracity. The treatment

requires specific technology and analytical tools, as to drive new value from unrestricted quantities of data [18].

3 Big Data and Security Context

Security as a way to maintain integrity embodies an element of defence and position conquering. The cross-pollination of these concepts' ideas and insights converge to the main basis: public order, harmony, and tranquillity of citizens [37]. Security is a sensitive sector therefore reactive, or proactive police strategies are subject to different stages of the evolution of society and criminality, which were exponentially catapulted with the advent of technology.

The increased intricacies of security concept through the decades [38], in different aspects – emotional, monetary, labour, environmental, health, State, global, and security in *stricto sensu* – stems from “technological development and electronic process, that form an inseparable triad of certainty and risk” [39]. On the other hand, the transformation in security also illustrates its impact on police information and intelligence [40]. The government uses police powers to reduce uncertainty and risk [40]; however, the rules made by the people and for the people, even though aren't an intrusion on liberty are scarce to keep up-to-date to the mutability of the security concept [22]. This context reveals some of the underlying difficulties that ultimately lie at the heart of traditional policing facing the new paradigm. Such a paradigm is based on the level of public scrutiny, operational and command strategies, the use of technology, responsibilities and accountability, material, and human resources [38]. Interestingly, the comprehensive approach to main elements in security, protection, and crime prevention opts for a proactive operational strategy, even in uncertainty, which is critical in the decision-making process [41]. The legal effectiveness of punishment refers to the extent that citizens' behaviour is really kept within predetermined limits. The measurement of effectiveness is often difficult and does not guarantee public security [42].

Ensuring the privileged link with democratic state and public security, the Police understood the need to adapt to new challenges. The critical mass of policing is sustained by scientific studies that determine the definition of Big Data. From this perspective, traditional policing, problem-oriented policing, or intelligence-led policing follow along on a well-trod not static approach. Moving beyond broad security theory, as it occurs is economic theory, there are no policing models in true form. According to some authors [43], policing focuses on a hybrid model, with a technological and holistic component. Within a prospectively set of macro set elements from models, in a hybrid model, there is no hierarchy between elements.

Before establishing the model, Big Data is a potential ally of police organisation and intelligence [44]. Even though the ambivalent feelings and the anxiety derived from its appliance to security [45], is a management dissension. For this reason, it is the subject of the debates to deter and apply the volume of data to intelligence to prevent crime, which is a consequence of economic, social, human changes alongside with the opportunity, and the existence of a natural crime rate [46]. Unable to be implemented faultless because of lack of adequate formation, and because it was not understood, the potential of Big Data is the main inhibitors of the use as a police resource [22, 44, 47]. It

has been suggested that the lack of objectivity in the analysis of data, together with the excess of data recollected, can be counterproductive and a disruptor element in being leverage and a solution for the problems [48]. The significant barrier to objectivity had been the employment of analytical methods for information analysis. The algorithms use a broad interpretation that contains discriminative elements [49], which conduct to individual's approach that Data defines as having deviant behaviour [50]. Accountability is also a problem. Critics claimed that the bias results from the externalities of action, sustained by the Big Data [51]. The onus of accountability has deviated to technology, which undermines the effect on the algorithm creator [50] and hedges the prejudice latent in the human being [49].

Big Data has been incorporated in policing. The ADN sample, video surveillance system, and predictive policing [52] are some of the examples. The analysis dynamics are such that problem-solving crime analysts flowed to present the best option for crime prevention [53].

The integration of big data in the police decision-making process requires that some requirements must be met: i) risk identification; ii) quality of information; iii) methodology; iv) knowledge of positive impact in police action and relation with the community; v) technology testing; and, vi) guaranteeing respect for citizens' rights [49]. The factors that make predictive policing interesting but challenging to the police are the efficiency that it is still to be proved [44], and the inability to capture the individual's individuality and circumstances of the behaviour in different moments and environments [6].

Information is not only created through discovery but through the compilation and organisation of existing data, upsurges, and assists the policing models of Portuguese National Public Police (PSP).

In fact, PSP also sustains the activity in the Strategic Information System, Management and Operational Control (SEI). Police use SEI to make rational decisions in terms of operational, tactical, and strategic operations. The system's implementation occurred in 2004, revealing characteristics and fundamental issues from Big Data [22].

4 Discussion and Conclusion

The paper discussed some general aspects of Big Data, with particular emphasis upon PSP. Over the last decades, Big data has emerged, and some questions raised. It is introduction as a tool in predicting behaviours incurred in police, administrative, and political complexities. Those complexities are subdued if the analysis is error-free and subjected to the actors and stakeholders' discretion.

Recognizing the challenges of a society technologically dependent, police organisations embrace this process by sustaining the decision-making process in the data that, after being collected, is processed, analysed until its information is upgraded to intelligence.

Despite the Big Data paradox's – transparency, identity, and power [54] – the ongoing process's irreversibility is unquestionable. The intelligence sustained by the Big Data allows quicker systematic responses, analysing contexts in a micro perspective, forecasting events, profiling, management of transit flows.

The present prominent role of Big Data, and the interconnection that may generate "data village", might enhance police productivity [55, 56] and the decision-making

process [56] and is richer if it can fulfil an interactive flow between individuals (citizen-police, and police-citizen) at one point in time and environment.

The widespread of the phenomena of crime implies its comprehension, the knowledge of transversal elements, the regularity, and the ability to predict the occurrence of deviant behaviour. Moving toward a larger perspective, it is crucial to understand that all this knowledge optimizes resources for the decision [57].

The fallacy of technological police operates over the assumption that technology will override Police's human side. The human side is the main core of Police action. Technology, rather than a substitute, is an essential complement to Police [58] because it requires human intervention in the analysis and evaluation and optimizes the intervention. Expanding the "law of the vital few" of Pareto [59], it can be advocated the police adage: that 20% of Big Data produce 80% of police action; that 20% of police action produce 80% of Big Data.

To grasp the complex process by which Big Data, stakeholders, police decisions are linked is useful to consider the analytical schema, the critical and clear vision of facts and events. Even if outliers may compromise the analysis, the micro and macro evolvement dynamics should be revisited in the normality of behaviours, thoughts, profiles, sustained in the Gauss curve that fundaments all the natural phenomena.

Big Data provides opportunities for the Police to transform data into contextualized information through analytical methods and techniques, leading to a new approach to smart policing. The characterization of deviant and criminal behaviour allows a proper interpretation of occurrences, for example through profiling, identification of patterns, appropriate use of force, enabling a more advantageous, rational, transparent and accountable police action. In short, Big Data is essential for the study of phenomena that collide with the security of people and the community. Knowledge increases the quality of each police officer's intervention because by decreasing the uncertainty and surprise of each situation, it ensures harmonisation of the standard of police action.

Big Data's effects can be observed at the organizational level by introducing new organizational models and work processes. It develops a collective framework and individual awareness, strongly supported by the technological component and scientifically validated practices. The analysis from a daily perspective of the benefits of using Big Data creates research opportunities in the near future.

As a final comment upon the future scope of Big Data in PSP, it is reasonable to speculate that it will be increased. Its appliance will turn out to develop the thinking out of the box as a way of successful and effective policing.

References

1. Morgado, S.M.A., Moniz, T., Felgueiras, S.: Facebook and polícia de segurança pública: an exploratory study if follower's engagement. In: Rocha, Á., Reis, J., Peter, M., Bogdanović, Z. (eds.) Marketing and Smart Technologies. Smart Innovation, Systems and Technologies, pp. 363–376. Springer, Cham (2020)
2. Kahneman, D.: Pensar depressa e devagar. Temas e debates, Lisboa (2002)
3. Rosling, H., Rönnlund, A., Rosling, O.: Factfulness. Temas e debates, Lisboa (2019)
4. Dick, P.D.: The Minority Report. Orion Books, London (2005)

5. Morgado, S.M.A., Anjos, O.: Qualitative methodology helping police sciences: building a model for prevention of road fatalities in São Tomé and Príncipe. In: Costa, A., Reis, L., Moreira, A. (eds.) Computer Supported Qualitative Research. WCQR 2018. Advances in Intelligent Systems and Computing, vol. 861, pp. 291–304. Springer, Cham (2018)
6. Pais, L.G.: Predictive policing: Is it really an innovation? In: European Law Enforcement Research Bulletin, Special Conference Edition: Innovations in law enforcement: Implications for practice, education and civil society, (4 SCE), pp. 125–131. CEPOL, Budapest (2018)
7. Lohr, S.: The origins of ‘big data’: an etymological detective story. [Bits]. <https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/>. Accessed 08 Nov 2020
8. Laney, D.: 3-D data management: controlling data volume, velocity and variety. META Group. <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Accessed 20 Oct 2020
9. Schroedl, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P.: Analytics: The Real-World Use of Big Data. https://www.informationweek.com/pdf_whitepapers/approved/1372892704_analytics_the_real_world_use_of_big_data.pdf. Accessed 20 Oct 2020
10. Dijcks, J.: Oracle White Paper: Big Data for the Enterprise. Oracle Corporation, Redwood Shores (2013)
11. Higdon, R., Haynes, W., Stanberry, L., Stewart, E., Yandl, G., Howard, C., Broomall, W., Koller, N., Kolker, E.: Unravelling the complexities of life sciences data. *Big Data* **1**(1), 42–50 (2013)
12. Intel IT Center: Big data analytics: Intel’s IT manager survey on how organizations are using Big Data. Intel IT Center, Intel Corporation, Santa Clara (2012)
13. Suthaharan, S.: Big data classification: problems & challenges in network intrusion prediction with machine learning. *Perform. Eval. Rev.* **41**(4), 70–73 (2014)
14. Beyer, M.A., Laney, D.: The importance of “Big Data”: A definition (Report No. G00235055) (2012)
15. Zaslavsky, A., Perera, C., Georgakopoulos, D.: Sensing as a service and big data. In: Proceeding of the International Conference on Advances in Cloud Computing (ACC), pp. 21–29. ACC, Bangalore (2013)
16. Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T., Lapis, G.: Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw, New York (2011)
17. Gebert, G.: What is Big Data?. (Paper presentation). In: Conference on Big Data in Forensic Science. Policia Judiciária, Lisboa (2014)
18. De Mauro, A., Greco, M., Grimaldi, M.: A formal definition of big data based on its essential features. *Libr. Rev.* **65**(3), 122–135 (2016)
19. Coyle, K.: Mass digitization of books. *J. Acad. Libr.* **32**(6), 641–645 (2006)
20. Somers, J.: Torching the modern-day library of Alexandria. The Atlantic. <https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/>. Accessed 08 Nov 2020
21. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work and Think. John Murray, London (2013)
22. Pereira, M.R.: Big Data: O caso do Sistema Estratégico de Informação, Gestão e Controlo operacional da Policia de Segurança Pública. (Unpublished Master Thesis). Instituto Superior de Ciências Policiais e Segurança Interna, Lisboa (2016)
23. Michel, J.B., Shen, Y.K., Aiden, A., Veres, A., Gray, M.K., Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M.A., Aiden, E.L.: Quantitative analysis of culture using millions of digitized books. *Science* **331**(6014), 176–182 (2011)
24. Rowley, J.: The wisdom hierarchy: Representations of the DIKW hierarchy. *J. Inf. Sci.* **33**(2), 163–180 (2007)

25. Jifa, G.: Data, information, knowledge, wisdom and meta-synthesis of wisdom - comment on wisdom global and wisdom cities. *Proc. Comput. Sci.* **17**, 713–719 (2013)
26. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, New York (2011)
27. Dhanklad, S.: A brief summary of apache hadoop: a solution of big data problem and hint comes from Google, towards data science. <https://towardsdatascience.com/a-brief-summary-of-apache-hadoop-a-solution-of-big-data-problem-and-hint-comes-from-google-95fd63b83623?fbclid=IwAR2qGnPIGhsgtoW4Za0IgJsD2uFMB5hv4xMLZaouy1OgdNhPUJdSJyW0b0>. Accessed 08 Nov 2020
28. Xiong, W., Yu, Z., Bei, Z., Zhao, J., Zhang, F., Zou, Y., Xu, C.: A characterization of big data benchmarks. In: Proceedings – 2013 IEEE International Conference on Big Data, Big Data, pp. 118–125. IEEE, Santa Clara (2013)
29. Moore, G.E.: Cramming more components onto integrated circuits. *IEEE Solid-State Circ. Newslett.* **11**(5), 33–35 (2006)
30. McAfee, A., Brynjolfsson, E.: Big data: the management revolution. *Harv. Bus. Rev.* **90**(10), 61–67 (2012)
31. Ward, J.S., Barker, A.: Undefined by data: a survey of big data definitions. <https://arxiv.org/pdf/1309.5821.pdf>. Accessed 12 Dec 2020
32. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.* **12**(4), 5–33 (1996)
33. Wang, R.Y., Ziad, M., Lee, Y.W.: Data Quality: Advances in Database Systems. Kluwer Academic Publishers, Dordrecht (2002)
34. Chen, H., Chiang, R., Storey, V.: Business intelligence and analytics: from big data to big impact. *MIS Q.* **36**(4), 1165–1188 (2012)
35. Dumbill, E.: Making sense of big data. *Big Data* **1**(1), 1–2 (2013)
36. Manovich, L.: Trending: the promises and the challenges of big social data. <https://manovich.net/content/04-projects/067-trending-the-promises-and-the-challenges-of-big-social-data/64-article-2011.pdf>. Accessed 09 Nov 2020
37. Felgueira, S., Machado, P.: Modelo de diagnóstico de Ordem Pública: uma abordagem metropolitana sincrónica. In: Rollo, M.F., Gomes, P.M., Rodríguez, A.C. (eds.) *Polícia (s) e Segurança Pública: História e Perspetivas Contemporâneas*, pp. 395–418. MUP, Lisboa (2020)
38. Bayley, D.H.: The complexities of 21st century policing. *Policing: J. Policy Pract.* **10**(3), 163–170 (2016)
39. Morgado, S., Mendes, S.: O futuro numa década: Os desafios económicos e securitários de Portugal. *Politeia – Revista do Instituto de Ciências Policiais e Segurança Interna Ano X-XI-XII: 2013–2014–2015* (1: Studio varia), pp. 9–35 (2016)
40. Mendes, S., Morgado, S.: Intelligence services intervention: constraints in Portuguese democratic state. In: Teixeira, N.S., Oliveira, C.S., Lopes, M., Sardinha, B., Santos, A., Macedo, M. (eds.) *International Conference on Risks, Security and Citizens: Proceedings/Atas*, pp. 285–297. Município de Setúbal, Setúbal (2017)
41. Felgueiras, S., Pais, L.G., Morgado, S.M.A.: Interoperability: diagnosing a novel assess model. In: European Law Enforcement Research Bulletin, Special Conference Edition: Innovations in law enforcement: Implications for practice, education and civil society, (4 SEC), pp. 1–6. CEPOL, Budapest (2018)
42. Nagin, D.S., Solow, R.M., Lum, C.: Deterrence, criminal opportunities, and police. *Criminology* **53**(1), 74–100 (2015)
43. Elias, L.: Desafios e prospetiva. Instituto Superior de Ciências Policiais e Segurança Interna, Lisboa (2018)

44. Chan, J., Moses, L.B.: Making sense of big data for security. *Br. J. Criminol.* **57**(2), 299–319 (2016)
45. Crawford, K.: The anxieties of Big Data. *The New Inquiry* (2014). <https://thenewinquiry.com/the-anxieties-of-big-data/>. Accessed 20 Oct 2020
46. Morgado, S.M.A.: Crime and socio-economic context: a framework approach. In: Proceedings in Advanced Research in Scientific Areas (ARSA 2013), pp. 139–142. EDIS, Slovakia (2013)
47. Babuta, A.: Big data and policing: an assessment of law enforcement requirements, expectations and priorities. Royal United Services Institute for Defence and Security Studies. https://rusi.org/sites/default/files/201709_rusi_big_data_and_policing_babuta_web.pdf. Accessed 13 Nov 2020
48. Ratcliffe, J.: *Intelligence-Led Policing*. Routledge, New York (2016)
49. Ferguson, A.G.: *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York University Press, New York (2017)
50. Speranza, A.: Big data and law enforcement: on predictive policing. Medium. <https://medium.com/@alishope/big-data-and-law-enforcement-on-predictive-policing-a16af882dd2>. Accessed 14 Oct 2020
51. Bovens, M.A.P., Schillemans, T., Goodin, R.E.: Public accountability in MAP. In: Bovens, M.A.P., Goodin, R.E., Schillemans, T. (eds.) *The Oxford Handbook of Public Accountability*, p. 18. Oxford University Press, Oxford (2014)
52. Joh, E.E.: Policing by numbers: big data and the fourth amendment. *Washington Law Rev.* **89**(1), 35–68 (2014)
53. Wyllie, D.: Rise of the crime analyst, PoliceOne. <https://www.policeone.com/police-products/software/Data-Information-Sharing-Software/articles/6396540-Rise-of-the-crime-analyst/>. Accessed 13 Nov 2020
54. Richards, N.M., King, J.H.: Three paradoxes of big data. *Stanford Law Rev.* **66**(41), 41–43 (2013)
55. Chen, H., Yan, Z.: Security and privacy in big data lifetime: a review. In: *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, pp. 3–15. Springer, Cham (2016)
56. Desai, P.V.: A survey on big data applications and challenges. In: *Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant*, pp. 737–740. Gnanamani College of Technology, Tamilnadu (2018)
57. Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., Liu, Q.: Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access* **7**, 106111–106123 (2019)
58. Perry, W.L., McInnis, B., Price, C.C., Smith, S.C., Hollywood, J.S.: *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation, Santa Monica (2013)
59. Pareto, V.: *Cours d'économie politique*. Librairie Droz, Geneva (1964)



Research Trends in Customer Churn Prediction: A Data Mining Approach

Zhang Tianyuan^(✉) and Sérgio Moro

Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisbon, Portugal

Abstract. This study aims to present a very recent literature review on customer churn prediction based on 40 relevant articles published between 2010 and June 2020. For searching the literature, the 40 most relevant articles according to Google Scholar ranking were selected and collected. Then, each of the articles were scrutinized according to six main dimensions: Reference; Areas of Research; Main Goal; Dataset; Techniques; outcomes. The research has proven that the most widely used data mining techniques are decision tree (DT), support vector machines (SVM) and Logistic Regression (LR). The process combined with the massive data accumulation in the telecom industry and the increasingly mature data mining technology motivates the development and application of customer churn model to predict the customer behavior. Therefore, the telecom company can effectively predict the churn of customers, and then avoid customer churn by taking measures such as reducing monthly fixed fees.

The present literature review offers recent insights on customer churn prediction scientific literature, revealing research gaps, providing evidences on current trends and helping to understand how to develop accurate and efficient Marketing strategies. The most important finding is that artificial intelligence techniques are obviously becoming more used in recent years for telecom customer churn prediction. Especially, artificial NN are outstandingly recognized as a competent prediction method. This is a relevant topic for journals related to other social sciences, such as Banking, and also telecom data make up an outstanding source for developing novel prediction modeling techniques. Thus, this study can lead to recommendations for future customer churn prediction improvement, in addition to providing an overview of current research trends.

Keywords: Telecom · Data mining · Customer churn prediction

1 Introduction

With the rapid development of computer and Internet technologies, people's lives have undergone earth-shaking changes. Changes in the form of communication have prompted the telecommunications industry to flourish (Sun 2018). In the “Big Data era” of information explosion, as one of the leading industries in the information age, the development of the telecom industry depends not only on communication technology, but also on the resource optimization and configuration capabilities of enterprises, and the management

of huge information and data resources becomes an enterprise. Massive data accumulation in the telecommunication (telecom) industry and the widespread application of data warehouse technology make it possible to gain insight into customer behavior characteristics and potential needs through systematic customer historical data records. It also provides prerequisites for targeted Marketing in the telecom industry (Wang et al. 2018).

Telecom operators have accumulated a large amount of customer information and consumption data during their development. These data truly and objectively reflect the behavior of consumers. Combining data mining technology with the rich data resources of the telecom industry can effectively help telecommunications companies predict customer churn and develop more accurate, efficient and effective Marketing strategies.

2 Overview

This research investigates 40 relevant articles published between 2010 and June 2020 and characterizes the customer churn prediction using data mining studies on their areas of research, main goals, dataset volume, techniques adopted and the outcomes according to each study. Most research areas of these articles are telecom. These studies are selected to represent different and recent literature analysis methodologies on research areas closely related to customer churn prediction, which is the focus of the proposed research. It is also taken into account that each of those studies mention the goal and method of research, expressed in the columns of Table 1, to enable comparing different approaches with the proposed method.

Customer churn prediction modelling is significantly affected by diverse factors, such as data mining techniques and their specificities, available data, data quality and data granularity. Other features such as modelling decisions conduct different operation of success and how it is evaluated. In terms of datasets, there is a big difference regarding source, volume, nature and quality. The data source used for customer churn studies is mostly originated from the big telecommunication companies or operators. There is a lot of explanatory features which could be found in literature, some researches use only a few features, but other researches make use of hundreds of features. The further investigation has been carried out through gathering the most popular features and divide them into distinct clustering groups, such as: demographic features, business features, industry features and SMS message features. Table 1 answers the following questions: Which are the most used techniques in the customer churn forecast? Thus, it is possible to verify that decision tree (DT), support vector machines (SVM) and Logistic Regression (LR) are the three most popular and useful method. By analyzing each method, it can be observed that these three methods are efficient techniques for extracting implicit information from the database and with high accuracy.

3 Literature Review

Nowadays, customer churn is one of the growing issues of today's competitive and rapidly growing telecom industry. The focus of the telecom industry has shifted from acquiring new customers to retaining existing customers owing to the associated high cost (Hadden et al. 2007). The telecom industry can save Marketing cost and increase

sales through retaining the existing customers. Therefore, it is essential to evaluate and analyze the customers' satisfaction as well as to conduct customer churn prediction activity for telecom industry to make strategic decision and relevant plan.

There are two popular algorithms with good predictive performance and comprehensibility in the customer churn prediction area: decision trees (DT) and logistic regression (LR) (Verbeke et al. 2012). But these two algorithms also have their shortcomings: decision trees are inclined to have problems to deal with linear relations between variables and logistic regression has problem with interaction effects between variables. Therefore, the logit leaf model (LLM) is proposed, which is a new algorithm that could better classify data. LLM tends to construct different models on segments of the data (not on the entire dataset), which could have better predictive performance while keeping the comprehensibility from the models. The LLM be composed of two stages: a segmentation stage and a prediction stage. Customer segments are recognized in the first stage and a model is formulated for each leaf of the tree in the second stage. After test and case study, we found some key advantage of the LLM compared to decision trees or logistic regression. (Caignya et al. 2018).

Customer churn problems could be solved from two different angles. One is to improve customer churn prediction models and boost the predictive performance (Verbeke et al. 2012). Another is trying to understand the most important factors that drive customer churn such as customer satisfaction. Customer churn prediction is considered as a managerial problem which is driven by the individual choice. Therefore, many researchers mention the managerial value for customer segmentation (Hansen et al. 2013). By considering the two research angles, customer churn prediction models need to create actionable insights and have good predictive performance.

Customer churn predication is part of customer relationship management since retaining and satisfying the existing customers is more profitable than attracting new customers for the following reasons: (1) Profitable companies normally keep long term and good relationships with their existing customers so that they can focus on their customer needs rather than searching new and not very profitable customers with a higher churn rate (Reinartz and Kumar 2003); (2) the lost customers can influence other customers to do the same thing using their social media (Nitzan and Libai 2011); (3) long-term customers have both profit and cost advantages. On the profit dimension, long term customers have tendency to buy more and they can recommend people to the company using positive words. On the cost dimension, they have less service cost since a company already masters information about them and understands their customer needs (Ganesh et al. 2000). (4) Competitive marketing actions have less effect on long term customers (Colgate et al. 1996); (5) Customer churn increases the demand and the cost to draw new customers and decreases the potential profits by the lost sales and opportunities. These effects lead to that retaining an existing customer has much smaller cost than drawing a new customer (Torkzadeh et al. 2006). Therefore, customer churn prediction is very necessary in a customer retention strategy.

Currently, Customer Relationship Management (CRM) is valued by many companies, since customer retention, which concentrate on developing and keeping long-term, loyal and profitable customer relationship, is an important factor for the company to win investment. Developing effective retention methods is critical for businesses, especially

for telecom operators since they lose 20% to 40% of customers per year (Orozco et al. 2015). Retaining existing customers doesn't have the cost of advertising, educating or creating new accounts as attracting new customers. Consequently, compared with attracting new customers, retaining an existing customer is five times cheaper (McIlroy and Barnett 2000). Decreasing customer churn rate from 20% to 10% can lead to annually saving about £25 million to the mobile operator Orange (Aydin and Özer 2005).

Predicting customer churn has been a subject for data mining. Compared with traditional surveys, using data mining is better at investigating customer churn (Huang et al. 2012). Traditional surveys suffer from high cost and limited access to the customer. However, data mining overcomes this kind of problem, which provides conclusion based on the analysis of historical data. Therefore, data mining becomes the most common method in customer retention to predict if customer will churn or not and identify patterns using customers' historical data (Liu and Fan 2014).

Many methods were used to predict customer churn in telecom companies. Most of these methods have applied data mining and machine learning. Most of the related work used only one method of data mining to obtain knowledge, and the other works tried to compare several different methods to predict churn (Ahmad and Aljoumaa 2019). (Bran-dusou et al. 2016) proposed an up-to-date data mining method to predict the prepaid customers' churn using 3333 customers' dataset with 21 features, and a dependent churn variable with two values: Yes/No. Some features consist of data about the number of customers' messages and voicemail. The author used "PCA" (the principal component analysis algorithm) to decrease data sizes. Tree machine learning algorithms including Bayes Networks and Neural Networks are used to predict churn factor. AUC is applied to measure the performance of the algorithms. The AUC values for Bayes Networks is 99.10%, for Neural networks is 99.55%. The dataset is small and there is no missing values in this study.

Makhtar et al. (2017) presented a telecom customer churn prediction model using rough set theory. The authors mentioned that, compared with other algorithms such as Decision Tree (DT), Linear Regression (LR), rough set classification algorithm achieves better predictive performance. Nevertheless, most approaches only focus on predicting customer churn with higher accuracy, very few approaches investigated the intuitiveness and understandability of a churn prediction system to recognize the customer churn reason (Bock and Poel 2012). However, (Idris et al. 2017) presented an advanced churn prediction method based on genetic programming (GP)'s strong searching ability supported by AdaBoost, which can recognize the factors leading to telecom customer's churn behavior. This study aims to apply the searching and learning ability of GP-AdaBoost method to design an intuitive and effective telecom customer churn prediction system.

4 Research Methodology

This study conducts a literature review on customer churn prediction. Therefore, the first task is to collect relevant literature on the domain being analyzed for building a comprehensive body of knowledge (Moro et al. 2020). The reason for this is to identify the research gap and see where this research may contribute to existing body of knowledge. Google Scholar is one of the most popular search engines to search academic articles and publications (Harzing 2013). The following search query: "Telecom"

OR “Customer churn forecast” OR “Data mining” OR “machine learning” was chosen for querying its database for articles. The filters used included setting the timeframe period for publications/articles from 2010 up to the present and keeping out patents. The number of hits is 17,800, and the 40 most relevant articles published in journals were gathered for a deeper analysis with roots on the famous Google’s search engine. Only articles/publications from experiments using data-driven approaches for customer churn forecasting, for example, empirical analyzes based on real data were considered. Each of the articles was checked carefully for investigating what method was used for data analysis, what was the timeframe and from which country were the data come. These three dimensions made up the three key element for the critical analysis and comparative analysis of the literature gathered. The study aspires to better understand the inherent laws of the telecom market business and obtain a control method for telecom customer management risk.

5 Results

The 40 articles gathered were published in a total of 30 different journals (Table 1 shows those from which more than one article was selected), corresponding to 11 different publishers (Table 2 for those publishers with more than one article selected). Such numbers prove that telecom forecasting is not totally limited to specific telecom literature, even though telecom gets the largest share, with 68 per cent of articles; on the contrary, the investigations found a bigger range of sciences, with a special emphasis on bank, energy and online social network literature (Table 3). The fact that big data in telecom industry are currently helpful for discovering using cutting edge information technologies makes it an interesting subject for empirical investigations to evaluate novel data modeling approaches and applications (Mikalef et al. 2019). Even so, it is a leading telecom journal such as Expert Systems with Applications that accommodates the highest number of publications focusing on telecom forecasting. From the perspective of the publisher, Elsevier and ieeexplore.ieee.org are currently the two publishers clearly ahead in telecom forecasting journal article publications. Based on the 40 articles analyzed, three main aspects were analyzed:

- (1) the main goal and outcome of each study;
- (2) the dataset (from where the data were extracted and data volume); and
- (3) the techniques adopted.

Since all the articles present empirical data-driven experiments, it is interesting to understand from which years are the data gathered for the experiments to evaluate if the periods are recent enough. It shows that most of the articles perform experiments based on data from the yearly 2011’s, with few articles before 2010. One of the key dimensions of data-driven knowledge discovery is the recency of data, especially considering that telecom customers’ behavior changes over the years. Therefore, using recent data decreases the risk of negatively influencing models built on these data for forecasting telecom business demand.

In additional, artificial intelligence techniques such as SVM (adopted 10 times) and NN (applied for 4 times) appear now as the dominant method, It would be attractive to

observe what future reserves for artificial intelligence applications to telecom customer churn forecasting.

Table 1. Journals from which more than one article was selected

Journal	No. of articles
<i>Expert systems with Applications</i>	10
<i>IEEE Transactions on Industrial</i>	4
<i>European Journal of operational research</i>	3
<i>Decision support systems</i>	2
<i>Neurocomputing</i>	2

Table 2. Publishers from which more than one article was selected

Publisher	No. of articles
<i>Elsevier</i>	26
<i>ieeexplore.ieee.org</i>	6
<i>Springer</i>	3
<i>researchgate.net</i>	2
<i>Citeseer</i>	1
<i>arxiv.org</i>	1

Table 3. Research domain from journals from which articles were selected

Research domain	No. of articles
<i>Telecom</i>	28
<i>Banking</i>	1
<i>Energy</i>	1
<i>Financial Service</i>	1
<i>Online Social Network</i>	1
<i>Newspaper</i>	1
<i>Online Gambling</i>	1

6 Conclusions

Forecasting telecom customer churn is a quite old problem where many researchers have focused on. However, since the telecom industry is under pressure for predicting future demand and if customer will churn or not, so it is one of the most important problems.

The present investigation is designed to provide a very recent literature review on data-based empirical researches for forecasting customer churn. Through providing a summary of the literature covering 40 relevant publications mostly after 2010 up to June 2019, thus a very recent timeframe. The present article offers a review on the most recent trends in this domain, focusing on what the future holds regarding customer churn prediction and trying to find the research gap.

The findings show that decision tree (DT), support vector machines (SVM) and Logistic Regression (LR) are the three most popular and useful method. Besides, artificial intelligence techniques are already demonstrating a significant use in what concerns to predicting customers' behavior. Especially, artificial NN are outstandingly recognized as a competent prediction method. In additional, the literature found is not limited to telecom journals, verifying that telecom themes are also of interest for a larger range of social sciences (e.g. Banking) and that telecom data comprises an important asset for evaluating novel for prediction modeling technologies. Based on the result of this study above described, a customer churn model to predict whether the telecom customer will be lost or retained will be establish. The model will combine data mining technology with the rich data resources of the telecom industry and the latest Marketing theories, which will not only maximize customer acceptance of telecom package within a manageable risk range, but also help increase the company's business volume and revenue. It would also be attractive to study that which trends will emerge on customer churn prediction in the future.

Acknowledgments. This work is partially funded by national funds through FCT - Fundação para a Ciência e Tecnologia, I.P., under the project FCT UIDB/04466/2020

References

- Abbasimehr, H.: A neuro-fuzzy classifier for Customer churn prediction. *Int. J. Comput. Appl.* **19**(8), 35–41 (2011)
- Ahmad, A.K., Aljoumaa, K.: Customer churn prediction in telecom using machine learning in big data platform. *J. Big Data* **6**, 28 (2019)
- Amin, A., et al.: Customer churn prediction in telecommunication industry: with and without counter-example. In: Mexican International Conference Artificial Intelligence, pp. 206–218 (2014)
- Amin, A., et al.: A prudent based approach for customer churn prediction. In: International Conference: Beyond Databases, Architectures and Structures, pp. 320–332 (2015)
- Amin, A., et al.: Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. *IEEE Access* **4**, 7940–7957 (2016)
- Amin, A., et al.: Customer churn prediction in the telecommunication sector using a rough set approach 2017. *Neurocomputing* **237**, 242–254 (2017)
- Au, W.H., et al.: A novel Evolutionary data mining algorithm With applications to churn prediction. *IEEE Trans. Evol. Comput.* **7**(6), 532–545 (2003)
- Awan, S.A., Said, M.: Loyalty Enhancing Communication for Telecom Customer Relationships: A Qualitative Study of Telecom Customers (2011). <https://www.diva-portal.org>. Accessed 21 Feb 2012
- Aydin, S., Özer, G.: The analysis of antecedents of customer loyalty in the Turkish mobile telecommunication market. *Eur. J. Mark.* **39**(7), 910–925 (2005)

- Ballings, M., Poel, D.: Customer event history for churn prediction: how long is long enough? *Expert Syst. Appl.* **39**, 13517–13522 (2012)
- Blattberg, R., et al.: Database Marketing: Analyzing and Managing Customers. Springer, New York (2010)
- Bock, K.W., et al.: An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Syst. Appl.* **38**, 12293–12301 (2011)
- Bock, K.W., Poel, D.: Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Syst. Appl.* **39**, 6816–6826 (2012)
- Brandusouiu, I., Toderean, G.: Churn prediction in the telecommunications sector using support vector machines. *Annals of the Oradea University* (2013)
- Brandusouiu, I., et al.: Methods for churn prediction in the prepaid mobile telecommunications industry. In: International Conference on Communications, vol. 11, pp. 97–100 (2016)
- Burez, J., Poel, D.: CRM at a pay-TV company: using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Syst. Appl.* **32**(2), 277–288 (2007)
- Burez, J., Poel, D.: Handling class imbalance in customer churn prediction. *Expert Syst. Appl.* **36**(3), 4626–4636 (2009)
- Caignya, A., et al.: A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *Eur. J. Oper. Res.* **269**, 760–772 (2018)
- Chen, Z.Y., et al.: A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *Eur. J. Oper. Res.* **223**, 461–472 (2012)
- Chitra, K., Subashini, B.: Customer retention in banking sector using predictive data mining technique. In: ICIT 2011 The 5th International Conference on Information Technology (2011)
- Coussement & Bock: Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning. *J. Bus. Res.* **66**, 1629–1636 (2013)
- Coussement, et al.: A comparative analysis of data preparation algorithms for customer churn prediction: a case study in the telecommunication industry. *Decis. Support Syst.* **95**, 27–36 (2017)
- Crittenden, V.L., et al.: Market-oriented sustainability: a conceptual framework and propositions. *J. Acad. Mark. Sci.* **39**, 71–85 (2011)
- Domingos, P.: A few useful things to know about machine learning. *Commun. ACM* **55**(10), 78–87 (2012)
- Fabrigar, L.R., Wegener, D.T.: Exploratory Factor Analysis. Oxford University Press (2011)
- Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936)
- Gallo, A.: Pediatric deceased donor renal transplantation: an approach to decision making I. Pediatric kidney allocation in the USA: the old and the new. *Pediatr. Transplant.* **19**, 776–784 (2015)
- Ganesh, J., et al.: Understanding the customer base of service providers: an examination of the differences between switchers and stayers. *J. Market.* **64**(3), 65–87 (2000)
- Gnanadesikan, R.: Discriminant Analysis and Clustering. National Academy Press (1988)
- Hadden, J., et al.: Computer assisted customer churn management: state-of-the-art and future trends. *Comput. Oper. Res.* **34**(10), 2902–2917 (2007)
- Hansen, H., et al.: The moderating effects of need for cognition on drivers of customer loyalty. *Eur. J. Mark.* **47**(8), 1157–1176 (2013)
- Hassouna, M., et al.: Customer churn in mobile markets: a comparison of techniques. *Int. Bus. Res.* **8**, 6 (2015)
- He, Y., et al.: A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 1, pp. 92–94 (2009)

- Huang, B.Q., et al.: A new feature set with new window techniques for customer churn prediction in land-line telecommunications. *Expert Syst. Appl.* **37**, 3657–3665 (2010)
- Huang, B., et al.: Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. *Expert Syst. Appl.* **37**, 3638–3646 (2010)
- Huang, B., et al.: Customer churn prediction in telecommunications. *Expert Syst. Appl.* **39**(1), 1414–1425 (2012)
- Huang, Y., et al.: Telco churn prediction with big data. In: ACM SIGMOD International Conference on Management of Data, pp. 607–618 (2015)
- Huang, Y., Kechadi, T.: An effective hybrid learning system for telecommunication churn prediction. *Expert Syst. Appl.* **40**, 5635–5647 (2013)
- Hung, S.Y., et al.: Applying data mining to telecom churn management. *Expert Syst. Appl.* **31**, 515–524 (2006)
- Idris, A., et al.: Genetic programming and adaboosting based churn prediction for telecom. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1328–1332 (2012)
- Idris, A., et al.: Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification. *Appl. Intell.* **39**, 659–672 (2013)
- Idris, A., et al.: Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Clust. Comput.* **22**, 7241–7255 (2017)
- Jadhav, R.J., Pawar, U.T.: Churn prediction in Telecommunication using data mining technology. *Int. J. Adv. Comput. Sci. Appl.* **2**(2), 17–19 (2011)
- Keramati, A., et al.: Improved churn prediction in telecommunication industry using data mining techniques. *Appl. Soft Comput.* **24**, 994–1012 (2014)
- Kim, S.Y., et al.: Customer segmentation and strategy development based on customer lifetime value: a case study. *Expert Syst. Appl.* **31**, 101–107 (2006)
- Kim, K., et al.: Improved churn prediction in telecommunication industry by analyzing a large network. *Expert Syst. Appl.* **41**, 6575–6584 (2014)
- Kirui, C., et al.: Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. *IJCSI Int. J. Comput. Sci. Issues* **10**, 1 (2013)
- Kisioglu, P., et al.: Applying Bayesian belief network approach to customer churn analysis: a case study on the telecom industry of Turkey. *Expert Syst. Appl.* **38**, 7151–7157 (2011)
- Lee, H., et al.: Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. *Decis. Support Syst.* **52**, 207–216 (2011)
- Li, et al.: Study on the segmentation of Chinese insurance market based on CHINA-VALS model. *J. Shanxi Finance Econ. Univ.*
- Lima, E., et al.: Monitoring and backtesting churn models. *Expert Syst. Appl.* **38**, 975–982 (2011)
- Liu, J., et al.: Market segmentation: a multiple criteria approach combining preference analysis and segmentation decision. *Omega* **110**, 324–326 (2018)
- Liu, D., Fan, S.: A modified decision tree algorithm based on genetic algorithm for mobile user classification problem. *Sci. World J.* (1) (2014)
- Long, X., et al.: Churn analysis of online social network users using data mining techniques. In: Proceedings of the International MultiConference of Engineers and Computer Scientists **2012**(1), 14–16 (2012)
- Lu, N., et al.: A customer churn prediction model in telecom industry using boosting. *IEEE Trans. Industr. Inf.* **10**, 2 (2014)
- Ma, J., et al.: Research on a triopoly dynamic game with free market and bundling market in the Chinese telecom industry. *Discret. Dyn. Nat. Soc.* **2018**, 1 (2018)
- Makhtar, M., et al.: Churn classification model for local telecommunication company based on rough set theory. *J. Fundam. Appl. Sci.* **9**(6), 854–868 (2017)
- McIlroy, A., Barnett, S.: Building customer relationships: do discount cards work? *Manag. Serv. Qual.* **10**(6), 347–355 (2000)

- Mikalef, P., et al.: Big data analytics and firm performance: findings from a mixed-method approach. *J. Bus. Res.* **98**, 261–276 (2019)
- Mikel, G., et al.: A review on ensembles for the class imbalance problem: baggingboosting- and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. C* **42**, 463–484 (2012)
- Moeyersoms, J., Martens, D.: Including high-cardinality attributes in predictive models: a case study in churn prediction in the energy sector. *Decis. Support Syst.* **72**, 72–81 (2015)
- Moro, S., et al.: Evaluating a guest satisfaction model through data mining. *Int. J. Contemp. Hospitality Manag.* (2020, in press).
- Nie, G., et al.: Credit card churn forecasting by logistic regression and decision tree. *Expert Syst. Appl.* **38**(11), 15273–15285 (2011)
- Nitzan, I., Libai, B.: Social effects on customer retention. *J. Marketing* **75**, 24–38 (2011)
- Orozco, J., et al.: A framework of IS/business alignment management practices to improve the design of IT Governance architectures. *Int. J. Bus. Manag.* **10**(4), 1–2 (2015)
- Owczarczuk, M.: Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Syst. Appl.* **37**, 4710–4712 (2010)
- Olle, G.D.O., Cai, S.: A hybrid churn prediction model in mobile telecommunication industry. *Int. J. e-Education e-Business e-Management e-Learning* **4**, 1 (2014)
- Petkovski, A.J., et al.: Analysis of churn prediction: a case study on telecommunication services in Macedonia. In: 2016 24th Telecommunications Forum (TELFOR) (2016)
- Qin, R., et al.: A pareto optimal mechanism for demand-side platforms in real time bidding advertising markets. *Inf. Sci.* **78**, 117–121 (2018)
- Qureshi, S.A., et al.: Telecommunication subscribers' churn prediction model using machine learning. In: Eighth International Conference on Digital Information Management (ICDIM 2013) (2013)
- Rahn: Iron deficiency and physical activity after a dietary iron intervention in female Indian tea pickers. *FASEB J.* (2013)
- Reinartz, W.J., Kumar, V.: The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* **67**, 1 (2003)
- Sanchez, B.U., Asimakopoulos, G.: Regulation and competition in the European mobile communications industry: an examination of the implementation of mobile number portability. *Telecommun. Policy* **36**, 187–196 (2012)
- Saradhi, W., Palshikar, G.K.: Employee churn prediction. *Expert Syst. Appl.* **38**, 1999–2006 (2011)
- Shaaban, E., et al.: A proposed churn prediction model. *Int. J. Eng. Res. Appl. (IJERA)* **2**, 693–697 (2012)
- Shaffer, G., Zhang, Z.J.: Competitive one-to-one promotions. *Manag. Sci.* **48**(9), 1143–1160 (2002)
- Sharma, A., Kumar, P.: A neural network based approach for Predicting Customer churn in cellular network services. *Int. J. Comput. Appl.* **27**(11), 26–31 (2011)
- Smith, W.R.: Product differentiation and market segmentation as alternative marketing strategies. *J. Market.* **21**, 3–8 (1956)
- Sun, Y.: The Reasons Why China's OBOR Initiative Goes Digital. Aalborg University and University of International Relations, Denmark (2017)
- Tang, L., et al.: Assessing the impact of derived behavior information on customer attrition in the financial service industry. *Eur. J. Oper. Res.* **296**, 624–633 (2014)
- Torkzadeh, G., et al.: Identifying issues in customer relationship management at Merck-Medco. *Decis. Support Syst.* **42**, 2 (2006)
- Tsai, C.F., Chen, M.Y.: Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Syst. Appl.* **37**, 2006–2015 (2010)
- Vadim, K.: Overview of different approaches to solving problems of Data Mining. *Procedia Comput. Sci.* **123**, 234–239 (2018)

- Vafeiadis, T., et al.: A comparison of machine learning techniques for customer churn prediction. *Simul. Model. Pract. Theory* **55**, 1–9 (2015)
- Verbeke, W., et al.: Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Syst. Appl.* **38**, 2354–2364 (2011)
- Verbeke, W., et al.: New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. *Eur. J. Oper. Res.* **218**, 211–229 (2012)
- Wang, Y., et al.: Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Chang.* **126**, 3–13 (2018)
- Wang, L., et al.: Research on financial advertisement personalized recommendation method based on customer segmentation. *Int. J. Wireless Mobile Comput.* **14**, 97–101 (2018)
- Winer, R.S.: A framework for customer relationship management. *Calif. Manag. Rev.* **43**, 89–105 (2001)
- Xu, H., et al.: Churn prediction in telecom using a hybrid two-phase feature selection method. In: Third International Symposium on Intelligent Information Technology Application. IITA 2009, pp. 576–579 (2009)
- Zhang, X., et al.: Predicting customer churn through interpersonal influence. *Knowl.-Based Syst.* **28**, 97–104 (2012)



Roll Padding and WaveNet for Multivariate Time Series in Human Activity Recognition

Rui Gonçalves¹(✉), Fernando Lobo Pereira¹, Vitor Miguel Ribeiro²,
and Ana Paula Rocha¹

¹ FEUP, Department of Electrical and Computer Engineering, Porto University, Porto, Portugal
{rj jpg, flp, arocha}@fe.up.pt

² FEP, Faculty of Economics of Porto, Porto University, Porto, Portugal
vssribeiro@fep.up.pt

Abstract. Padding is a process used for the border treatment of data before the convolution operation in Convolutional Neural Networks. This study proposes a new type of padding designated by roll padding, which is conceived for multivariate time series analysis when using convolutional layers. The Human Activity Recognition raw time distributed dataset is used to train, test and compare four Deep Learning architectures: Long Short-Term Memory, Convolutional Neural Networks with and without roll padding, and WaveNet with roll padding. Two main findings are obtained: on the one hand, the inclusion of roll padding improves the accuracy of the basic standard Convolutional Neural Network and, on the other hand, WaveNet extended with roll padding provides the best performance result.

Keywords: HAR · WaveNet · Padding · Time series · Classification · Deep learning

1 Introduction

This article concerns the implementation of a roll padding technique in Convolutional Neural Networks (CNN) and variant models such as the WaveNet targeting the analysis of Multivariate Time Series (MTS). A time series, by definition, is a continuous sequence of observations taken repeatedly over time normally with equidistant intervals. The relationship between past and future observations can be stochastic, which implies that the conditional probability distribution of a vector of inputs $x \in \mathbb{R}^n$ as a function of past observations is given by:

$$P(x_{t+d}|x_t, x_{t-1}, \dots) = f(x_t, x_{t-1}, \dots)$$

In some MTS studies, train and test datasets are composed by observations (i.e., examples) of independent time series segments with available time distributed information such that changes in the context across examples can be easily observed. Although the Long Short-Term Memory (LSTM) [1], which is a particular type of Recurrent Neural Network (RNN) model, is more suitable for segmented MTS problems from a

theoretical point of view, CNN have gained popularity to analyze these specific problems, which suggests that using Deep Learning (DL) models with memory cells to track information between examples may be neglected in time series that exhibit either frequent or periodic changes of context.

A remarkable example that reflects this trend is the WaveNet model from Google DeepMind [2], which was initially applied to audio signal generation. One important component to accomplish the predictive task is the WaveNet sound classifier, which is based on uni-dimensional convolutional layers. By considering it as the base architecture for the case study under analysis, the present research extends the WaveNet component to become fully functional with bi-dimensional inputs (i.e., time steps \times variables).

Padding is a relevant method used in convolution operations because it ensures that the border treatment of the input space is preprocessed by a convolution kernel to retain as much original and meaningful information as possible at the output level. This study presents a new padding method to be incorporated in CNN and variant models such as the WaveNet to understand whether their performance can be improved in the context of MTS problems.

The key idea behind the inclusion of roll padding is to ensure that the variables dimension of the input feature map is properly interconnected by the bi-dimensional convolution kernel. This innovation constitutes an alternative to standard padding methods that fill the bi-dimensional convolution kernel with zeros at the boundaries level, which increases the likelihood of loosing valuable information.

The main results are summarized as follows. First, the incorporation of roll padding has a positive impact on the accuracy of a CNN model used for the MTS classification task. Second, the WaveNet extended with roll padding provides the best performance result. The rest of the study is organized as follows. Section 2 presents the case study. Section 3 describes convolution and padding operations in detail. Section 4 presents DL architectures and the inclusion of roll padding. Results and conclusions are clarified in Sect. 5.

2 Case Study

The Human Activity Recognition (HAR) dataset from the University of California Irvine (UCI) Machine Learning repository is used to train, test and compare our methodology. This is a well-known and competitive dataset focused on smartphone data, which contains 3-axial gravitational acceleration, 3-axial body acceleration and 3-axial body gyroscope readings captured at a constant rate 50 Hz, totalizing 9 variables over 128 time steps. Readings were taken from 30 volunteers holding a smartphone to record six different types of activities: walking, walking upstairs, walking downstairs, sitting, standing, and laying. Overall, the UCI HAR dataset consists of 10 299 examples.

It is important to highlight that the UCI repository provides the separation of data into train and test datasets. The training set contains 7352 examples, while the testing set has 2947 examples. This working setup is referred to as 21-9, which means that 21 subjects are used for training and 9 subjects are used for testing. Models of this type of working setup are said to fall into the category of impersonal models [3]. As clarified

in Table 1, another relevant point is the number of examples *per* class. Any unbiased comparison of test results between different studies requires a persistent consistency on the adoption of these values.

Table 1. Number of examples per class in the train and test HAR dataset provided by UCI.

	Walking	Upstairs	Downstairs	Sitting	Standing	Laying
Test set	496	471	420	491	532	537
Train set	1226	1073	986	1286	1374	1407

In addition to the Raw Time Distributed (RTD) dataset, UCI also provides a Feature Engineered (FE) dataset that transforms the RTD data into 561 non-temporal features (e.g., average, max, min, etc.). Since the FE dataset maintains the order and number of examples equal to the RTD dataset, the performance of models that use the FE dataset can be compared to the performance of models that use the original RTD dataset.

Table 2. Most relevant studies using the UCI HAR dataset with 21-9 working setup.

Study	Method	UCI HAR Dataset Type	ACC (%)
[4]	OVO SVM Ensembling Voting	FE	96.40
[5]	Multiclass SVM	FE	96.37
[6]	Kernel variant of LVQ	FE	96.23
[7]	tFFT + CNN	RTD	95.75
[8]	DFT + CNN	RTD	95.18

Table 2 summarizes test results of some prominent studies. [4] proposes a One-vs-One (OVO) multi-classification Support Vector Machine (SVM) with a linear kernel for the classification task. The method uses majority voting to find the most likely activity for each test sample from an arrangement of 15 binary classifiers. [5] introduces the HAR dataset and obtain results exploiting a multi-classification SVM. [6] employs a sparse kernelized matrix Learning Vector Quantization (LVQ) model. Their method is a variant of LVQ in which a metric adaptation with only one prototype vector for each class is defined. [7] presents a tFFT + CNN model that uses the temporal fast Fourier transform concept from [9] to process information that subsequently feeds a CNN. Similarly, [8] applies a bi-dimensional Discrete Fourier Transform (DFT) to the MTS raw inputs followed by the use of a CNN.

The best result in the Kaggle competition was 98.01% in the private dataset (i.e., internal dataset used for the final ranking of competitors) and 97.18% in the public test dataset (i.e., the one used by competitors for testing and development). However, results from these highly problem-dependent architectures are not comparable with results from studies presented in Table 2 since the train and test partition of public and private datasets is not equal to the original 21-9 working setup.

Similar concern is applied to [10] reporting an accuracy of 97.63% because it is not clear the type of partition considered by the author. After running the available code in his GitHub repository, we not only observe that the test dataset contains an excessive number of examples, namely 2993, but also the number of examples *per class* is different from the canonical UCI partition. Other works with variants of this dataset include [11–13]. As such, based on Table 2, we consider the current state-of-the-art accuracy, for studies that maintain the original UCI 21-9 working setup unchanged, stands at 96.40%.

3 Convolution and Padding

3.1 Convolution

CNN architectures are accurately examined in [14]. Let us consider a bi-dimensional input feature map x^l in the layer l of size (H, W) , where W is the width and H the height of the feature map, and a stride S of $(1, 1)$. The simple mathematical formalization for the computation of a convolutional layer l to obtain the output feature map y^l with kernel K of size (k_H, k_W) is expressed as:

$$y^l = \phi\left(\sum_{i=0}^{H-k_H} \sum_{j=0}^{W-k_W} K \cdot x^l_{i,j}\right) \quad (1)$$

where ϕ is the activation function. This equation represents the application of kernel K in the input map x at coordinates i, j . A bias term b is normally added to $y^l_{i,j}$, which is omitted for the sake of a clearer presentation. An example of this computation without applying the activation function considering:

$$x^l = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, K = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

and $S = (1, 1)$ is given by:

$$\begin{bmatrix} 1 \times 1 & 1 \times 0 & 1 \times 1 & 0 \\ 0 \times 0 & 1 \times 1 & 1 \times 0 & 1 \\ 0 \times 1 & 0 \times 0 & 1 \times 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & \dots \\ \dots & \dots \end{bmatrix}, \begin{bmatrix} 1 & 1 \times 1 & 1 \times 0 & 0 \times 1 \\ 0 & 1 \times 0 & 1 \times 1 & 1 \times 0 \\ 0 & 0 \times 1 & 1 \times 0 & 1 \times 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 3 \\ \dots & \dots \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 \times 1 & 1 \times 0 & 1 \times 1 & 1 \\ 0 \times 0 & 0 \times 1 & 1 \times 0 & 1 \\ 0 \times 1 & 0 \times 0 & 1 \times 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 3 \\ 2 & \dots \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 \times 1 & 1 \times 0 & 1 \times 1 \\ 0 & 0 \times 0 & 1 \times 1 & 1 \times 0 \\ 0 & 0 \times 1 & 1 \times 0 & 1 \times 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 3 \\ 2 & 4 \end{bmatrix}$$

One observes that the output is smaller than the input when the convolution kernel is larger than $(1, 1)$. If the input has size (H, W) and the kernel has size (k_H, k_W) , then the convolution outcome has size $(H - k_H + 1, W - k_W + 1)$, which is smaller than the original input. Usually, this is not a concern for inputs with large dimension (i.e.,

images) and small filter size. However, it can constitute a problem for inputs with small dimension or for a high number of stacked convolutional layers. As such, the practical effect of large filter sizes and very deep CNN on the size of the resulting feature map entails the potential loss of information such that the model can simply run out of the data upon which it operates. The padding operation is conceived to tackle this concern.

3.2 Traditional Padding Methods

The standard procedure to avoid the border effect problem consists in applying same padding (i.e., inclusion of zeros outside the input map). For every channel of the bi-dimensional input x , we insert zeros $\frac{k_H-1}{2}$ rows above the first row and $\frac{k_H}{2}$ rows below the last row, and $\frac{k_W-1}{2}$ columns to the left of the first column and $\frac{k_W}{2}$ columns to the right of the last column. The convolution output size will be (H, W) , thus, having the same extent of the input.

Note, however, that if the goal of research is to analyze a MTS problem, the bi-dimensional input feature map has a relatively small size in the variables dimension compared to image processing problems. Therefore, the inclusion of zeros through the same padding method implies a weaker learning capability since the learned kernel is affected by the dot product operation with the included zeros, which has the potential to promote erroneous generalizations. Other padding methods are commonly used in image processing [15]. As exemplified in Table 3, these make use of the information in the input x to fill in the borders.

Table 3. Padding examples of size 4 for uni-dimensional input.

Method	Example of padded info		
	Pad	Input	Pad
Valid (None)		a b c d e f	
Same (Zero)	0 0 0 0	a b c d e f	0 0 0 0
Reflect (Mirror)	d c b a a b c d e f	f e d c	
Reflect101	e d c b a b c d e f	e d c b	
Constant n	n n n n	a b c d e f	n n n n
Tile 2	a b a b a b c d e f	e f e f	
Causal (Zero Left)	0 0 0 0	a b c d e f	
Wrap	c d e f a b c d e f	a b c d	

3.3 Roll Padding

Roll padding is an extension of wrap padding that is conceived for MTS problems. Wrap copies information from opposite sides of the image, effectively mapping the image into a torus. This operation corresponds to four copies of information under a bi-dimensional input: wrapped rows (columns) above the top (on the left) of the input are a

copy of bottom rows (right columns), respectively and reciprocally. Although wrapping is not typically useful for natural images, it is appropriate for computed images such as Fourier transforms and polar coordinate transforms where pixels in opposite borders are computationally adjacent.

As observed in Fig. 1, similar to wrap padding, roll padding copies information from opposite sides in the MTS input 2D feature map that consists of time steps \times variables, but only in one dimension: the variables dimension. In turn, the time steps dimension remains without padding (i.e., valid padding), from which one obtains a cylinder instead of a torus.

The reduction of the time steps dimension after several convolutions is not a source of concern due to the presence of a high number of time steps in MTS problems. Nevertheless, as clarified in Fig. 2 and Subsect. 4.1, roll padding can be combined with other padding methods applied to the time steps dimension (e.g., causal).

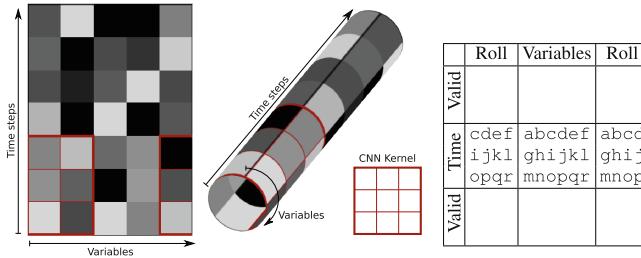


Fig. 1. Roll padding scheme in MTS analysis.

4 Tested Architectures

Four DL architectures are used based on:

1. Staked 3 LSTM layers;
2. CNN, LeNet-5 based, using same padding;
3. CNN, LeNet-5 based, using roll padding;
4. WaveNet with convolutional 2D layers using causal and roll padding.

The first architecture serves as the baseline. The second and third ones highlight the potential of roll padding in CNN. The last proposal emphasizes how roll padding is introduced in the WaveNet to obtain bi-dimensional input data in the context of MTS problems.

The LSTM [1] network considered in this study is constituted by four layers, three of which use bidirectional LSTM plus a dense layer with softmax activation for classification. The skeleton of both CNN is derived from the basic LeNet [16]. They are constituted by two bi-dimensional convolutional layers, two dense layers with Rectified Linear Unit (ReLU) and output dense layer with softmax. Both CNN use the same

hyperparameters, so that the only difference is the type of padding method adopted. This allows to infer implications of roll padding at the performance level. For the CNN where roll padding is introduced, valid padding is used for the time steps dimension, while roll padding is considered for the variables dimension. At this point, it is important to highlight that the error function considered in this study is the categorical cross-entropy and accuracy is the metric used to compare performances.

4.1 WaveNet with Roll Padding

A key point in the standard WaveNet is the use of causal padding before convolutions to ensure that the order of data is not violated. The prediction $P(x_{t+1}|x_t, x_{t-1}, \dots)$ generated by the model at time step $t + 1$ cannot depend on future time steps. For uni-dimensional data such as audio signal, the causal convolution is implemented by shifting the input a few time steps behind. Since models with causal convolutions are absent of recurrent connections, they are typically faster to train than RNN. A problem of causal convolutions is that they require either many layers or large filters to increase the receptive field [2].

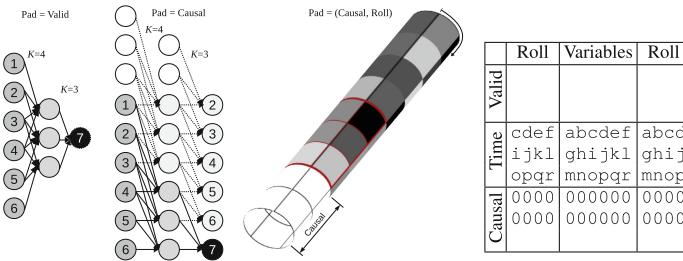


Fig. 2. Combination of causal and roll padding scheme in MTS analysis.

In the first two left subplots of Fig. 2, one can observe the output difference between resorting or not to the use of causal padding in convolutions. Although being transformed, causal padding allows to preserve the time steps of past information for the next layer. Formally, this requires that the number of zeros added before the begin of the sequence is given by $k_H - 1$.

Another important characteristic of the WaveNet model is the use of dilated convolutions to gradually increase the receptive field. A dilated convolution is a convolution where the filter K is applied over an area larger than its length k_H by skipping input values with a certain step. Hence, it is equivalent to a convolution with a larger filter derived from the original filter by dilating it with zeros. As a result, when using a dilation rate dr (i.e., for $dr > 1$), the causal padding has size given by $dr \times (k_H - 1)$.

Last but not least, the residual block [17] is the heart of a WaveNet. It is constituted by two convolutional layers, one using *sigmoid* activation and other using *tanh* activation, which are multiplied. Then, inside the block, the result is passed through into another convolution with $k_H = 1$ and $dr = 1$ in the time steps dimension. Normally, this allows to downsample input channels and control the number of feature

maps, thereby justifying why this technique is often referred to as a projection operation or channel pooling layer. Both residual and parameterized skip connections are used throughout the network to speed up convergence and enable the training of deeper models. This block is executed a given number of times in the depth of the network, with $N = \{1, \dots, \text{depth}\}$. The dilatation dr increases exponentially according to the formula $dr = (k_H)^N$.

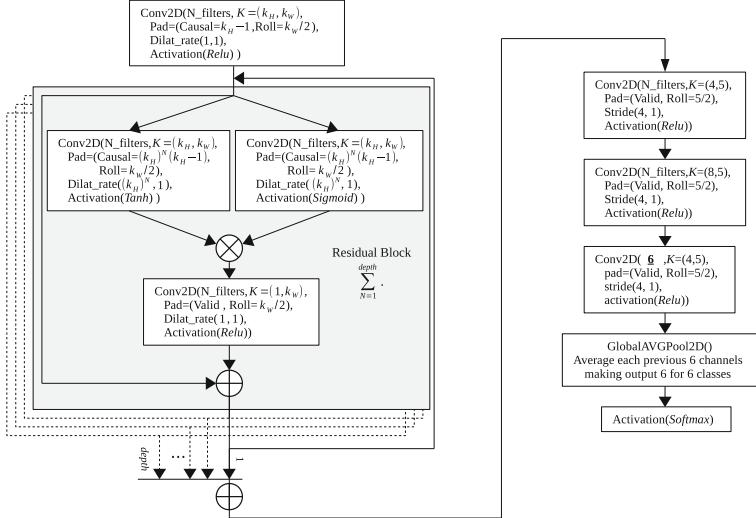


Fig. 3. WaveNet architecture for MTS classification, using 2D convolutions with causal padding in the time steps dimension and roll padding in the variables dimension.

Figure 3 shows the extended WaveNet working with a bi-dimensional input map. The architecture maintains the standard WaveNet processing in the time steps dimension by using causal padding, while roll padding is introduced in the variables dimension. The combination of both padding methods is clarified in the last two right subplots of Fig. 2. The use of roll padding is illustrated in the variables dimension, whose size depends on k_W . We establish a roll padding of size $\frac{k_W}{2}$ by copying opposite $\frac{k_W}{2}$ columns information of the input map. For the sake of simplicity, one assumes odd fixed sizes in k_W . No dilation rate is considered in the variables dimension (i.e., $dr = 1$).

Finally, after adding skip connections, three bi-dimensional convolutional layers are considered. In this scheme, a stride $S = (S_H, S_W)$ with $S_H > 1$ and $S_W = 1$ is used to downsample only the time steps dimension rather than considering pooling layers. The last convolution has 6 filters to generate 6 feature maps in which a global average pooling is applied. In this way, one can directly use softmax in the 6 resulting values for classification, thereby avoiding the use of dense layers. In summary, the methodology follows the standard WaveNet philosophy in the time steps dimension, while everything is processed as normal convolutional layers with roll padding in the variables dimension.

5 Results and Conclusions

Figure 4 presents the evolution of accuracy in the train and test datasets as well as the epoch threshold where the highest accuracy is achieved, which corresponds to the point where the best model is found.

One observes that the CNN with same padding has a similar evolution to the LSTM. On average, the CNN with roll padding has a slightly superior accuracy relative to the previous models both in the train and test datasets. Therefore, the first relevant conclusion is the improvement of accuracy with the introduction of roll padding in the CNN. Moreover, the bi-dimensional WaveNet model with roll padding surpasses all

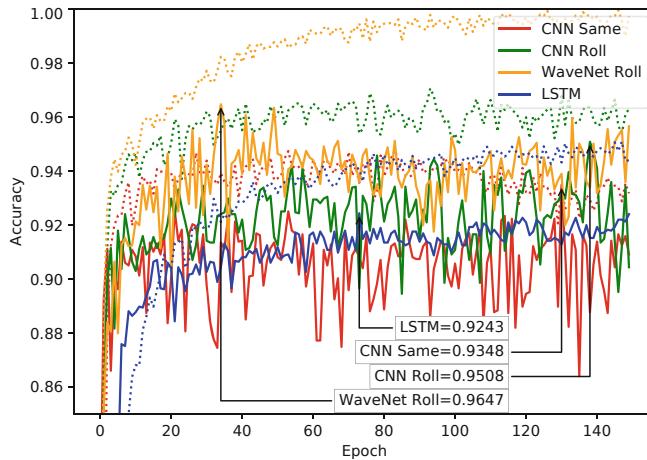


Fig. 4. Accuracy evolution during learning process. Doted lines refer to the train dataset, while solid lines refer to the test dataset.

Table 4. Confusion matrix for the bi-dimensional WaveNet model with roll padding on the HAR RTD test dataset as originally provided by UCI.

		Predicted						Recall(%) ACC (%)	
		Walking	Upstairs	Downstairs	Sitting	Standing	Laying		
Real	Classes								
	Walking	474	15	4	1	1	0	95.77	96.47
	Upstairs	2	462	0	5	2	0	98.09	
	Downstairs	2	3	415	0	0	0	98.81	
	Sitting	0	0	0	458	33	0	93.28	
	Standing	1	0	0	33	498	0	93.61	
	Laying	0	2	0	0	0	537	99.63	
Precision (%)		98.96	95.85	99.05	92.15	93.26	100		

the alternative options in terms of accuracy. As such, the second relevant conclusion is that the incorporation of roll padding in the WaveNet allows to achieve the best performance result. The accuracy reaches near 100% in some epochs in the training dataset, which seems to suggest there is some room for some improvement by choosing the right amount of regularization. Table 4 summarizes results by class of the best DL model.

This study trains, tests and compares the performance of four DL architectures using the UCI HAR dataset. UCI provides two types of datasets for this case study. Instead of using the FE dataset, we consider the RTD dataset since our architectures are particularly conceived for this type of data. All models were tested without any type of data preprocessing or feature engineering methodology. The WaveNet extended with roll padding provides an accuracy of 96.47% for this case study, which is greater than the value reported in [4]. One can conclude that the present analysis constitutes another case where DL methodologies outperform alternative Machine Learning (ML) and classical techniques. Interestingly, little or nothing has been made by the specialized literature to fall into a problem-dependent architecture design, which seems to suggest that we developed a new architecture with potential for generalization in other types of MTS problems.

References

1. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
2. van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: Wavenet: a generative model for raw audio (2016)
3. Lockhart, J.W., Weiss, G.M.: Limitations with activity recognition methodology & data sets. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 747–756. ACM (2014)
4. Romera-Paredes, B., Aung, M.S.H., Bianchi-Berthouze, N.: A one-vs-one classifier ensemble with majority voting for activity recognition. In: ESANN (2013)
5. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: ESANN (2013)
6. Kaden, M., Strickert, M., Villmann, T.: A sparse kernelized matrix learning vector quantization model for human activity recognition. In: ESANN (2013)
7. Ronao, C.A., Cho, S.-B.: Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **59**, 235–244 (2016)
8. Jiang, W., Yin, Z.: Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM International Conference on Multimedia, MM 2015, pp. 1307–1310. ACM, New York (2015)
9. Sharma, A., Lee, Y., Chung, W.: High accuracy human activity monitoring using neural network. In: 2008 Third International Conference on Convergence and Hybrid Information Technology, vol. 1, pp. 430–435 (2008)
10. Ignatov, A.: Real-time human activity recognition from accelerometer data using convolutional neural networks. *Appl. Soft Comput.* **62**, 915–922 (2018)
11. Qi, L., Xu, X., Wan, S., et al.: Deep learning models for real-time human activity recognition with smartphones. *Mobile Netw. Appl.* **25**, 743–755 (2020)
12. Qin, Z., Zhang, Y., Meng, S., Qin, Z., Choo, K.K.R.: Imaging and fusing time series for wearable sensor-based human activity recognition. *Inf. Fus.* **53**, 80–87 (2020)

13. Slim, S., Atia, A., Elfattah, M., Mostafa, M.S.M.: Survey on human activity recognition based on acceleration data. *Int. J. Adv. Comput. Sci. Appl.* **10**(3), 84–98 (2019)
14. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pp. 253–256 (2010)
15. Hamey, L.G.C.: A functional approach to border handling in image processing. In: *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8 (2015)
16. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. *CoRR*, [abs/1512.03385](https://arxiv.org/abs/1512.03385) (2015)



Automatic Classifier of Scientific Contents

Samuel Machado and Jorge Oliveira e Sá

Algoritmi Center, University of Minho, Braga, Portugal

jos@dsi.uminho.pt

Abstract. The growth of scientific production, associated with the increase in the complexity of scientific contents, makes the classification of these contents highly subjective and subject to misinterpretation. The taxonomy on which this classification process is based does not follow the scientific areas' changes. These classification processes are manually carried out and are therefore subject to misclassification. A classification process that allows automation and implements intelligent algorithms based on Machine Learning algorithms presents a possible solution to subjectivity in classification. Although it does not solve the inadequacy of taxonomy, this work shows this possibility by developing a solution to this problem. In conclusion, this work proposes a solution to classify scientific content based on the title, abstract, and keywords through Natural Language Processing techniques and Machine Learning algorithms to organize scientific content in scientific domains.

Keywords: Taxonomy · Machine Learning · Natural Language Processing

1 Introduction

Humans learn to classify things at a very young age. Categorizing fills a need of human nature, that is, to impose order and find hidden relationships. However, we are not very good at classifying because we organize empirically, based on intuition or experience. It is simple to classify a set of ten black and white balls into two classes: black and white. But as we increase the number of characteristics, so does the complexity of the task. Classification allows us to understand diversity better.

A Text Classifier is an abstract model, which describes a set of predefined classes generated from a collection of labeled data or training set. The classifier is used to correctly classify new texts for which the class label is unknown [1].

Real-world raw data is usually unsuitable for direct use in classifier training, so some cleaning and preprocessing steps are generally applied before the classification task. Thus, scientific contents must go through a Natural Language Processing (NLP) techniques for the data to be ready for classification [2].

Classification in science adds several challenges, some of which can result in biased models when we try to understand feature like:

- The actual content of the document. It is sometimes classified into an existing class even when it does not fit in an emerging research field.

- The person that decides the classification can be either the author, the designated person who submits the publication, or a committee of peers.

With the increase in publications, the human factor, especially under the pressure of numbers or information overload, is most likely to make mistakes and fail to identify correctly and consistently. Humans often prone to errors during analysis or when trying to establish relationships between multiple features. Machine Learning algorithms can be applied to solve or mitigate these problems while improving efficiency.

2 Concepts and Subjects

To provide a suitable solution to the problem under study, we needed to address some concepts and subjects. Regardless of the classification system, the variety of Machine Learning (ML) classification techniques is wide and constitutes this core. Thus, in this section, we will address two: classification systems and how to develop automatic classifiers based on ML algorithms.

2.1 Classification Systems

As scholarly research becomes increasingly interdisciplinary, an essential purpose for a classification system is to facilitate multidisciplinary research and information sharing [3]. Comte [4] proposed a schema of science classification. He argues that the division of intellectual labor is necessary and that the scientific domains would have to be cultivated separately. He also stressed that the sciences all belonged to a larger whole and that any division is artificial.

A classification system should contain, amongst other features [5]:

- Breadth - defined as either a typology or a taxonomy based on classes where the subjects would be classified or grouped;
- Meaning - supporting the rational use of the selected classification method and classes should be a philosophical foundation;
- Depth – as close as possible to support the diversity of real-life phenomena;
- Recognizability – must mirror the real world.

To better understand a classification system, we need to understand the concepts of taxonomy, ontology, and thesaurus [6] finally, how it can be applied to the classification of science results (for example, articles). For example, taxonomy allows to define groups of biological organisms based on shared characteristics and to name these groups. Thus, it groups the organisms in a taxonomic classification; groups of a given class can be aggregated to form a higher-level supergroup, thus creating a taxonomic hierarchy [7]. A taxonomy typically has some hierarchical relations incorporated in its class classifications. Thesaurus can be understood as a taxonomy extension: it takes taxonomy as described above, allowing subjects to be arranged in a hierarchy. Besides, it adds the ability to enable other statements to be made about the topics. Both the taxonomy and the thesaurus can fall into the Knowledge Organization Schemes (KOS) class because they

provide the set of structured elements to be used for describing and indexing objects, browsing collections, etc. Ontology, originally from the philosophical domain, has been given a new definition with the development of Artificial Intelligence as a formal, explicit specification of a shared conceptualization [8]. They represent the set of objects, properties, and relationships we can use in a specified domain of knowledge. By defining the terms and their relationships, ontology encodes a knowledge domain so that a machine can understand it. The W3C standard for defining ontologies is OWL, a key component of semantic web technologies [9]. Ontologies are also often interpreted as the classification mechanism itself. A controlled vocabulary is a closed collection of terms that have been explicitly grouped and can be used for classification. It is controlled because the list is limited, and there is control over who can add terms to the list, when, and how (Fig. 1).

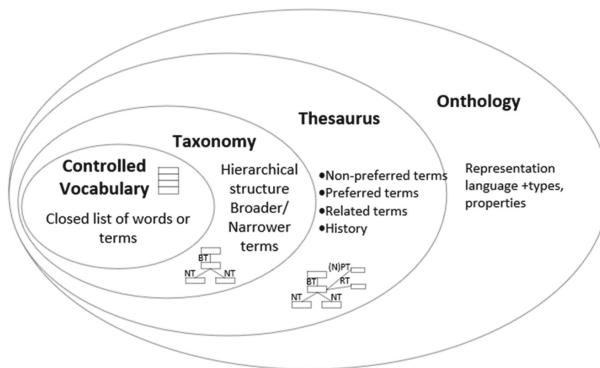


Fig. 1. Classification categories, adapted from [10]

2.2 Machine Learning

Artificial Intelligence (AI) can be used in Texts and Knowledge Discovery Databases using NLP techniques. For example, this serves to annotate automatically, and index texts through text corpora classification, which requires external data support in the form of ontologies, thesaurus, etc. [11]. However, there are restrictions on applying new patterns not yet discovered, often in innovative scientific publications [12].

ML aims to provide automated extraction of insights from data. Standard learning systems (like neural networks or decision trees) operate on input data after they have been transformed into feature vectors. The data vectors or points can be separated by a surface, clustered, interpolated, or otherwise analyzed. The resulting hypothesis will then be applied to test points in the same vector space to make predictions or classifications [13]. This approach loses all the word order information, only retaining the terms' frequency in the document by removing non-informative words (stop words) and replacing words with their stems or stemming [14]. NLP, in its many aspects, is illustrated in Fig. 2. On the left side are represented the requirements to develop an NLP system. The first big

challenge is to get enough data as a word dictionary to provide the system with enough linguistic and semantic knowledge of each possible class in taxonomy to use.

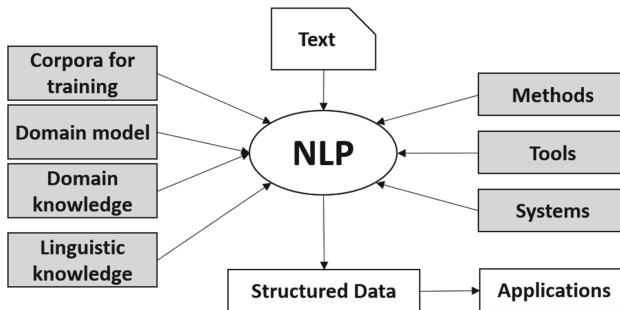


Fig. 2. Aspects of NLP, adapted from [15].

The right side of Fig. 2 represents NLP's operationalization with the methods, systems, and tools. The output with structured data can then feed an ML (or other) system. We can split natural language understanding at a word level, and concept level approaches as Syntax-centered NLP and Semantics-based NLP, respectively. NLP has excellent potential to be used as a preprocessing step on a classification ML or classifier itself. Recent investigations show that NLP's use as a pre-processor for neural networks or in a more advanced fashion Convolutional Neural Networks, with multiple levels and stages of perceptrons[16], and supported by a Thesaurus and a useful Ontology can achieve good classification results. There would still be some limitations for the discovery of new classes, though. This preparation of the texts is relevant to apply a Taxonomy capable of dealing with science's complexity, i.e., scientific documents present interdisciplinarity of scientific domains. [17]. Thus, the classification of scientific documents includes an additional complexity factor in applying a scientific taxonomy [15].

3 Application Scenario

The ALGORITMI Research Center is a research unit of the University of Minho, Portugal, that develops R&D activity in Information and Communications Technology and Electronics (ICT&E) and it is divided into four research fields [18]:

1. Electrical Engineering, Electronics, and Nanotechnology.
2. Operations Research, Statistics, and Numerical Methods.
3. Information Systems, Software, and Multimedia.
4. Communications, Computer Networks, and Pervasive Computing.

ALGORITMI includes 9 R&D groups, divided into 14 R&D domains, the number of integrated researchers at Algorithm Center is 102, but the total number of researchers (integrated and collaborators) are approx. 500.

We can start to ask if the taxonomy in place can deal with these multidisciplinary publications. ALGORITMI internally uses a taxonomy equivalent to that adopted by governmental institutions of science, as one would expect. Which in turn follows a taxonomy recommended by the OECD, called the Frascati Manual. This taxonomy suffers from reduced depth levels, tending to generalize more and, therefore, to be somewhat limiting or reductive, causing an increase in overlap or high aggregation of domains or subjects.

The scientific publications, produced in ALGORITMI, cover the four research domains and the 14 existing R&D domains. The increase in the number of coauthors per publication may or may not belong to different R&D domains, leading to a rise in publications belonging to various scientific research communities, causing an increase in publications covering several R&D domains. For example, the scientific article produced in ALGORITMI “Calado, A., Leite, P., Soares, F., Novais, P., & Arezes, P. (2018). Design of a Framework to Promote Physical Activity for the Elderly. In International Conference on Human Systems Engineering and Design: Future Trends and Applications (pp. 589–594). Springer, Cham.” apparently belongs to the scientific domain linked to Health, but the article reports the development of a UI that allows to show and compute real-time results of the Boccia game. From here, two points are clear:

- Cross-domain research, which shares disciplinary knowledge by investigating a phenomenon, presents additional complexity to the classification system;
- A classification method requires an increasing effort to maintain consistency to cope with existing complexity, making existing classification systems unable to allow correct classifications.

The complexity and dynamics of science make existing taxonomies, which, as a rule, are static, into inaccurate classification results. To make taxonomies more dynamic, i.e., the ability to arise new disciplines through an iterative interdisciplinarity cycle [19].

The problems identified in the Manual Frascati taxonomy were also verified in other studied taxonomies, e.g., Scopus, Microsoft Academic, CORDIS, among others. Classification inconsistencies, different approaches, and scalability are some of the additional problems identified. Therefore, from the results obtained in the taxonomies analysis process, it was possible to accomplish a Frascati Manual taxonomy adaptation with major identified problems fixed. This adapted taxonomy consists of 15 scientific knowledge domains and 447 scientific knowledge subdomains and was implemented in the developed classification system.

4 An Automatic Classifier

The hardware used for this study, namely to training the classification algorithms, was a CPU Intel(R) Core(TM) i7-7700HQ 2.80 GHz, with 16 GB of RAM and a 250 GB SSD, and the dataset used in this study contains scientific publications produced by ALGORITMI researchers. In total, there are 2,665 scientific documents created between the years 2008 and 2017. Of these 2,665 documents, 2,389 are coauthored. All these documents were classified manually by a librarian by using the Frascati Manual taxonomy. Therefore, the documents were manually reclassified according to the adapted taxonomy.

Table 1 presents the structure of the dataset. It contains ten fields. The goal is to classify the fields “knowledge domain” and “knowledge subdomain”, and the training set includes the previous manual reclassification.

Table 1. ALGORITMI dataset fields.

#	Field	Sample
1	Author	S. Azevedo
2	Publication	Systematic Use of Software Development Patterns through a Multilevel and Multistage Classification
3	Type of publication	Book Chapter
4	Knowledge Domain	Computer and Information Science
5	Knowledge Subdomain	Computer Sciences
6	Date of publication	2011
7	Weblink	https://www.scopus.com/record/display.uri?eid=...
8	Coauthors	A. Bragança; R. J. Machado, H. Ribeiro
9	Abstract language	English

Python programming language is becoming very popular in ML applications. The justification is because Python includes several ML libraries, and there are packages ready to use, for instance, Anaconda. It turns out that we can find some top-rated scientific computing tools, including Deep Learning (DL) virtual environments. Anaconda provides integrated end-to-end tools to manage libraries, dependencies, and environments to develop and train ML and DL models and analyze data, including data visualization tools. Through Anaconda, the Jupyter Notebook served as a virtual Python environment, and Python 3 kernel (version 3.7.4) was used for this task. Because it provides easy-to-use APIs for a wide variety of text preprocessing methods, Python’s Natural Language ToolKit (NLTK) was installed, providing predefined NLP tasks. It is one of the most used libraries for NLP and computational linguistics. It consists of a suite of program modules, data sets, and tutorials supporting research and teaching in computational linguistics and NLP. NLTK contains several corpora and includes a small selection of texts from Project Gutenberg, which provides 25,000 free electronic books. The toolkit Stopwords Corpus package enables the removal of redundant repeated words. To do data analysis, the platform used is Pandas. Pandas provide high-performance, easy-to-use data structures and data analysis in Python programming language, allowing fast analysis and data cleaning and preparation. Pandas’ alternative would be Numpy or Scipy, but Pandas works well with labeled data, hence the root of Pandas name: Panel Data. Numpy could be more helpful for the numerical data type (Num).

We need to disambiguate the meanings of the sentences by eliminating the punctuation. It introduces noise and adds little value to the analysis capacity based on a text’s word vectors, which in this study case. The punctuation is removed by running a function through each character in the sentence and removes it. Removing punctuation from a

text makes it unstructured. The tokenization process separates this text into units, such as phrases or words, by giving structure to a previously unstructured text. For example, the sentence “Modeling Software Product” is divided into tokens [modeling, software, product]. This task is useful to prepare the text to be handled by a lexical analyzer, which is the next step. After the text’s tokenization, we can feed a lexical analyzer to remove “stop words”. These are generally the most common words used in a given language and do not add any value to the data. The NLTK contains a list of irrelevant words in English, so it is necessary to process the text using a lexical analysis function that compares each word with the items in this list and removes them. The remaining text was properly tagged using Part-of-Speech tagging and since it still contains several derived words two approaches can be followed: Stemming, to eliminate words inflected (or sometimes derived) to the word stem, base, or root form. This is useful for simplifying words in the text without losing their meaning (except in a semantic analysis, which is not the case); Lemmatization reduces the words “modelling”, “modelled”, and “modeller” to the root word, “model”. We find that Stemming’s approach cuts the end of words. In this way, the words are meaningless as “sourc” or “emiss”. Although the process is fast, it is not very useful and can reduce the model’s accuracy. On the other hand, the Lemmatization approach is based on a dictionary to make a morphological analysis of the word to determine its root form.

After text processing, the next step is to test a collection of classifiers to assess the speed and accuracy of each algorithm used. For all the algorithms used, the resulting models will be built based on vectorization data. A TF-IDF is applied, and the relative count of each word is stored in a sparse matrix. TF-IDF differs from the standard TF calculation that counts only the frequency of terms and would give more weight to longer documents than shorter documents. The IDF calculates the term frequency times the inverse frequency of the document. For the algorithms training and testing process, the data was split into two different blocks, the training block having 70% of the total data and the testing block having the remaining 30%. To the initial data, it has added the abstract text of the article, extracted from the location “weblink” in the original dataset. One possible approach for using ML to classify documents could be the author field. In a scarce dataset, the model is highly biased by the author’s affiliation to a particular school or domain. Therefore, the author’s names were disregarded as classification features. It is possible that, with a better dataset, the attributes author and affiliation can be used to improve the accuracy of the model. Therefore, given the low quantity of data available and the fact that several potential good features were ignored with the intent not to influence the model (like author’s name or affiliation), the scores were very promising ,

with NB and SVM models scoring 80% accuracy. However, the obtained accuracy was also achieved since the data used to train the algorithms was unbalanced, which resulted in a biased model. Hence, it was necessary to proceed to the data balancing resorting to oversampling technics to verify if the models accuracy will improve. After the data balancing process, were also implemented features to optimize the hyperparameters of the ML algorithms automatically, using the GridSearchCV module from the sci-kit learn library. Therefore it was selected a set of values for each hyperparameter for each ML algorithm to allow the optimization module to find the optimal set of hyperparameters.

This work used the algorithms: Support Vector Machine (SVM), namely SVC, LinearSVC, and NuSVC; Naïve Bayes (NB), specifically MultinomialNB, BernoulliNB, and ComplementNB; and Neural Network (NN) using the MLP classifier. To make the comparison between algorithms were performed one hundred hyperparameters optimizations for each algorithm. Thus, metrics related to the precision of the algorithms were collected, namely, the optimization score and elapsed time. Figure 3 presents the accuracy and optimization score of the used algorithms. Table 2 shows the elapsed time divided into four columns: the first two columns contain the training time of the best model achieved and the average training time of the algorithms, and the remaining columns present the average and the total time of the algorithms hyperparameters optimization.

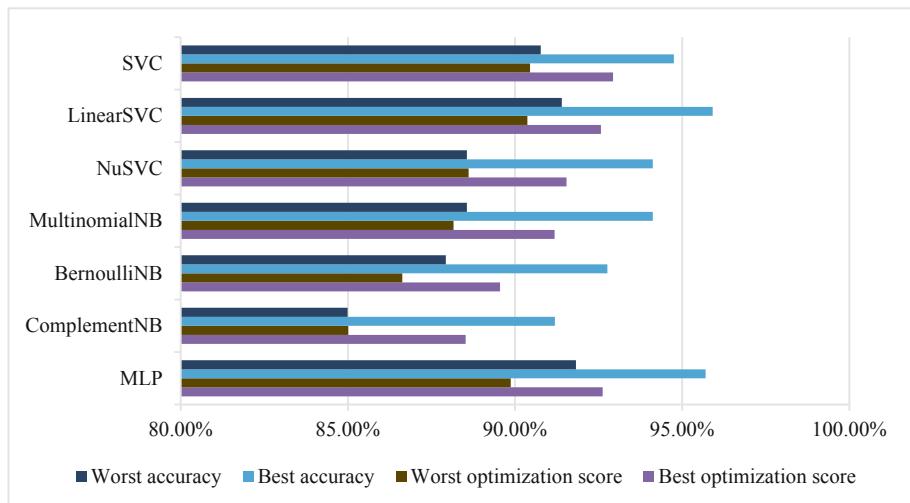


Fig. 3. Accuracy achieved with the algorithms adopted

Table 2. Algorithms training and optimization time comparison

Algorithms	Best model training time (s)	Average training model time (s)	Average optimization time (s)	Total optimization time (m)
SVC	33.51	33.67	135.73	222.66 (≈ 3.71 h)
LinearSVC	0.07	0.06	28.26	47.42
NuSVC	2.32	2.36	115.38	191.7 (≈ 3.2 h)
MultinomialNB	≈ 0.002	≈ 0.002	0.69	1.22
BernoulliNB	≈ 0.003	≈ 0.003	0.79	1.39
ComplementNB	≈ 0.002	≈ 0.002	0.80	1.40
MLP	16.98	16.80	484.58	844.36 (≈ 14 h)

5 Conclusions

The algorithm with the best accuracy result obtained was the LinearSVC algorithm, belonging to the class of SVM algorithms, with an accuracy of 95.91%. In the class of NB algorithms, the MultinomialNB algorithm reached an accuracy of 94.12%, and the MLP algorithm, belonging to the NN algorithms class, got the second-best precision value in the total set of ML algorithms with 95.70%.

However, the training time is also a relevant factor in the algorithm implementation, since, ideally, the implemented algorithms should be able to learn continuously. Therefore, depending on the requirements of the implementation, it is necessary to consider if it is worth it, a higher training time for better accuracy. For example, the training time of the best accuracy MLP classifier took 16.98 s while to train the best MultinomialNB classifier took 0.002 s, which means that for an accuracy improvement of 1.58%, the time needed to train got 8490 times higher. With this low amount of data, the time difference is already substantial, but with the continuous learning of the algorithms, the training time could get unbearable.

The training time of the algorithms, to be able to make an automatic classification with high accuracy, can be long. Still, the time necessary for manual classification of scientific contents is much more significant and subject to errors. Thus, the need arises to verify the result of automatic classifications with the result of manual classifications. Therefore, it was verified whether the “wrong” classifications made by the algorithms to the test dataset were analyzed to understand if they were wrong or if the scientific content was manually classified in the wrong way.

An ML text classifier based on supervised learning is highly dependent on the amount of training data available. The results obtained in this work can improve with the increase in the amount of training data, as well as in terms of quality. For example, authors identification, authors affiliation were not used for this purpose. Another attribute relevant is the keywords, but it would be useful to use keywords supported on controlled vocabulary from a taxonomy. However, a future automatic classifier tool should validate the keywords through an ML algorithm to detect emerging areas of knowledge or alert for

misuse of keywords. To increase the classification accuracy, we propose to editors (conferences and journals) to limit the keywords used in an article to a controlled vocabulary based on taxonomic classes.

To be explored is also the integration of more complex ontology-based knowledge in classification. The development of more efficient non-associative classification algorithms that integrate taxonomy information in classifier training and DL's use, the more data you give and the more computational time you provide, the better accuracy classification is obtained.

Finally, there is a need to classify into multiple knowledge domains correctly, and a classification tool must consider this.

Acknowledgments. This work has been supported by IViSSEM: POCI-01-0145-FEDER-28284.

References

1. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: a review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **160**, 3–24 (2007)
2. Romanov, A., Lomotin, K., Kozlova, E.: Application of natural language processing algorithms to the task of automatic classification of Russian scientific texts. *Data Sci. J.* **18**(1), 1–17 (2019). <https://doi.org/10.5334/dsj-2019-037>
3. Jones, K.S.: Some thoughts on classification for retrieval. *J. Docum.* **61**(5), 571–581 (2005)
4. Comte, A.: Introduction to Positive Philosophy. Hackett Publishing (1988)
5. Rich, P.: The organizational taxonomy: definition and design. *Acad. Manag. Rev.* **17**(4), 758–781 (1992). <https://doi.org/10.2307/258807>
6. Brewster, C., Wilks, Y.: Ontologies, taxonomies, thesauri: learning from texts. In: Proceedings of the Use of Computational Linguistics in the Extraction of Keyword Information from Digital Library Content Workshop 32 (2004)
7. Frodeman, R., Klein, J.T. (eds.): *The Oxford Handbook of Interdisciplinarity*. Oxford University, Press (2012)
8. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. *Data Knowl. Eng.* **25**(1–2), 161–197 (1998). [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
9. OWL - Semantic Web Standards (2012). <https://www.w3.org/OWL/>
10. Kopácsi, S., Hudak, R., Ganguly, R.: Implementation of a classification server to support metadata organization for long term preservation systems. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* **70**(2), 225–243 (2017)
11. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI Mag.* **17**(3) (1996)
12. Atkinson-abutridy, J., Mellish, C., Aitken, S.: A semantically guided and domain-independent evolutionary model for knowledge discovery from texts. *IEEE Trans. Evol. Comput.* 546–560 (2003)
13. Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., Watkins, C.: Text classification using string kernels. *J. Mach. Learn. Res.* **2**(Feb), 419–444 (2002)
14. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: European Conference on Machine Learning, pp. 137–142 (1998)
15. Friedman, C., Rindflesch, T.C., Corn, M.: Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. *J. Biomed. Inform.* **46**(5), 765–773 (2013). <https://doi.org/10.1016/j.jbi.2013.06.004>

16. Otter, D.W., Medina, J.R., Kalita, J.K.: A survey of the usages of deep learning for natural language processing. *IEEE Trans. Neural Netw. Learn. Syst.* (2020)
17. Manda, P., Ozkan, S., Wang, H., McCarthy, F., Bridges, S.M.: Cross-ontology multi-level association rule mining in the gene ontology. *PLoS ONE* 7(10), e47411 (2012)
18. ALGORITMI – University of Minho (2020). <https://algoritmi.uminho.pt>
19. Frodeman, R., Klein, J.T. (eds.): *The Oxford Handbook of Interdisciplinarity*. Oxford University Press (2012)



Crowdsourced Data Stream Mining for Tourism Recommendation

Fátima Leal^{1(✉)}, Bruno Veloso^{1,2}, Benedita Malheiro^{2,3},
and Juan C. Burguillo⁴

¹ Universidade Portucalense, Porto, Portugal

{fatimal,brunov}@upt.pt

² INESC TEC, Porto, Portugal

mbm@isep.ipp.pt

³ School of Engineering, Polytechnic Institute of Porto, Porto, Portugal

⁴ School of Telecommunication Engineering, University of Vigo, Vigo, Spain

J.C.Burguillo@uvigo.es

Abstract. Crowdsourced data streams are continuous flows of data generated at high rate by users, also known as the crowd. These data streams are popular and extremely valuable in several domains. This is the case of tourism, where crowdsourcing platforms rely on tourist and business inputs to provide tailored recommendations to future tourists in real time. The continuous, open and non-curated nature of the crowd-originated data requires robust data stream mining techniques for on-line profiling, recommendation and evaluation. The sought techniques need, not only, to continuously improve profiles and learn models, but also be transparent, overcome biases, prioritise preferences, and master huge data volumes; all in real time. This article surveys the state-of-art in this field, and identifies future research opportunities.

Keywords: Crowdsourced data streams · Data stream mining · Profiling · Recommendation · Tourism

1 Introduction

Tourism crowdsourcing platforms have revolutionised both the tourist behaviour and the tourism industry. Platforms such as AirBnB, Booking or TripAdvisor are popular online intermediaries between tourism businesses and tourists and, as a result, continuously accumulate large amounts of data shared by the tourists about their tourism experiences. They adopt a business model where stakeholders play predefined roles: (*i*) businesses pay to have their services on display; and (*ii*) tourists search for services of interest at no cost and provide feedback on their customer experience for free. According to Leal *et al.* (2018) [9], depending on the main type of data shared by the crowd, crowdsourcing tourism services can be classified as evaluation-based, map-based, wiki-based, and social network-based.

While the processing of crowdsourced data can be performed off-line, using data mining, or on-line, using data stream mining, this review article addresses exclusively the challenge of the on-line processing of tourism crowdsourced data. Specifically, the application of data stream mining techniques to crowd inputs is more demanding due to real-time and transparency requirements.

This paper surveys existing techniques and recognises the most promising research trends in tourism crowdsourced data stream recommendation. The adopted method analyses the tourism data stream mining pipeline to identify techniques and technologies for real-time predictions driven by the accountability, responsibility and transparency design principles. To this end, the review of the stream-based processing pipeline covers: (*i*) profiling, (*ii*) recommendation, (*iii*) explanation, (*iv*) evaluation and (*v*) support technologies, such as blockchain or High Performance Computing (HPC). Figure 1 illustrates this approach. Tourism data stream mining is event-driven and implements, in real time, a profiling, recommendation and evaluation loop. In this context, the continuous arrival of crowd-originated events triggers, first, the update of the involved profile and prediction models and, then, the suggestion and evaluation of personalised self-explainable recommendations.



Fig. 1. Review proposal

The remaining contents of this document details the data stream recommendation *status quo*, challenges and support technologies (Sect. 2); identifies future research trends (Sect. 3); and draws the conclusion (Sect. 4).

2 Data Stream Mining

Data stream mining explores methods and algorithms for extracting knowledge from data streams, which are data sequences occurring continuously and independently. By applying learning algorithms to crowdsourced data streams, *i.e.*, performing tourism crowdsourced data stream mining, it is possible to predict the tourist behaviour based on the associated digital footprint. However, due to the intrinsic dynamic nature of these heterogeneous data streams, they require on the fly techniques to perform automatic model learning and updating, concept drift identification and recovery; as well as cope with preference changes over time, uncurated crowdsourced data and extremely large volumes of data. In this context, automatic model learning refers to the selection of a suitable predictive model (or combination of models), whereas concept drift describes unforeseeable

changes in the underlying distribution of streaming data overtime, which need to be addressed to prevent poor learning results [18].

Given the natural evolution of user interests over time, data stream recommendation needs to reflect current rather than outdated interests and, evaluation-wise, requires specialised evaluation protocols [6] and metrics. Finally, crowdsourced data are potentially unreliable and accumulate in huge volumes, dictating the adoption of technologies, which monitor traceability and authenticity, and create the need to perform parallel processing [21].

2.1 Profiling

Entity profiling, *i.e.*, the creation and maintenance of entity models, is central to generate personalised tourism recommendations. Using crowdsourced tourism data, it is possible to model the stakeholders according to the corresponding digital footprint stored in tourism crowdsourcing platforms. Resource (item) profiling can be based on intrinsic characteristics, crowdsourced information and semantic enrichment. Tourist (user) profiles are mainly based on crowdsourced data, which can be classified as entity-based or feature-based. While entity-based profiles are directly associated to tourism resources; feature-based rely on intrinsic characteristics, *e.g.*, category, location, theme, *etc.* Based on the contents of crowdsourced data, the literature identifies further types of profiles.

Rating-based profiles rely on ratings to express, quantitatively, opinions concerning multiple services aspects. In evaluation-based crowdsourcing platforms, users can classify tourism resources using multiple service dimensions. Leal *et al.* (2019) [11] provide a survey of single and multiple criteria rating-based profiles approaches, whereas Nilashi *et al.* (2015, 2017) [22, 23] present a stream-based multiple criteria approach adopting ensemble techniques. Recently, Veloso *et al.* (2018, 2019) [34, 36] have used hotel and restaurant evaluations to create stream-based profiles adopting incremental updating.

Review-based profiles are created from textual reviews. These reviews generally include qualitative comments and descriptions. In this context, a collection of reviews, rather than being perceived as static, constitutes an ongoing stream [29], leading to opinion stream mining.

Context-based profiles use context information, which can be personal context data, social context data, and context-aware information data [15]. Gomes *et al.* (2010) [7] propose a context-aware system with data stream learning to improve existing drift detection methods by exploiting available context information. Similarly, Akbar *et al.* (2015) [1] explore context-aware stream processing to detect traffic in near real-time.

Quality profiles model tourism entities using quality related parameters. It has been used mainly to model tourism wiki pages and corresponding publishers. Wiki publishers originate continuous data streams in the form of content revisions. However, scant research has been conducted to construct quality-based profiles employing wiki-based information as data streams. Leal *et al.* (2019) [14] use this approach to model the quality of publishers and pages, using wiki streams of publisher-page-review triplets.

Popularity profiles use views, clicks and related-data to model the popularity of tourists and tourism resources. These profiles are frequently used to avoid the cold start problem in collaborative filtering. Leal *et al.* [14] rely on a page view data stream to model wiki publishers and pages in terms of popularity.

Trust and Reputation profiles model reliability. Trust defines the reliability of stakeholders based on direct one-to-one relationships. Reputation is based on third party experiences, *i.e.*, many-to-one relationships. Leal *et al.* (2018) [10] propose trust and reputation modelling for stream-based hotel recommendation, and Leal *et al.* (2019) [12] employ incremental trust and reputation models for post-filtering, improving the accuracy of recommendations in both cases. Recently, Leal *et al.* (2020) [13] recommended chaining¹ trust and reputation models for reliability and explainability purposes.

Hybrid profiles combine multiple types, leading to richer and more refined profiles and, in principle, to higher quality recommendations. Hybrid-based profiles are indicated for heterogeneous data environments, which have been explored using ensembles [25]. However, building hybrid-based profiles from crowdsourced tourism data streams remains unexplored.

Regardless of the contents or the type of profiling used, crowdsourced data streams allow the continuous updating of tourism stakeholder profiles.

2.2 Recommendation

Recommendation engines play an important role in the tourism domain, providing personalised recommendations before a large variety of options. They rely mostly on data filtering techniques, ranging from pre-recommendation, recommendation and post-recommendation filters. Standard recommendation filtering techniques include:

Content-based filters match tourists with tourism resources. They create tourist profiles based on past interactions with the system, and make recommendations based on the similarity between the content of the tourist and resource profiles, *i.e.*, regardless of other tourist profiles [17].

Collaborative filters recommend unknown resources to tourists based on other like-minded tourists, using memory or model based algorithms, and building profiles based on the crowdsourced data. While memory-based approaches combine the preferences of neighbours with identical profiles to generate recommendations, model-based algorithms build models based on the tourist profile to make predictions. Collaborative filters may implement tourist-based or resource-based variants by computing the similarity between tourists or between resources. These techniques have been adapted with success to data stream recommendation [10, 12, 23, 34, 36].

Hybrid filters combine content-based and collaborative counterparts to eliminate frailties and reinforce qualities and, thus, improve the quality of recommendations. Hybrid filters, aggregating multiple mechanisms in parallel, have been explored by session-based recommendation systems [8, 28, 30].

¹ Storage in a blockchain.

A priori and *a posteriori* filtering aims to refine the recommendations reducing the search space. Pre-recommendation and post-recommendation filters have been explored mainly using context-based profiles.

Pre-Recommendation filters are applied beforehand to select appropriate tourist data [40], *e.g.*, weekdays recommendations, business or leisure travels. They increase recommendation relevance by analysing context-aware data.

Post-Recommendation filters remove or reorder the recommendations generated by the recommendation filter. In tourism domain, the value-for-money, the sentiment-value and the pairwise trust have been used, among others, as post-recommendation filters. Value-for-money confronts the price, the crowd overall rating and the resource official star rating to establish the crowd-sourced value for money. The sentiment-value of textual reviews is computed using sentiment analysis [34, 36]. Finally, the pairwise trust and similarity have been used to reorder the generated predictions [12].

Data stream recommendation enables the continuous updating of the users and items models and contributes to improve the quality of real-time recommendations. While data stream tourism recommendation has been able to adapt standard recommendation techniques, mainly collaborative filters, to real-time processing, it still needs to address:

Concept drifts in collaborative filters can be detected by focusing on the recency, temporal dynamics or time period partitioning. In stream-based environments, concept drifts can be identified using window-based monitoring, accuracy-based model monitoring, and ensemble-based methods. Alternatively, an incremental adaptive unsupervised learning algorithm for recommendation systems that uses k -means clustering to detect drifts has been explored [38]. In the case of stream-based tourism recommendation, concept drifts has been explored using monitoring accuracy metrics [2, 33].

Model learning is mandatory for data stream recommendation. In the tourism domain, Nilashi *et al.* (2017) [22] explores automatic model learning by proposing a hybrid ensemble. Alternatively, Veloso *et al.* (2018) [33] perform a controlled exploration of the model search space. Recently, Al-Ghossein (2019) [2] propose dynamic learning model methods to generate stream-based context-aware recommendations.

Preference evolution must be monitored in data stream mining contexts to ensure that recommendations reflect only the current interests of the user. Matuszyk *et al.* (2018) [20] identify standard deviation-based user factor fading as the best forgetting strategy to improve the results of the Biased Regularised Incremental Simultaneous Matrix Factorisation algorithm. Wang *et al.* (2018) [39] describe a probabilistic matrix factorisation model based on Bayesian personalised ranking to keep relevant long-term user interests.

However, only the works reported in [2, 22] address the tourism domain.

2.3 Explanations

Given the highly influential nature of recommendations, there are growing concerns about the principles behind recommendation algorithms. In this regard, Dignum (2017) [5] recommends that the development of such algorithms should be guided by the following design principles: accountability – explain and justify decisions; responsibility – incorporate human values into technical requirements; and transparency – describe the decision-making process and how data is used, collected, and governed. This means that data stream recommendation must explain and justify the rationale behind all recommendations, increasing the confidence of the users and the transparency of the system.

An explanation is any additional information which clarifies why a system arrived at a particular decision. Specifically, in the case of recommendations, explanations justify why an item has been recommended, adding transparency and supporting decision making. An explainable and transparent system helps the user understand whether the output is based on his/her preferences rather than third party interests. Explanation models can use multiple sources of information, ranging from entity-based, feature-based, text-based, visual-based to social-based. In this regard, Veloso *et al.* (2019) [35] suggest exploring trust and reputation profiles to explain recommendations in tourism crowdsourcing platforms while, at the same time, storing these profiles in a blockchain to ensure authenticity and integrity. This proposal was implemented by Leal *et al.* (2020) [16]. They incrementally update and store trust models of the crowd contributors in the blockchain as smart contracts and, then, use them to derive reputation models and generate stream-based explainable recommendations.

2.4 Evaluation

Stream-based evaluation has two main components: the evaluation protocol and the evaluation metrics. An online evaluation protocol has three main constraints: (*i*) space, where the available memory is limited; (*ii*) learning time, when the time required to learn is equal that the rate of incoming events; and (*iii*) accuracy or the capacity of the model capture the data variations. The most used online evaluation protocol is the prequential protocol [6], which adopts sliding windows or fading factors to forget less relevant examples. It has three steps: (*i*) produce a prediction for an unlabelled instance in the stream; (*ii*) assess the prediction error; and (*iii*) update the model with the most recently observed error.

In terms of evaluation metrics, there are predictive, classification, and statistical metrics. Prediction metrics describe the accuracy in the accumulation of predictive errors [31]. In terms of classification metrics, Cremonesi *et al.* (2010) [4] present a three-step methodology: (*i*) generate the predictions of all items not yet classified by the active user; (*ii*) select randomly 1000 of these predictions plus the active user real value; and (*iii*) sort this list of 1001 item values using the post-filter. Finally, concerning statistical metrics, Souza *et al.* (2018) [27] have recently suggested a new evaluation measure (Kappa-Latency), which takes into account the arrival delay of actual instances. Alternatively, Vinagre

et al. (2019) [37] propose, for recommendation algorithms, the adoption of the k -fold validation framework together with McNemar and Wilcoxon signed-rank statistical tests applied to adaptive-size sliding windows.

2.5 Support Technologies

Real time processing requirements of stream-based recommendation, and the uncurated nature of crowdsourced data poses infrastructural challenges. This review highlights two key technologies to address them: blockchain and HPC.

Blockchain is a distributed ledger technology maintained by a peer-to-peer network of nodes where blocks, containing validated transactions, are sequentially chained through cryptographic hashes. The network validates new transactions concurrently, using consensus mechanisms. Once validated, they are committed to a block granting security, authenticity, immutability, and transparency. Moreover, it ensures end-to-end verification, which can be used to record data and track sources over time in a trusted manner. Blockchain has been explored, in stream-based environments, for auditable purposes [26] and to store tourism smart contracts and transact cryptocurrencies [3, 24].

High Performance Computing and, in particular, cloud computing infrastructures, underpin the algorithmic analysis of large amounts of data, becoming a *de facto* pillar of scalable data analytics [32]. In the tourism domain, Veloso *et al.* (2018) [34] explores the scalability of crowdsourced data stream recommendation using HPC.

3 Research Trends

The most relevant research trends in the crowdsourced data stream recommendation for tourism encompass reliable profiling, automated model learning, including the detection of concept drifts, preference evolution, processing transparency as well as the identification of support technologies that meet the data authenticity and traceability (blockchain) and seamless scalability (HPC) requirements.

Reliable profiling – Crowdsourced data streams are unfiltered and uncurated by default, meaning that they are exposed to malicious manipulation. This suggests the need to build reliable models of data contributors, and to trace data contributions back to contributors themselves. To this end, trust and reputation profiling approaches has been explored [16, 35].

Concept drift detection – On the fly concept drift detection relies on constant monitoring of relevant metrics and on model learning [2, 22, 33].

Model learning – The processing of tourism crowdsourced data streams demands dynamic model learning to continue generating meaningful recommendations over time [2, 22, 33].

Preference evolution – Stream-based tourism recommendation requires techniques that ignore outdated user preferences [19, 39].

Transparency – Personalised recommendations require the explanation of the underlying reasoning and data, particularly when they are based on crowdsourced data. In this context, trust and reputation models of contributors have been explored to explain recommendations [16, 35].

Blockchain – Crowdsourced data is prone to manipulation. Blockchain provides data quality control for data authenticity and traceability [16, 35].

HPC – Smart tourism produces crowdsourced data at a high rate and volume, demanding agile mechanisms to profile and filter information in real time and, consequently, efficient computational infrastructures [34].

4 Conclusion

The research on tourism crowdsourced data stream recommendation presents multiple algorithmic and technology challenges. On the one hand, the algorithmic design needs to address further concept drift identification, crowd reliability, distributed processing, model learning, preference evolution and transparency. As shown, these research directions are beginning to be explored, but there is still a long way to go. On the other hand, the data reliability, pace and volume and the near real time operation impose extremely demanding requirements for supporting technologies. Nevertheless, blockchain and HPC appear as two promising pillars. The adoption of blockchain grants data traceability, authenticity and, when integrated with trust and reputation modelling, provides algorithmic transparency, whereas HPC contributes with a computational infrastructure solution for the real time performance requirements.

Acknowledgments. This work was partially financed by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UIDB/50014/2020, and also from Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2019–2022) and the European Union (European Regional Development Fund - ERDF).

References

1. Akbar, A., Carrez, F., Moessner, K., Sancho, J., Rico, J.: Context-aware stream processing for distributed IoT applications. In: 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), pp. 663–668. IEEE (2015)
2. Al-Ghossein, M.: Context-aware recommender systems for real-world applications. Ph.D. thesis, Paris Saclay (2019)
3. Calvaresi, D., Leis, M., Dubovitskaya, A., Schegg, R., Schumacher, M.: Trust in tourism via blockchain technology: results from a systematic review. In: Information and Communication Technologies in Tourism 2019, pp. 304–317. Springer (2019)
4. Cremonesi, P., Koren, Y., Turrin, R.: Performance of recommender algorithms on top-n recommendation tasks. In: Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys 2010, pp. 39–46. ACM, New York (2010)

5. Dignum, V.: Responsible artificial intelligence - how to develop and use AI in a responsible way. *Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer (2019)
6. Gama, J., Sebastião, R., Rodrigues, P.P.: Issues in evaluation of stream learning algorithms. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 329–338 (2009)
7. Gomes, J.B., Menasalvas, E., Sousa, P.A.C.: CALDS: context-aware learning from data streams. In: Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques, pp. 16–24 (2010)
8. Guo, L., Yin, H., Wang, Q., Chen, T., Zhou, A., Quoc Viet Hung, N.: Stream-based session-based recommendation. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1569–1577 (2019)
9. Leal, F.: Recommendation of tourism resources supported by crowdsourcing. Ph.D. thesis, University of Vigo (2018)
10. Leal, F., Malheiro, B., Burguillo, J.C.: Trust and reputation modelling for tourism recommendations supported by crowdsourcing. In: World Conference on Information Systems and Technologies, pp. 829–838. Springer (2018)
11. Leal, F., Malheiro, B., Burguillo, J.C.: Analysis and prediction of hotel ratings from crowdsourced data. *Wiley Interdisc. Rev.: Data Min. Knowl. Discov.* **9**(2), e1296 (2019)
12. Leal, F., Malheiro, B., Burguillo, J.C.: Incremental hotel recommendation with inter-guest trust and similarity post-filtering. In: World Conference on Information Systems and Technologies, pp. 262–272. Springer (2019)
13. Leal, F., Veloso, B., Malheiro, B., González-Vélez, H.: Trust and reputation smart contracts for explainable recommendations. In: Trends and Innovations in Information Systems and Technologies, vol. 18, pp. 124–133. Springer (2020)
14. Leal, F., Veloso, B.M., Malheiro, B., González-Vélez, H., Burguillo, J.C.: Scalable modelling and recommendation using wiki-based crowdsourced repositories. *Electron. Commer. Res. Appl.* **33**, 100817 (2019)
15. Leal, F., Malheiro, B., Burguillo, J.C.: Context-aware tourism technologies. *Knowl. Eng. Rev.* **33**, e13 (2018)
16. Leal, F., Veloso, B., Malheiro, B., González-Vélez, H., Burguillo, J.C.: A 2020 perspective on “scalable modelling and recommendation using wiki-based crowdsourced repositories:” fairness, scalability, and real-time recommendation. *Electron. Commer. Res. Appl.* **40**, 100951 (2020)
17. Lops, P., Jannach, D., Musto, C., Bogers, T., Koolen, M.: Trends in content-based recommendation - preface to the special issue on recommender systems based on rich item descriptions. *User Model. User-Adapt. Interact.* **29**(2), 239–249 (2019)
18. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G.: Learning under concept drift: a review. *IEEE Trans. Knowl. Data Eng.* **31**(12), 2346–2363 (2019)
19. Matuszyk, P., Spiliopoulou, M.: Stream-based semi-supervised learning for recommender systems. *Mach. Learn.* **106**(6), 771–798 (2017)
20. Matuszyk, P., Vinagre, J., Spiliopoulou, M., Jorge, A.M., Gama, J.: Forgetting techniques for stream-based matrix factorization in recommender systems. *Knowl. Inf. Syst.* **55**(2), 275–304 (2018)
21. Nguyen, H.L., Woon, Y.K., Ng, W.K.: A survey on data stream clustering and classification. *Knowl. Inf. Syst.* **45**(3), 535–569 (2015)
22. Nilashi, M., Bagherifard, K., Rahmani, M., Rafe, V.: A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. *Comput. Ind. Eng.* **109**, 357–368 (2017)

23. Nilashi, M., Jannach, D., bin Ibrahim, O., Ithnin, N.: Clustering- and regression-based multi-criteria collaborative filtering with incremental updates. *Inf. Sci.* **293**, 235–250 (2015)
24. Ozdemir, A.I., Ar, I.M., Erol, I.: Assessment of blockchain applications in travel and tourism industry. *Qual. Quant.* **54**, 1549–1563 (2020)
25. van Rijn, J.N., Holmes, G., Pfahringer, B., Vanschoren, J.: The online performance estimation framework: heterogeneous ensemble learning for data streams. *Mach. Learn.* **107**(1), 149–176 (2018)
26. Shafagh, H., Burkhalter, L., Hithnawi, A., Duquennoy, S.: Towards blockchain-based auditable storage and sharing of IoT data. In: Proceedings of the 2017 on Cloud Computing Security Workshop, pp. 45–50 (2017)
27. Souza, V., Pinho, T., Batista, G.: Evaluating stream classifiers with delayed labels information. In: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS), pp. 408–413. IEEE (2018)
28. de Souza Pereira Moreira, G., Jannach, D., da Cunha, A.M.: Contextual hybrid session-based news recommendation with recurrent neural networks. *IEEE Access* **7**, 169185–169203 (2019)
29. Spiliopoulou, M., Ntoutsi, E., Zimmermann, M.: Opinion Stream Mining, pp. 938–947. Springer, Boston (2017)
30. Sun, S., Tang, Y., Dai, Z., Zhou, F.: Self-attention network for session-based recommendation with streaming data input. *IEEE Access* **7**, 110499–110509 (2019)
31. Takács, G., Pilászy, I., Németh, B., Tikk, D.: Scalable collaborative filtering approaches for large recommender systems. *J. Mach. Learn. Res.* **10**, 623–656 (2009)
32. Talia, D.: Clouds for scalable big data analytics. *IEEE Comput.* **46**(5), 98–101 (2013)
33. Veloso, B., Gama, J., Malheiro, B., Vinagre, J.: Self hyper-parameter tuning for stream recommendation algorithms. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 91–102. Springer (2018)
34. Veloso, B., Leal, F., González-Vélez, H., Malheiro, B., Burguillo, J.C.: Scalable data analytics using crowdsourced repositories and streams. *J. Parallel Distrib. Comput.* **122**, 1–10 (2018)
35. Veloso, B., Leal, F., Malheiro, B., Moreira, F.: Distributed trust & reputation models using blockchain technologies for tourism crowdsourcing platforms. *Proc. Comput. Sci.* **160**, 457–460 (2019)
36. Veloso, B.M., Leal, F., Malheiro, B., Burguillo, J.C.: On-line guest profiling and hotel recommendation. *Electron. Commer. Res. Appl.* **34**, 100832 (2019)
37. Vinagre, J., Jorge, A.M., Rocha, C., Gama, J.: Statistically robust evaluation of stream-based recommender systems. *IEEE Trans. Knowl. Data Eng.* (2019)
38. Wanás, N.M., Farouk, A., Said, D., Khodeir, N., Fayek, M.B.: Detection and handling of different types of concept drift in news recommendation systems. *Int. J. Comput. Sci. Inf. Technol.* **11**, 87–106 (2019)
39. Wang, W., Yin, H., Huang, Z., Wang, Q., Du, X., Nguyen, Q.V.H.: Streaming ranking based recommender systems. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 525–534 (2018)
40. Zheng, Y., Burke, R., Mobasher, B.: Differential context relaxation for context-aware travel recommendation. In: International Conference on Electronic Commerce and Web Technologies, pp. 88–99. Springer (2012)



W-core Transformer Model for Chinese Word Segmentation

Hai Lin¹, Lina Yang^{1(✉)}, and Patrick Shen-Pei Wang²

¹ School of Computer, Electronics and Information, Guangxi University,
Nanning 530004, People's Republic of China

lnyang@gxu.edu.cn

² Computer and Information Science, Northeastern University,
Boston, MA 02115, USA
patwang@ieee.org

Abstract. Chinese word segmentation is an important research content in the field of Natural Language Processing (NLP). In this paper, we combine the Transformer model to propose the Window Core (W-core) Transformer for the tasks. In this model, W-core can preprocess sentence information according to the characteristics of Chinese and incorporate features extracted by the Transformer model. Experimental results show that the W-core Transformer model can improve the effect of the original Transformer model on Chinese word segmentation. Finally, we improve the performance of W-core Transformer by increasing the number of encoder layers and oversampling.

Keywords: Transformer · NLP · Chinese word segmentation · W-core

1 Introduction

In NLP, the tasks of word processing include word feature analysis [1,2], new word recognition [3] and so on. Word segmentation tasks is a very special research direction, as not all languages need to split their words from sentences. For Chinese, on the one hand, it is different from Indo-European language family such as English. As we all know English separates each independent word by a space, while the word structure of Chinese requires readers to combine context and judge based on experience. On the other hand, for the same Chinese sentence, different word segmentation forms will bring different semantic results. Therefore, Chinese word segmentation is very necessary.

Chinese word segmentation is usually used to preprocess Chinese sentences and extract key information [4]. In recent years, with the rise of neural networks, neural networks have become more and more widely used in NLP tasks. By using the word vector space, the neural network can learn the features of the text in an automatic learning manner. In addition, different types of parameters and activation functions inside the neural network make it closer to the real way of human learning, and closer to a complex nonlinear function in theory. Therefore,

neural network-based models can do better in NLP tasks compared to traditional methods.

Traditional NLP models use Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) [5] for task processing. These two different neural networks have improved the machine's ability to analyze natural language in different degrees and directions. The CNN model can quickly and efficiently extract the relationship between words in the text, and because of its parallel design concept, the task is processed very quickly. However, the CNN model is not ideal for extracting information from long texts. The appearance of the LSTM model largely solves the problem of long text information loss. The LSTM model is better at processing long texts. It can record information on a long time step very well. But because of its linear design structure, the training time of the model becomes longer. In recent years, everyone was trying to combine the CNN model with the LSTM model to give play to their respective advantages, thereby achieving a better neural network model structure such as Sequence to sequence (Seq2seq) model.

Since the Attention mechanism was proposed, the Seq2seq model with attention has been improved in each task, so the current Seq2seq model refers to the model that combines Recurrent Neural Networks (RNN) and Attention. Until Google proposed a Transformer model to solve the sequence to sequence problem, replacing LSTM with a full attention structure, and achieved better results in NLP tasks. In recent years, there are many models using Bidirectional Encoder Representations from Transformers (BERT) [6,7] to process Chinese word segmentation tasks. This paper is based on Transformer model, proposing W-core Transformer model.

Our main contributions include

- introduced the W-core layer, and the result is improved without increasing too much calculation.
- optimized Transformer network for Chinese word segmentation task.
- contrasting the experimental results, put forward relevant optimization ideas for the Chinese word segmentation task.

2 Related Work

The earliest research on Chinese word segmentation is based on the rule matching method based on a fixed dictionary. In order to overcome the problem that the dictionary matching method cannot handle ambiguity words and unregistered words, in 2003, some researchers proposed to convert Chinese word segmentation into a sequence labeling task based on character annotation. In 2004, some researchers used Conditional Random Field (CRF) to solve this problem [8,9]. In 2011, Collobert proposed the use of neural networks to deal with sequence labeling problems.

There has been a lot of research work in the field of neural networks applied to Chinese word segmentation tasks. Many researchers have found that adding an attention mechanism to the original model can easily improve the performance of

the model. Self-attention mechanism can make generate information interaction from sequence data in the network computing layer. This information interaction enables the network computing layer to establish feature connections between different sequences in the sequence. Compared with CNN, the advantage of the network using the self-attention mechanism is that it has a stronger ability to encode the entire sequence. Moreover, the convolution calculation layer and the self-attention calculation layer can be used simultaneously in the network to take into account the short-distance information relationship dependence and the long-distance information relationship dependence. Therefore, we use Transformer model based on the self-attention mechanism to process the Chinese word segmentation task [10, 11].

2.1 Encoder in Transformer Model

Like most Seq2seq models, the structure of the Transformer model is also composed of encoder and decoder. Transformer model uses an encoder to extract sequence features in sentences, and then uses the encoder to combine the features of the input sentence to predict the output sentence. For the Chinese word segmentation task, we need a model to extract the relationship between the words in the sentence, so the encoder in the Transformer model is selected as the feature extraction model.

The encoder consists of an Embedding layer, Multi-head self-attention layer and Position-wise feed-forward networks layer. The model structure of Encoder is shown in Fig. 1:

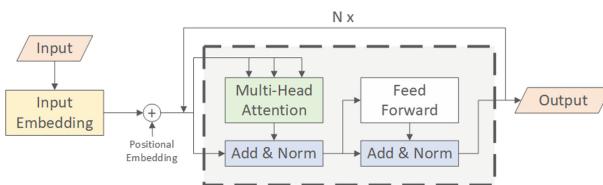


Fig. 1. Encoder model structure.

Encoder consists of N identical layers. Layer refers to the unit on the left side of the above figure. There is an $N \times$ on the far left. Each Layer is composed of two sub-layers, called multi-head self-attention mechanism and a fully connected feed-forward network. Each sub-layer adds residual connection and normalization, so the output of the sub-layer can be expressed as:

$$\text{sub-layer-output} = \text{LayerNorm}(x + \text{SubLayer}(x)) \quad (1)$$

The embedding layer is used to map a single word to a multi-dimensional word space vector. In this way, the computer can better understand the meaning of the word. But the self-attentive mechanism has no temporal characteristics.

Therefore, when dealing with the embedding layer, a location information coding layer called Positional Encoding was also added. The position coding layer is generated as Method 2.

Positional embedding is calculated directly with sine and cosine functions of different frequencies:

$$\begin{aligned} PE_{(pos,2i)} &= \sin(pos/10000^{2i/d_{model}}) \\ PE_{(pos,2i+1)} &= \cos(pos/10000^{2i/d_{model}}) \end{aligned} \quad (2)$$

Assuming that $d_{model} = 512$ and sentence length = 50 are used for position encoding.

After the embedding layer is completed, the generated word vector is passed into the Multi-head self-attention layer. It is an upgraded Attention layer. The operation of Attention is as formula (3).

$$attention\text{-}output = Attention(Q, K, V) \quad (3)$$

Multi-head attention is to project Q, K, V through h different linear transformations, and finally splice different attention results, as formula (4):

$$\begin{aligned} MultiHead(Q, K, V) &= Concat(head_1, \dots, head_h)W^o \\ head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \quad (4)$$

Self-attention is the same as Q, K and V. In addition, the calculation of attention uses scaled dot-product, as formula (5):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

Another kind of calculation method with similar complexity, additive attention, is similar to dot-product when d_k is small. When d_k is large, it performs better without scaling. However, the calculation speed of dot product is faster and the influence of scaling is reduced.

The Position-wise feed-forward networks layer mainly provides nonlinear transformation in order to ensure that the data format received by each sub-layer is consistent.

2.2 Window-core Layer

W-core is specially designed according to the characteristics of the Chinese word segmentation task, which is also the focus of this paper. Words in Chinese are generally divided into single-word, double-word, three-word, and four-word, and the most commonly used words in daily life are double-word words. The original Encoder layer can well deal with the relationship between single-word and single-word, and infer the word segmentation structure of the sentence through this relationship [12]. But relatively long words, such as three-word or four-word,

cannot be well divided. Therefore, we introduced the W-core layer in order to better discover the relationship of long words in sentences.

W-core's design idea is similar to the convolution kernel, but its calculation method is more portable than the convolution operation. Usually a convolutional layer will be followed by a Max-pooling layer to reduce the amount of data and extract features. But for the Chinese word segmentation task, the most important thing is the small distance relationship between specific words. The TextCNN [13] model will lose a lot of metadata information. And we have used the efficient feature extraction network in the Transformer model—Encoder, so here we use the idea of TextCNN to propose the W-core layer. Its calculation process is shown in Fig. 2.

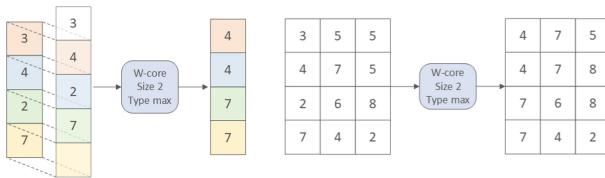


Fig. 2. W-core with Size 2 , Type ‘max’.

The calculation method used by the W-core layer in Fig. 2 is *MAX*, and the window size is 2. Compare the squares of the same color, and take the maximum value. The operation rule is to compare the data in the window and assign the maximum value as the final value to the new word vector. The calculation formula is 6:

$$\begin{aligned} X_{i,j} &= \text{MAX}(X_{i,j}, X_{i+1,j}, \dots, X_{i+s-1,j}) \\ X_{i,j} &= \text{AVG}(X_{i,j}, X_{i+1,j}, \dots, X_{i+s-1,j}) \end{aligned} \quad (6)$$

Where i is the word position, j is the word vector dimension, and s is the window size. *AVG* is an alternative method type under this circumstance, and the window size can also be adjusted according to specific tasks.

Another parameter of the W-core layer is padding. If *SAME* is used, the shape of the data will not be changed after the operation. When the data is insufficient, only the data in the window is judged, not by adding data 0. Just like the last yellow data 7 in Fig. 2, it is not $\text{MAX}(7, 0)$, but only data 7 in the window, so the final value is 7.

3 Proposed Method

Combined with the design idea and purpose of the W-core layer, we designed and analyzed the following five experimental models.

- (1) Encoder: The original encoder layer in the Transformer model. According to the data configuration in Table 2, the word vectors are directly passed into the model for training, as show in Fig. 3.

- (2) W-core Encoder: After embedding the words, a W-core layer is used to extract the features of the data. In this paper, the W-core layer with a window size of 2 and an extraction method of *MAX*, as show in Fig. 4.
- (3) TextCNN Encoder: Replace the W-core extraction method in the W-core Encoder with the TextCNN layer, and use this more complex processing method to make a perfect comparison of the experimental results, as show in Fig. 5.
- (4) W-core TextCNN Encoder: The W-core layer and the TextCNN layer are used to extract features, and then merged into the Encoder layer for operation. This model design is mainly used for comparison with the W-core layer and two Encoder layer networks, as show in Fig. 6.
- (5) W-core double Encoder: The Encoder layer is a very efficient and useful feature extraction layer. Therefore, the W-core layer is used to connect an Encoder layer, and then the metadata is connected to the Encoder layer. The result of the operation is integrated to obtain the final word segmentation result, as show in Fig. 7.

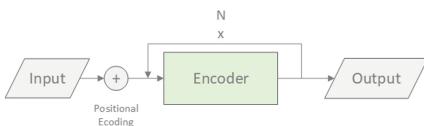


Fig. 3. Encoder model.

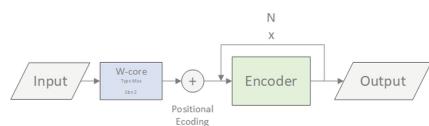


Fig. 4. W-core Encoder model.

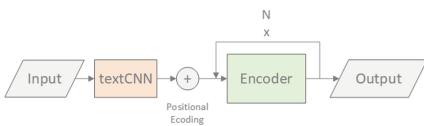


Fig. 5. TextCNN Encoder model.

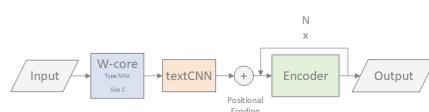


Fig. 6. W-core TextCNN Encoder model.

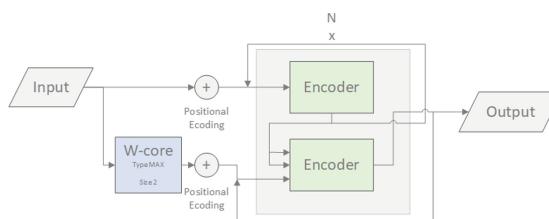


Fig. 7. W-core double Encoder model.

4 Experimental Analysis

4.1 Experimental Data

There are not many experimental corpus data sets for Chinese word segmentation. This paper selects the PKU data and MSR date of the four corpus data sets in SIGHAN2005. Table 1 lists the summaries for PKU data and MSR data.

Table 1. Experimental data

Name	Train	Test	Average length
PKU	19056	1944	96
MSR	86924	3985	47

There are two points that need special attention for all Chinese word segmentation data sets including but not limited to the PKU data set and MSR data set used in the experiments in this article:

- If certain countermeasures are not done in advance, multiple corpus data should not be directly combined for training. There is no unified standard for the segmentation basis of these corpus data sets. Training directly together may cause harm to the model's word segmentation performance.
- Even in the same data set, sentence segmentation results of similar (or even the same) contexts may be different. That is, the internal consistency of the data set itself is not 100%. Therefore, there is an upper limit on the results of the word segmentation performance test on each data set.

4.2 Model Framework Settings

In order to ensure that the experiment is only used to analyze the effect of W-core in the model, the unified structure and parameters of the experimental model are set uniformly.

The first is the parameter setting of the Encoder layer in the Transformer model, as shown in Table 2.

Table 2. Parameter settings

Parameter	Value
The number of bur-layer	4
Word vector length	128
Feed-forward networks layer size	512
Vocab size	6887
Dropout rate	0.1
Learning rate	Customized Schedule

The learning rate Customized Schedule is a function that can dynamically adjust the learning rate. Its change curve is shown in Fig. 8.

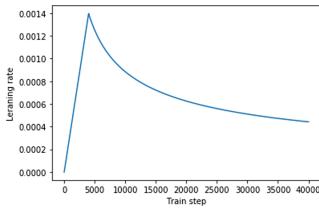


Fig. 8. Learning rate of Customized Schedule.

4.3 Results and Analysis

The model proposed in this paper is trained on PKU data set. F1 is used to comprehensively judge the performance of the model, P is the accuracy of the model, and Rough-L R is the calculated recall rate for the sentence result after word segmentation. The F1 , P and Rough-L R of models on this data set is shown in Table 3.

Table 3. Experimental results

Model	F1	P	Rough-L R
Encoder	91.11	91.58	90.20
W-core Encoder	91.27	92.14	90.40
TextCNN Encoder	91.37	91.56	90.62
W-core TextCNN Encoder	90.77	91.03	90.12
W-core double Encoder	91.45	92.38	90.47

As can be seen from Table 3, the best F1 value and the best P value appear in the W-core double Encoder model, and the best Rough-L R appear in the TextCNN model. Compared with the Encoder model, after adding the W-core layer, the three numerical results are improved. It can be seen that by adding the W-core layer, the performance of the model can be improved to a certain extent. However, it can see that if only the TextCNN model and the W-core TextCNN model are compared, its indicators become worse. In general, using more feature extraction layers can retain information better and achieve better results. However, the decrease in the effect here is due to the W-core layer ‘highlight’ the information of the two adjacent words, and passing the processed data directly to TextCNN, which will not get a good and correct relationship

between words. It can be seen from this that the role of the W-core layer is to highlight the data relationship in the window. By combining the Encoder layer based on the Attention mechanism, it can play a better role. If use the CNN model based on the window mechanism like W-core layer after W-core layer, it will get worse results.

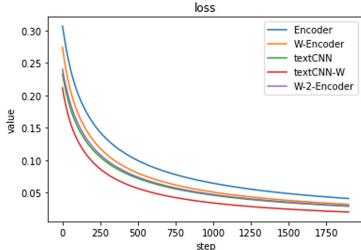


Fig. 9. Loss value for each model.

W-core double encoder	循环 神经 网络 在 自然语言 处理 领域 中 , 应用 比较 广泛 。
W-core encoder	循环 神经 网络 在 自然语言 处理 领域 中 , 应用 比较 广泛 。
others	循环 神经 网络 在 自然语言 处理 领域 中 , 应用 比较 广泛 。

Fig. 10. Word Segmentation results.

At the same time, in Fig. 9 it give a line chart of the five model loss values. As can it can see, the loss value (red line) of the W-core TextCNN Encoder model is the fastest, and is the closest to the data fitting. However, combined with the F1, P and Rough-L R values, after the W-core layer and the TextCNN layer, the word features of the data are extracted excessively, resulting in poor robustness of the model. The W-core Encoder(orange line) model's loss value drops faster and fits better than the Encoder model (blue line). The TextCNN Encoder model(green line) is similar to the W-core double Encoder model(purple line).

The results of Chinese word segmentation, as shown in Fig. 10. Most models do not recognize four-word words. The W-core Encoder model and the W-core double Encoder model have different recognition results of four-word-length words.

4.4 The Role of W-core

Based on the results and analysis mentioned above, it can deduce that the role of the W-core layer is to highlight the characteristics of adjacent sequences and properly ‘highlight’ the original information of the data. This behavior is a method specially designed for the Chinese word segmentation task and can be combined with different existing model structures.

During the experiment, it also tested the *AVG* method in the W-core layer, which is not as effective as the *MAX* method in the Chinese word segmentation task. Because *AVG* will ‘smooth’ the data, making the features more obscure. Corresponding to Table 4.

Table 4. Different method

TYPE	<i>F1</i>	<i>P</i>	<i>Rough-L R</i>
MAX	91.27	92.14	90.40
AVG	90.92	91.65	90.16

Table 5. Different window

SIZE	<i>F1</i>	<i>P</i>	<i>Rough-L R</i>
2	91.27	92.14	90.40
3	90.33	90.44	89.11
4	90.13	90.46	89.07

At the same time, different window sizes will also affect the experimental results. The experimental procedure was tried three windows of 2, 3, and 4. Only when the window size is 2, the result is the best. This may be related to the fact that most Chinese words are 2 characters long. Corresponding to Table 5.

5 Conclusion

This paper mainly describes the design principle of the W-core layer and its effect in combination with different networks. For the Chinese word segmentation task, the Transformer model can achieve better results than the traditional RNN model and CNN model. It analyzed the essential goals of the Chinese word segmentation task, and proposed a specific optimization layer that can be used for this task. The experimental results show that, to a certain extent, the W-core layer can be added to the model structure to optimize the results. But still, there are many problems that need to be improved in the course of future experiments. For example, the model combining the W-core layer and the TextCNN layer, from the perspective of loss value, this model can achieve better experimental results. But the facts are not satisfactory, and later experiments will try to solve this problem.

By refining the needs of specific tasks, and then optimizing the model structure. This is the design idea of this paper. It hopes that on this basis, we can design a better and more robust model structure for the Chinese word segmentation task.

References

1. Jiang, T., Lu, Y., Zhang, J., Hong, J.: Application of unsupervised word segmentation algorithm in new word recognition. *Small Microcomput. Syst.* **41**(04), 888–892 (2020)
2. Zhang, D.Y., Cui, Z.J., Li, Y.X., Zhang, W., Lin, H.F.: Noun metaphor recognition based on transformer and BERT. *Data Anal. Knowl. Discov.* **4**(04), 100–108 (2020)
3. Yu, C., Wang, M., Lin, H., Zhu, X., Huang, T., An, L.: Comparative study of vocabulary representation models based on deep learning. *Data Anal. Knowl. Discov.* 1–19
4. Huang, D.: Chinese word segmentation and keyword extraction model based on deep learning. Master, Beijing University of Posts and Telecommunications (2019)

5. Ren, W., Xu, H.Y., Feng, S.L., Zhou, H., Shi, J.: Sequence annotation Chinese word segmentation based on LSTM network. *Comput. Appl. Res.* **34**(05), 1321–1324+1341 (2017)
6. Li, F., Jin, Y., Liu, W., Rawat, B.P.S., Cai, P., Yu, H.: Fine-tuning bidirectional encoder representations from transformers (BERT)-based models on large-scale electronic health record notes: an empirical study. *JMIR Med. Inf.* **7**(3), e14830 (2019)
7. Yang, B., Li, D., Yang, N.: Intelligent judicial research based on BERT sentence embedding and multi-level attention CNNs. In: Proceedings of 2019 2nd International Conference on Information Science and Electronic Technology (ISET 2019), p. 7. International Informatization and Engineering Associations, Taiyuan (2019)
8. Zhang, Z.R., Liu, Y.: based on BI-LSTM-CRF model of Chinese word segmentation. *J. Changchun Univ. Sci. Technol. (Nat. Sci. Edn.)* **40**(04), 87–92 (2017)
9. Mu, R.W.: Research and analysis of Chinese word segmentation based on GRU neural network and CRF. Master, South China University of Technology (2018)
10. SiChen, L.: A neural network based text classification with attention mechanism. In: Proceedings of IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT 2019), p. 6. IEEE Dalian Jiaotong University (Dalian Jiaotong University) Ritsumeikan University, Japan Shenyang University of Technology Liaoning Province Software Industry School-Enterprise Alliance, Dalian (2019)
11. Xiong, X., Yan, P.M.: Chinese classification method of combining multi-head self-attention mechanism. *Electron. Meas. Technol.* **43**(10), 125–130 (2020)
12. Huang, T., Deng, Z.-H., Shen, G., Chen, X.: A window-based self-attention approach for sentence encoding. *Neurocomputing* **375**, 25–31 (2020)
13. Guo, B., Zhang, C., Liu, J., Ma, X.: Improving text classification with weighted word embeddings via a multi-channel TextCNN model. *Neurocomputing* **363**, 366–374 (2019)

Computer Networks, Mobility and Pervasive Systems



Implementation and Evaluation of WBBR in ns-3 for Multipath Networks

Thalia Mijas-Abad¹, Patricia Ludeña-González^{1,2}, Francisco Sandoval¹, and Rommel Torres¹

¹ Departamento de Ciencias de la Computación y Electrónica,
Universidad Técnica Particular de Loja,
San Cayetano Alto s/n, 11-01-608, Loja, Ecuador
{tlmijas,pjludena,fasandoval,rovitor}@utpl.edu.ec

² Universidad Politécnica de Madrid, Madrid, Nikola Tesla s/n, 28031 Madrid, Spain
<http://www.utpl.edu.ec>
<http://www.upm.es>

Abstract. Today's networks carry a large amount of data, it is common for traffic to exceed the capacity of the network links, this event is called congestion. Therefore, the development of congestion control algorithms has become a key task in improving the performance of networks. BBR is a congestion control algorithm that continuously measures bottlenecks and RTT to establish the size of the congestion window. WBBR is an algorithm that incorporates a weighting factor to BBR to achieve fairness. The present work proposes an implementation of WBBR in ns-3 and performs a comparative evaluation with BBR since it has not been done before. The implemented module was validated with the previous results available in the literature. The results show that WBBR is fairer than BBR achieving load balancing, without increasing the RTT values.

Keywords: BBR · Computer networks · Congestion control · ns-3 · WBBR

1 Introduction

According to the International Telecommunication Union (ITU) currently, more than half of the world's population is connected to the internet [8]. The growth in demand and the number of connected devices increases the probability of congestion. Congestion occurs because network resources as memory queues in routers and bandwidth (BW) of the links are limited. Congestion control algorithms (CCAs) are currently a topic of research interest [9]. CCAs often use packet loss as an indicator of congestion [1]. However, other algorithms propose other congestion detection mechanisms such as Round-Trip propagation Time (RTT) values [19].

Traditionally Transmission Control Protocol (TCP) tries to mitigate congestion using mechanisms like Additive-Increase/Multiplicative-Decrease

(AIMD) [7], which decrease network performance. For this reason, CCAs are developed with various approaches, such as avoiding queue saturation, reducing packet discarding, and reducing the delay in data transmission [20]. CCAs can be reactive and proactive. Proactive algorithms, like B-Neck [14], UFA [2], and s-PERC [10], offer congestion control mechanisms with lower convergence time than reactive algorithms. However, these algorithms do not guarantee compatibility with applications with older TCP versions that are widely implemented up to now [21]. For this reason, in this article, we will focus on reactive algorithms, specifically Bottleneck Bandwidth and Round-trip propagation time (BBR) and Weighted BBR (WBBR).

A very important requirement for CCAs are the compatibility with the operation of TCP [17] and in the literature, we can find several algorithms that satisfy this requirement, for example wVegas [19], New Reno [7] and BBR [1]. Multipath TCP (MPTCP) is an emerging transport protocol that provides resistance to network failures and improves performance by splitting a data stream into multiple sub-streams across all available multiple paths [12]. However, [13, 22, 23] showed that MPTCP can lead to an injustice with regular TCP flows. For this reason, within CCAs there is a great variety of algorithms developed for MPTCP whose main objective is load balancing and fairness. For example, BELIA [22], which allows congestion control based on the BW estimate. However, these options focus primarily on fairness without guaranteeing good performance in most scenarios [13].

BBR achieves congestion control by keeping RTT values low. This characteristic has led to the proposal of modifications focused on the core of its algorithm to achieve its adaptability to specific environments [6, 11, 13, 20]. Of special interest is the WBBR [23] algorithm that is applied to a multipath environment and introduces weight factors to control the convergence rate. A lack in the literature is the comparative evaluation between BBR and WBBR to determine the level of improvement that exists.

Simulation is a valid tool for research development since it does not have the complexity or cost that a real implementation would imply [5]. For CCAs evaluation, ns-3 [16] is a popular network simulator because it allows adding dynamism in the input and output of flows and has TCP implementations [3]. In the literature there are several CCAs implementations in ns-3, for example: Scalable, Vegas, Veno, and Yeah [15], MPTCP [12], and BBR [4]. Finally, our module is to our knowledge the first implementation of WBBR in ns-3.

In this paper, we presented the design and implementation of WBBR in ns-3 based on the description available in the literature. It presents a comparison between the WBBR algorithm and its predecessor BBR. The results show WBBR achieves fairness and load balancing for multipath flows.

The rest of this paper is organized as follows. In Sect. 2 we review the congestion control and present the operation of BBR and WBBR. Section 3 gives the implementation details of our proposed module. In Sect. 4 we describe the metrics, scenarios and simulation results to compare the performance difference between BBR and WBBR. Finally, we concluded in Sect. 5.

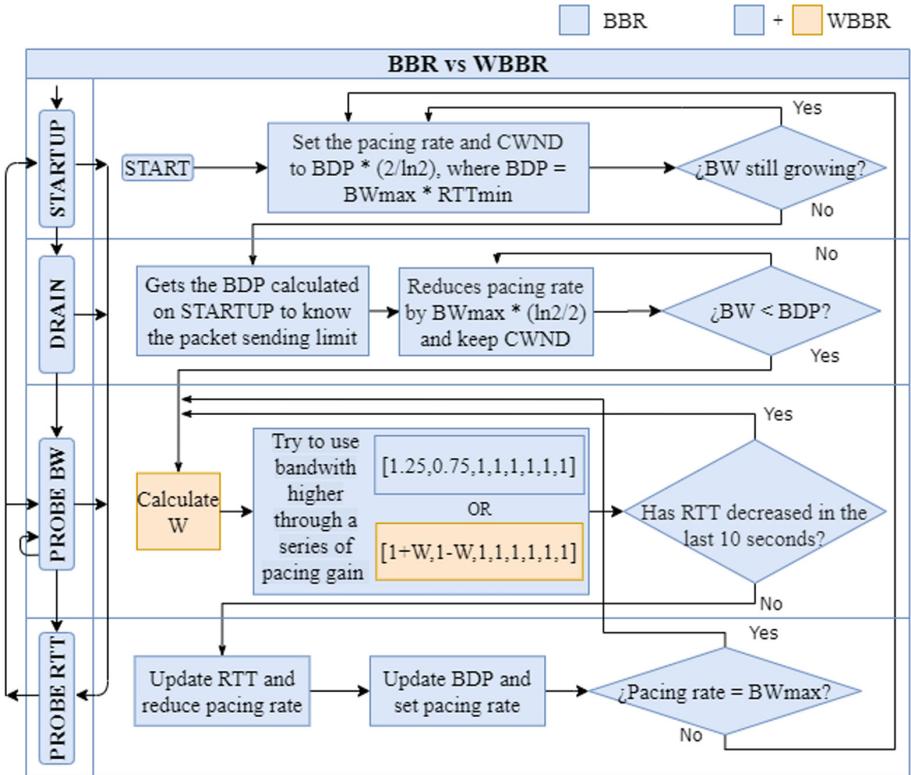


Fig. 1. BBR and WBBR operation based on [4,23]

2 Background

The high traffic of today's networks frequently generates congestion, whose effects are router queue saturation, packet discarding, and increased data transmission delay. CCAs are used to avoid or mitigate the congestion effects.

BBR is an algorithm developed by members of Google's "make-tcp-fast" project. It is based on two fundamental parameters: bottleneck BW and RTT. It continuously estimates these values, resulting in a distributed congestion control algorithm that reacts to actual congestion [1].

BBR maintains two windows that allow it to store the last ten values of both BW and RTT. Then, it selects the highest BW value and the lowest RTT value to calculate the number of bytes in flight to control the window growth and avoid the congestion. BBR has four transition states to perform congestion control: StartUp, Drain, Probe BW and Probe RTT. Figure 1 details the operation of BBR in each of its states.

WBBR is based on the BBR algorithm and uses the estimation of bottlenecks for multipath TCP [23]. The state machine used by BBR is maintained in

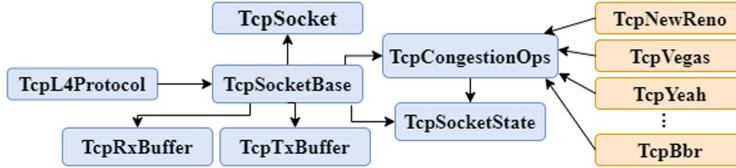


Fig. 2. TCP class diagram for ns-3 based on [16]

WBBR, but, it introduces a weighting factor w in `Probe` BW state to control the convergence BW of each subflows, see Fig. 1.

To ensure balancing load and fairness, WBBR uses a weighting function $f(w_i) = \frac{C_i}{A_i w_i + 1}$. Where C_i the capacity of the bottleneck link, w_i the weight factor for the i th subflow, x_i the BW convergence for the i th subflow, and A_i depends on the subflows sharing the same bottleneck and is computed with $A_i = \sum_{j=1, j \neq i}^N \frac{1}{w_j}$. WBBR complies with the Congestion Equality Principle by maintaining $\frac{C_i - x_i}{A_i} = \sum_{i=1}^N x_i$. This algorithm has been implemented and evaluated in MATLAB.

3 Implementation

WBBR algorithm is implemented re-using existing BBR code. A BBR version was implemented in ns-3.27 by Claypool [4], the code is available in GitHub¹. Figure 2 shows the classes used to CCA implementation in TCP module in ns-3. The BBR model is based on theses classes. The BBR model includes `tcp-bbr.cc`, `ttcp-bbr.h`, `tcp-bbr-state.cc` and `tcp-bbr-state.h`. Additionally, this implementation uses the following classes: `TcpCongestionOps`, `TcpSocket`, `TcpSocketBase`, `TcpSocketState`, `TcpL4Protocol`, `TcpRxBuffer`, `TcpTxBuffer`, and `RttEstimator`.

In Fig. 1 we showed WBBR has the same states as BBR. Thus, the model adaptation focuses on the `Probe` BW state. The proposed adaptation includes a new method in `tcp-bbr.cc` to compute w , named `getW()`.

Also, in `tcp-bbr.cc` the `getW()` method is added, which includes the Zhu code adaptation for ns-3 and it will be used in the state machine. In order to keep the w values between $\{0,1\}$ and for compatibility with BBR implementation, $w_0=0.25$ (`bbr::PROBE_FACTOR`) and $w_u=0.95$ (`bbr::PACING_FACTOR`). In line 7 in the method `CalcutateW()` we replicate the weight-bandwidth model proposed in [23]. The equations from Sect. 2 are added in the `TcpBbr::PktsAcked()` method (Fig. 3).

Finally, in `BbrProbeBWState::execute()` method of the `tcp-bbr-state.cc` file is replaced the value of `bbr::PROBE_FACTOR` and `bbr::DRAIN_FACTOR` by the value obtained in `getW()`.

¹ <https://github.com/mark-claypool/bbr>.

```

1 double TcpBbr::getW() const {
2     Time now = Simulator::Now();
3     double w0 = bbr::PROBEFACTOR;
4     double wU = bbr::PACINGFACTOR;
5     double x,w,BBP,BTP,xp,wp;
6     x = getBW();
7     w = CalculateW();
8     BTP = (xp-x)/((x*w)-(xp*wp));
9     BBP = ((xp*x*(wp-w))/((x*w)
10      -(xp*wp)));
11     double bw_mp = Sum of BWs;
12     double st_sz = 0.1;
13     double bw_bg = (4*(BBP-x))/(BTP);
14     if(bw_mp < bw_bg){w = (w0*w)/(w0+w*st_sz)};
15     else if(bw_mp > bw_bg){w = (w0*w)/(w0-w*st_sz)};
16     if(w < w0){w = w0};
17     if(w >= 1){w = wU};
18     return w;
}

```

Fig. 3. Algorithm used to calculate w factor based on [23]

For the validation of the proposed module, the scenario available in [23] is implemented and the results for multipath flow are evaluated. The values obtained for the given intervals have a desirable behavior similar to the results in the literature, therefore the functionality of our implementation in ns-3 is valid.

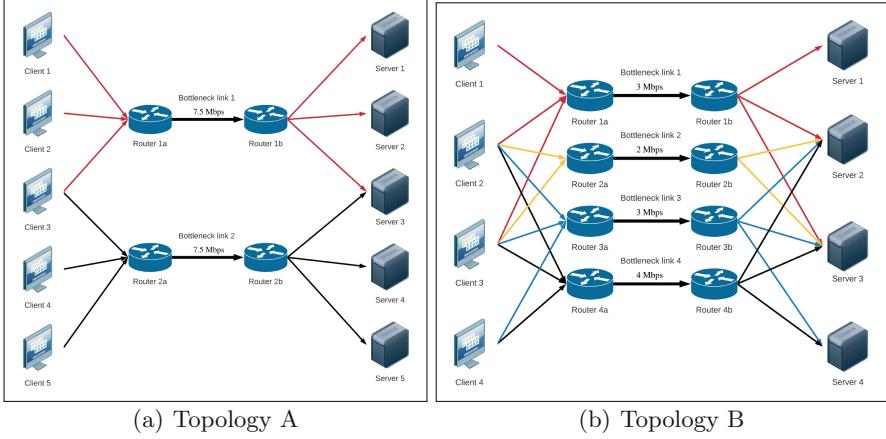
4 Results

In this section, we present the results obtained from both BBR and WBBR algorithms.

The metrics used to evaluate the performance of the proposed algorithm are Throughput, RTT and Gini Coefficient to measure the fairness level.

- **Throughput:** It is the BW at which each session can transmit data. It is calculated by dividing the total received bits by the total simulation time. The utilization is the ratio between the sum of the throughput of each session and the total BW that can be carried through the whole network.
- **RTT:** It is the time requires for a data packet to be sent from a source to a specific destination plus the time it takes for an ACK to be sent back. It includes propagation delay and processing time.
- **Gini coefficient:** This metric is a value between 0 and 1, the lower values being more equitable than the higher ones [18].

There are two network topologies used in the simulations which are specified in Fig. 4. Topology A, from Zhu et al. [23], has two bottlenecks, and Topology B has 4 bottlenecks. The simulation parameters are in Table 1.

**Fig. 4.** Topologies for scenarios used in simulations**Table 1.** Simulation parameters for scenarios

Parameter	Scenario 1	Scenario 2	Scenario 3
Topology	A	B	A
Number of nodes	14	16	14
Number of flows	5	4	4
Queue size [packets]	100	100	100
Simulation time [s]	300	50	50
End links capacity [Mbps]	60	60	60
Bottleneck 1 capacity [Mbps]	7.5	3	7.5
Bottleneck 2 capacity [Mbps]	7.5	2	7.5
Bottleneck 3 capacity [Mbps]	—	3	—
Bottleneck 4 capacity [Mbps]	—	4	—

Scenario 1. It models the dynamism of in/outflows, thus, the flows 1–5 are active in the interval 50–100, 0–300, 0–250, 0–300, and 150–200 seconds, respectively. The results for this scenario are shown in Fig. 5. As we can see in Fig. 5(a) flow 3 achieves greater BW using WBBR, on the contrary in Fig. 5(b) it is noted that WBBR controls the growth of this flow to assign more BW to flows 2 and 4. WBBR has a Gini coefficient of 0.0674 for the intervals 50–100 and 150–200 s versus the value of 0.1684 for BBR in the same intervals, thus WBBR is fairer without significantly reducing network utilization. Additionally, WBBR achieves to keep RTT levels low, at values similar to those achieved by BBR (see Fig. 5(c) and Fig. 5(d)).

Analyzing the throughput for the multipath flow, which has two subflows (Fig. 6), BBR gets the same BW for both subflows. While WBBR in the interval

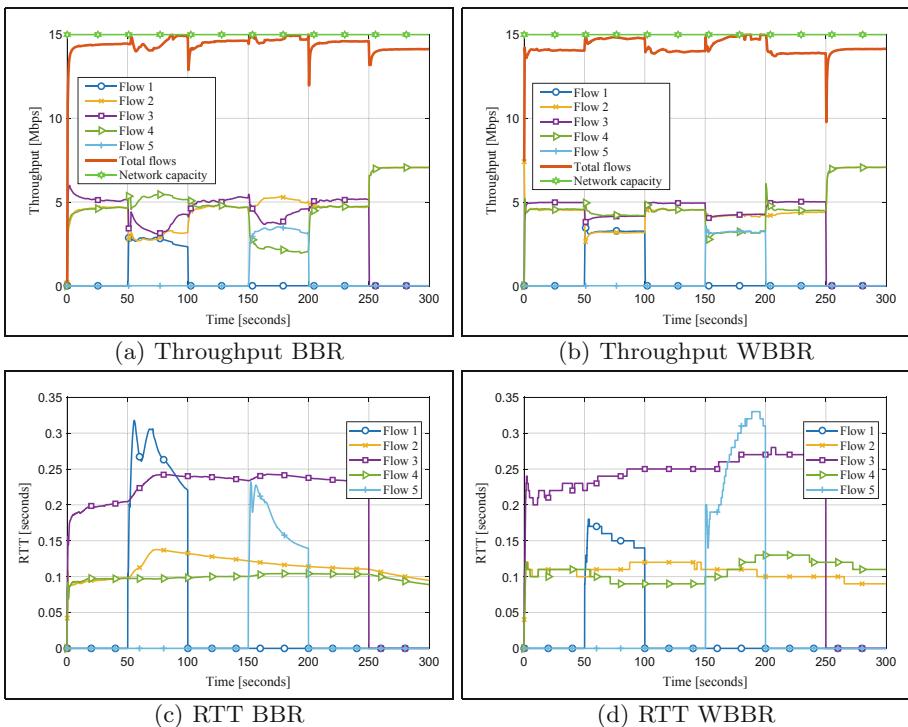


Fig. 5. Throughput and RTT of BBR and WBBR in scenario 1

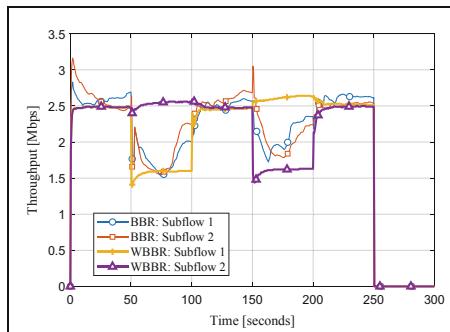


Fig. 6. Throughput BBR and WBBR for subflows for flow 3 in scenario 1

50–100 s achieves less BW for subflow 1, because flow 1 shares bottleneck 1 with it, and in the 150–200 s interval subflow 2 has less BW because flow 4 shares the capacity of bottleneck 2 with it. This shows WBBR performs load balancing.

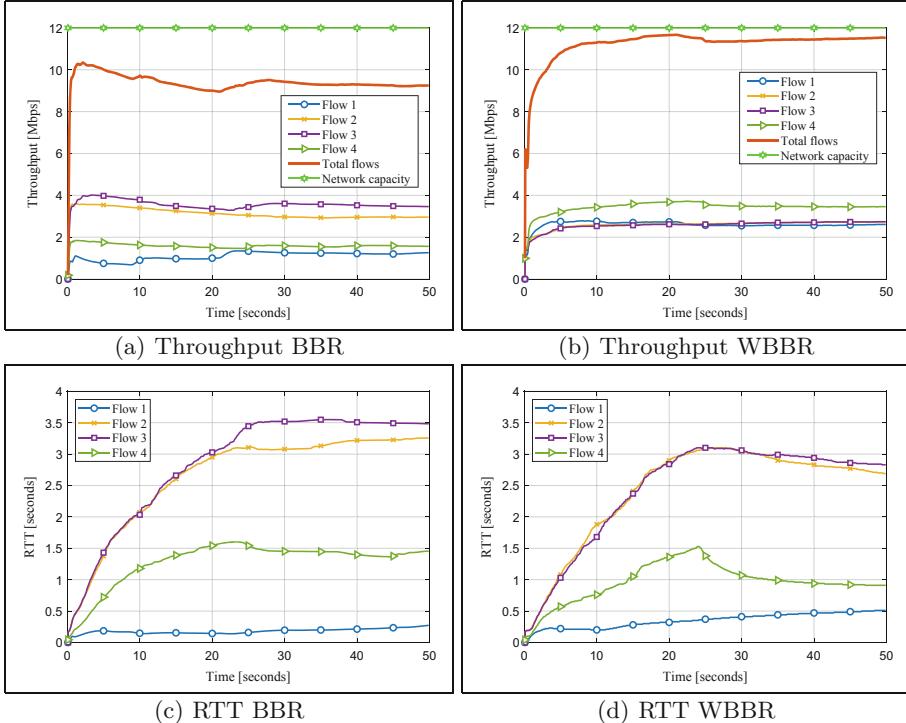


Fig. 7. Throughput and RTT of BBR and WBBR in scenario 2

Scenario 2. It aims to evaluate the behavior of the algorithms when there is more than one multipath flow, thus, flow 1 is singlepath, flows 2 and 3 have four subflows, and flow 4 has two subflows. It can be seen in Fig. 7(a) BBR gets more BW for flows 2 and 3 because they have more subflows than flows 1 y 4, therefore its Gini coefficient is high 0.2159. On the other hand, WBBR balances the BW of all the flows (Fig. 7(b)) obtaining a Gini coefficient of 0.0556, with an increase in the total throughput of the network from 9 Mbps to 11.4 Mbps which is equivalent to 78.3% and 95.8% respectively. Figure 7(c) shows the RTTs obtained for BBR, the average value is 8.47 s, while, WBBR has an average RTT of 6.94 s, see Fig. 7(d).

Scenario 3. It is used to evaluate the weight factor impact in the WBBR performance, thus, all flows are active during the simulation time. WBBR bases its operation on the value w_0 , where $0 < w_0 < 1$. In this scenario, we evaluate the impact that the selection of w_0 has on network utilization. Table 2 shows the throughput and utilization values for different values of w_0 . Thus we can see that when the value of w_0 decreases, the use of the network also decreases because w_0 reflects the degree of greed for BW.

Table 2. Network utilization in scenario 3

Weighting factor	Throughput [Mbps]	Utilization [%]
$w_0 = 0.250$	14.7	98.6
$w_0 = 0.125$	14.4	96.6
$w_0 = 0.083$	13.9	93.0

5 Conclusions

We presented the congestion problem and some approaches to congestion control, focusing on the BBR and WBBR algorithms. We described the implementation of WBBR protocol in network simulator ns-3. Our proposed module was validated with the results presented in the paper where WBBR was proposed. Therefore, it can be used as a tool to evaluate the performance of other algorithms or to explain the congestion control and load balancing concepts.

In order to verify the improvements proposed by WBBR about BBR, the results were compared and discussed. The used metrics were throughput, RTT, and Gini coefficient. The results show that WBBR achieves a fair allocation of BW, without increasing the RTT values. Furthermore, since WBBR assigns the new flows added to the network to the less congested paths, it achieves load balancing in the whole network.

References

- Cardwell, N., Cheng, Y., Gunn, C.S., Yeganeh, S.H., Jacobson, V.: BBR: congestion-based congestion control. *Queue* **14**(5), 20–53 (2016)
- Carrión, J., Ludeña-González, P., Sandoval, F., Torres, R.: Evaluation of utility function algorithm for congestion control in computer networks. In: *Information and Communication Technologies*, pp. 453–467. Springer (2020). https://doi.org/10.1007/978-3-030-62833-8_33
- Casoni, M., Patriciello, N.: Next-generation TCP for ns-3 simulator. *Simul. Model. Pract. Theory* **66**, 81–93 (2016). <https://doi.org/10.1016/j.simpat.2016.03.005>
- Claypool, M., Chung, J.W., Li, F.: BBR'. In: *Proceedings of the 10th Workshop on ns-3 - WNS3 2018*, pp. 1–8. ACM Press, New York (2018). <https://doi.org/10.1145/3199902.3199903>. <http://dl.acm.org/citation.cfm?doid=3199902.3199903>
- Coudron, M., Secci, S.: An implementation of multipath TCP in ns3. *Comput. Netw.* **116**, 1–11 (2017). <https://doi.org/10.1016/j.comnet.2017.02.002>
- Grazia, C.A., Patriciello, N., Klapez, M., Casoni, M.: BBR+: improving TCP BBR performance over WLAN. In: *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, vol. 2020-June, pp. 1–6. IEEE (2020). <https://doi.org/10.1109/ICC40277.2020.9149220>. <https://ieeexplore.ieee.org/document/9149220/>
- Idwan, H., Ihsanuddin, I.: Analisis optimasi round trip time (RTT) pada Jaringan transmission control protocol (TCP) new reno. *Jurnal JTIK (J. Teknol. Inf. Komunikasi)* **4**(2), 92 (2020). <https://doi.org/10.35870/jtik.v4i2.143>
- ITU: Informe sobre la medición de la sociedad de la información 2018. Technical report, International Telecommunication Union (2018). <https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2018/MISR2018-ES-PDF-S.pdf>

9. Jiang, X., Zhu, J., Jin, G.: DPC: a delay-driven prioritized TCP congestion control algorithm for data center networks. In: International Conference on Frontiers in Cyber Security, pp. 459–474. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-9739-8_35
10. Jose, L., Ibanez, S., Alizadeh, M., McKeown, N.: A distributed algorithm to calculate max-min fair rates without per-flow state. In: Proceedings of the ACM on Measurement and Analysis of Computing Systems, vol. 3, no. 2, pp. 1–42 (2019)
11. Kim, G.H., Cho, Y.Z.: Delay-aware BBR congestion control algorithm for RTT fairness improvement. IEEE Access **8**, 4099–4109 (2020). <https://doi.org/10.1109/ACCESS.2019.2962213>. <https://ieeexplore.ieee.org/document/8943219/>
12. Luzuriaga-Jiménez, J., Torres-Tandazo, R., Ludeña-González, P., Rohoden-Jaramillo, K.: MPTCP multipath protocol evaluation in packet networks. In: International Conference on Applied Technologies, pp. 69–81. Springer (2019)
13. Mahmud, I., Lubna, T., Song, Y.J., Cho, Y.Z.: Coupled multipath BBR (C-MPBBR): a efficient congestion control algorithm for multipath TCP. IEEE Access **8**, 165,497–165,511 (2020). <https://doi.org/10.1109/ACCESS.2020.3022720>
14. Mozo, A., López-Presa, J.L., Fernández Anta, A.: A distributed and quiescent max-min fair algorithm for network congestion control. Expert Syst. Appl. **91**, 492–512 (2018). <https://doi.org/10.1016/j.eswa.2017.09.015>
15. Nguyen, T.A.N., Gangadhar, S., Rahman, M.M., Sterbenz, J.P.: An implementation of scalable, Vegas, Veno, and YeAH congestion control algorithms in ns-3. In: Proceedings of the Workshop on Ns-3, WNS3 2016, p. 17–24. Association for Computing Machinery, New York (2016). <https://doi.org/10.1145/2915371.2915386>
16. Nsnam: TCP models in ns-3 – Model Library. <https://www.nsnam.org/docs/models/html/tcp.html>. Accessed 22 July 2020
17. Raiciu, C., Handley, M., Wischik, D.: Coupled congestion control for multipath transport protocols. Technical report, IETF RFC 6356, October 2011. <https://doi.org/10.17487/rfc6356>. <https://www.rfc-editor.org/info/rfc6356>
18. Sitthiyot, T., Holasut, K.: A simple method for measuring inequality. Palgrave Commun. **6**(1), 1–9 (2020)
19. Cao, Y., Xu, M., Fu, X.: Delay-based congestion control for multipath TCP. In: 2012 20th IEEE International Conference on Network Protocols (ICNP), pp. 1–10. IEEE (2012). <https://doi.org/10.1109/ICNP.2012.6459978>
20. Zhang, S., Lei, W., Zhang, W., Guan, Y., Li, H.: Congestion control and packet scheduling for multipath real time video streaming. IEEE Access **7**, 59758–59770 (2019). <https://doi.org/10.1109/ACCESS.2019.2913902>
21. Zhang, T., Huang, J., Chen, K., Wang, J., Chen, J., Pan, Y., Min, G.: Rethinking fast and friendly transport in data center networks. IEEE/ACM Trans. Netw. **28**(5), 2364–2377 (2020). <https://doi.org/10.1109/tnet.2020.3012556>
22. Zhu, M., Wang, L., Qin, Z., Ding, N., Fang, J., Liu, T., Cui, Q.: BELIA: bandwidth estimate-based link increase algorithm for MPTCP. IET Netw. **6**(5), 94–101 (2017). <https://doi.org/10.1049/iet-net.2016.0102>. <https://digital-library.theiet.org/content/journals/10.1049/iet-net.2016.0102>
23. Zhu, T., Qin, X., Chen, L., Chen, X., Wei, G.: wBBR: a bottleneck estimation-based congestion control for multipath TCP. In: 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), vol. 2018-August, pp. 1–5. IEEE (2018). <https://doi.org/10.1109/VTCFall.2018.8690919>



Trends, Challenges and Opportunities for IoT in Smallholder Agriculture Sector: An Evaluation from the Perspective of Good Practices

C. Alexandra Espinosa¹(✉) Jhon Pineda¹(✉) Oscar Ortega¹(✉) Astrid Jaime Author²(✉) Román Sarmiento¹(✉) and George Washington Archibald Taylor³

¹ Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia
{mespinosa3,jduarte890}@unab.edu.co

² Universidad El Bosque, Bogotá, Colombia
majaime@unbosque.edu.co

³ Del Laboratorio al Campo, Bogotá, Colombia

Abstract. In addition to being considered as a food pantry by different entities, such as the FAO, the Colombian agriculture industry represents 6.7% of the country's GDP and 7.4% of total exports. This shows a clear inclination towards agriculture and the urgent demand for technicizing the countryside. Even more when considering the increase of almost 70% in food production, projected for 2050, to provide food for the people who will inhabit the world. However, even with government efforts, work in the small-scale Colombian agricultural sector is rudimentary and traditional. This leads to land degradation and low production, which prevents small farmers from competing in national and international markets on equal terms. This paper proposes a review of the literature using the DANDELION methodology, focused on the use of IoT in agriculture since this technology has proven to have interesting results in increasing food production by optimizing the amount of land, water, and other resources used by the agricultural sector. Evaluating each of the approaches in the context of good practices in IoT in agriculture, by comparing the definitions provided by the ITU, FAO, OECD, and World Bank since they are considered as recognized organizations offering their respective definitions of good practices both in IoT and in the agricultural sector. This work establishes a basis for standardizing the development of IoT solutions in agriculture at the national level and the state of IoT solutions in farming and several challenges to be solved in small-scale agriculture.

Keywords: Good practices · IoT · Agriculture · Smallholders · Literature review

1 Introduction

Talking about the Internet of Things (IoT) is to recognize the innovative technological advances focused on making people's lives simpler on any of their aspects [1]. Rural

lifestyles are no exception. From farming alternatives, going through the enhancement of productivity [2], to architectures designed for precision agriculture [3], the inevitable inclusion of these tools in day-to-day operations is evident by considering that the size of the IoT market in agriculture. According to the Allied Market Research portal [4], this market was valued at \$16,330 million in 2017. It could reach \$48,714 million by 2025, growing at a CAGR of 14.7% from 2018 to 2025.

Nevertheless, the introduction of IoT in Colombia has been a slow process: The scenario for 2018 (18 years after IoT term emerged [1]) revealed that, although 13% of the companies in the country had adopted these technologies, 56.3% of the entrepreneurs were unaware of the concept. A strategic plan for 2022 was established, projecting that 50% of the Colombian companies would depend on these technologies in their production chains [5]. Despite problems related to food security¹, the lack of infrastructure, and sufficient knowledge about the topic makes it impossible to achieve the established objectives: Colombia's first (and only) IoT laboratory ignores innovations focused on the smallholder farmers [6], which represents a problem when considering that small-scale agriculture is responsible for 15.8% of employment in Colombia [7].

Based on the current solutions (models, prototypes, applications, and platforms), a methodology is required to identify IoT practices better suited for the Colombian rural context, characterized by small plots and low income. For this reason, this research focuses on defining good IoT practices in agriculture, which will enable the evaluation of different IoT solutions described in selected papers to provide a review of the current strategies, trends, challenges, and barriers identified for their implementation.

This document has three segments: first, it presents the methodology used to define good IoT practices in agriculture, the second introduces a discussion about the findings from a classification phase, and for the last part, it presents recommendations for future work related to the findings of this research procedure.

2 Methodology

The research procedure uses a two-stage methodology: (1) developing the concept of good practices on IoT in agriculture: Using the definitions provided by globally recognized organizations concerning outstanding IoT practices - from World Bank and the ITU as references- and agriculture -from OECD, FAO, and World Bank-, identify the necessary aspects to be considered as a good practice. Then, compare them to find similarities and differences. Table 1 shows the result of this comparison.

This comparison allows to propose the following definition: “Good practices in IoT solutions for agriculture are those that satisfy the following criteria:

1. Effective and successful: Refers to the achievement of the objective of the solution proposed generating a positive impact and reliability concerning the agriculture development using IoT solutions.
2. Sustainable: Satisfies current requirements without compromising the ability to face future needs. It can refer to the scalability of the embedded IoT solution in the agricultural environment.

¹ In 2016, Colombia had 43% of its inhabitants in chronic malnutrition [7].

Table 1. Criteria of definitions of good IoT practices in agriculture

Criteria	FAO [1]	ITU [2]	World Bank-BPA [3]	World Bank-IoT [5]	OCDE [6]
Effective and successful	x	x	x	x	x
Sustainable	x	x	x		x
Gender sensitive	x		x	x	x
Technically possible	x	x		x	x
Resulting from a participatory process	x		x	x	x
Replicable, flexible, and adaptable	x	x		x	x
Reduction of risks of natural disasters	x	x			x
Innovative		x	x		x
Transversality - integrity		x		x	x
Systematization of experiences		x		x	x
Data use				x	x

3. Gender-sensitive: A solution that benefits a target audience, by understanding the role of women in agricultural technification².
4. Technically possible: easy to learn and apply within the community with materials, instruments, equipment, and human capital available in nearby urban areas.
5. Replicable and adaptable: Has a potential for replication and adaptability to agricultural environments, and;
6. Reduces the risk of disasters or crises by promoting their mitigation: Referring to catastrophes related to climate change.”³

For the identification of IoT solutions in agriculture, the DANDELION methodology [8] is used, which employs the benefits of systematic literature review and is complemented by bibliometric techniques that encourage the approach to scientific research. The methodological process for scientific article selection employed tools such as Publish or Perish and VOSViewer, which were used to identify the research context, inclusion and exclusion criteria, search equation selection (<< architecture AND (IoT OR “Internet of Things”) AND agriculture > >) and search database selection (private databases: IEEE, ACM, ScienceDirect, Scopus, Springer Link; and public databases: DIMENSION and LENS). Figure 1 shows the inclusion and exclusion criteria selection, quality, and study of articles concerning IoT solutions used in agriculture.

² According to the 2013 national agricultural census, in rural areas, 60% of peasant women complete sixth grade, compared to 50% of peasant men who reach fifth grade [21].

³ In 2016, climate change was the principal cause of natural disasters in Colombia [22].

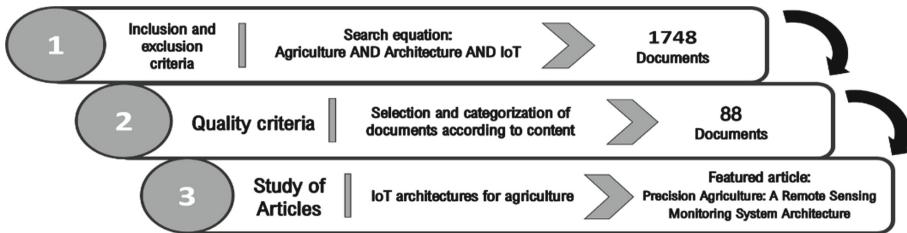


Fig. 1. DANDELION methodological process for selection of scientific papers

3 Analysis of Results

As a result, eighty-eight papers represent the IoT trends used in agriculture. They are categorized according to six criteria established as evaluation tools (see Table 2). For this purpose, each one of the scientific articles was classified according to the information provided concerning the definition of good IoT practices in agriculture indicated above related to the project's objective in the agriculture sector.

Table 2. Categorization of papers by tendency

Label	Description	Definition	Cant
GIP	Good IoT practices	Documents with relevant information that articulates the criteria of good practices developed, in addition to its implementation	7
IPI	IoT pilot initiatives	Documents containing IoT designs in TRL1 stage	10
IPO	IoT practices in another sector	Documents with IoT implementations with significant contributions in agriculture, but not implemented for the agricultural sector	11
IPN	IoT practices not aimed to the project	Documents with IoT designs in TRL1-TRL2 stage, which have been selected for their theoretical and technical contribution	37
NP	IoT relevant information	Documents that may contain IoT implementation and have relevant information on projects related to ICT in agriculture	6
RGP	Reference for the elaboration of good practices	Documents with both theoretical and practical content whose considerations must be implemented when a IoT solution in agriculture is developed	11

From this classification, given their potential as a reference in the IoT design of solutions for the agricultural sector, RGP label documents were selected. By developing analysis in these articles, specific topics were identified that respond to similarity and difference between them (see Table 3).

Table 4 presents the characterization process performed by the authors. Topics A, B, C, E, and H are the most developed the selected papers, while embedded systems with less frequency of appearance, related to IoT solutions in agriculture.

Table 3. Labels used for the characterization of documents.

Characteristics	Label
Contains an architecture related to agriculture	A
Contains definitions related to IoT technologies	B
Contains information related to sensorics	C
Contains information related to embedded systems	D
Contains information related to communication technologies	E
Contains information related to communication protocols	F
Contains information related to IoT implementations in agriculture	G
Contains information related to IoT platforms	H

Table 4. Characterization of the selected papers

Authors	Ref.	Characteristics							
		A	B	C	D	E	F	G	H
M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour and M. Aggoune	[9]	x	x	x		x		x	x
M. Farooq, S. Riaz, A. Abid, K. Abid, And M. Naeem	[10]	x	x	x		x	x	x	x
M. Abbasi, M. Yaghmaee, And F. Rahnama	[11]	x	x	x	x	x			x
M. Sheikhalishahi, A. R. Biswas, And T. Bures	[12]	x	x	x	x	x	x		x
A. Triantafyllou, P. Sarigiannidis, And S. Bibi	[13]	x	x	x	x	x	x	x	x
J. R. Celestrini, R. N. Rocha, E. B. Saleme, C. A. S. Santos, J. G. Pereira Filho, And R. V. Andreão	[14]	x	x	x	x	x	x	x	x
C. Verdouw, H. Sundmaeker, B. Tekinerdogan, D. Conzon, And T. Montanaro	[15]	x	x	x		x	x	x	x
K. Yelamarthi, Md. S. Aman, And A. Abdalgawad	[16]	x	x	x	x	x	x	x	x
P. Lavanya, And R. Sudha	[17]	x	x	x	x	x	x		x
M. C. Vuran, A. Salam, R. Wong, And S. Irmak	[18]	x	x	x		x			x
H. Cadavid, W. Garzón, A. Pérez, G. López, C. Mendivelso, And C. Ramírez	[19]	x	x	x		x	x	x	x

4 Discussion

By performing a co-occurrence analysis based on data text on the selected articles using VOSViewer, two trends are identified: one focused on the implementation potential of IoT in the agricultural sector to increase productivity and improve crop and cultivation processes (red cluster), and another one, the green cluster (which also sets a trend in 2019, as shown in the overlay visualization map of Fig. 3), that recognizes the difficulties of implementation. The most common technological topics are cloud computing, generic reference architecture, smart farming, energy consumption restriction, and important non-functional (see Fig. 2).

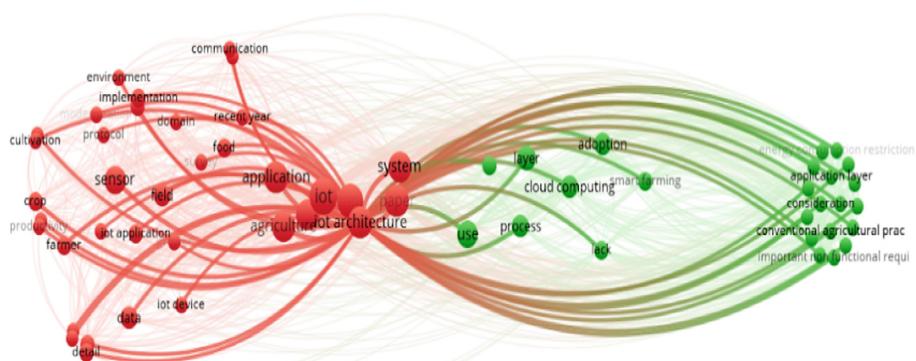


Fig. 2. Trend analysis from eleven selected documents using VOSViewer

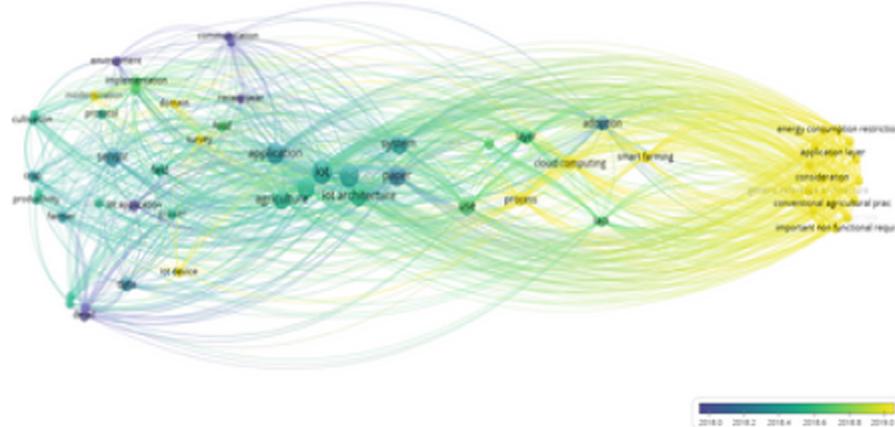


Fig. 3. Time-trend analysis from the eleven selected documents using VOSViewer

Table 4 shows the result of an exhaustive analysis of each of the documents, determining their potential for the development of an IoT solution based on good practices.

The articles identified are those that contained all the features established in Table 3 to be the core of the research.

Taking into account the proposed definition of good IoT practices in agriculture, the authors consider it necessary to focus on the development of frameworks that aim to guide the process of designing and implementing IoT solutions in agricultural areas considering the restriction of energy consumption [17], the implementation of a sensor layer and the redefinition of the network, service and application layers [13] allowing modularity depending on the system needs [16].

However, the interaction of small farmers and the system is not evident since they are the primary users of the infrastructure. Although there are documents that are concerned with the design of low-cost solutions [12, 17] that meet the socio-economic needs of the population, the use of outdoor devices [10], and the lack of Internet connections in remote areas [10], these reviews are focused on identifying hardware [11, 18] and software [10, 19] elements that are not sufficiently accessible for developing countries, as well as other technological elements for which implementation is difficult (for example the use of IPv6). Also, these architectures (that mostly use the OSI model) do not consider human factors as a systematized element that allows, through expertise, to aggregate more information from the soils in the physical layer of the architecture. Therefore, the human perspective in farmers, technicians, and agricultural engineers, provides the opportunity to validate the acquired sensor data; or even supply it when sensors lack the characteristic to provide it [20].

There is also a lack of analysis of interoperability among the different elements that compose the layer in the architectures proposed by the selected authors, or their relationship with the set of layers suggested, which may serve as a useful decision-making tool, especially when the development budget is low so that decisions are entirely at the discretion of the designer's expertise.

Finally, some documents identify the importance of apps focused on agriculture [9] about the possibilities of available IoT solutions [19]. However, no one indicates the establishment of interfaces or projects focused on encouraging mental models based on rural communities, or studies conducted about the most effective way to provide information about the architecture that would allow local people to be able to make optimal decisions.

5 Conclusions and Future Work

With the gradual implementation of IoT in the agricultural sector, the tendency in academic circles has focused on developing architectures in agricultural environments that support their proposed solutions, such as projects aimed at crops monitoring through the web or mobile platforms, whether in an area that belongs to a small farmer or even a more industrial environment. However, their incorrect deployment may lead to unforeseen difficulties, given the lack of awareness about the needs of a long-neglected sector such as agriculture (intermittent Internet connection, that becomes non-existent as the area gets farther away from the urban area, a high percentage of technological illiterates and deficient public services [19]).

The evaluation of the proposed selected solutions from the perspective of good practices reveals a series of challenges that need a better understanding to establish viable

alternatives available to the communities to modernize their farming and production techniques. This conceptualization formed the basis for the design of an IoT architecture.

References

1. FAO: Buenas prácticas en la FAO: Sistematización de experiencias para el aprendizaje continuo. 13:12 (2013)
2. Sanou, B., Grindeanu, S.: GSR-18 best practice guidelines on new regulatory frontiers to achieve digital transformation. In: ITUGSR, Geneva 2018, pp. 1–8 (2018)
3. Mundial, B.: Alcanzando a los pobres rurales. Nueva estrategia de desarrollo rural (2001)
4. Kadam, A.: IOT in Agriculture Market (2017). <https://www.alliedmarketresearch.com/internet-of-things-iot-in-agriculture-market>. Accessed 12 Dec 2020
5. Group, W.B.: Internet of Things: the new government to business platform: a review of opportunities, practices, and challenges. Lecture Notes Computer Science (including Subseries Lecture Notes Artificial Intelligence Lecture Notes Bioinformatics), vol. 7768, pp. 257–282 (2017). <https://doi.org/10.1007/978-3-642-41569-2-13>
6. OECD: The Internet of Things - seizing the benefits and addressing the challenges. In: OECD Digital Economy Papers, pp. 4–11. OECD Publishing, Paris (2016)
7. Food and Agriculture Organization of the United Nations (FAO), World Food Programme (WFP): Monitoring food security in countries with conflict situation (2016)
8. Espinosa, C.M.A., Romero, R.E., Flórez, G.L.Y., Guerrero, C.D.: DANDELION: Propuesta metodológica para recopilación y análisis de información de artículos científicos. Un enfoque desde la bibliometría y la revisión sistemática de la literatura. RISTI. Iber J. Inf. Syst. Technol. 110–123 (2020). <https://search.proquest.com/openview/e3b85a7260c758fd943bc4d5a0447f13/1?pq-origsite=gscholar&cbl=1006393>
9. Ayaz, M., Ammad-Uddin, M., Sharif, Z., et al.: Internet-of-Things (IoT)-based smart agriculture: toward making the fields talk. IEEE Access 7, 129551–129583 (2019). <https://doi.org/10.1109/ACCESS.2019.2932609>
10. Farooq, M.S., Riaz, S., Abid, A., et al.: A survey on the role of IoT in agriculture for the implementation of smart farming. IEEE Access 7, 156237–156271 (2019). <https://doi.org/10.1109/ACCESS.2019.2949703>
11. Abbasi, M., Yaghmaee, M.H., Rahnama, F.: Internet of Things in agriculture: a survey. In: Proceedings of 3rd International Conference on Internet Things Applications IoT 2019, pp. 1–12 (2019). <https://doi.org/10.1109/IICITA.2019.8808839>
12. Dupont, C., Sheikhalishahi, M., Biswas, A.R., Bures, T.: IoT, big data, and cloud platform for rural African needs. In: 2017 IST-Africa Week Conference on IST-Africa 2017, pp. 1–7 (2017). <https://doi.org/10.23919/ISTAFRICA.2017.8102386>
13. Triantafyllou, A., Sarigiannidis, P., Bibi, S.: Precision agriculture: a remote sensing monitoring system architecture. Information 10, 348 (2019). <https://doi.org/10.3390/info10110348>
14. Celestrini, J.R., Santos, C.A.S., Rocha, R.N., et al.: An architecture and its tools for integrating IoT and BPMN in agriculture scenarios. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing Part F1477, pp. 824–831 (2019). <https://doi.org/10.1145/3297280.3297361>
15. Verdouw, C., Sundmaeker, H., Tekinerdogan, B., et al.: Architecture framework of IoT-based food and farm systems: a multiple case study. Comput. Electron Agric. 165, 104939 (2019). <https://doi.org/10.1016/j.compag.2019.104939>
16. Yelamarthi, K., Aman, M.S., Abdelgawad, A.: An application-driven modular IoT architecture. Wirel. Commun. Mob. Comput. (2017). <https://doi.org/10.1155/2017/1350929>

17. Lavanya, P., Sudha, R.: A study on WSN based IoT application in agriculture. In: Proceedings of 3rd International Conference on Communication Electronic System ICCES 2018, pp. 1046–1054 (2018). <https://doi.org/10.1109/CESYS.2018.8724020>
18. Vuran, M.C., Salam, A., Wong, R., Irmak, S.: Internet of underground things: sensing and communications on the field for precision agriculture. In: IEEE World Forum Internet Things, WF-IoT 2018 - Proceedings 2018-January, pp. 586–591 (2018). <https://doi.org/10.1109/WF-IoT.2018.8355096>
19. Cadavid, H., Garzón, W., Pérez, A., et al.: Towards a smart farming platform: from IoT-based crop sensing to data analytics. *Commun. Comput. Inf. Sci.* **885**, 237–251 (2018). https://doi.org/10.1007/978-3-319-98998-3_19
20. Rossel, R.A.V., Adamchuk, V.I., Sudduth, K.A., et al.: Proximal Soil Sensing : An Effective Approach for Soil Measurements in Space and Time, 1st edn. Elsevier Inc. (2011)
21. DANE: 3er Censo Nacional Agropecuario (2016)
22. Francescutti, C., IFAD: Investing in rural people in Colombia, Colombia (2016)



Optical Wireless Communication Applications and Progress to Ubiquitous Optical Networks

Simona Riurean¹ , Monica Leba¹ , Andreea Ionica¹ , and Álvaro Rocha²

¹ University of Petrosani, 332006 Petrosani, Romania

{simonariurean, monicaleba, andreeaionica}@upet.ro

² University of Lisbon, Lisbon, Portugal

amr@iseg.ulisboa.pt

Abstract. Nowadays, we are forced to face off, more than ever before, an exponential growth of tremendous quantities of data necessary to be local or remotely transferred within cabled or wireless communication networks. We present in this work a methodology used for a short survey and the main findings following queries in important scientific databases. As a result of this research in databases, a sort presentation of achievement related to data rates and optical link, are chronological presented. The Optical Wireless Communication (OWC), complementary to the classical wireless transmission technologies into the radiofrequency (RF) spectrum, consists of Visible Light Communication (VLC), Optical Camera Communication (OCC), Light Fidelity (LiFi), Free Space Optics (FSO) and Infrared (IR). All these technologies and applications are briefly described here with their key characteristics. Various implementations of OWC in different areas so far, as well as the challenges and achievements reached during the last decade of intensive research, are also addressed in this work.

Keywords: VLC · OCC · LiFi · FSO · IR

1 Introduction

The OWC technologies aim to relieve the overcrowded wireless communication links in RF domain. The rapidly growth of the number of smart devices with a huge “hunger for data” as well as the need for greater bandwidth and spectral relief, drove to intensive research efforts to develop mature, alternative technologies for wireless communication. The LED’s light that piggybacked data is foreseen to provide ubiquitous wireless communication, being a reliable partner to the current RF wireless transmission technologies.

The LEDs are used, beside illumination, in many electronic devices such as mobile phone displays, computer monitors, TV screens, advertising billboards, and many kinds of digital signing displays with the primary function of illumination and/or content’s visualization. Is forecast that LEDs will replace in few years, the present illumination systems with bulbs/lamps (incandescent bulbs, fluorescent or halogen lamps) and become ubiquitous lighting fixtures, for their advantages (long lifetime—up to 50,000 h,

high colour rendering capability, high energy conversion efficiency, low heat generation, high tolerance to humidity and high/low temperatures, eco - friendly, mercury free, and compact size) [1].

Gradually, the electromagnetic (EM) spectrum have been discovered and defined by scientists, since the 18th century, as it seen in Fig. 1.

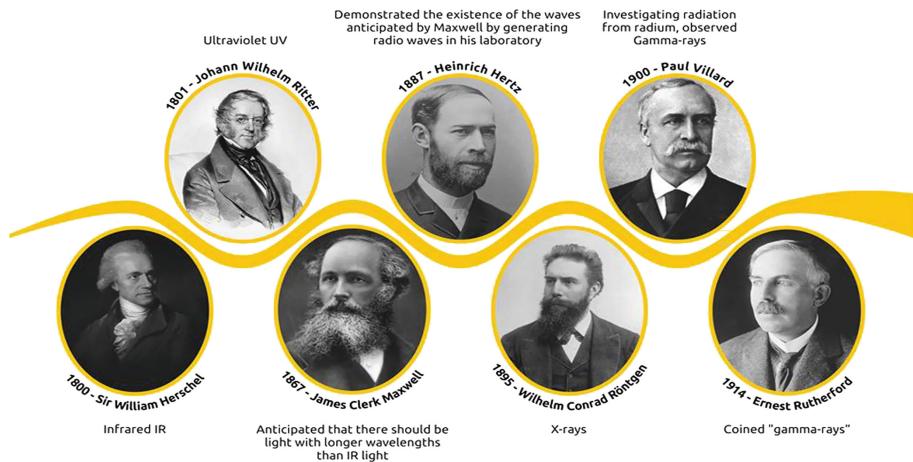


Fig. 1. The electromagnetic spectrum discovery timeline

OWC uses wavelengths in IR and the visible light regions of the spectrum. The visible light region in the EM spectrum starts at 380 nm and ends at 750 nm covering more than 300 THz [2]. The first milestone for indoor OWC has been set by *Gfeller and Bapst* in 1979 who demonstrated an experimental communication at 1 Mbps [3] data rate. They used the On Off Keying (OOK) modulation at 950 nm wavelength in IR spectrum. In 1996, using OOK and IR, *Marsh and Kahn* prove an indoor data rate transfer of 50 Mbps [4].

The paper is structured in five sections. The first section of the paper presents the methodology used to perform a broad survey of the OWC's technologies and applications, databases and keywords used to search the literature as well as bibliographic resources found. The second section addresses the of R&D in OWC resulting in promising perspective regarding increasingly higher data rates and distances of transmission achieved in OWC links. The third section describes the key characteristics of the OWC's technologies and applications developed so far (VLC, OCC, LiFi, FSO and IR) with their main concepts and applications already implemented. The fourth section provides a short description of miscellaneous implementations of OWC in different areas, from indoor positioning systems to mature applications in industry, health, vehicle to vehicle communication, underwater or underground communication. The fifth section sums up the literature review, discuss the key characteristics with specific limitations to be mitigated for mature OWC systems as well as further development trends of OWC's technologies and applications.

2 Methodology

Our major aim is to provide a transparent analysis based on the existing research, describe the OWC's technologies with their key characteristics and applications developed so far. Our research methodology addresses three steps [5]. We first planned the review, then conducted the review and, at the end, we presented the results of the review. As part of planning step, the research area as well as relevant keywords related to OWC technologies are identified [6]. The most relevant database, Web of Science Core Collection has been used with indexes in SCI-EXPANDED, CPCI-S, BKCI-S and ESCI. Other databases, such as Scopus or IEEE Xplore Digital Library or sources as ResearchGATE or Google Scholar have also been searched in. Using Web of Science Core Collection, the search query since 1995 to date, for topic with the following acronyms, returned unexpected outcomes: OWC 1,572 results, VLC 5,134 results, OCC 2,468 results, Li-Fi 239 results, LiFi 236 results, FSO 4,020 results and IR 444,644 results. This search has yielded a huge amount of results, most of them irrelevant for our study. Therefore, the next search strategy was to expand the acronyms and refer to all words in it and the results ("optical+wireless+communication" 2,173 results, "visible+light+communication" 5,183, "optical+camera+communication" 191 results, "light+fidelity" 246 results, "free+space+optics" 1,933 results and "infrared+communication" 400 results), proved to be, most of them, relevant for our search seeing that engineering electrical and electronics, optics, computer science or telecommunications are some of the main domains displayed.

Considering both the IEEE 802.15.7 standardisation of VLC [7], as a milestone for OWC (since the R&D activity worldwide has been noticed to be intensive), we resumed the same research conducted above for a limited period time, since 2011 to date. The results proved to be quite closed to previous findings ("optical+wireless+communication" 1,792 results, "visible+light+communication" 5,067 results, "optical+camera+communication" 190 results, "light+fidelity" 246 results, "free+space+optics" 1,309 results and "infrared+communication" 154 results), therefore the research novelty during the last decade of the OWC's technologies and applications is obvious.

In order to find the most appropriate works for our goal, we planned to refine our search and extract relevant papers for the same time (since 2011 to date) using additional keywords as the search query. LED's nonlinearity, light dispersion, modulation bandwidth or data rate enhancement are some of the targeted difficulties to be overcome for suitable OWC systems, hence we refined our search query using multiple keywords, wildcards (* \$?), as well as operators (AND, OR, NOT, NEAR, SAME). For example, "visible+light+communication" AND "LED* nonlinearity" (145 findings), "Infrared" AND "data rate" (111), "visible+light+communication" AND "data rate?" (293 findings), "optical wireless communication" AND "application?" (178 findings) (Mbps) OR (Mb/s) OR (Mbit/s) as well as (Gbps) OR (Gb/s) OR (Gbit/s), "LED key characteristics" "LED*+VLC" and so on. Finally, 412 articles and proceedings paper resulted, 11 highly cited, 296 open access in fields as Optics (195), Engineering Electrical Electronic (182) or Physics Applied (35) to be read and analysed in full. Articles and papers considered in the end regarding increasingly higher data rates and optical length link, proved to bring innovation either regarding the transmitter (oTx) and/or receiver (oRx) electrical and/or

optical modules (PCBs, LEDs and/or PDs), and/or in modulation techniques. Following the overview of content findings in line with the research objectives, the OWC's technologies and applications (developed so far worldwide) with their main key characteristics and challenges, as well as their main areas of applications in miscellaneous domains are also addressed in this survey.

3 Data Rates in VLC Optical Wireless Links

Due to fast technological growth and development in optics as the white Light Emission Diode (WLED), RGB LED, μ LED and electronics, the visible domain of the EM spectrum was for the first time considered by *Tanaka et al.* to convey data indoor using a LED with two functions at once: illumination and data communication. The first VLC system using OOK was thus demonstrated in laboratory, in 2003, at Keio University in Nakagawa in Japan, with a 400 Mbps data rate [8]. Short time after the standard has been published, in 2011, the Light Fidelity (LiFi or Li-Fi), was coined by professor Harald Haas at TED Global talk [9]. The obstacle of Mbps has been overpassed in 2012 by *Khalid et al.*, at 1 Gbps with a single WLED [10]. The highest data rate with longest transmission distance (over 800 mm) using a RGB LED and a pre-equalization circuit has been achieved by *Huang et al.* in 2015. *Manousiadis et al.*, demonstrated, in 2015, a reliable VLC transmission with a data rate of 2.3 Gbps using GaN μ LEDs [11].

Cossu et al. for the first time, in 2015, developed a VLC duplex communication with a length of more than 1500 mm with four LEDs they achieved 5.6 Gbps downlink and with one IR LED for uplink with 1.5 Gbps [12]. The 2015 year was a prolific one for VLC, with a significant number of scientific research published and brilliant ideas. Using a hybrid post equalizer and a RGYB LED, a data rate of 8 Gbps has been reported by *Wang et. al.* at the end of the year [13].

Based on the previous research regarding data rate achieved and distance between LED and PD, at the end of 2016, an imaging Multiple Input Multiple Output (MIMO) system became available. *Hsu et al.* built a high speed 3×3 imaging MIMO VLC system with 3 WLEDs that reached 1 Gbps data rate at 1-m distance [14]. In March 2017, *Islim et al.* (a team of fifteen UK researchers) achieved 11.95 Gbps data rate using 400 nm violet GaN μ LEDs [15]. A record 20.231 Gbps data rate with bidirectional communication over 1 m with tricolor RGB laser diode (LD) based VLC system supporting signal re-modulation has been experimentally demonstrated in 2018 [16].

A high data rate reported reached by end of March, 2019 for LED based VLC systems of 15.73 Gbps over a 1.6 m link was achieved by a team of scientists in University of Edinburgh, using four LEDs (red, blue, green and yellow) and dichroic mirrors [17]. A VLC system using polarization multiplexing has been demonstrated in 2019 that achieved a data rate of 40.665 Gbps using RGB LDs, over a 2 m distance [18].

4 OWC's Various Technologies and Applications

OWC refers to any type of wireless communication based on *light* as wireless transmission medium. It uses the superior modulation bandwidth of LEDs to convey data

simultaneously with illumination. Visible Light Communication (VLC), Optical Camera Communication (OCC), Light Fidelity (LiFi), Free Space Optics (FSO) and Infra-Red (IR) communication are all part of OWC (Fig. 2). OWC is useful in many wireless communication applications, from millimetres range interconnected within integrated circuits (ICs) up until outdoor kilometres links [19].

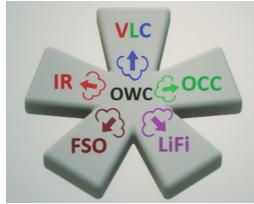


Fig. 2. OWC with its optical range and applications

4.1 VLC

VLC technology refers to an illumination source, a LED embedded into an optical transmitter (oTx) that uses illumination to send data piggybacked by the optical signal. The optical signal is received by a photodetector (PD), as the “main actor” of the optical receiver (oRx). The PD converts the optical signal received, into an electrical one, being furthermore processed by different other modules, in order to become as close as data sent by oTx. Using ultra-high frequency band, VLC enables extremely fast data transmission [20]. The communication technique used in VLC system is typically referred to as light Intensity Modulation/Direct Detection (IM/DD). The most important challenges in the VLC links are the limited bandwidth of the LEDs (about 20 MHz), their non-linear behaviour and the multi-path propagation of the light emitted by LED. Key characteristics of the oTx and oRx refer to the optical spectral response, electrical modulation bandwidth, light radiation as well as light’s detection patterns, the LED’s optical power, the PD’s photosensitive area and its noise configuration. The optical wireless channel is linear, independent of time with an impulse response of a limited duration. Higher data rates can be achieved with complex modulation schemes with high spectral efficiency and robustness against the intersymbol interference (ISI) [8].

The PD collects all or part of the desired optical signals together with other optical signals (from ambient light) known as Additive White Gaussian Noise (AWGN) that has a negative effect on the communication quality. The ambient light consists on both natural light (coming from the sun) and different other artificial lights (fluorescent, incandescent, LEDs or halogen lamps). In front of the PD are usually positioned an optical filter and/or a non-imaging concentrator. The optical filter aims to decrease the effect of AWGN produced by ambient light. The optical non-imaging concentrator aims to acquire as much as possible of the optical signal send, in order to keep the quality of data received as high as possible, expressed as Bit Error Ratio (BER). Characterization of a VLC channel is done by its optical Channel Impulse Response (CIR), that is used to

investigate the channel quality and attempt to attenuate, as much as possible, the effects of distortions [21].

4.2 OCC

The OCC technology uses the camera function embedded into smart mobile device (smart phone, tablet) as receiver (oRx). OCC systems depend on the type of image sensor, the method by which the exposure is made and the light source used. OCC can be classified based on the two technical considerations: (i) the characteristics of modulated light, as observed by the human eye; (ii) the demodulation performed by the camera, as receiver (oRx) [22]. OCC has a lower response than a VLC system, since it receives data using (instead of PIN or an APD), an image sensor that cannot recognize high-frequency signals. OCC is preferred, when mobility is important since smart devices with embedded camera allow fast and easy connection to internet, and, due to today's smart devices universality. Most of the Indoor Positioning Systems (IPS) developed so far, rely on camera with a dedicated software for indoor navigation [23]. In case that different systems of VLC and OCC are used into the same area indoor, they interfere with each other, since both use the same visible light emitted by LED, that works as noise at the other's receiver.

OCC data rate can be improved using the rolling-shutter mechanism of complementary metal-oxide-semiconductor (CMOS) image sensors [24]. The amount of received light at the image sensor is controlled by opening and closing of the shutter. The image sensor receives light only for a limited period of time that is so-called exposure time (t_e). The inter-frame gap (IFG) is also important in OCC. The exposure time is inversely proportional to the frame rate [25]. OCC has also gain importance due to its two important main advantages over VLC systems: mobility and worldwide already availability of smart phones.

4.3 LiFi

LiFi, is a fully networked high speed, full duplex, MIMO wireless communication that uses the lighting fixture network for communication. Full duplex communication is possible due to download on visible light and upload on IR spectrum. Multiple mobile users and wireless handoff from one LiFi access point to another is also possible in early deployed LiFi systems. Since LiFi allows multiple gigabits transmission, promises to solve challenges faced by the “spectrum crunch” and 5G wireless technology, due to its strengths: uniquely more secure, virtually interference free and more reliable than currently wireless technologies based on RF [26–28]. Although not standardized or fully developed yet at its entire potential, LiFi technology has already been deployed on the market in 2018 (as academic evaluation kits, first) and promises to exponentially grow in the near future. Today, a LiFi system works with arbitrary LED light fixtures being currently tested in different industries that need high security, such as schools or in airplanes [8].

4.4 FSO

The range of FSO link starts from few meters indoor to a few kilometres in outdoor environments. FSO communication, similar to VLC, covers additionally the IR spectrum and it does not have an illumination requirement, thus, communication links between buildings is a good example of a proper application [29]. FSO, from idea to a proper implementation, also encounter few challenges since additional noises as fog, dust, rain or snow attenuate the optical signal outdoor. This is the reason why, intensive research focus on the CIR, improved oTx and oRx components and advanced modulation schemes, improved spectral efficiency of the link, in order to mitigate different atmospheric weather conditions that negatively affect the entire system [30].

4.5 IR Communication

As in VLC systems, the commonly used modulation technique in IR links is IM/DD. IM is achieved by changing the bias current of the light source. This change will result in a varying intensity of optical carrier. At the oRx, the PD (PIN or APD) will detect the optical power and, producing a photocurrent, converts it back into the electrical domain. The amount of photocurrent depends on the PD's surface area, its responsivity, the strength of electric field and the additional noises. The first IPS applications based on IR were Olivetti Active Badge system in 92, Xerox ParcTab in 96 and the Cyberguide project in 97 [31]. Today, IR is widely used in different devices, from TV remote control to smart security gadgets.

5 Applications and Challenges in OWC

Different domains, where wireless communication is based on radio frequency, are suitable for OWC implementations as additional or replacement technology in different outdoor and indoor topologies and scenarios.

The VLC embedded in IPS has raised academic community attention and early implementations of different companies worldwide, have been noticed since 2012, mainly in retail industry and museums. Different types of methods - already applied for IPS based on RF - are also used in IPS based on VLC and OCC with various techniques: Received Signal Strengths (RSSs), Angle-of-Arrival (AoA), Time of Arrival (ToA), Time-Difference-of-Arrival (TDoA), Image, or a combination of these [32].

The medical area, both medical facilities and devices have huge potential of numerous application of the OWC technologies, especially due to the main drawbacks of the wireless transmission based on RF: interference with medical devices (RM scanners, for example or implantable devices), lack of security and many concerns related to negative effects on human health [33–35]. The most fitted application in underground mining is positioning of personnel with additional data regarding the underground environmental air quality and even information regarding predictive harmful situations brought with the support of Augmented Reality (AR) technology and Artificial Neural Networks (ANN) [36, 37]. The applications in aeronautic industry (airplanes) have frontloaded a bit the other domains since important R&D projects and investments have been already been

noticed [8]. Following years of intense R&D with many innovative ideas of various implementations in automobiles, the automotive industry is also eager to embed this technology into become well known for both academic community and companies [38]. Underwater OWC research also takes shape with promising results [39].

Important steps, with significant results, for intense R&D efforts, have been done in the last decade in all OWC technologies and applications. Still, there are some challenges in front of reliable, high speed and solid ubiquitous wireless optical net-works OWC network. When we consider the oTx, the LED's nonlinearity and its limited bandwidth, and multi-path propagation that constrain the potential of the optical spectrum that is about 300 GHz (more than 1000 the bandwidth of the entire RF spectrum). The environment, where the optical wireless system is intended to be setup, has to be carefully studied to apply the best suited topology to achieve the optimum CIR, considering the route mean square delay spread (RMS DS) and the path loss (PL). As for the oRx, the main issues in IM/DD technique is the detector technology, related to the PD's sensitivity situated around -40 dBm. The PD's sensitivity is proportional to the number of photons that have to be collected on its active area in order to achieve the targeted signal-to-noise ratio (SNR). Increasing the PD's active area is not a solution since it reduces the field of-view (FoV) according to 'etendue law, and decreases the bandwidth due to the higher capacitance. Possible solutions are the use angular diversity receivers or segmented (fly-eye) receivers [40].

6 Conclusions

The last decade, the academic community together with researchers worldwide changed our everyday living due to astonishing discoveries and achievements with useful applications in fast and reliable local and remotely wireless communication. Following queries completed in different relevant scientific databases and sources, a timeline of the world-wide scientific achievements in OWC's systems, emphasizing constantly increased data rates transmitted wireless and distances achieved so far, has been presented in this paper. A comprehensive description of OWC systems with their various, so far developed applications and technologies, VLC, OCC, LiFi, FSO and IR have also been addressed.

Due to the scientific discoveries and fast industrial improvements of the components integrated in the OWC systems, we expect to have, in the coming years, new, reliable technologies and applications for optical wireless transmission.

References

1. Homepage. <https://www.juliantrubin.com/encyclopedia/electronics/led.html>. Accessed 05 Jan 2021
2. Bapst, R.F., Gfeller, U.: Wireless in-house data communication via diffuse infrared radiation. In: Proceedings of the IEEE, November 1979, vol. 67, no. 11, pp. 1474–1486 (1979)
3. Kahn, J.M., Marsh, G.W.: Performance Evaluation of Experimental 50-Mb/s Diffuse Infrared Wireless Link Using On–Off Keying with Decision-Feedback Equalization, November 1996, vol. 44, no. 11, pp. 1496–1504 (1996)
4. Tanaka, Y., Komine, T., Haruyama, S., Nakagawa, M.: Indoor visible light data transmission system utilizing white LED lights. In: IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC), vol. 2 (2001)
5. Tranfield, D., Denyer, D., Smart, P.: Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manage. **14**(3), 207–222 (2003)
6. Rahimli, N., See-to, E.W.K.: Exploring open innovation technologies in creative industries: systematic review and future research agenda. In: Antipova, T., Rocha Á. (eds.) Information Technology Science. MOSITS 2017. Advances in Intelligent Systems and Computing, vol. 724. Springer, Cham (2018)
7. IEEE Standard for Local and Metropolitan Area Networks--Part 15.7: Short-Range Wireless Optical Communication Using Visible Light. In: IEEE Std 802.15.7–2011, pp. 1–309, 6 September 2011. <https://doi.org/10.1109/IEEEESTD.2011.6016195>
8. Dimitrov, S., Haas, H.: Principles of LED Light Communications. Towards Networked Li-Fi. Cambridge University Press, Cambridge (2015)
9. Homepage Wireless Data from Every Light Bulb. TED Talk, August 2011. <https://bit.ly/tedvlc>. Accessed 05 Jan 2021
10. Khalid, A.M., Cossu, G., Corsini, R., Choudhury, P., Ciaramella, E.: 1Gbit/s visible light communication link based on phosphorescent white LED conference. IEEE Photonics J. **4**(5), 1465–1473 (2012). <https://doi.org/10.1109/JPHOT.2012.2210397>
11. Manousiadis P., et al.: Demonstration of 2.3 Gb/s RGB white-light VLC using polymer based colour-converters and GaN micro-LEDs, Nassau, Bahamas. In: IEEE Summer Tropical Meeting, Visible Light Communications (VisC) (2015)
12. Cossu, Y., Ali, W., Corsini, R., Ciaramella, E.: Gigabit-class optical wireless communication system at indoor distances (1.5–4 m). Opt. Express **23**, 15700–15705 (2015)
13. Wang, Y., Tao, L., Huang, X., Shi, J., Chi, N.: 8-Gb/s RGBY LED-based WDM VLC system employing high-order CAP modulation and hybrid post equalizer. Photonics J. IEEE **7**(6), 1–7 (2015)
14. Hsu, C.W., Chow, C.W., Lu, I.C., Liu, Y.L., Yeh, C.H., Liu, Y.: High speed imaging 3×3 MIMO phosphor white-light LED based visible light communication system. IEEE Photonics J. **8**(6), 1–6 (2016)
15. Islim, M.S., Ferreira, R.X., He, X., Xie, E., Videv, S., Viola, S., Watson, S., Bamiedakis N., Penty, R.V., White, I.H., Kelly, A.E., Gu, E., Haas, H., Dawson, M.D: Towards 10 Gb/s orthogonal frequency division multiplexing-based visible light communication using a GaN violet micro-LED. Photon Res. **5**(2), A35–A43 (2017)
16. Wei, L.Y., Hsu, C.W., Chow, C.W., Yeh, C.H.: 20.231 Gbit/s tricolor red/green/blue laser diode based bidirectional signal remodulation visible-light communication system. Photon. Res. **6**, 422–426 (2018). <https://doi.org/10.1364/PRJ.6.000422>
17. Bian, R., Tavakkolnia, I., Haas H.: 15.73 Gb/s visible light communication with off-the-shelf LEDs. J. Lightwave Technol. **37**(10), 2418–2424 (2019)

18. Wei, L.-Y., et al.: Tricolor visible-light laser diodes based visible light communication operated at 40.665 Gbit/s and 2 m free-space transmission. *Opt. Express* **27**(18), 25072–25077 (2019). <https://doi.org/10.1364/OE.27.025072>
19. Uysal, M., Capsoni, C., Ghassemlooy, Z., Boucouvalas, A., Udvary E.: Optical Wireless Communications. An Emerging Technology. Springer, Switzerland (2016). ISBN 978-3-319-30201-0
20. Leba, M., Riurean, S., Ionica, A.: Li-Fi -the path to a new way of communication. In: CISTI'2017 12^a Conferência Ibérica de Sistemas e Tecnologias de Informação. IEEE Xplore Digital Library (2017)
21. Udvary, E.: Visible light communication survey. *Infocommun. J.* **XI**(2) (2019).
22. Marcu, A.E., Dobre, R.A., Datcu, O., Suciu, G., Oh, J.: Flicker free VLC system with automatic code resynchronization using low frame rate camera. In: 42nd International Conference on Telecommunications and Signal Processing (TSP), Hungary, pp. 402–405 (2019)
23. Jiao, J., Li, F., Deng, Z., Ma, W.: A smartphone camera-based indoor positioning algorithm of crowded scenarios with the assistance of deep CNN. *Sensors* **17**(4), 704 (2017)
24. Danakis, C., Afgani, M., Povey, G., Underwood, I., Haas, H.: Using a CMOS camera sensor for visible light communication Proc. In: IEEE Globecom Workshops (GC Wkshps), pp. 1244–1248, December 2012
25. Nguyen, D.T., Park, Y.I.: Data rate enhancement of optical camera communications by compensating inter-frame gaps. *Opt. Commun.* **394**, 56–61 (2017)
26. Homepage. <https://purelifi.com/>. Accessed 05 Jan 2021
27. Wu, X.P., O'Brien, D.C., Deng, X., Linnartz, J.P.M.G.: Smart handover for hybrid LiFi and WiFi networks. *IEEE Trans. Wirel. Commun.* **19**(12), 8211–8219 (2020). <https://doi.org/10.1109/TWC.2020.3020160>
28. Wu, X.P., O'Brien, D.C.: Parallel transmission LiFi. *IEEE Trans. Wirel. Commun.* **19**(10), 6268–6276 (2020). <https://doi.org/10.1109/TWC.2020.3001983>
29. Hulea, M., Ghassemlooy, Z., Abadi, M.M., Rajbhandari, S., Tang, X.: Fog mitigation using SCM and lens in FSO communications. In: 2nd West Asian Colloquium on Optical Wireless Communications (WACOWC), Tehran, Iran, pp. 46–50 (2019)
30. Avătămăniței, S.A., Căilean, A.-M., Done, A., Dimian, M., Prelipceanu, M.: Noise resilient outdoor traffic light visible light communications system based on logarithmic transimpedance circuit: experimental demonstration of a 50 m reliable link in direct sun exposure. *Sensors* **20**, 909 (2020)
31. Alresheedi, M.T., Hussein, A.T., Elmirghani, J.M.H.: Uplink design in VLC systems with IR sources and beam steering. *IET Commun.* **11**(3) 311–317 (2017)
32. Do, T.-H., Yoo, M.: An in-depth survey of visible light communication based positioning system. *Sensors* **16**, 678 (2016)
33. Riurean, S.M., Leba, M., Ionica, A.: VLC embedded medical system architecture based on medical devices quality' requirements. In: Quality-Access to Success, Romania (2018)
34. Riurean, S., Antipova, T., Rocha, Á., Leba, M., Ionica, A.: Li-Fi embedded wireless integrated medical assistance system. In: Rocha, Á., Adeli, H., Reis, L., Costanzo, S. (eds.) New Knowledge in Information Systems and Technologies. WorldCIST 2019. Advances in Intelligent Systems and Computing, vol. 931, Springer, Cham (2019)
35. Riurean, S., Antipova, T., Rocha, Á., Leba, M., Ionica, A.: VLC, OCC, IR and LiFi reliable optical wireless technologies to be embedded in medical facilities and medical devices. *J. Med. Syst.* **43**, 308 (2019)
36. Riurean, S., Olar, M., Ioničă, A., Pellegrini, L.: Visible light communication and augmented reality for underground positioning system. In: MATEC Web of Conference, vol. 305, p. 00089 (2020)

37. Riurean, S., Stoicuta, O., Leba, M., Ionica, A., Rocha, Á.: Underground channel model for visible light wireless communication based on neural networks. In: Rocha, Á., Adeli, H., Reis, L., Costanzo, S., Orovic, I., Moreira, F. (eds.) Trends and Innovations in Information Systems and Technologies. WorldCIST 2020. Advances in Intelligent Systems and Computing, vol. 1160. Springer, Cham (2020)
38. Marcu, E., Dobre, R.A., Vlădescu, M.: Key aspects of infrastructure-to-vehicle signalling using visible light communications. In: Future Access Enablers for Ubiquitous and Intelligent Infrastructures. FABULOUS 2017. Lecture Notes of the Institute for Computer Science, Social Informatics and Telecommunication Engineering, vol. 241, pp. 212–217. Springer, Cham (2018)
39. Zhu, S., et. al.: Recent progress in and perspectives of underwater wireless optical communication. *Prog. Quantum Electron.* **72** (2020)
40. Haas, H., Cogalan, T.: LiFi opportunities and challenges. In: 16th International Symposium on Wireless Communication Systems (ISWCS) (2019). <https://doi.org/10.1109/iswcs.2019.8877151>



The Perspective of Cyclists on Current Practices with Digital Tools and Envisioned Services for Urban Cycling

Inês Fortes¹ , Diana Pinto¹ , Joana Vieira² , Ricardo Pessoa³ , and Rui Jose¹

¹ Algoritmi Research Centre, University of Minho, Braga, Portugal

{ines.fortes,dianapinto}@algoritmi.uminho.pt, rui@dsi.uminho.pt

² Centro de Computação Gráfica, Guimarães, Portugal

Joana.Vieira@ccg.pt

³ Bosch Car Multimedia, Braga, Portugal

Ricardo.Pessoa@pt.bosch.com

Abstract. As cycling becomes increasingly important in sustainable mobility policies, there is also an urge for new digital applications and services for urban cycling. This new generation of cycling applications should be able to connect cyclists with their local cycling ecosystem, promote cycling, and empower cyclists to become active agents of urban mobility. In this work, we aim to explore the new opportunity space of digital tools and applications designed specifically for urban cycling. We pursue this goal by trying to uncover current practices associated with digital tools that are already available and also by trying to uncover new information needs, even those that cyclists are not yet able to fully express. To explore these topics, we conducted 2 focus group sessions and 10 interviews with cyclists. The result is a set of design opportunities for the development of new applications, tools and methods for improving the cycling experience in the context of urban mobility. We expect this contribution might help to better define the design space of innovative digital tools for urban cyclists.

Keywords: Digital practices · Cycling · User research

1 Introduction

Cycling is assuming an increasing importance in sustainable mobility policies. Leading cities and central governments all over the world are making significant investments to bring cycling, and other micro-mobility modes, to the forefront of their mobility strategies. This transition is being fueled by a combination of sustainability [1, 2], public health [3, 4], urban life [5] and economic [6, 7] agendas. It is also happening in a context of major technology trends and new mobility paradigms, such as shared, electric, and connected bicycles, which are reshaping our perception of bicycles as a core element of urban mobility.

This disruptive change is strongly driven by the increasingly pervasive presence of digital platforms and mobile applications in cycling systems, which is likely to become a decisive element for the success of cycling as a modern urban mobility mode. Smartphones are already playing a key role in these new connected cycling paradigms, especially in bike sharing systems. They can be a valuable resource for cyclists, and a plethora of mobile applications is now available to offer cyclists a diverse set of services. They explore the huge convenience associated with the immediate availability of advanced interaction, computation, communication, positioning and sensing capabilities, and their unique capability to scale deployment of new applications.

However, current applications are mainly conceived for the cyclist as an individual (e.g., quantified self) and for cycling as a leisure or sports activity (e.g., performance goals). They do not usually give much consideration to the role of cyclists as agents of urban mobility, or to the role of cycling as being primarily a mobility mode for reaching from A to B as safely, smoothly, and efficiently as possible.

Urban cycling calls for a new generation of cycling applications and tools, designed to connect cyclists with their local cycling ecosystem, promote cycling, provide information about safe paths, and empower cyclists to contribute to mobility policies, by expressing their preferences, reporting problems, or simply share their cycling data to feed local mobility services. While many of the features from current applications may also migrate to urban cycling tools, a design mindset focused on urban cycling would certainly call for new specific features or redesigned versions of existing ones.

1.1 Objectives

In this work, we aim to explore the new opportunity space of digital tools and applications designed specifically for urban cycling. We pursue this goal through two complementary paths. The first path explores in more detail what is already available, and the emerging practices associated with digital tools. This path is particularly relevant to identify elements that may be appropriated by urban cycling, in their current form or with only minor adjustments. It may also enable us to identify ways in which people use technology that was not primarily designed for that purpose. Such cases may provide alternative mindsets throughout the design process [8].

The second path is to uncover new information needs, even when cyclists are not able to fully express them. Some of these are not supported by current tools and might correspond to the more utilitarian perspectives of cycling as an urban mobility mode. The result is a set of design opportunities for the development of new applications, tools and methods for improving the cycling experience in the context of urban mobility.

2 Related Work

In recent years, cycling and other soft mobility modes are being increasingly recognized as a key element for sustainable mobility policies of the future [9]. At the same time, bicycle technology has improved significantly and is now much more capable of offering quality solutions to different cyclists profiles [10].

Smartphones are a powerful tool for large-scale data collection [11]. They already integrate a very vast range of sensors, enabling the collection of substantial data about people and their movements. Using data provided by urban cyclists through the smartphone sensors can enable the generation of collective knowledge to improve the quality of cycling mobility. BeCity [12] is an example of a mobile application that allows riders to share their tracks and comments, working as a distributed data collection system. It also includes the ability to recommend routes, considering factors such as distance, presence of bike paths and even the attractiveness of those paths. Another example is the BikeNet, a mobile application that gathers data about the rides to provide cyclists with a general perspective of their experience and performance. This system is able to obtain information about the environment and the entire experience along the way, such as pollution levels, noise and floor condition [13].

Meireles and Ribeiro [14] explored the use of digital platforms and smartphone applications as behavioral change tools that may help to promote the growth of the bicycle as a means of transport, especially for mid-sized starter cycling cities. Based on a survey targeting cyclists, the authors concluded that even though most cyclists (77%) used at least one cycling application, there is a lack of digital solutions to promote cycling. This is also suggested by the fact that most respondents used generic cycling apps such as Strava (39%), mainly to track their daily bicycle trips, and Google Maps (51%), mainly for navigation. Regarding what could be added to a cycling application or platform, cyclists referred a compilation of features of already existing solutions, and their integration into a single platform.

3 Method

To expand our knowledge on the perspective of cyclists regarding digital services, we conducted two focus groups and ten individual interviews. Next, for each method, the participants and procedure are described.

3.1 Focus Groups

A total of 10 participants (all men) aged between 24 and 59 years ($M = 36.20$, $SD = 10.82$) participated in the focus-group sessions, divided into two groups of 5. Participants were internally recruited BOSCH employees at Braga, Portugal, and the only criteria to participate in the focus group was to own and ride a bicycle ($n = 9$) or a standing scooter ($n = 1$). Two cyclists used the bicycle only to commute, 4 used it for leisure and 4 used in both contexts. For most participants ($n = 6$), rides usually take less than one hour.

Both focus groups explored: (1) the experience of riding a bicycle and (2) the use of digital technologies for cycling. On each focus group, there was a moderator and a note-taker. At the beginning of each session, we explained that the objective of the focus group was to gather information about the current practices and needs of cyclists and that there were no right or wrong answers. We explained that the session was going to be recorded and that all video and audiotapes were confidential and would only be used by researchers of the project. After that, all participants signed an informed consent. The moderator started the session following the script. The sessions took 60 to 90 min.

3.2 Interviews

To gain a deeper insight into the topics, we also conducted interviews with cyclists. Participants were recruited via LinkedIn and Facebook. A total of 10 participants (7 men and 3 women) aged between 23 and 53 years ($M = 35.70$, $SD = 8.96$) were interviewed. All except one cyclist used the bicycle to commute, and for those, the rides were usually short, taking less than one hour.

The interviews were semi-structured and focused on several topics including (1) the use of digital technologies for cycling, and (2) the ideal mobile application for urban cyclists. The interviews were online and were image and sound recorded with Zoom recording tools, for later analysis. To start the interview, participants were asked some demographic questions, and a verbal consent was made to record the session. After that, the recording started, and the interview script was followed. The interview took approximately 40 min.

3.3 Data Analysis

For the focus groups, sessions were transcribed from audio to text and a qualitative content analysis was implemented, where materials with similar meaning were classified into categories [15].

For the interviews, the recordings were listened, and detailed notes and partial transcriptions were made. Similarly to the focus groups, the results for each topic in the interview from all participants were aggregated and summarized. At the end, the results of the focus groups and interviews were aggregated because they addressed the same topics and shared the same general objectives. The analyses were conducted using the qualitative analysis software MAXQDA version 10 [16].

4 Results

The aggregated results of the qualitative analysis of the focus groups and the interviews, generated the following set of main themes: (1) Current practices in digital tools, wearables, and sensors, (2) Technological difficulties and needs, and (3) Useful features of a mobile application.

4.1 Current Practices in Digital Tools, Wearables, and Sensors

One of the themes that emerged from the results of both focus groups and interviews was how cyclists currently use digital tools, wearables, and sensors to support the cycling activity. Several cyclists reported using mobile applications and forums or websites (e.g., Reddit, Facebook) related to cycling. The latter are used to search for specific information, such as where to buy or how to modify a bike, to look for trails or information regarding some brands, and to share experiences and doubts. The use of apps, however, is more prevalent and covers more needs. Participants use Strava (to register routes and activities with several statistics), Garmin and TomTom (GPS or smartwatches connected to the smartphone), the iPhone Find My app (to share location with family and friends),

Google Maps (to navigate), and Wikiloc and AllTrails (to save, find, and share trails). Cyclists also referred using COBI (from BOSCH), Bike Citizens, See.sense, and a local app to register the routes. Some of these apps are more suitable for leisure purposes, while others are specifically designed for city riders.

One of the most used applications, Strava, provides several features that cyclists appreciate: routes for leisure, slope, distance, speed, time, heart rate (when paired with a band or smartwatch), calories burned, and effort. Another referred feature is gamification, which ranks cyclists according to their time in given route segments. Some cyclists like this competitive feature, while others use it more to challenge themselves by setting goals for riding distances, number of days per week, etc.

Google maps is another popular app, especially for navigation. Unfortunately, in most starter cycling cities, including all starter cities in Portugal, Google Maps is not yet optimized for cyclists. Thus, cyclists select either the “by car” or “by foot” option. The by-foot view may be more appropriate when riding downtown, and the car option may be more useful for longer rides. However, some maps do not even identify existing cycling paths, so sometimes using google maps is more useful to understand distances and not so much to choose the route:

P14, Female, 29 years-old: *“It is more to know where the places are, not so much to use with the bicycle.”*

But how do cyclists follow the route? To learn a new route, cyclists usually try to memorize it (e.g., take the third turn on the left and then the first turn on the right). When in doubt, most cyclists prefer to stop to look at the map and then proceed to the route. Another interesting feature of Google Maps is that it continuously registers the GPS position, and that position can be shared with a family member. Its main advantage is that it is not necessary to turn on the app because this position sharing is automatic.

One of the devices that has several advantages for city riders with an electric bicycle is the COBI system. With this system, the smartphone can be used as a bicycle computer that shows instant speed, allows the rider to take a call, choose a song, etc. It also communicates and transfers data to other apps, such as Strava, but one of the users reported to not take advantage of this because the routes were almost always the same, and there was no advantage of sharing the data with other apps:

P15, Male, 53 years-old: *“The only interesting thing for us is to know how many Km I rode in a week, how many Km per month, what were the average hours, that is interesting. But with time this all dilutes and ... it does not change much.”*

Even though sharing the routes for those who always do the same route may not be beneficial for themselves, it can be extremely useful for other cyclists. One of the apps that uses this premise is the Bike Citizens app. This app records the routes, creating a cycling flow city map to show the most used routes. The major benefit of this app is to provide data that may be useful to others:

P15, Male, 53 years-old: *“At the end of the day it feels like volunteer work, that is, I use this application to contribute to route log, because I think that if everyone who uses the bicycle used an application like this, it was possible to see exactly*

which routes people use and then better plan the routes and serve a larger number of people.”

This app also provides detailed and up-to-date maps that can be purchased or acquired in exchange for points earned by the riding.

Regarding wearables or sensors, cyclists reported using smartwatches, heart-rate bands, power banks for the mobile phone, cycling computers such as Nyon (from BOSCH), earphones, and sensors in the bicycle to measure cadence or a barometer to measure altitude. Note, however, that not all cyclists are in favor of technology:

P13, Female, 36 years-old: “*I have some aversion to adhere to some digital tools unless they are really necessary. Regarding the bicycle, no, I never joined.*”

The key reasons why participants justified not using technology are as follows: they are usually designed for leisure cyclists, they are sometimes inaccurate (e.g., google maps not identifying cycling paths), and there is no need because cyclists know their cities and feel they can identify the best routes by themselves. Some cyclists referred that one disadvantage of several apps is the need to turn it on and off manually. However, one of the participants also referred to using an app that could activate automatically, but he preferred to control it manually. Another disadvantage of some apps is the battery consumption and occupying memory. To solve the battery problem, urban cyclists could use their electric bicycle to charge it, or a system could use the cycling activity to charge an external battery. Finally, another hindrance to use some apps is that some useful features are only available when paid.

4.2 Technological Difficulties and Needs

On several occasions, cyclists refer that there are many sensors on the market that can be added to the bicycle. However, these are usually expensive and different brands have different sensors for different purposes, so there is a need of integrated solutions that serves several purposes at once. Participants report that they would rather have one single economical platform or device:

P10, Male, 42 years-old: “*Centralized on something, a single tool or device... I'm not going to buy a sensor, a locator, ... no, the cyclist already spends a lot of money ...*”

P01, Male, 45 years-old: “*But we basically already have it all on the smartphone, but then we need duplication. If there were such a thing, it shouldn't be a cell phone, I don't want to have two... but if I had an interface that communicated with my cell phone... it already has music, the applications, it's everything there, that interface.*”

Participants recognize the smartphone as an interesting approach for achieving this type of integration, despite the battery problem that has been mentioned. A different alternative would be to use a bicycle display to mirror the smartphone, where the displayed information could be chosen, but again that would be one more, and possibly

expensive, item to acquire. To use the smartphone while on the bicycle, the interaction mode should be adapted and fully compatible with the reality of the riding experience.

A smartphone associated with a sensors pack could enable security, safety, and comfort features. Concerning safety, cyclists would like to have an automatic emergency call in case of an accident. To detect the accident there could be sensors in the bicycle, where a sudden brake followed by inactivity or a decrease pressure in the saddle would trigger the alarm. In terms of security, the bicycle could have a sensor and a locator and only the owner could unlock it; also, in case it was stolen, it would send an alarm. In terms of comfort, the bike sensors could help the cyclist adjust the position of the handlebar and saddle or give tips according to the way the person rides the bicycle.

4.3 Useful Features of a Mobile Application

Cyclists referred several needs and features that could be integrated in a mobile application for urban cycling. Clearly, the most important feature of an app would be a navigation system with tracking. The ideal mobile application for cyclists should also have social features where cyclists could get together and share information. Another suggestion is that this app should integrate several services within the city.

Table 1 shows the main features that cyclists referred as important for an urban cycling mobile app.

Table 1. User needs: main features for an urban cycling mobile app.

Category	Specific need (Cyclists need/want/like to ...)
Safety	Be seen by drivers Inform others that they are braking or changing direction Receive alerts of dangerous situations Inform their real-time position Quickly inform others in case of an accident
Security	Be alerted and alert the authorities in case the bicycle is stolen
Comfort	Get tips on how to adjust the bicycle or increase comfort
Communication	Communicate within a group while riding Communicate with others (i.e., share experiences or doubts) Communicate with other entities (e.g., alert a bus driver that a cyclist needs to carry the bicycle in the bus) Communicate with other platforms (e.g., Facebook or Instagram)
HMI Interaction	Interact with the cell phone or other device while riding
Bicycle status	Tutorials on regular check-ups and minor repairs Obtain information on the bicycle status, when is electric or has sensors
Gamification system	Compare cycling metrics across time, and/or with others
Bike sharing	Have an easy access to a bike sharing platform
Associations, groups, activities	Find other cyclists and participate in cycling activities
Navigation system	Plan and choose the route

Table 2 shows the information that cyclists would like to obtain and share from the navigation system. These information range from utilitarian features, such as indication the route type, to more social features such as media sharing with GPS tracking.

Table 2. User needs: types of information to be provided by a navigation map.

Navigation category	Details
Route planner	Select a predefined route or draw route passing by specific places
Road type	E.g., pedestrian zone, cycling path, inside a park, etc.
Road condition	E.g., holes or pavement in bad condition
Type of pavement	E.g., tar floor, cobblestone, etc.
Type of traffic flow	E.g., shared with other vehicles? One way vs. Two way?
Traffic	Suggest routes to avoid traffic. Provide the average speed of vehicles
Frequency of use of roads	Indicate the traffic flow of cyclists. Consider those frequencies, when suggesting a route. Share that information with the City Hall
Safety alerts	Static dangers (e.g., dangerous crossings) and dynamic dangers (e.g., approaching vehicle)
Weather	Indicate the weather along the route. Route suggestion depending on the weather
Location of interest points and shops	E.g., drinking fountains, viewpoints, workshops, restaurants, cafes, diet and health food stores, and highlight the bike-friendly shops (i.e., with parking or discounts for those arriving by bike)
Suggestion of places or things to do	Suggest places (e.g., museum) or things to do (e.g., a theatre play) along the route or in specific locations
Location of parking and resting zones	Location of parking with the type of parking (bike racks, bike lockers, covered parking, etc.). Location of benches or resting zones
Estimated time of arrival and distance	Provide several route options with their distance and estimated duration
Media sharing	Share photos, videos or other contents associated with a geographic place or route/track (share with everyone or with a close group)

Concerning routes, these could be ranked from 1 to 5 according to: how cyclable they are, beauty, effort, satisfaction, slope, difficulty and distance. Also, they could be categorized according to the type of route (e.g., leisure, sports, and daily use). After following a route, cyclists would like to know: total duration, calories burned, see the route on the map, and CO₂ consumption if the route had been made by car. Importantly, all these suggestions should be up-to-date and change according to the person's location.

5 Conclusion

In this research, we have studied how current practices with existing digital tools and specific needs expressed by cyclists may inform the design of a new class of cycling tools and applications to address the broader challenges of urban cycling. Based on the results we presented several insights that summarize the key topics expressed by cyclists, which constitute the main contribution of this work to inform the design of new cycling tools and applications.

Beyond those insights, there is one major issue that deserves some discussion. In this study, cyclists seemed to be very willing to identify and describe ideas for improving the applications and tools they already know. This seems to suggest the existence of a relevant opportunity space for the evolution of current digital tools for cycling. However, most of the features suggested, either explicitly or implicitly, seem to be somewhat incremental, and do not necessarily correspond to what could be described as a new class of tools and

applications. Whether this represents a general satisfaction with current applications, or it is just that those prevailing applications are already shaping our perception and our expectations of what a cycling application should be is a question that remains to be answered.

One limitation of the present work was the underrepresentation of women. As this group presents a cycling behavior different than men [17], they may also have different needs. Future studies should assure a more representative sample of cyclists, to assure that services are created to suit every type of cyclist. As future work, we also plan to explore the development and evaluation of new digital tools for urban cyclists. We expect that a different set of design principles may help to satisfy and exceed currently envisioned needs and enable digital tools and applications to assume their role as a key enabler for urban cycling.

Acknowledgements. This work is supported by: European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project n° 039334; Funding Reference: POCI-01-0247-FEDER-039334].

References

1. UN [United Nations]: The 2030 agenda for Sustainable Development. sustainabledevelopment.un.org (2015)
2. Pucher, J., Buehler, R.: Cycling towards a more sustainable transport future. *Transp. Rev.* **37**, 689–694 (2017). <https://doi.org/10.1080/01441647.2017.1340234>
3. Pucher, J., Buehler, R.: Walking and cycling for healthy cities. *Built Environ.* **36**, 391–414 (2010). <https://doi.org/10.2148/benv.36.4.391>
4. Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B., Kohlberger, T.: Health benefits of cycling: a systematic review. *Scand. J. Med. Sci. Sports.* **21**, 496–509 (2011). <https://doi.org/10.1111/j.1600-0838.2011.01299.x>
5. Salat, S., Ollivier, G.: Transforming the urban space through transit-oriented development: the 3V Approach (2017). <https://doi.org/10.1596/26405>
6. Blondiau, T., Van Zeebroeck, B., Haubold, H.: Economic benefits of increased cycling. *Transp. Res. Proc.* (2016). <https://doi.org/10.1016/j.trpro.2016.05.247>
7. Arancibia, D., Savan, B., Ledsham, T., Bennington, M.: Economic impacts of cycling in dense urban areas: literature review. In: Transportation Research Board 94th Annual Meeting (2015)
8. Ljungblad, S., Holmquist, L.E.: Transfer scenarios: grounding innovation with marginal practices. In: Conference on Human Factors in Computing Systems – Proceedings, pp. 737–746 (2007). <https://doi.org/10.1145/1240624.1240738>
9. Bulc, V.: Cycling: green and efficient transport for the future. https://ec.europa.eu/commission/commissioners/2014-2019/bulc/blog/cycling-green-and-efficient-transport-future_en. Accessed 31 Oct 2019
10. Stamatiadis, N., Pappalardo, G., Cafiso, S.: Use of technology to improve bicycle mobility in smart cities. In: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 86–91 (2017). <https://doi.org/10.1109/MTITS.2017.8005636>.
11. Srivastava, M., Abdelzaher, T., Szymanski, B.: Human-centric sensing. *Philos. Trans. Roy. Soc. A Math. Phys. Eng. Sci.* **370**, 176–197 (2012). <https://doi.org/10.1098/rsta.2011.0244>

12. Torres, S., Lalanne, F., Del Canto, G., Morales, F., Bustos-Jimenez, J., Reyes, P.: BeCity: sensing and sensibility on urban cycling for smarter cities. In: 34th International Conference of the Chilean Computer Science Society (SCCC) (2016). <https://doi.org/10.1109/SCCC.2015.7416587>
13. Eisenman, S.B., Miluzzo, E., Lane, N.D., Peterson, R.A., Ahn, G.S., Campbell, A.T.: BikeNet: a mobile sensing system for cyclist experience mapping. *ACM Trans. Sensor Netw.* **6**, 1–39 (2010). <https://doi.org/10.1145/1653760.1653766>.
14. Meireles, M., Ribeiro, P.J.G.: Digital platform/mobile app to boost cycling for the promotion of sustainable mobility in mid-sized starter cycling cities. *Sustainability* (2020). <https://doi.org/10.3390/su12052064>
15. Onwuegbuzie, A.J., Dickinson, W.B., Leech, N.L., Zoran, A.G.: A qualitative framework for collecting and analyzing data in focus group research. *Int. J. Qual. Methods.* **8**, 1–21 (2009). <https://doi.org/10.1177/160940690900800301>
16. VERBI Software, MAXQDA 10 [computer software]. maxqda.com (2010)
17. Beecham, R., Wood, J.: Exploring gendered cycling behaviours within a large-scale behavioural data-set. *Transp. Plan. Technol.* **37**, 83–97 (2014). <https://doi.org/10.1080/03081060.2013.844903>



Enhancing the Motorcycling Experience with Social Applications: A Study of User Needs

Inês Fortes¹ , Diana Pinto¹ , Emanuel Sousa² , Vera Vilas-Boas³ , and Rui José¹

¹ Algoritmi Research Centre, University of Minho, Braga, Portugal

{ines.fortes,dianapinto}@algoritmi.uminho.pt, rui@dsi.uminho.pt

² Centro de Computação Gráfica, Guimarães, Portugal

Emanuel.Sousa@ccg.pt

³ Bosch Car Multimedia, Braga, Portugal

Vera.VilasBoas@pt.bosch.com

Abstract. For many riders, the motorcycle is much more than a transportation mode. Riding a motorcycle can be a pleasurable experience in itself, and the motorcycle is frequently a tool of socialization. An evidence for that can be found in the numerous motorcycling communities around the world. However, those communities may not be accessible to everyone, or they may not satisfy everyone. We aim to explore how social applications for motorcyclists could reach more riders and serve a broader range of their needs. To pursue this idea, we conducted focus groups with motorcyclists, exploring their current practices with digital tools as well as any related needs. The results hinted on several aspects that need to be considered in the design of a social application for motorcyclists. Generally, motorcyclists are willing to receive and share information with others, and referred several needs in terms of communication and trip planning. However, they also expressed concerns regarding the use of digital tools while riding, both for safety reasons and because it could disrupt the riding experience. These results are a contribution to inform the design of new social concepts for motorcyclists.

Keywords: Social applications · Motorcycling · User research

1 Introduction

Riding a motorcycle can be much more than traveling from A to B. It is common to find motorcyclists who grew up around motorcycles and soon developed a fascination for these vehicles. Others may only embrace this world later in their lives, but they end up making it an integral part of their lifestyle. For many of them, riding a motorcycle almost defines them as a person, giving them the feeling of being part of a community of riders and a sense of belonging that only members can fully understand and share [1]. In fact, for many riders this social component is a defining element of being a rider [2, 3]. However, buying a motorcycle does not buy a social experience, and many riders may never really find the right context to experience that sense of belonging and

companionship that others often praise as one of the main reasons why they love being motorcyclists [1].

One might place riders in a continuum of social engagement: at one extreme there are those (social riders) who mainly ride for leisure, prefer to ride in group, and like to share their passion for riding; at the other extreme, there are those for whom a motorcycle is mainly a utilitarian vehicle for commuting (commuting riders), and for whom the social experience is not seen as very important. Despite these differences, both these profiles, as well as the many others in between these two extremes, should be able to benefit from social applications that could enhance their riding experiences.

Digital technologies can be key enablers for many new forms of social engagement between motorcyclists. However, so far, the digital element of that experience has been limited to the pre-ride and, especially, the post-ride phase. Most social applications used by motorcyclists are essentially generic services that allow users to share content related to their events and their passion for motorcycles. To improve the motorcyclists' experience, we should also consider how engagement between riders might occur, in the context of a continuous social experience that can span all the way from the cosy environment of the living room (when a ride is still just a plan), to the most thrilling moments of riding along with other motorcyclists. However, the level of concentration and the physical control needed to safely ride a motorbike represent huge hampering factors for common forms of digital interaction [4]. For obvious safety reasons, most interaction possibilities are either too limited, too intrusive or simply not acceptable when riding, making them technological hindrances to the riding experience. Moreover, even when adequate solutions may exist, an unequal access to them (e.g., due to price) has a negative impact on social collaborations while riding.

In this research, we aim to understand the needs of motorcyclists regarding digital features that could connect them, in any possible way, with other motorcyclists and the broader motorcycling ecosystem. Our goal is to inform the design of a new generation of social applications conceived specifically for the motorcycling experience. These social applications could significantly enhance the social experience of many riders, while hopefully helping to create services that benefit the whole motorcycling community.

2 Related Work

Recent years have witnessed a strong growth in communication, navigation and entertainment systems for vehicles, including motorcycles [5]. Current research is addressing almost any element of the riding experience, including, for example, safety technology based on the interaction of vehicle-to-driver/environment [6], communication between motorcyclists [7], and communication between motorcyclists and other vehicles to signal the presence of each other to prevent accidents [7]. Additionally, studies are also focusing on vehicle-to-vehicle or vehicle-to-infrastructure communication, to share information related to routes, weather, traffic and restaurants [8]. Importantly, there have also been efforts to develop technology for promoting social interaction between motorcyclists [4, 9].

Some motorcycle communities promote all sorts of social interactions, for purely recreational purposes or more formal goals, such as, fundraising, competitions, political protests or community services [10]. However, it is important to note that online

communities may differ from the physical ones. For instance, when bikers engage in online communities, they tend to communicate more with photos (e.g., of their vehicle) or videos, rather than with written text [11]. Despite these differences, both community types share the same rituals and traditions between members [12].

Therefore, for many riders, riding a motorcycle is also a social experience, and that social component can significantly enhance the riding experience [2]. For example, motorcyclists passing by each other on a road may not be seen as a social event. However, even these brief and spontaneous encounters with other like-minded motorcyclists can be very valued. Indeed, sometimes motorcyclists show their appreciation by waving each other. Other forms of social interaction include riding more to promote those encounters, and engaging online [4].

However, there are also major risks associated with the use of digital tools while riding. For example, mobile phone usage during motorcycle riding constitutes a risky behavior associated with accidents and fatalities, although highly prevalent among motorcyclists [13–15]. In fact, research shows that bikers perform several operations on their mobile phones while riding, such as dialing, talking, texting, or searching for information [13, 14, 16]. For that reason, it is crucial that new technologies incorporate adapted and safe modes of interaction.

3 Research Methodology

To address the research goals of this study, we conducted four focus group sessions, in which motorcyclists were invited to discuss ideas, needs and concerns related to different perspectives of social applications. A total of 25 participants (24 men and one woman) between 24 and 54 years-old ($M = 36.28$, $SD = 10.15$) participated in these four sessions, each with 5 to 8 people.

Two discussion guides were created to conduct the focus groups, and participants were randomly assigned to one of them. All groups started by exploring (1) the experience of riding a two-wheeled vehicle, and (2) the use of connected technologies. Then, two groups discussed (3) the willingness to share information and (4) the expected trade-offs of that sharing; the other two groups discussed topics not explored in the present study. Before the focus group session, participants were asked to respond to a short questionnaire about their riding routines, their preferred brands and the use of applications.

On each session, there was a moderator and a note-taker. At the beginning, participants were informed about the objectives of the focus group, and were told that there were no right or wrong answers, and that they could stop the session anytime. It was additionally explained that the session was going to be recorded and that all video and audiotapes were confidential and could only be used by researchers of the project. After that, all participants signed an informed consent. The moderator explained the session rules (e.g., one person talks at a time) and started the session following the respective guide. The sessions took between 60 and 90 min.

The audio data collected during the sessions was transcribed from audio to text. We then applied qualitative content analysis to the transcripts [17]. A first reading of the transcriptions and session notes allowed us to reach a general perspective on the topics

addressed by participants. Then, one of the researchers followed a coding protocol to classify transcriptions segments into categories with similar meaning. During the coding process, however, emerging categories that were not previously included in the initial protocol were also added. Therefore, the analysis was conducted following an inductive approach, by including categories that were identified in participants' speech, as well as a deductive approach that considered the coding protocol developed for this study [17]. The coding protocol was thus constantly reorganized throughout the data analysis process. Towards the end of the process, we started an aggregation effort in which each category was organized into higher-order specific dimensions, and then into more general dimensions [18].

4 Results

The questionnaire filled before the focus group showed that most riders used their motorcycle on both recreational and daily/commuting contexts (64%) and only 36% of the participants reported riding for recreational purposes only. The most common brand was Honda (32%), followed by Yamaha (24%), and BMW (16%).

Eleven participants have been riding for less than 5 years, and, among them, four were new riders, riding for less than one year. Five participants reported riding for between 5 and 10 years; five between 10 and 20 years; and four for over 20 years.

Twelve participants reported to ride the motorcycle at least four times per week, five between 1 and 3 times per week and eight a maximum of once per month. Most of the participants (60%) reported that their rides are usually short, taking on average between 15 min and one hour. Only 32% ($n = 8$) of the participants use a navigation system, with Google Maps being the most preferred application ($n = 6$). Most participants (64%) were not members of any community or motorcycle related group.

The results of the focus-group sessions were organized around three major themes that emerged during the qualitative analysis of the data: (1) planning a trip, (2) sharing information, and (3) communicating. For each theme we report the user's current practices and their needs.

4.1 Planning a Trip

Planning a motorcycle ride depends on several factors, such as whether it is a solo or group trip, whether it is a short or a long trip, and on the weather conditions. Current practices for trip planning seem to involve three major steps: (a) planning the route, (b) checking the motorcycle, and (c) preparing themselves (without a predefined sequence). When riding with friends the routes are chosen together, either in person or using communication applications, specially WhatsApp. To choose the route, riders rely on friends' knowledge of the surrounding areas, on web forums, and on their own searches on the map (physical or a mobile app). The most frequently used tools to plan the routes are mobile apps such as Google Maps to define interest points, restaurants, and meeting points. Some participants reported that the downside of using GPS-enabled navigation apps (as opposed to following the signs) is that sometimes travel planning is less detailed than the ideal because riders assume the app will be available at any moment:

P01, Male, 52 years-old: “*Yes, and it is funny, because I rode a motorcycle for so many years, and in longer trips when there was no GPS or mobile phones. It was a map, and things happened, and the trip was made. It was better prepared than now, we planned better. Now we rely on the GPS and then get used to it.*”

Using a GPS device can interfere with safety when riding a motorcycle, because riders have to either look away from the road to look at the device, or they have to follow audio cues, and both can potentially disrupt the riding experience. Moreover, some motorcyclists prefer not to use maps at all during the trip. Also, one mountain rider preferred using a paper-map, to avoid the risk of a technological failure:

P24, Male, 33 years-old: “*No, I don't use technologies much. (...) Or sometimes, when I go to more dangerous trails, I take a compass and a map, but they are physical because I do not want to fall, the cell phone breaks and then I end up with nothing. So, if the map tears, I can still see something ... See where I am located. But technology ... no. For what I do I don't want to depend on technology*”.

For those willing to use technologies, the navigation system seems to be the most important one, as far as it ensures safety:

P17, Male, 42 years-old: “*More than contextualizing specific solutions, I would like something that would make my trip especially safer, that is, all those alerts about potential problems on the road... more convenient, if I want to go to a given destination, and if I want to go fast, it provides the best route; But if I want to go for leisure, it provides the most scenic route.*”

4.2 Sharing Information

The second topic, willingness to share information with others, was especially important for our focus on social applications. Three major questions have emerged regarding information sharing: (1) who to share with, (2) what to share, and (3) what not to share.

Regarding whom, riders are generally available to share information with the community, so everyone can provide and receive useful information and benefit from it. Sharing information with manufacturers and brands would also be possible, but in those cases, participants expressed that they would be consider to be fair if they could also benefit from it:

P14, Male, 54 years-old: “*If there are companies that can benefit, I think they should also give something in return, e.g. a little check-up [for the motorcycle] or something.*”

On the one hand, companies could benefit by collecting data on motorcycle use, and by creating closer relationships with the customers. On the other hand, users could benefit from special offers or premium services.

Regarding what type of information motorcyclists would like to share and receive from others, navigation and safety were the most common topics. Motorcyclists believe that if all road users are connected, they should all benefit from receiving alerts of

proximity of other vehicles. Participants would also like to receive recommendations about the speed limit according to the type of pavement, points of interest along the route, closed roads, accidents, transit, etc. It would also be useful to have real-time weather information along the route. Concerning more specifically the roads, they could be categorized based on their scenery, and pleasure to drive (based on curves and straight roads). Also, these routes could be shared:

P05, Male, 25 years-old: *"If it could be connected to Facebook, or something. That there were people you know and who ride motorcycles too ... Yes, I don't know, if they were around, and your friend is riding a motorcycle or something, if you knew his location, things like that, even the itineraries, creating itineraries, sharing itineraries with others, or even executing those itineraries ...".*

When riding in groups, the app could also allow sharing photos within the group:

P16, Male, 29 years-old: *"It would be a sensor pack together with the application. Not for the driving experience itself but more for after [driving], to see the average consumption we had during the whole trip, the stops... But this is statistical data collection for the application itself, which is then good to add, one thing that is interesting to see... and there's even that part of google where you can share the photos with, if we have a group we can share if... look, we stopped at this viewpoint or something. You have some pictures, even of the motorcycle or something like that. And even share with the people who went on that trip, so everyone has access to everyone's photos.".*

When riding in groups, locating everyone would also be extremely useful:

P11, Male, 29 years-old: *"Share the location with friends, perhaps. Go meet a group of friends who also ride a motorcycle. They say, "we are here" and our GPS directly gives the indications to reach them."*

Despite all the interest around information sharing, participants have also expressed some concerns and some situations where they would prefer not to share at all. They mentioned that information should be anonymous, shared content should be curated, and should not depend exclusively on the community as information providers. Ideally, part of this could be accomplished with sensors in the vehicles and in the infrastructures, so data would be less biased, and riders would not have to manually add information while driving:

P23, Male, 26 years-old: *"The problem with Waze is that it needs to be the users to put the warning there. When we are riding a motorcycle or driving car it is not recommended to... [do it]."*

Also important, information sharing should always be optional:

P05, Male, 26 years-old: *"It could be our choice. Share or not."*

Fundamentally, participants referred two types of information they do not want to share: location and speed. Sharing the location should be done cautiously and only in

certain situations. In particular, when riding alone it should only be shared with trusted people. Also, the home location should never be exposed for security reasons. In general, participants do not want to share information that can be traced back to them.

P16, Male, 29 years-old: “*But with GPS... He left this point, went there and returned to this point, you're at home. (...) If you have GPS data, you usually can't identify who the person is, but if the starting point and the ending point are the same, it is immediately known that the person lives there. Or keeps the motorcycle there, at least.*”

Concerning speed, it could be shared only if it were anonymous and no negative consequences could come from that. Participants also emphasized that too much information can be distracting and especially on leisure trips, it can also interfere with the driving experience:

P16, Male, 29 years-old: “*A lot of information when you are taking a longer trip is going to distract of what you want.*”

4.3 Communicating

Participants often referred to various communication contexts, which mainly seem to address three types of communicating needs: (1) with people that are not riding with them (e.g., a family member at home), (2) with friends riding with them in a group, and (3) with the passenger in the motorcycle. Usually, in the first two cases conversations are short (e.g., acknowledging that she/he will be late, or signaling to stop for gas, respectively). In the last case, communication may be more continuous:

P18, Male, 50 years-old: “*If you go in the car with your wife or your friend or whatever, you like to talk, look at the landscape. And in the motorcycle is the same. A person going there quietly for 200 km is a bit boring.*”

Communicating while riding can lead to dangerous situations. For example, when riding in a group, motorcyclists may try to communicate with the motorcycle light signals (e.g., using the right light signal to communicate the intent to stop) or take over everyone and shout a short message (e.g. “Stop!”). Both situations can be dangerous:

P17, Male, 42-years-old: “*When it is a group tour, sometimes it is like, the guy who comes back needs to put gasoline and either gives light signals or have to use the turn lights because there is no easy way to communicate. Or he overtakes us all and then says “Hey, I want to stop”.*”

P14, Male, 54 years-old: “*Yes, but it ends up creating a risky situation.*”

P17, Male, 42 years-old: “*Communication is not effective.*”

P14, Male, 54 years-old: “*We shouldn't be doing this when traveling in a group. We should maintain our position.*”

Motorcyclists also referred communication difficulties while driving. On the one hand, manipulating a device is impossible or dangerous while driving.

P02, Male, 31 years-old: “*I already did that, using applications to connect mobile phones via Bluetooth (...) for example with the passenger. For example, making a call ... There are applications for calls and a person speaks, but it ends up being distracting at the same time. So, it works, but at the same time, it distracts us.*”.

On the other hand, the existing communication devices, such as hands-free kits, sometimes are too expensive and when riding in groups not everyone has one:

P15, Female, 26 years-old: “*The problem is that with some people having intercoms and others not, there is always someone who is left behind, and others that go a little bit faster, and we end up being out of sync.*”

One possible solution to ease communication could be to have voice interaction while riding:

P17, Male, 42 years-old: “*I don't want to see the messages while driving because it is dangerous, but I would like to have someone reading them for me or something like that... check if I have a call waiting, something like that (...) By sound. Something like that.*”.

4.4 Other Results

This study has also identified several other relevant insights for the design of a social application. The first is the central role of safety. Even though the topic of the study was presented as being social applications, the issue of safety remained clearly amongst the top priorities. Also, many participants seemed to exhibit a general disbelief in digital technologies, at least for the riding context. While part of the problem may be associated with safety risks derived from technology use during the ride, there also seems to be some sort of disillusion with current products, particularly when compared with the thriving industry market of digital technologies for the car. Several participants have even expressed that they would just prefer not to use any digital technology at all, simply for fear that it will ruin the pleasure of riding.

5 Conclusion

The concept of social applications for motorcyclists can be a good fit for the socializing component of being a rider. In this study, we have explored the motorcyclists' perception about this type of digital tools. The results revealed several challenges, such as interaction limitations, the possible impact on the riding experience, the very diverse set of motorcyclist profiles, or simply a general disbelief in digital tools altogether. On several occasions, participants have shown signs of resistance to technology and emphasized that information should be limited, not only for safety reasons, but also because it could interfere with the pleasure of driving. Altogether, the present results may inform future participatory design studies, pointing the main directions for social applications and highlighting the major challenges that need to be considered to ensure acceptability. Importantly, in the present study most participants were men. Even though most

motorcyclists are men, the underrepresentation of women in the present study can have produced biased results. Thus, prior to developing a digital solution for motorcyclists, it is fundamental to understand the needs of all types of motorcyclists.

The general conclusion is that the exploration of design opportunities for a social application for motorcyclists must be much more than a mere identification of relevant features or a simplistic attempt of recreating, in this context, the principles and concepts of other social networks. Like social networks in general, social applications for motorcyclists need to leverage the collective and socializing elements of motorcycling in a way that fits the different needs of various motorcyclists' profiles. However, their specific design should also be closely aligned with the micro-contexts of the riding experience in a way that blends smoothly with riding itself.

Acknowledgements. This work is supported by: European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project n° 039334; Funding Reference: POCI-01-0247-FEDER-039334].

References

1. Tunnicliff, D., Watson, B., White, K.M., Lewis, I., Wishart, D.: The social context of motorcycle riding and the key determinants influencing rider behavior: a qualitative investigation. *Traffic Inj. Prev.* (2011). <https://doi.org/10.1080/15389588.2011.577653>
2. Jderu, G.: Motorcycles, body and risk: the motorcyclists' social career. *J. Sociol.* (2015). <https://doi.org/10.1177/1440783312474081>
3. Maxwell, A.H.: Motorcyclists and community in post-industrial urban America. *Urban Anthropol.* **27**, 263–279 (1998)
4. Esbjörnsson, M., Juhlin, O., Östergen, M.: Motorcycling and social interaction -design for the enjoyment of brief traffic encounters. In: Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work, pp. 85–94 (2003)
5. Coninx, P.: Riding in big brother's automobile: vehicle technology and consumer privacy. *Automobile Consumer Coalition, Car Help Canada* (2011)
6. Spelta, C., Manzoni, V., Corti, A., Goggi, A., Savaresi, S.M.: Smartphone-based vehicle-to-driver/environment interaction system for motorcycles. *IEEE Embed. Syst. Lett.* **2**, 39–42 (2010). <https://doi.org/10.1109/LES.2010.2052019>
7. Lin, H.M., Tsai, H.M., Boban, M.: Scooter-to-X communications: antenna placement, human body shadowing, and channel modeling. *Ad Hoc Netw.* **37**, 87–100 (2016). <https://doi.org/10.1016/j.adhoc.2015.09.006>
8. Silva, R., Iqbal, R.: Ethical implications of social internet of vehicles systems. *IEEE Internet Things J.* **6**, 517–531 (2019). <https://doi.org/10.1109/JIOT.2018.2841969>
9. Esbjörnsson, M., Juhlin, O., Östergren, M.: Traffic encounters and Hocman: associating motorcycle ethnography with design. *Pers. Ubiquitous Comput.* **8**, 92–99 (2004). <https://doi.org/10.1007/s00779-004-0260-4>
10. Bagozzi, R.P., Dholakia, U.M.: Antecedents and purchase consequences of customer participation in small group brand communities. *Int. J. Res. Mark.* **23**, 45–61 (2006). <https://doi.org/10.1016/j.ijresmar.2006.01.005>
11. Laroche, M., Habibi, M.R., Richard, M.-O., Sankaranarayanan, R.: The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. *Comput. Human Behav.* **28**, 1755–1767 (2012). <https://doi.org/10.1016/j.chb.2012.04.016>

12. Madupu, V., Krishnan, B.: The relationship between online brand community participation and consciousness of kind, moral responsibility, and shared rituals and traditions. In: Advances in Consumer Research, pp. 853–854 (2008)
13. De Gruyter, C., Truong, L.T., Nguyen, H.T.T.: Who's calling? Social networks and mobile phone use among motorcyclists. *Accid. Anal. Prev.* **103**, 143–147 (2017). <https://doi.org/10.1016/j.aap.2017.04.010>
14. Truong, L.T., De Gruyter, C., Nguyen, H.T.T.: Calling, texting, and searching for information while riding a motorcycle: a study of university students in Vietnam. *Traffic Inj. Prev.* **18**, 593–598 (2017). <https://doi.org/10.1080/15389588.2017.1283490>
15. Truong, L.T., Nguyen, H.T.T., De Gruyter, C.: Correlations between mobile phone use and other risky behaviours while riding a motorcycle. *Accid. Anal. Prev.* **118**, 125–130 (2018). <https://doi.org/10.1016/j.aap.2018.06.015>
16. Phommachanh, S., Ichikawa, M., Nakahara, S., Mayxay, M., Kimura, A.: Student motorcyclists' mobile phone use while driving in Vientiane. Laos. *Int. J. Inj. Contr. Saf. Promot.* **24**, 245–250 (2017). <https://doi.org/10.1080/17457300.2016.1166141>
17. Schilling, J.: On the pragmatics of qualitative assessment. *Eur. J. Psychol. Assess.* **22**, 28–37 (2006). <https://doi.org/10.1027/1015-5759.22.1.28>
18. Vaismoradi, M., Turunen, H., Bondas, T.: Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. *Nurs. Health Sci.* **15**, 398–405 (2013). <https://doi.org/10.1111/nhs.12048>



Maximizing Sensors Trust Through Support Vector Machine

Sami J. Habib^(✉) and Paulvanna N. Marimuthu

Computer Engineering Department, Kuwait University, P. O. Box 5969, 13060 Safat, Kuwait
sami.habib@ku.edu.kw

Abstract. Trust has become a major concern in wireless sensor networks (WSN), since many WSNs are deployed in data sensitive applications, especially in health-care and surveillance. Factors, such as the sensitivity to internal and external noises, constraints on sensing devices, varying deployment platforms and network topologies, generate uncertainty in data accuracy and consistency. With the advent of smart sensing solutions, the data accuracy is increased to a greater extent. However, explicit discrimination of uncertainty from the sensor data is still remain challenging, as it is difficult to quantify the individual impact. The environmental uncertainty is one of the difficult parameter to quantify and to predict; moreover, it greatly influences the received signal strength (RSS) in outdoor sensor networks, leveraging frequent data inconsistency ended up with sensor distrust. We propose a framework to maximize sensors' trust by classifying the level of impact under the existence of a few environmental uncertainties, such as temperature, humidity and wind speed. We have applied multiclass support vector machine (SVM) classifier to analyze RSS of sensor under the individual and combined presence of defined environmental uncertainties and to classify the dataset into four groups; moreover, the penalty corresponding to the level of uncertainty is added to boost the sensor trust. We have selected Quadratic SVM to train the dataset, as the data varied non-linearly. The experiment shows 97% accuracy during training and 96.2% accuracy during testing with 3.8% misclassifications. With these predicted level of uncertainties and corresponding boosting in RSS, the framework is found to move 42% of sensors from uncertain to trusted category.

Keywords: Supervised learning · Support vector machine · Environment uncertainties · Maximization · Sensor trust · Wireless sensor networks

1 Introduction

The widespread applications of sensors in location monitoring or surveillance demand accurate and reliable data. Accuracy is defined as the deviation of measured value from true value, whereas reliability reflects the trust on the source of information. Data accuracy is affected by sensor's calibration, aging of components, poor maintenance and so on. However, the inconsistent behavioural changes, due to unpredicted weather or malfunctioning of a part, may affect sensor reliability. Ferson and Ginzburg [1] classified uncertainty into two: objective and subjective. The objective uncertainty is due to

the stochastic behaviour of sensing system, whereas the subjective uncertainty is due to incomplete knowledge of the system. It is difficult to avoid the objective uncertainty, since the volatility in sensing output due to environmental uncertainties cannot be predicted completely in advance, as the environmental behaviour is not in human control. Such stochastic variations in the reporting data, leveraging false belief about the source. By carefully analyzing the possible sources of uncertainties and quantifying their impact on sensing process, the sensors trust value may be maximized.

Sensors trust is defined using trust factors directly associated to sensor's observed behaviour or indirectly estimated from third party recommendations. The existing trust models classify sensor's reliability as trusted or untrusted. For a sensor deployed for outdoor monitoring, harsh weather conditions play a vital role in producing erratic sensing data, thereby leading to misclassification of sensor trust. There are uncountable sources of uncertainty in the surroundings that may cause variations in sensing data; it is very challenging to quantify the impact from the measured data. In this work, we have considered temperature, humidity and an unexpected occurrence of strong wind (storm) are the sources of uncertainties. The variation in temperature, and humidity increase the resistance of the transmission medium by producing variations in received signal strength (RSS) [3]. The strong wind changes the orientation of receiving or transmitting antenna, which causes change in polarization that impact the performance of the wireless sensor networks (WSN) [2]. A relative measure of angle of deviation with respect to normal position is defined using the changes in RSS measurement.

Many strategies are being followed to minimize the uncertainties. Machine learning is one of the strategy, which utilizes various learning algorithms to understand the trends and patterns present in the given dataset and to classify the dataset under various categories. It will be helpful for the sensors deployed in unmanned outdoor environment to analyze the changes in the performance without human interventions. Thus, sensors may be embedded with a suitable machine learning algorithm to comprehend the changing pattern in their performance from normal. The training dataset comprising of uncertainties and the impact on sensor performance may be generated using prior experimental data combined with human experience. We selected support vector machine (SVM), a supervised training algorithm to train data, as it works well with unstructured data. By using a labelled dataset, SVM may be trained to realize the presence of defined uncertainties and their impact on RSS, and to classify the dataset into multiple categories accordingly, which is then utilized to maximize sensors' trust values.

In this paper, we have utilized the changes in RSS (data consistency) and the retransmission history as two measured parameters to derive trust factors under the selected uncertainties, such as temperature, humidity and wind speed. We have generated the dataset by extrapolating data from previous experimental studies [3, 21], which demonstrated variations in RSS against orientation angle deviation, and against varying temperature and humidity respectively. We have proposed an intelligence framework to classify the uncertainties, which utilizes a multiclass Quadratic support vector machine to train and to analyze the behaviour pattern; the dataset is classified into four groups based on the uncertainties level. Class 4 is categorized as the worst case, which includes the presence of all uncertainties and class 1 is defined as the best with no uncertainties. The trust values calculated using the defined trust factors are improvised by adding a

RSS value proportional to the severity level of uncertainties. Our experimental results demonstrated that the quadratic SVM produced 96.2% accuracy during testing, and by predicting the uncertainty levels, the framework is able to move 42% of sensors from uncertain to trusted category.

2 Related Work

The increase in the deployment of sensors in various autonomous systems and data sensitive applications raise issues about the trust of the source. The traditional trust management are based on subjective and objective belief [4] and are focusing authenticating user's identity, validation of host and so on, from third party recommendations. With the introduction of sensors in monitoring industries, the trust computation is extended to include specialized measured parameters, such as calibration, group deviation, true deviation [5, 6] and so on. Habib and Marimuthu [7] developed a trustworthy self-adaptive framework, where they utilized trust as a control parameter to trigger the adaptive techniques. Guo et al. [8] analyzed the trust evaluation methods for managing services in IoT, whereas Bansal and Kohli [9] carried out website evaluation for ensuring the trust.

Uncertainty measurements were mainly focused on calibration error, measurement error, and transmission error. Li et al. [10] carried out a survey on theory and practices of uncertainty, whereas Movahednia et al. [11] developed an optimized energy management scheme for Microgrids with uncertainties.

Supervised learning is employed in trust evaluation to separate malicious nodes from trusted nodes [12, 13], where the authors carried out binary classification using support vector machine (SVM). Lopez and Maag [14] used a multiclass classification SVM to develop a generic trust management framework. Eziamah et al. [15] performed a comparative analysis using Naïve Bayes, Decision tree, radical SVM, and random forest machine learning techniques to distinguish malicious and trusted nodes. Furthermore, machine learning algorithms were used to test the trust of social networks [16], and to locate the node misbehaviour [17–20].

In this work, we have exploited quadratic SVM to analyze the existence of selected environmental uncertainties and to perform multiclass classification of the uncertainties to improve sensors' trust.

3 Maximizing Trustworthiness

There are innumerable source of uncertainties existed for WSNs; data trust is still remain challenging as weather, calibration, interferences, obstacles, sensing hardware failure, aging and so on, causing error in sensor data, thereby reducing the trustworthiness of the sensors. It is really challenging to distinguish the uncertainties directly from the sensors' outputs, as the uncertainties are inseparably mixed with the sensor output. One of the strategy is by employing specialized monitoring sensors, the impact of uncertainties on the performance may be quantified. We have derived a mathematical model to maximize sensors trustworthiness, which attempts to learn the uncertainties and their impact on sensing data, and to classify the dataset into multiple groups accordingly. We have selected the uncertainties, whose impacts on signal strength are quantifiable.

We have considered three uncertainties, such as temperature, humidity, and wind speed with direction of flow, causing deviations in sensor measurements. We have assigned a binary mapping to classify their existence into four categories, as listed in Table 1. Here, the first column lists the presence of possible combinations of uncertainties and the second column lists the classification of dataset according to the uncertainties. The last column tabulates the assigned penalties corresponding to uncertainties. We have assigned a penalty for the presence of uncertainty, which is estimated from the change in signal strength from normal, as shown in Eq. (1), and the last column lists the improved trustworthiness after deducting the penalties from the trust value calculations. Here, the indices i, j and k represent the defined uncertainties, the term N represents the normal weather, and the parameter $P(s_i)$ represents the penalty for sensor s_i . The term β represents the weightage to each uncertainty, and we have assigned 40%, 30% and 30% to wind speed, humidity and temperature respectively from the observation of the experimental data.

$$P(s_i) = \beta_1 * \frac{\partial RSS_{U_i}}{\partial RSS_N} + \beta_2 * \frac{\partial RSS_{U_j}}{\partial RSS_N} + \beta_3 * \frac{\partial RSS_{U_k}}{\partial RSS_N} \quad (1)$$

Initially, the transmitter and the receiver antennas are placed inline to receive the data and the deviation in orientation due to strong wind has substantial impact on the received signal strength (RSS), according to [2]. A strong wind in a direction either normal to or at oblique incidence may tilt the antenna position, thus causing error in measurement, and the sensor data deviates from its neighbours leading to reduction in performance and trust.

Table 1. Presence of uncertainties and the corresponding penalties.

Uncertainties			Category	Penalty
Temperature	Humidity	Wind speed		
0	0	0	Normal	0
0	0	1	Semi-normal	P1
0	1	0		P1
1	0	0		P2
0	1	1	Semi-worst	P1+P2
1	0	1		P1+P2
1	1	0		P1+P2
1	1	1	Worst-case	P1+P2+P3

3.1 Selection of Uncertainties

We have explored the uncertainty subspaces to figure out the individual and cumulative contribution of uncertainties to sensors' untrustworthiness. It is observed from

the experimental data that the signal strength is inversely proportional to the humidity and temperature. According to [3], increase in temperature increases the resistance of the transmission medium, whereas increase in humidity increases the refraction and reflection (scattering), thereby reducing the signal strength.

The change in orientation of either the transmitter or receiver will affect the polarization of the deflected antenna, which decreases the received signal strength. The wind flowing normal to the antenna will produce certain angular deviation in the orientation based on its speed; the angle of deviation ranges from 90 to -90° , where the positive sign represents the direction towards the transmitter and the negative sign represents the direction away from it. A sample of antenna deviations against the wind speed along with its direction of flow based on generalized observations is illustrated in Fig. 1.

3.2 Selection of Trust Factors

We have selected two measured parameters, such as retransmission history and data consistency to compute the sensor trust as shown in Eqs. (2) and (3), which are closely related to the selected environmental uncertainties in Eq. 1. The increase in number of retransmissions and data inconsistency are observed during strong winds, high humidity and high temperatures. In Eq. (2), the parameter $\text{Re}_i(ts)$ represents the number of retransmissions of i th sensor at sampling time instance ts , and m is the sample length. The retransmission history and the data consistency are checked for a selected timespan, which is defined using the sample length m . In Eq. (3), the parameter $RSS_i(ts)$ represents the received signal strength of i th sensor at sampling instance ts .

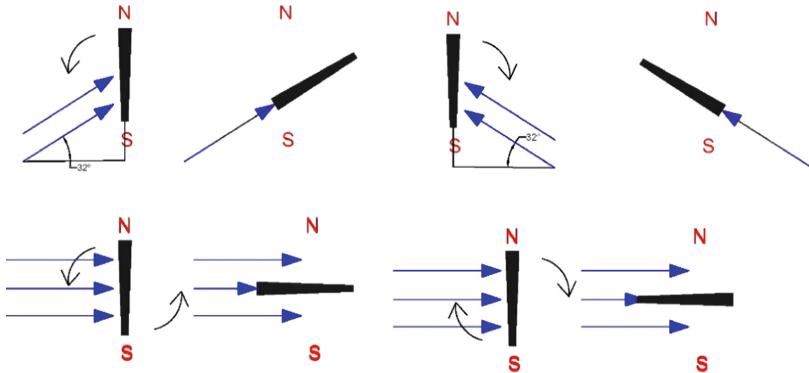


Fig. 1. Wind directions and its impact on antenna orientation.

The trust factor is represented as

$$tr_1 = \alpha_1 * \left(\frac{\sum_{j=1}^m \text{Re}_i(t_s-j)}{\text{Re}_i(t_s)} \right) \quad (2)$$

And the second trust factor is

$$tr_2 = \alpha_2 * \left(\frac{RSS_i(t_s) - \frac{\sum_{j=1}^m RSS_i(t_s-j)}{m}}{RSS_i(t_s)} \right) \quad (3)$$

By combining the two defined trust factors, the trust value of a sensor is given by Eq. (4).

$$Tr(s, ts) = (tr_1) + (tr_2) \quad (4)$$

4 Proposed Methodology

The proposed methodology to maximize sensor's trust is illustrated in Fig. 2, which is comprised of three procedures: generation of training and testing dataset, uncertainties classification using supervised learning, and reevaluation of trust by adding a boost value in proportion to the impact on RSS. The dataset is comprised of the following features: temperature, humidity, and the angle of deviation, as uncertainty parameters, the corresponding RSS values with the presence and absence of uncertainties, and a label to the uncertainties to classify the dataset into four categories as illustrated in Table 1. The framework checks the log of parameters, which facilitated the awareness of environmental uncertainty, and the trust factor is estimated prior to and after the presence of uncertainties. Then, the penalty corresponding to the uncertainties are added to increase the trust of the sensor.

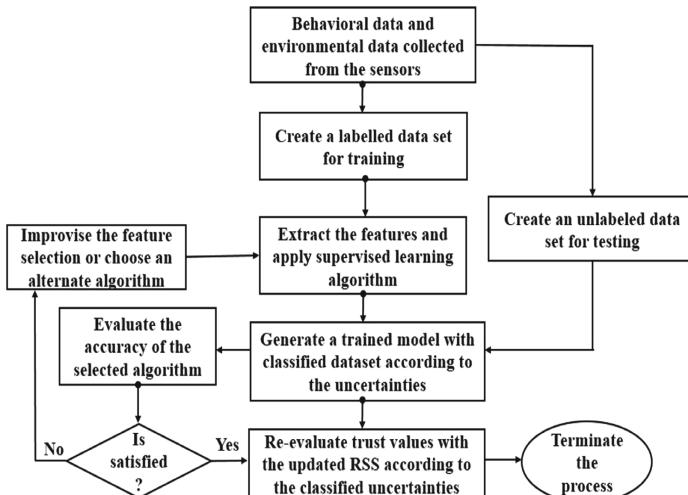


Fig. 2. Proposed framework.

The SVM utilizes one-against-one construction method to solve multiclass classification problem, where each classifier separates data of two different classes and all classifiers (based on the number of classification) together generate a multiclass classification. SVM uses a hyper-plane to separate the points in the dataset, as represented in Eq. (5). Here, the x_i is input vector, T is the transpose operator, w is the weight vector and b is the bias.

$$w^T * x_i + b = 0 \quad (5)$$

5 Results and Discussion

We coded the proposed trust maximization problem in Matlab platform utilizing machine learning toolbox. We have considered a set of sensor nodes within WSN deployed for outdoor monitoring; we have utilized the datasets [3, 21] to study the changing pattern of RSS against the uncertainties, and generated an augmented dataset with the combination of following features: temperature, humidity, wind speed, orientation of the antenna, and received signal strength (RSS). We categorized the uncertainties into four levels, according to Table 1 and assigned penalties according to Eq. (1), to increase RSS. We selected optimized support vector machine to train the dataset and to classify the dataset according to four level of uncertainties, where we managed with 97% of accuracy. Figure 3 demonstrates the given uncertainty class (column_1) against the predicted class (column_2), where the colors of the data points represents the label (class1 to class 4, as listed in Table-1) of the predicted class. Class-1 represents RSS with no uncertainties, class-2 with the presence of any one and class-3 with the presence of any two and class-4 with all three uncertainties.

For the given a list of expected values (true class) and a list of predictions (predicted class) from SVM model, the confusion matrix in Fig. 4 shows the performance of SVM for the training data.

The training dataset has a input matrix of size (300×6) , where the row specifies the number of observations and column specifies the features used to train the model. In the confusion matrix, the diagonal elements represents the correct classifications and the remaining elements in each row represents the misclassification against each class.

The predictions on test dataset of size (52×5) is shown by the confusion matrix in Fig. 5. Here, the target class is the actual classification and the output class in the predicted classification by SVM. The diagonal elements represents the correctly classified percentage of each class and the last column lists the percentage of correct and incorrect classification. After classifying the uncertainty, RSS signal strengths are boosted based on the predicted level of uncertainty, as listed in Table 2; we found a total of 42% shift in semi-trusted and uncertain states to higher level, where 23% of sensors are shifted to level-1.

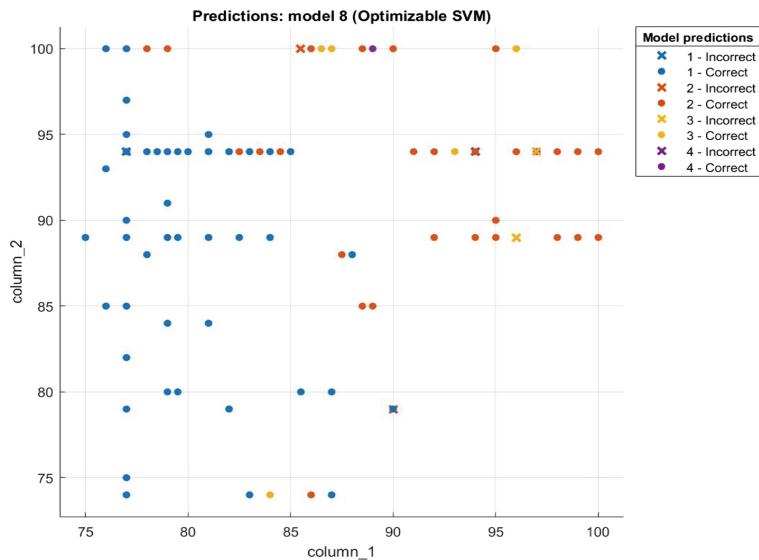


Fig. 3. Input vector against the predicted uncertainty level.

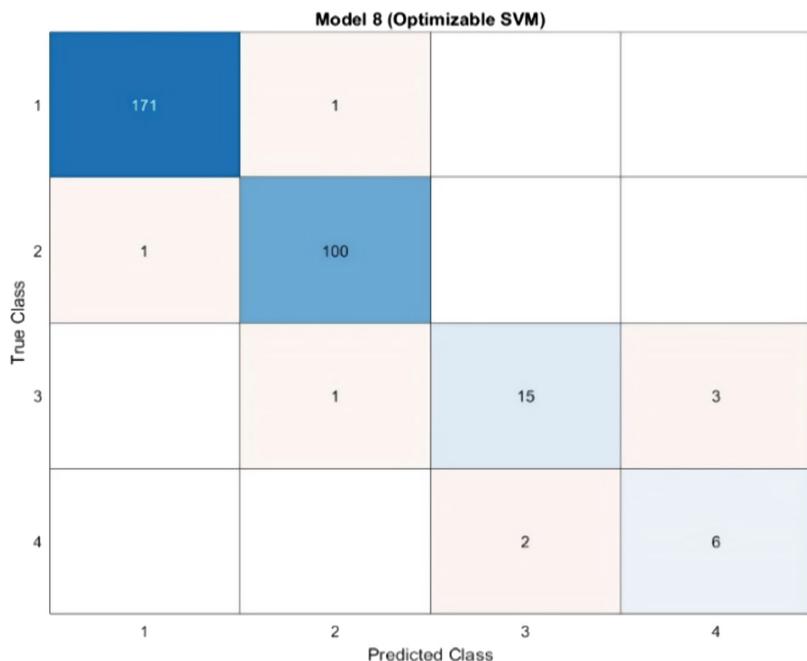


Fig. 4. Confusion matrix (training data).

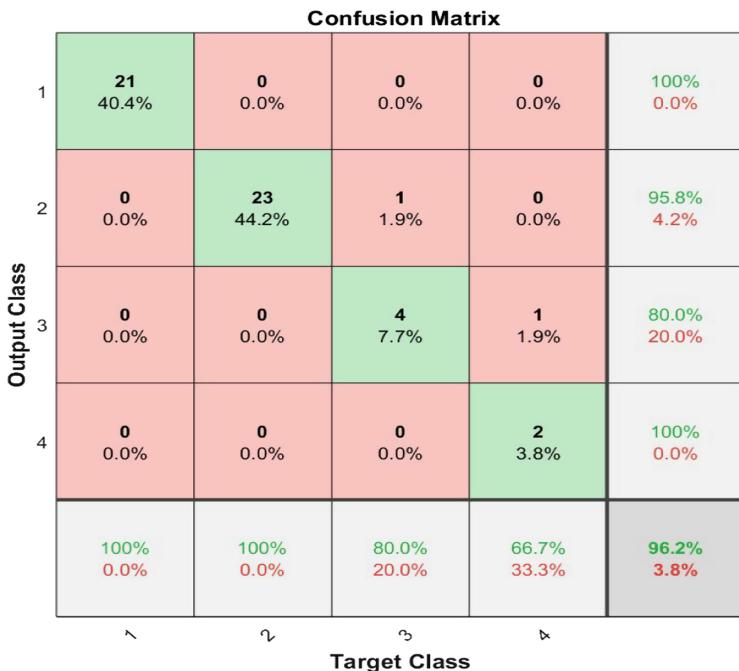


Fig. 5. Confusion matrix (testing data).

Table 2. Sensors and their trusted state.

RSS range	Trust-state	Number of sensors (Actual)	Number of sensors (boosted)
80–100	Trusted	9	21
60–80	Semi-trusted	23	23
40–60	Uncertain	16	5
<40	Untrusted	4	3

6 Conclusion

We proposed a framework to maximize sensors' trust by classifying the level of a few environmental uncertainties, such as temperature, humidity and wind speed based on their impact on received signal strength (RSS). A multiclass quadratic support vector machine (SVM) classifier is utilized to study the individual and combined presence of defined environmental uncertainties and to classify the dataset into four levels. With these predicted level of uncertainties and corresponding boosting signal strength to actual RSS, the framework is found to move 42% of sensors from uncertain and semi-trusted to trusted category. The experiment shows 97% accuracy during training and 96.2% accuracy during testing with 3.8% misclassifications.

We are continuing our research with more observations and features, and training using other supervised learning algorithms to improve the precision.

Acknowledgement. This work was supported by Kuwait University under a research grant no. QE02/17.

References

1. Ferson, S., Ginzburg, L.R.: Different methods are needed to propagate ignorance and variability. *Reliab. Eng. Syst. Saf.* **54**, 133–144 (1996)
2. Wadhwa, M., Song, M., Rali, V., Shetty, S.: The impact of antenna orientation on wireless sensor network performance. In: Proceedings of IEEE International Conference on Computer Science and Information Technology, Beijing, China, 8–11 August (2009)
3. Luomala, J., Hakala, I.: Effects of temperature and humidity on radio signal strength in outdoor wireless sensor networks. In: The Proceedings of Federated Conference on Computer Science and Information Systems, Lodz, Poland, 13–16 September (2015)
4. Khalid, O., Khan, S.U., Madani, S.A., Hayat, K., Khan, M.I., Min-Allah, N., Kolodziej, J., Wang, L., Zeadally, S., Chen, D.: Comparative study of trust and reputation systems for wireless sensor networks. *Secur. Commun. Netw.* **6**, 669–688 (2013)
5. Habib, S.J., Marimuthu, P.N.: Reputation analysis of sensors' trust within tabu search. In: The Proceedings of World Conference on Information Systems and Technologies, Madeira, Portugal, 11–13 April (2017)
6. Boudriga, N., Marimuthu, P.N., Habib, S.J.: Measurement and security trust in WSNs: a proximity deviation based approach. *Ann. Telecommun.* **74**(5–6), 257–272 (2019)
7. Habib, S.J., Marimuthu P.N.: Development of trustworthy self-adaptive framework for wireless sensor networks. In: The Proceedings of World Conference on Information Systems and Technologies, Budva, Montenegro, 7–10 April 2020
8. Mohammadi, V., Rahmani, A.M., Darwesh, A.M., Sahafi, A.: Trust-based recommendation systems in internet of things: a systematic literature review. *Hum. Cent. Comput. Inf. Sci.* **9**(21) (2019)
9. Bansal, H., Kohli, S.: Trust evaluation of websites: a comprehensive study. *Int. J. Adv. Intell. Paradigms* **13**(1–2), 101–115 (2019)
10. Li, Y., Chen, J., Feng, L.: Dealing with uncertainty: a survey of theories and practices. *IEEE Trans. Knowl. Data Eng.* **25**(11), 2463–2482 (2013)
11. Movahednia, M., Karimi, H., Jadid, S.: Optimal hierarchical energy management scheme for networked microgrids considering uncertainties, demand response, and adjustable power. *IET Gener. Transm. Distrib.* **14**(20), 4352–4362 (2020)
12. Li, W., Joshi, A., Finin, T.: ATM: automated trust management for mobile ad hoc networks using support vector machine. In : the Proceedings of IEEE 12th International Conference on Mobile Data Management, Lulea, Sweden, 6–9 June (2011)
13. Fu, J., Xue, J., Wang, Y., Liu, Z., Shan, C.: Malware visualization for fine-grained classification. *IEEE Access* **6**, 14510–14523 (2018)
14. López, J., Maag, S.: Towards a generic trust management framework using a machine-learning-based trust model. In: The Proceedings of IEEE Trustcom/BigDataSE/ISPA, Helsinki, Finland, 20–22 August, pp. 1343–1348 (2015)
15. Eziamaka, E., Jaimes, L.M.S., James, A., Nwizege, K.S., Balador, A., Tepe, K.: Machine learning-based recommendation trust model for machine-to-machine communication. In: The Proceedings of IEEE International Symposium on Signal Processing and Information Technology, Louisville, KY, USA, 6–8 December, pp. 1–6 (2018)

16. Khadangi, E., Bagheri, A.: Comparing MLP, SVM and KNN for predicting trust between users in facebook. In: ICCKE 2013, Mashhad, Iran, 31st October–1 November, pp. 466–470 (2013)
17. Parihar, R., Jain, A., Singh, U.: Support vector machine through detecting packet dropping misbehaving nodes in MANET. In: The Proceedings of International Conference of Electronics, Communication and Aerospace Technology, 20–22 April, Coimbatore, India, pp. 483–488 (2017)
18. Chkirkene, Z., Erbad, A., Hamila, R., Gousssem, A., Mohamed, A., Guizani, M., Hamidi, M.: Weighted trustworthiness for ML based attacks classification. In: The Proceedings of IEEE Wireless Communications and Networking Conference, Seoul, South Korea, 25–28 May (2020)
19. Zhang, C., Chen, K., Zeng, X., Xue, X.: Misbehaviour detection based on support vector machine and dempster-shafer theory of evidence in VANETs. *IEEE Access* **6**, 59860–59870 (2018)
20. Chang, C.-C., Chien, L.-J., Lee, Y.-J.: A novel framework for multiclass classification via ternary smooth support vector machine. *Pattern Recogn.* **44**(6), 1235–1244 (2011)
21. Anderson, E.W., Phillips, C.: CRAWDAD dataset/cu/antenna (v.2009-05-08), traceset: rss. <https://crawdad.org/cu/antenna/20090508/rss>



Spreading Factor Analysis for LoRa Networks: A Supervised Learning Approach

Christos Bouras^{1(✉)}, Apostolos Gkamas²,
Spyridon Aniceto Katsampiris Salgado^{1(✉)}, and Nikolaos Papachristos^{1(✉)}

¹ Computer Engineering and Informatics, University of Patras, Patras, Greece
bouras@cti.gr

² University Ecclesiastical Academy of Vella, Ioannina, Greece

Abstract. Today, the Internet of Things (IoT) has been introduced in our lives, giving a variety of solutions and applications. The critical requirements for devices connected to the IoT are long battery life, long coverage, and low deployment cost. Some applications require the transmission of data over long distances, thus Low Power Wide Area Networks (LPWAN) have emerged, with LoRa being one of the most popular players of the market. In order, to improve energy consumption and connectivity problems, machine learning can be used in LoRa networks. In this paper, we intend to improve the energy consumption of end nodes by using machine learning models. For this reason, we present a comparison of classification algorithms, specifically, the k-NN, the Naïve Bayes, and Support Vector Machines (SVM), for the Spreading Factor (SF) assignment in LoRa networks. The simulation results indicate that, both energy efficiency and reliability in IoT communications could be significantly improved using the proposed learning approach. These promising results, which are achieved using lightweight learning, make our solution favorable in many low-cost low-power IoT applications.

Keywords: LPWAN · LoRa · IoT · Machine learning

1 Introduction

As the Internet of Things (IoT) market is growing with fast rates, intending to solve major problems such as, climate change e.g. by monitoring crucial areas, or to provide Search and Rescue Systems [1] new technologies have been introduced. One of these technologies is Low Power Wide Area Networks (LPWAN). LPWAN comes to solve the problem of transmitting data to long distances, with very small energy consumption. Some examples of LPWAN technologies are LoRa, Narrowband IoT (NB-IoT), SigFox, Weightless [2]. Each technology, has its advantages and disadvantages, tries to provide energy-efficient, long-distance, low-cost solutions, sacrificing high throughput, and low latency similar to what cellular technologies provide. As mentioned before, IoT tries to cope with different parameters in the context of the application. One of the most important challenges that should be taken into consideration during the development of a system, is the appropriate resource allocation. Resource allocation can be focused on energy

consumption, latency, throughput, packet loss, etc. Many techniques have been proposed for resource allocation, one of them being the use of Machine Learning (ML). The rest of this work is organized as follows: In the next section related work is being described. We briefly discuss the LoRa technology in Sect. 3. Section 4 refers to the architecture of our ecosystem while Sect. 5 describes and analyzes the Machine Learning Integration perspective in our research study. Section 6 includes the performance evaluation of the proposed mechanism. Section 7 concludes our study and presents our future work.

2 Related Work

LPWAN networks have recently been attracting great attention in the IoT community. As energy consumption and the connectivity with Network Server (NS) should be of great importance, we should be led to the integration of machine learning solutions that could offer improvement to the specific issue. Several studies have been proposed and studied network simulators and tools to replicate real network operations without the need for real hardware. In [3], the authors present the most important LoRa simulation environments available in the literature and after that a comparative evaluation of LoRa simulation environments. The benefits, the disadvantages, and the highlights of each LoRa simulation environment are also presented. The reference above led us to the choice the FLORA simulator for this research work. In [4] authors have implemented a classification IoT system that benefits from the use of LoRa and embedded ML using k-NN (k-Nearest Neighbors) [5] classification. The scope of this research work is to reduce power consumption and increase battery life of IoT devices. Artificial neural networks for wearable devices can greatly improve detection and data analyses. Moreover, [6] aims to push beyond the current power walls for neural networks and move toward a micro-power neural network. This requires working on algorithms, architecture, circuits.

In [7], authors have conducted a survey of machine learning algorithms importance for IoT data analysis. They suggest that the big data generated by the IoT devices can lead to a tremendous benefit to human, when machine learning is combined, making the IoT applications smarter. Also, a comprehensive explanation of supervised, unsupervised, and reinforcement learning algorithms is presented, and in which application each algorithm is better suited is presented as well. Moreover in [8], an irrigation system is proposed focusing on the advantages of LoRa technology with ML. The results showed that this combination can lead to a smart and accurate irrigation system that can be widely used in agriculture. Another domain that ML can be combined with LoRa technology is geolocation. In paper [9], ML techniques such as Random Forests and neural networks are introduced to deal outdoor geolocation. The results were promising.

This paper focuses on LPWAN and LoRa, which provides good performance in terms of reliability and energy consumption. For this reason, we examine the feasibility of using ML classification algorithms, such as k-NN for the assignment of Spreading Factor (SF) in LoRa networks. From the literature, it is known that SF is a very important parameter for LoRa operation. SF value increase leads to an increase of the airtime and an energy consumption [10]. Using ML, we will be able to extract the appropriate - ideal SF to be used by the NS. A network architecture contains end-devices, gateways, and a NS, forming a star topology. It operates at unlicensed frequency ISM (Industrial,

Scientific, and Medical) bands of 868 MHz and 915 MHz in Europe and the U.S., respectively. In this research three classification algorithms are examined: the k-Nearest Neighbors (k-NN), the Naïve Bayes classifier, and the Support Vector Machine (SVM). The classifiers' learning phase is assumed to occur in the NS of the LoRa network. This assumption is realistic, as the NS is responsible for the network configuration. Also, the de facto LoRaWAN policy for the SF assignment called Adaptive Data Rate (ADR) is running on the NS [11], too.

3 LoRa Technology

LoRa is a physical layer LPWAN solution, which is a derivative of Chirp Spread Spectrum (CSS). LoRa constitutes a spread spectrum technique designed to work in 433 MHz, 868 MHz, and 915 MHz. LoRa has shown resistance against the Doppler Effect and multipath fading.

In a typical LoRa deployment there are three main devices: LoRa end-nodes, which acquire data from sensors at the application layer and send these data LoRaWAN; one or more LoRa Gateways (GWs) that receive the LoRa frames and cast them to be forwarded through a wired network and one or more Network Servers, usually in the cloud, which are responsible to process the received and are likely in charge of decision-making.

LoRa's physical layer uses CSS modulation over a variety of frequency bands in Europe, USA. The value of 868MHz is one of the common values in most regions. There are multiple factors that characterize the LoRa communication between the end-nodes and the GWs such as SF, Transmission Power (TP), Carrier Frequency (CF), Coding Rate (CR) and of course the Bandwidth (BW). The SF is defined as the ratio between the symbol rate and chip rate [12]. The number of chips per symbol is defined as 2^{SF} . The SF values vary from SF7 to SF12, where higher SF values achieve higher ranges. The relation between the data rate and the SF is defined by Eq. 1, where R_b signifies the bit rate.

$$R_b = \frac{2^{SF}}{BW} \quad (1)$$

On the other hand, TP usually ranges from -4dBm to 20dBm. This parameter sets the intensity in which LoRa end-nodes transmit the LoRa data frames to the GW. Theoretically as SF and TP increases, the LoRa coverage area is larger. CR provides security against interferences, where higher values provide higher protection (4/5, 4/6, 4/7 and 4/8). BW is the frequency width in the transmission band [13].

In general, the communication between LoRa end-nodes and GWs can be unidirectional or bidirectional. LoRaWAN, on the other hand specifies the architecture, layers and protocols operating over LoRa. Mesh or stars are the two possible topologies supported in LoRa [14].

4 Architecture

In this section the general simulation architecture is presented, in order to define the assumption that were made. The simulation environment which was used is the FLoRa

simulator [15], because it is quite comprehensive, and many parts of a real network are simulated.

First and foremost, the Log Normal Shadowing model [16] was used. The model that is presented in Eq. 2:

$$PL(d) = PL(d_0) + 10n\log_{10}\left(\frac{d}{d_0}\right) + X_\sigma \quad (2)$$

- $PL(d_0)$: the mean path loss for d_0 distance
- n : path loss exponent
- X_σ : zero mean Gaussian distributed random variable with deviation σ .

Moreover, for a successful LoRa transmission, the value of the received signal power needs to be higher than the threshold related to the sensitivity of the receiver. The power of the received signal is affected also by the transmission power and the losses that are occurred as a result of the shadowing as presented in Eq. 2 and the signal attenuation. Moreover, the phenomenon of signal interference is taken into account in the simulation. Thus, it is assumed that two signals that are orthogonal (have different SFs) do not interfere, while contrariwise, in case of being non-orthogonal they display collisions when there is an overlap in the time domain. The capture effect is also taken into consideration. Capture effect is the phenomenon observed in real LoRa networks, where even in case of collision of two transmissions, the strongest signal (the power difference of these signals is greater than a threshold) succeeds to be correctly received by the GW [17]. As far as the energy consumption model is concerned, the values related current power consumption and the voltage are based on the SX 1272 transceiver [18], made by Semtech. The energy expenditure is based on the measurement of the nodes in the states of the transmission. It is assumed that there are three states a) transmit b) sleep and c) receive. The node is supposed to be in sleep mode, apart from the cases of transmitting and receiving messages.

As far as the general architecture of the examined system is concerned, the system consists of nodes that communicate to the Gateway (GW) using LoRa. The GW transmits the uplink packet and sends it through the Internet to the Network Server. The transmission from the GW to the NS is conducted using the Internet Protocol (IP) (as far as the Network layer is concerned). For the simulation of the physical layer of the GW – NS communication (Wireless-Fidelity) Wi-Fi technology is assumed. The implementation is based on INET's Wi-Fi modules.

5 Machine Learning Integration

Nowadays, ML is very promising in helping to solve several problems. ML consists of different approaches such as supervised learning, unsupervised learning and reinforcement learning [7]. Supervised learning refers to the cases where the training data are labeled, and on the other hand, the unsupervised learning refers to the techniques where the training data are not labeled. Reinforcement learning refers to the techniques, where the goal is to find an equilibrium between the previous knowledge and how learning new

things is feasible. In this paper three supervised learning techniques are investigated a) the k-NN, b) the Naïve Bayes, and c) the SVM.

The k-NN algorithm assumes that similar nodes exist in proximity, or similar nodes are near to each other and have common behavior. Its main characteristic is its ease and simplicity of implementation and high enough accuracy. Specifically, after the learning process, the new unseen data is fed to k-NN algorithm as a d-dimensional point. Then the minimum distance of the input point from the points of the training is calculated. The distance has an important role in the k-NN algorithm; thus, the appropriate distance metric selection is vital. The metrics used are the Euclidean, Minkowski, Manhattan, Mahalanobis, and Chebysev distance. Then, the decision in which class should be assigned is made according to the predominant class of the k nearest points, to the unknown data point. In this work, the predominant class is computed by a simple majority vote.

Naïve Bayes in reality, is not one classifier, but a class of probabilistic classifiers. The basic idea is that, given an input vector, that represents the unseen data point, a Naïve Bayes classifier, applies the Naïve Bayes theorem, assuming independence between the features of the given input vector. The likelihood of the features can be assumed to follow a distribution. Thus, the Naïve Bayes classifier can use the Gaussian or Bernoulli distribution, etc. In this work, the Gaussian variant of the Naïve Bayes classifier is used. The main advantage of the Naïve Bayes classifier is that can achieve high accuracy with small data, in contrast to more complex models, such as the neural networks.

The SVMs are in contrary to the Naïve Bayes, a non-probabilistic family of classifiers. The main idea behind the SVMs, is that the objective is to find a hyperplane that splits the classes of the training set with the largest margin. When the new unseen data is fed to the SVM, the prediction of the label is occurred based on which part of the hyperplane it falls. The SVMs can handle both binary and multi class problems and are supposed one of the best classification algorithms [7].

The classification process involves accumulating several samples from the path-loss evaluation. The collected samples are then used to extract a set of commonly used features using the correlation, and statistical tests, such as chi-square, in different topologies. It is important to note here that the position of the end-node is excluded, and other features were investigated, as the LoRa localization can have a distance error of 300 m [19]. In this work, we have concluded to use the TP, total energy consumed, and the total packet sent as the feature vector.

Finally, it is important to explain how ML is considered in the framework of this research work. The goal is to assign a value to the SF. As explained in the previous section, the values of SF vary from 7 to 12. So, the problem of SF assignment can be considered as a classification problem. Due to the fact that the target values range from 7 to 12, the classes are numbered to 6. Hence, the problem of the SF assignment should be considered as a multi-class classification problem.

6 Performance Evaluation

6.1 Simulation Parameters

Regarding the needs of the results' presentation, we conducted the following experiment in the FLORA simulator environment. The necessary simulation parameters for the conduction of experiments are presented in Table 1. The LoRa topology consists of multiple end-nodes varying from 100–700 with a 100-node step, for two different cases.

Table 1. Simulation Configuration.

Parameter	Urban	Suburban
Network Size	480 m*480 m	9800 m*9800 m
Number of Nodes	100–700	100–700
σ	0	0
Spreading Factors	7–12	7–12
Code Rate	4	4
Number of GWs	4	4
Bandwidth	125 KHz	125 KHz

In our simulations, we considered a network of urban and suburban setup. For this reason, we used two different models derived from papers [12] and [20] for both cases. Two different areas examined 480 m*480 m and a topology based on Oulu town with coverage area of 9800 m*9800 m. The deployment of the end-nodes was determined randomly in the topology. In this simulation, the stationary mobility model was used. Moreover, the energy consumption metric is considered as the ratio of the energy consumption of all LoRa nodes and the cardinality of the messages received by the NS.

6.2 Simulation Results

In this section, the simulation results are presented. The goal in this case is to extract the suitable SF that could be used for the transmission between NS and the end-nodes. For this reason, we start by the integration of the classifiers for SF assignment, knowing the TP, total energy consumed, and the total packet sent. The dataset in which the classification algorithms were applied was created by running the LoRa simulations. After checking about missing values and checking the different values, we standardized the data.

When the dataset is created, the dataset was split into a training, validation and test dataset. To evaluate the classifiers in the context of LoRa and SF selection, the K-Fold

cross-validation technique was also used. In K Fold cross-validation, the data is divided into k subsets. This method is repeated k times, such that each time, one of the k subsets is used as the test set/validation set and the other k-1 subsets are put together to form a training set. The error estimation that comes from, is averaged over all k trials to get the total effectiveness of our model from FLoRa.

Furthermore, as far the k-NN classifier is concerned, it was necessary to find the optimal k (neighbors) number for our dataset. The results showed that with k = 3 and cross validation of 10-fold, an accuracy of 95 percent was achieved. In Fig. 1: left diagram the best number of k, in terms of misclassification error is presented, while in the right the accuracy of each fold in the cross-validation process in the urban and suburban scenario is presented.

$$D = \left(\sum_1^n |x_i - y_i|^p \right)^{1/p} \quad (3)$$

As shown in the Eq. 3 in order to classify the SF, we calculate the distance between the features, using $p = 2$, in order to have the Euclidean distance.

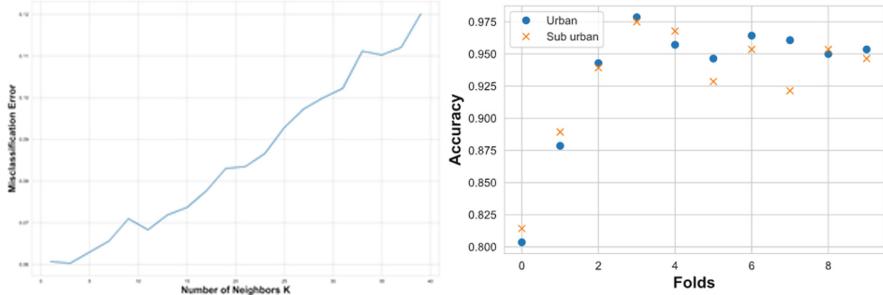


Fig. 1. Left diagram: The optimal number of neighbors. Right diagram: Results of the accuracy on 10-foldcross-validation.

As far as, the Naïve Bayes classifier is concerned, the Gaussian, Multinomial, the Bernoulli (for multi-class problem) variants of Naïve Bayes were examined, with 10-fold cross validation. As it turned out the Gaussian variant with an accuracy around 80% in the cross validation. As far as the SVM is concerned, the SVM parameters were selected according to the Grid Search and Random Search methods. Specifically, after limiting the parameters by the Random Search method, the Grid Search was conducted to the parameters. The parameters that achieve high scores are a) the linear function as the kernel function, and b) $c = 10$. The mean validation score is 0.946.

After the training phase, the classifiers were evaluated in the test dataset, thus the confusion matrix (cm) of each classifier in each scenario (urban and suburban) is presented in Fig. 2. Specifically, x axis of the cm is the actual class of the dataset, and in y axis is the class that has been predicted by the classifier. Cm also shows the class that the classifier gave wrong answer. Cm is a very important metric, should be highly taken into consideration regarding the evaluation. In the first row the cms of the k-NN, Naïve Bayes, SVM (in this order) for the urban case, while in the second row the cms of the suburban case are presented (in the same order as previously).

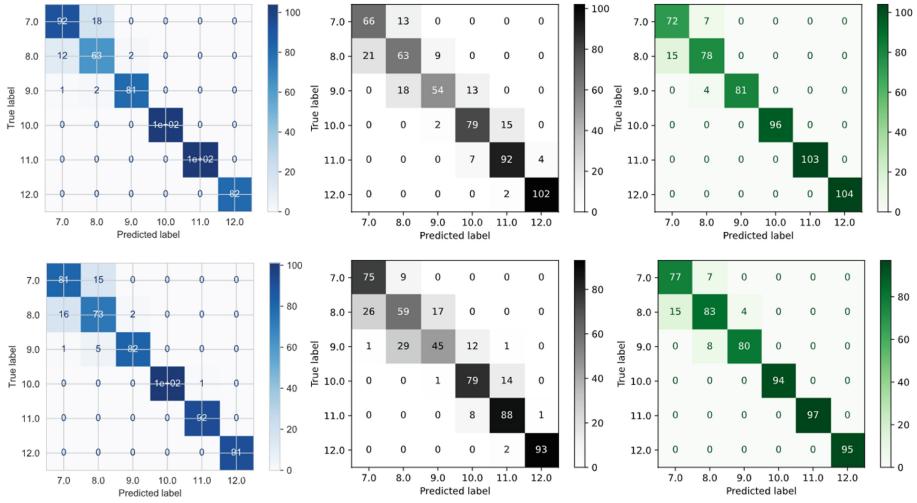


Fig. 2. The classifiers' confusion matrices first row the urban scenario, second row the suburban scenario, from the right to the left the k-NN, Naïve Bayes, SVM

All three classifiers are very accurate in their prediction. The k-NN, predictions are correct in most of the cases, and only in the classes of SF 7, 8, and 11, some errors may be found in both urban and suburban cases. The Naïve Bayes performed the least good as some misclassification errors occurred in the classes SF 7–12 in both cases, but in acceptable level. The SVM, performed very well, as some errors occurred only in classes SF7 and 8.

Finally, we present the final metrics included in the tests. The classifiers were evaluated by the metrics of accuracy, precision, recall, and F. Accuracy is the ratio of the correct predictions. Precision per class is the ratio of the number indicating the correctly predicted answers by the overall answers that predicted this class. Recall per class is the ratio of the number indicating the correctly predicted answers of the class by the number of the actual instances of the class. Last but not least, the F1 metric is defined in Eq. 4.

$$F1 = 2 \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

In Table 2, the above-mentioned scores are presented. As the table shows, the k-NN and SVM algorithms achieve high scores on all 4 metrics, with the highest being the SVM in the urban case for a small margin to the k-NN, while the k-NN achieved higher in the suburban case. Naïve Bayes scored the least in the urban case, while in the suburban case scored less than the other classifiers but with smaller margin. These high scores imply that k-NN and SVM may be used effectively for the SF assignment in LoRa networks.

Table 2. Metric Scores

Metric	k-NN		Naïve Bayes		SVM	
	Urban	Suburban	Urban	Suburban	Urban	Suburban
Accuracy	0.9446	0.9429	0.8000	0.9171	0.9679	0.9357
Precision	0.9459	0.9430	0.8012	0.9107	0.9670	0.9357
Recall	0.9458	0.9391	0.7941	0.9243	0.9658	0.9343
F1	0.9456	0.9401	0.7944	0.9391	0.9663	0.9344

7 Conclusions

In this paper, LoRaWAN and ML has been studied in terms of classification for SF assignment in order to save module's energy requirements. The proposed mechanism achieves an improvement on performance accuracy by using ML and k-NN in order to extract the suitable SF factor for the transmission of the data. In order to have a reliable and good classifier, it is necessary to study the metrics that allow classifier evaluation. In this work, Accuracy, Precision, Recall, and F1 metrics were used showing that the k-NN and the SVM classifiers can be promising, as the scores in terms of these metrics were high. Finally, the effectiveness of the classifiers is also presented in the confusion matrices, where in both urban and suburban cases.

Our future work includes the population of a configuration file after the ML study which will be suitable to configuration network server in OPNET¹ topology on the fly based on SF recommendation of the ML. This library will be implemented as an external tool able to be integrated for future works and will be able to set various parameters that will improve the performance accuracy of the system based on a ML model. Specifically, the NS with the data obtained could update the node's parameters, whenever downlink window is opened. Moreover, it is intended to evaluate the above accuracy improvements in real life scenarios such as for Search and Rescue (SAR) operations, in the framework of WeSAR project. The evaluation will be conducted using hardware such as the Pycom modules (e.g. LoPy, FiPy²) and the Dialog DA14861 wearable module.

Acknowledgment. This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project code: T1EDK-01520).

References

1. <https://www.wesar-project.upatras.gr/>
2. Buurman, B., Kamruzzaman, J., Karmakar, G., Islam, S.: Low-power wide-area networks: design goals, architecture, suitability to use cases and research challenges. IEEE Access **8**, 17179–17220 (2020)

¹ <https://opnetprojects.com/opnet-network-simulator/>.

² <https://pycom.io/>.

3. Bouras, C., Gkamas, A., Salgado, S.A.K., Kokkinos, V.: Comparison of LoRa simulation environments. In: Lecture Notes in Networks and Systems, pp. 374–385. Springer (2019)
4. Suresh, V.M., Sidhu, R., Karkare, P., Patil, A., Lei, Z., Basu, A.: Powering the IoT through embedded machine learning and LoRa. In: 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, pp. 349–354 (2018)
5. Zhou, L., Wang, L., Ge, X., Shi, Q.: A clustering-based KNN improved algorithm CLKNN for text classification. In: 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010), Wuhan, pp. 212–215 (2010)
6. Magno, M., Pritz, M., Mayer, P., Benini, L.: DeepEmote: towards multi-layer neural networks in a low power wearable multi-sensors bracelet. In: 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), Vieste, pp. 32–37 (2017)
7. Mahdavinejad, M.S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., Sheth, A.P.: Machine learning for internet of things data analysis: a survey. *Digit. Commun. Netw.* **4**(3), 161–175 (2018)
8. Chang, Y., Huang, T., Huang, N.: A machine learning based smart irrigation system with LoRa P2P networks. In: 20th Asia-Pacific Network Operations and Management Symposium (APNOMS), Matsue, pp. 1–4 (2019)
9. Carrino, F., Janka, A., Abou Khaled, O., Mugellini, E.: LoRaLoc: machine learning-based fingerprinting for outdoor geolocation using LoRa. In: 6th Swiss Conference on Data Science (SDS), Bern, Switzerland, pp. 82–86 (2019)
10. Bor, M., Roedig, U.: LoRa transmission parameter selection. In: 13th International Conference on Distributed Computing in Sensor Systems (DCOSS), Ottawa, ON, pp. 27–34 (2017)
11. <https://www.thethingsnetwork.org/docs/lorawan/adaptive-data-rate.html>
12. Bor, M.C., Roedig, U., Voigt, T., Alonso, J.M.: Do LoRa low-power wide-area networks scale. In: Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems - MSWiM 2016 (2016)
13. Bouras, C., Kokkinos, V., Papachristos, N.: Performance evaluation of LoraWan physical layer integration on IoT devices. In: Global Information Infrastructure and Networking Symposium (GIIS), Thessaloniki, Greece (2018)
14. Huh, H., Kim, J.Y.: LoRa-based mesh network for IoT applications. In: IEEE 5th World Forum on Internet of Things (WF-IoT), Limerick, Ireland, pp. 524–527 (2019)
15. <https://flora.aalto.fi/>
16. Ruggieri, M., Graziosi, F., Santucci, F.: Modeling of packet cellular networks with log-normal shadowing. In: Proceedings of PIMRC 1996 - 7th International Symposium on Personal, Indoor, and Mobile Communications, Taipei, Taiwan, pp. 296–300 (1996)
17. Croce, D., Gucciardo, M., Mangione, S., Santaromita, G., Tinnirello, I.: LoRa technology demystified: from link behavior to cell-level performance. *IEEE Trans. Wirel. Commun.* **19**(2), 822–834 (2020)
18. <https://www.semtech.com/products/wireless-rf/lora-transceivers/sx1272>
19. Daramouskas, I., Kapoulas, V., Paraskevas, M.: Using neural networks for RSSI location estimation in LoRa networks. In: 10th International Conference on Information, Intelligence, Systems and Applications (IISA), PATRAS, Greece, pp. 1–7 (2019)
20. Petajajarvi, J., Mikhaylov, K., Roivainen, A., Hanninen, T., Pettissalo, M.: On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology. In: 14th International Conference on ITS Telecommunications (ITST), Copenhagen, pp. 55–59 (2015)

Ethics, Computers and Security



Web Guard

Google Chrome Extension for Malicious Web Content Detection

Mohamed Haoud, Raid Djehiche, and Lalia Saoudi^(✉)

Mohamed Boudiaf University of M'sila, M'Sila, Algeria
lalia.saoudi@univ-msila.dz

Abstract. Malicious content on the web became a global risk at web users, its huge spreading made users vulnerable to all types of cyber attacks that can be performed behind websites. Many recent researches were proposed to detect malicious content. In this work, we propose a chrome extension to defeat the majority types of malicious web content by analyzing URL and page content. The extension is based on a relevant list of malicious features geared with a machine learning classifier also it integrated Google safe browsing for more precision. The obtained experimental results demonstrate the effectiveness of our extension for detecting the malicious content which proved by the perfect result scored on several classifiers with 99.5% accuracy.

Keywords: Web browser extension · Malicious content · Machine learning · HTTP request and response analysis

1 Introduction

Nowadays, the web became a more essential part of human life, it responds to all the needs of people of communicating, business, entertainment, banking, shopping, and many other activities are done via web applications. The content of the web is evolved from simple static HTML web pages to complex dynamic web application, which is the result of the evolution on the web activities. With this huge and fast evolution on the Internet and the exchange of sensitive data, the risk of steal or lose of user information is increasing day by day due to the increase of malicious content on the web and the evolution of Cyber-attacks. This makes the security of web application more important than any time before. This risk made people worry about their information and data which lead them to find methods or techniques to protect it. The detection of malicious web content is a very sensitive topic which is always in research to keep pace with development of the web, to prevent this content there are many techniques that have been developed. We can distinguish two categories of prevention techniques:

Server-side prevention techniques: those techniques are implemented on the website most of them related to the configuration of files and source code to prevent the execution of scripts or explore the database.

Client side prevention techniques: those techniques are implemented on the client machines as proxies, firewalls, and Browser extensions.

In this paper we propose a Google Chrome browser extension (WebGuard) which based on analyzing both the URL and the page content with a minimized features list in order to accelerate the detection process and inform the user as soon as possible, WebGuard also uses Google safe browsing service to check the malicious web content.

The rest of this paper is organized as follows. Section 2 provides related works on malicious web content detection, Sect. 3 presents our proposed approach in details. Section 4 explains the experimental results. Finally, Sect. 5 contains a conclusion and some directions for future work.

2 Related Works

There are many approaches that have the same purpose of our work:

Sirageldin et al. [1] proposed an approach to secure web content based on a model of three components: Feature Extractor, Learning & Model Selector, and the Detector. It begins with the extraction of features, then training on a data set, and finally the detection of malicious content.

- Feature Selector: The main step in this work is the selection of features, this model uses 39 features, 21 of them are new or modified and the others are obtained from previous works, the set of features is divided into two categories: URL Lexical Features, Web page content features.
- Learning & Model Selector: It needs a data set to train the model and select the classification algorithm that gives the best results.
- Detector: It uses the training model from the previous step to make the classification.

Anand Desai et al. [2] proposed another approach to secure web content against phishing; they use the UCI Dataset (University of California Irvine) to train a classifier after the extraction of features from the URL. This approach is based on three major steps: Obtaining Dataset, extraction of features and classification.

- Obtaining Dataset: The Dataset is obtained from UCI-Machine Learning Repository which contains 11055 sites which are classified as phishing sites and benign sites, and each site has 30 features. As some features uses standard database and some of them are impossible to extract, the dataset is restructured to contain 22 features.
- Features extraction: The model is based on 22 major features such as: the length of the URL if the length of the URL is more than 52 characters the site is considered as a suspect phishing site. Google Index to check whether the site is found in Google Index or not, generally phishing sites are not indexed by Google.
- Classification: This approach is based on three Machine Learning Algorithms which are: K-Nearest neighbor (KNN), Support Vector Machines (SVM) and Random Forest.

M. Aldwairi et al. [3] proposed an extension named MALURLs, This extension detects malicious websites based on the lexical features of the URL and the features of the host, it uses genetic algorithms and the Naive Bayes classifier to classify the sites.

- Features extraction: The features used in this approach are divided into three categories: The lexical features of URL which are: the length of the main domain, domain, hostname, length of the URL, number of dots in the URL, also tokens in hostnames and tokens in the URL path. The features of the host as: IP address, geographic properties, DNS properties, time to live (TTL), DNSA, DNS PTR, DNS MX records, WHOIS information and dates. Special features as: JS enable/disable, HTML Title tag content (<title> </title>), 3-4-5 grames, Term Frequency and Invers Document Frequency (TF - IDF)
- Classifier: It is a component based on the Naive Bayes algorithm, it uses the features extracted from the URL of the previous step to classify the sites are malicious or not.

Immadisetti et al. [4] proposed an approach which is based on machine learning model which is based on the Convolutional Neural Networks (CNN) algorithm, to analyse the extracted Features and classify the sites.

2.1 Comparison of Previous Works

In this subsection we would make a comparison of the works that we reviewed in this section. Our comparison based on the common characteristics of those works.

The following Table 1 gives a comparison between all presented approaches.

Table 1. Approaches comparison

	Approach 1	Approach 2	Approach 3	Approach 4
URL analysis	×	×	×	×
Web page analysis	×			
Type	Framework	Google Chrome extension	Lightweight System	Classification model
Detected content type	Web pages	phishing	Websites	Websites

3 Our Approach

In our work we have implemented a web browser extension named Web Guard, it is a Google Chrome extension that intercepts any traffic between web server and Chrome browser. WebGuard analyses the requests and responses from web sites in order to detect the malicious content of websites. WebGuard has three major components: Feature Extractor, training module and the classifier (Fig. 1):

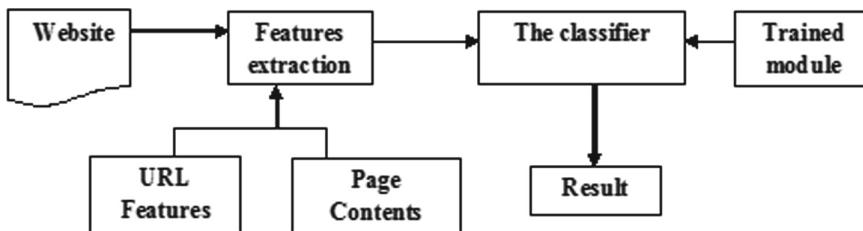


Fig. 1. General architecture of WebGuard extension.

3.1 Features Extraction

We choose specific features extracted from URL and Web Page that help our classifier to classify the web site, the web sites have a huge number of features so we reduce the number of features in order to optimize the execution time of the extension, as it is mentioned in Table 2.

- URL Lexical Feature: Lexical features are the textual properties of the URL itself such as URL length, number of dots etc.
- Page Content Based Features: the source code of the web page is an important source of feature for identifying malicious web pages because it contains several malicious features such external links, hidden scripts and more.
 - DHTML or HTML features: DHTML or HTML feature includes word count, distinct word count, and size of I-frame. These features are exploited by the attacker for obfuscating the malicious code or script into the web page.
 - JavaScript features: JavaScript is one of the popular scripting languages which is used as the validation code for the web page[5]. Some functions in the JavaScript are quite vulnerable and are exploited by the attacker for injecting the malicious script. These malicious scripts get executed just by clicking on the page. Some of JavaScript functions which are most exploited are eval(), escape(), unescape(), exec(), ubound() etc.

For example the unescape() function used to decode escaped sequences to the original set of characters.

So the attacker can mask malicious code with escape() function then use unescape() to decode it then execute it with eval() function.

In this example this instruction:

```

<script>
eval(unescape('%61%6C%65%72%74%28%27%68%61%78%21%27%29'))
</script>
Is in real Alert ('hax!');

```

- Google safe browsing: Safe Browsing API [7] is a Google service that lets client applications check URLs against Google's constantly updated lists of unsafe web resources.

Table 2. List of selected features

Features	Explanation
Double slash redirection	Detecting if there are double slash redirecting in the URL
Longest domain length	Calculate the Length of the Domain
Number of dots	Calculate the number of dots in the URL
URL length	Calculate the Length of the URL
Host length	Calculate the length of the Host
URL Scripts	Detect the scripts and commands in the URL
Sub-domain (www)	Check the type of the Sub-Domain
HTTPS	Check the HTTPS protocol
Domain (.com,.net...)	Check type of the Domain
Numbers in URL	Detect if there is numbers in the URL
Count number	Count numbers in The URL
Count letter	Count letters in the URL
Rank host	Check the rank of the Host
Google search engine	Check the URL by Google Search Engine
JavaScript in URL	Detect JAVA Script in the URL
Malicious script in content	Detect the malicious scripts in the web page
External links	Detect if there are external links in the web page
Iframe count	Count the I frames in webpage

3.2 Learning Phase

In order to train the classifier we have prepared a dataset composed of 600 sites. Our dataset contains 300 safe web sites which extracted from Alexa top sites and 300 malicious web sites extracted from previous researches. In this phase we generate the trained module which stores the data of the training phase so in new classification operation the classifier predicts directly from the module, it doesn't need a new training session which is a positive point to reduce the process time.

3.3 The Classifier

For good classification results we test our extension with four machine learning algorithms which are: Support Vector Machine (SVM), Naive Bayes (NB), K Nearest Neighbors (KNN), and Random Forest Algorithm (RF). The selection is based on the best results of precision detected by the previous algorithms.

4 Experiment and Results

4.1 How It Works

Our extension starts with analyzing the visited URL, through the following steps:

- Read the URL as String.
- Extract the selected features.
- Compare to the trained model.
- Get the result.
- If the URL is malicious alert the client.
- Else analyze the content.

The second stage of content analysis passes through the following steps:

- Read the HTML and JS content of the WEB.
- Extract the selected features.
- Compare to the trained model.
- Get result.
- If it is benign, permit access to the website.
- Else alert the user.

4.2 Experimental Procedure

The experiment is passed via four phases. Firstly, we prepare the data set of 600 entries. Secondly we extract the relevant features from the websites. Thirdly, we generate the trained model. And finally, the generated model is tested using 10 cross folds method which devise the dataset to 10 parts for the training and the test at the same time.

4.3 Results

After the experimentation, we present the obtained results of the accuracy, recall and precision of our extension using four algorithms of classification, which are the Random Tree, Naïve Bayes, SVM, and KNN. The obtained results shown in the Table 3:

Table 3. Results metric of our extension (Accuracy, Precision, Recall)

Classifier	Accuracy	Precision	Recall
Random Tree	99.5%	99.7%	99.3%
SVM	98.8%	99.3%	98.3%
KNN	98.5%	100%	97%
NB	98.5%	98.7%	98.3%

During the test our extension scored a high accuracy, precision and recall, these scores are result of the right selection of features set which improved this results and make our extension a powerful tool against the web malicious content.

4.4 Discussion

The following table (Table 4) shows comparison between our extension and the closest one from the previous works (Approach1) using the common classifiers between them.

The found results show that our approach achieves better results than approach 1, using the same data set.

Table 4. Comparison between WebGuard and Approach 1

Classifier	Accuracy		Precision		Recall	
	WebGuard	Approach 1	WebGuard	Approach 1	WebGuard	Approach 1
SVM	100%	93.57%	100%	92.8%	100%	93.85%
KNN	99.7%	92.9%	100%	98.4%	99.7%	88%
NB	99.7%	88.47%	99.7%	89.6%	100%	90.7%

5 Conclusion and Future Work

In our project we have studied existent approaches of detecting the malicious web content. The study shows that those approaches still incomplete and have several limits which led us to developed a Google chrome extension "Web Guard" to defeat the malicious content on the web by analyzing the URL and page content from websites to detect the malicious web pages. The obtained experimental results demonstrate that our extension detects perfectly the majority of malicious websites during the test phase, which prove the effectiveness of our selected features list to detect malicious content with 99.5% Accuracy. Our approach focused on analyzing URL more than the page content, it can be improved by focusing more at the analyzing of the content, and adding a database with user feedback to improve the process of classification and support all web browsers.

References

1. Sirageldin, A., Baharudin, B.B., Jung, L.T.: Malicious, Web Page Detection. A machine learning approach. Computer & Information Science Department, University Technology Pertonas Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia (2014)
2. Anand, D., Janvi, J., Rohit, N., Nataasha, R.: Malicious web content detection using machine learning. In: IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), India, 19–20 May 2017
3. Aldwairi, M., Alsalmam, R.: ALURLS: a lightweight malicious website classification based on URL features. *J. Emerg. Technol. Web Intell.* **4**(2) (2012)
4. Immadisetty, N., Venkata Durga, N., Manamohana, K., Rohit, V.: Detection of malicious URLs using machine learning techniques. In: International Journal of Innovative Technology and Exploring Engineering (IJITEE), **8**(4S2), 389–393 (2019)
5. Doyen, S., Chengho, L., Steven, H.: Malicious URL detection using machine learning: a survey, **1**(1) (2019)

6. Jitendra, A., Shikha, A., Anurag, A., Sanjeev, S.: Malicious web page detection through classification technique: a survey. *IJCST* **8**(1), 74–79 (2017)
7. Google safe browsing. <https://developers.google.com/safe->. Accessed Aug 2020



Filters that Fight Back Revisited: Conceptualization and Future Agenda

Sampsaa Rauti^(✉) and Samuli Laato

Department of Computing, University of Turku, Turku, Finland
{sjprau,sadala}@utu.fi

Abstract. Online scams, unsolicited advertisements, messages containing malicious files and other forms of spam continue to be a nuisance in today’s internet, wasting users’ time and causing financial damage to companies and organizations. There have been many proposals on how spam should be stopped, from various kinds of spam filters to legislative measures. One of the more extreme suggestions is fighting back by bombarding spammers’ servers with a deluge of HTTP requests. In the current study, we revisit this idea “filters that fight back” originally proposed by Graham in 2003, and investigate why the approach has received little attention recently. We also showcase an example solution that automatically sends false information back to spammers by filling forms on their websites or replying to mail addresses they have provided. We offer a conceptualization and future agenda of filters that fight back, and discuss the ethical and technical challenges related to this solution.

Keywords: Spam filters · Filters that fight back · Offensive defense · Cybersecurity · Offensive security

1 Introduction

Unsolicited messages sent to a large number of recipients, prominently referred to as spam [8], have been a major nuisance in the internet almost ever since its conception. These messages can be, for example, commercial advertisements, attempts to obtain users’ personal information or messages related to financial fraud. Spam messages regularly contain links to dubious web pages built with the goal of phishing or distributing malicious software.

Spam can be annoying to deal with and wastes recipients’ time [20]. Thus, significant efforts have been devoted to prevent spam messages from reaching their destination, for example, by using spam filters and restricting spam with legislation. During the 90’s these filters were mostly rule-based, but today more complex solutions such as statistical spam filters [19] and filters based on artificial neural networks [1] are widely used by prominent email service providers and sometimes also by end users. Despite these countermeasures, users still react and respond to spam messages in high enough numbers to make spamming a

profitable business. Botnets such as Necurs [4] have been employed by spammers to effectively and effortlessly send large numbers of spam messages.

During the 90's, the number of spam emails increased steadily, amounting for as much as 90% of all email in 2008 [11]. More recently, unsolicited messages have gone down to 50–60% of all email traffic [15, 23]. Today, spam campaigns have been moved to other mediums besides email, for example, to instant messaging applications, phone calls and social media platforms [2].

While spam filters can analyse spam messages and use them as data for training machine learning models to better detect and block spam [9], most spam filters simply settle for deleting spam messages after they have been detected, taking no further action against the perpetrator. This paper explores the idea of punitive spam filters that aim to incur a cost to a spammer each time they send junk messages. We explore the idea of filters that put a strain on spammers' servers by bombarding their servers with HTTP requests [13]. This idea was originally proposed by Graham in 2003 [13], but contains several ethical and technical challenges which make using such techniques non-straightforward.

In this study, we first talk about Grahams' proposal in detail and go through what has been studied in the academic field on filters that fight back (FFB) back since then. Subsequently, we present an example solution created for this manuscript that illustrates the idea of filters that fight back. This solution fills forms on spammers' websites with fake information or alternatively provides automated bogus answers to spam messages. We follow this example with discussion on the technical, ethical and legal challenges of FFB as well as the benefits of such solution. We conclude the work by providing a future agenda for researchers and engineers interested in FFB.

2 Filters that Fight Back - The Current View

The original concept of FFB which Paul Graham talked about on his website [13] proposed the idea of punishing spammers by sending bogus data back to them. The solution would be implemented by adding a “punish mode” feature to spam filters [13, 24]. When turned on, this mode launches a counterattack on spammers by opening all the URLs in a spam message N times (N is 0 or greater, chosen by the user). The web pages are crawled, that is, all the links on the found pages are followed (which can be repeated for k levels of links) [12]. Consequently, sending huge amount of spam messages now works against spammers, flooding their servers with HTTP requests increasing the bandwidth usage and inflating the costs of maintaining a webpage. The deluge of requests can also make the spammer's servers unavailable to those users who would otherwise have responded to the spammer in good faith and fallen into the scam. If the spammer churns out a million messages an hour, they will potentially receive millions of hits an hour on their servers. This would make operating scams through spamming unremunerative.

Although a URL sent to millions of people is likely to be an address of a spam page, it is important to ensure that HTTP requests are only launched against

spam pages. Graham [12] suggested only crawling sites that are on a special blacklist. Web pages are blacklisted only after being inspected by humans. As a spam message has a lifetime of couple of hours at least, the blacklist can be updated in time to ensure counter-spam measures can be activated against the adversary.

The challenge from the spammers' perspective is the fact that to reach a few gullible recipients who will reply, the spammer needs to send messages to tens of thousands of recipients if not more. FFB has the benefit of enabling the non-gullible majority to make it more difficult for most ductile users to fall for scams. Here one additional potential positive consequence for users is, that in order for spammers to protect their servers against a deluge of HTTP requests, they could be prompted to provide their victims an "unsubscribe" option. This would free spammers from counter-spam and enable recipients to free themselves from the spammers' mailing list [12]. The problem here of course is that as the tech-savvy spam respondents unsubscribe, the more gullible victims would remain in spammers' mailing lists.

An idea similar to Graham's scheme was used in practice when Lycos Europe web portal launched a screensaver that sent HTTP requests to websites that were known to be promoted in unsolicited mail messages [14]. An advertisement on Lycos Europe's urged the users to "annoy a spammer now!". As the percentage of spam mails has decreased over the years, it appears these kinds of campaigns have became less commonplace. One of the reasons for this could be that modern machine learning-based spam filters have evolved to be so efficient in screening and deleting spam emails [9] that no further action is needed. An additional reason could be that spammers are choosing to host their websites and spamming operation on servers with a flat rate charge. As an extreme example, if an FFB solution would be implemented on an web hosting company X email service, we could see a case of web hosting company X filters sending counter-spam to a web hosting company X server hosted by spammers. Therefore, it is not in the interest of web hosting company X to use FFB.

One of the most recent examples of FFB technical implementations comes from a bachelor project carried out at the Delft University of Technology. In this project, Bansagi et al. [3] created a system that recommends replies that are sent back to spammers, making it easy to waste a scammer's time and money. The authors developed a Google Chrome plugin to enable quick replies to spam emails. This solution, however, does not include offensive defense in the form of crawling the scammer's website.

Spam emails are typically sent using botnets consisting of infected machines. That is why it is often difficult to directly target this infrastructure built by spammers. By targeting the spammers' dubious websites and feeding them fabricated information, their operations can be disrupted. Although tech companies such as Microsoft have managed to take down huge botnets such as Necurs [4], smaller scale operations can still be beneficial. Hence, here we focus on attacking spammers' websites.

3 Our Example Solution

For the purpose of illustrating how FFB could work in the modern online ecosystem, we created our own conceptual solution that wastes spammers' time by automatically sending fake information to them. After all, spammers usually aim to collect information about their victims. In what follows, we discuss a conceptual solution for sending fallacious data through web forms and email. Figure 1 shows the general idea of our solution.

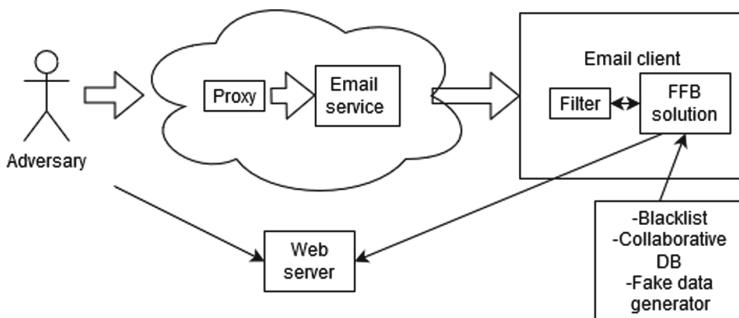


Fig. 1. An overview of how FFB operates. The FFB module responds by targeting the malicious website to which the scammer tries to lure gullible victims.

3.1 The Algorithm

A high abstraction level skeleton of a general algorithm for FFB implementation that is invoked when an email is identified as spam, and which crawls weblinks given in spam messages, is described below in pseudocode:

```

If the mail is a spam message
  If the mail contains URLs
    For each URL
      If the URL is on the blacklist
        Crawl each subpage of the website K levels deep
        Load the subpage N times
        If the subpage contains a form
          Fill in fake information M times
      If the mail contains a form
        Respond with fake information
  
```

If the spam filter classifies a received mail as spam, the mail is checked for URLs. Each URL is then tested against a blacklist, and if the URL is on the

list, it is chosen for further inspection. The subpages under the main address are crawled and loaded N times, following Graham's scheme. This is done K levels deep, meaning all possible link chains (with the length of $K - 1$ or smaller) from the main page are followed. However, if there are a huge number of subpages, this process could be stopped after a specific number of pages to avoid needlessly wasting our own time and resources. Also, the punitive functionality suggested by Graham could be completely turned off by setting N to 0.

Each crawled subpage is also checked for forms. If a form for collecting a user's information is found, it is filled with fake information. This can be made several (M) times, but depending on the checks implemented on the spammer's server, the form might only be accepted once. The process of generating fake information and filling in the form is further discussed in the next section.

Finally, the body of the received email can also be checked for forms. This is not usually a HTML form but a list of details that the spammer wants the user to fill in and reply to an email address. Using the same fake data generation functionality as previously on the spammer's website, fallacious information is created and sent to the email address provided by the scammer.

It is worth noting that the conceptual solution we have presented employs many other components: an email client, a spam filter, a blacklist, and a fake data generator. Still, our scheme is independent of how these other components have been implemented. The solution could be added to an email client like Mozilla Thunderbird as an extension. The spam filter and the blacklist can use any of the currently available approaches as long as they are accurate enough so that they do not produce a significant number of false positives. The fake data generator can be a part of the implementation or a component implemented by a third party. It has to be able to generate believable bogus names, addresses, phone numbers etc. Our solution can also be applied to other types of spam messages such as SMS spam or spam in Facebook or Twitter.

3.2 Filling in the Forms

An important part of the discussed conceptual solution is filling in the forms on spammers' websites. Web forms are typically included in phishing websites which aim to steal victims' credit card information or other details. The entities in the form (such as name address, phone number) can be recognized by using relatively simple rules such as looking at the descriptions and names of the form fields seeing if they match known entity names. If the meaning some form field cannot be recognized, we can simply fill it with random content. The filters could also use human assistance for finding the type of some specific field and then collaboratively share this information to other filters that are also completing the same form.

After the form has been found and the types of the fields have been decided, the fake data generator will generate an appropriate input for each field. This kind of fake data generating component has to use a large list of possible values in order to make the fake details convincing [16, 17]. It has to be able to generate

wide variety of valid values such as addresses, phone numbers and social security numbers.

If the form submission fails, we can retry submitting it a certain number of times. In many cases, the red text indicating an error near a form field can be used to guess which field was not accepted, and a different value can be chosen for that field. The filters can also share information on what kinds of values were successfully accepted for a specific form.

One challenge is the fact that many forms which require credit card information. A seemingly valid fake credit card number can be generated but if the spammer's system immediately attempts to charge the credit card, deceiving the fraudster will not succeed. Checking the validity of the provided card number wastes spammer's resources (computational power or human effort).

There is also an interesting side effect when feeding false information to spammers. If some unique piece of fake information – also called honeytoken [5, 22] – is included in the data given to spammers, it could later resurface somewhere else. Consequently, planting honeytokens can help in tracking and attributing spammers [18, 21], as well as finding out where they sell the information.

4 Discussion

Supplying spammers with fake information poisons their database and wastes their time. Perpetrators can no longer be sure which entries are fake and which are not. This resembles scambaiting, which aims to waste the scammers time and resources by exchanging messages with them, but in our scheme the whole process is automatic. We summarize the key findings of the potential damage FFBs cause spammers below:

- wastes computational resources and bandwidth
- wastes spammer's time (when the obtained information is processed manually)
- prevents gullible users from becoming victims of phishing
- may cause software development related costs when the spammers have to fix their website or information gathering model so that poisoning becomes more difficult
- damages the spammer's reputation as a business partner if poisoned low-quality data is sold.

4.1 Technical Challenges

A potential challenge with our scheme is the fact that clicking links in spam mails and replying to spammers often causes them to send more spam mail. However, when this happens, the punitive filter will punish them even more, and as long as the spam filter works, the user does not see increasing amount of spam. The increased bandwidth consumption should not be a problem with modern broadband connections.

Another potential weakness of the solution is that blacklists are prone to abuse. While our scheme is independent of the implementation of blacklist, it is important to ensure that untrustworthy individuals or spammers themselves cannot easily poison the blacklist with entries that should not be there. Also, simply being on the blacklist does not cause a website any problems, it only gets hit when it is already blacklisted and a new spam message arrives with a link to the site arrives [12].

4.2 Ethical and Legal Issues

One can argue that this kind of offensive defense and “striking back” is immoral or illegal [6]. It is not completely clear whether loading spammer’s webpages a few times means participating an organized denial of service attack or whether automatically filling spammers’ web forms constitutes any kind of offense. In some jurisdictions, however, the user might be rendered legally liable.

Spammers may not be likely to take the matter to court, but if some innocent party was accidentally targeted, things might be different. Still, the solution we propose is very different from “hacking back”, that is, which would mean tracing back to the attacker and invading their system. There are probably very few, if any, precedents pertaining to this kind of offensive defense. Of course, what is permissible from a legal or ethical points of view, also depends on who is doing it. Some kind of authority could also take care of striking back against spammers.

4.3 Benefits

Our solution has some additional benefits compared to Graham’s original solution. It not only increases the load for spammers’ servers, it also deteriorates the quality of the data they receive. Several spammers and other cybercriminals may try to use the poisoned data if the data is sold. Moreover, the data is potentially made traceable with honeytokens. Our solution is also more likely to waste time of human perpetrators, as the information spammers receive may be manually examined.

If the functionality of loading webpages repeatedly is turned off, our version of a FFB also does not have the problem of launching denial of service attacks in the same sense Graham’s solution does. For example, If the spammers website shares a host with some other innocent customers, continuously bombarding the spammer’s website can cause needless collateral damage. Our solution, when used without repeated page loads, avoids this problem. Then again, if this functionality is turned off, then the filter does not protect gullible users from falling for the scam as effectively.

The conceptual solution we have presented does not depend on the environment where spam needs to be combated. Along with the email system, our solution can also be used for counter spam messages in social media or text messaging (SMS) spam. Another application is fighting search engine spam [10]. For all these types of spam, filters have been build and our solution could be combined with those filters.

5 Conclusions and Future Agenda

In this paper, we revisited the idea of FFBs introduced by Paul Graham in 2003 [13] that has since been dormant and seen minimal attention in the scientific community. To see whether FFBs are still a viable security measure, we developed Graham's idea further and provided an example case of a spam filter that provides a form of offensive defense by replying to spammers with fake information and wasting their time. The solution works on individual cases, but large-scale application of this approach remains untested. In addition to this example, other versions of FFB adjusted to the modern online ecosystems have been presented (e.g. [3]) but these filters have not been widely adopted in practice either. With regards to evidence as to why FFBs are currently not used, the following main reasons emerge:

- Ethical concerns related to the justification of offensive defense and its large scale operation.
- Security concerns related to misuse (intentional or unintentional) of FFBs.
- Advances made in other spam filters and other technologies for curbing rampant spam messages.
- Spammers' utilization of 3rd party flat rate online services for their schemes, where FFBs would not in fact cause major damage to the spammer.
- Other concerns related to the feasibility and effectiveness of FFBs such as FFBs alerting the scammer that their operation has been detected, and FFBs enabling scammers to reverse engineer ML-based filters.

While these certainly seem to explain the lack of use of FFBs, it might still be early to throw away the idea of offensive defense against online spam. In fact, recently we have seen the rise of scambaiting, that is, online streamers and content creators making fun of scammers and wasting their time [25]. This activity takes a harmful activity (scamming) and turns it into popular entertainment. While scambaiting also has obvious ethical concerns, in principle it serves to educate people about scamming, cause harm on scamming as a business and protecting potential scam victims, in addition to being entertaining. For these reasons, we believe that FFB should also deserve further attention from the scientific community to see whether it can be applied to make the internet a safer environment. Here we would like to provide a future agenda related to interesting research topics and questions in the domain of FFB.

First, an empirical longitudinal analysis of the consequences of using FFBs in the large scale could be carried out. This analysis needs to focus on both the impact FFBs have on spammers and scammers, and the impact FFBs have on service providers. To this end, a FFB implementation of our solution and field work testing of it is required. Here we identify legislative and ethical challenges in addition to the technical. Second, further analysis is required on whether FFBs could be abused by making them target innocent or trusted parties. Third, the possibility of sending honeypot tokens to the scammers' system and being able to track information that the scammers have is another promising future research

avenue. For example, this could enable discovering which scammers are connected to one another. Fourth, the ethics and lawfulness of offensive defense deserves attention with regards to FFBs, scambaiting and other forms of means to retaliate. Online vigilantism, sometimes discussed as digilantism [7] which includes FFBs, remains in many regards problematic. While in the ideal situation authorities would take care of cybercrime, there are many reasons as to why this is not currently the case. These reasons include lack of technical skills of the authorities, the global scale of the cyberworld where people operate in the same environment under different sets of laws and the rapidly changing and evolving nature of the internet.

References

1. Alghoul, A., Al Ajrami, S., Al Jarousha, G., Harb, G., Abu-Naser, S.S.: Email classification using artificial neural network (2018)
2. Almeida, T.A., Silva, T.P., Santos, I., Hidalgo, J.M.G.: Text normalization and semantic indexing to enhance instant messaging and SMS spam filtering. *Knowl.-Based Syst.* **108**, 25–32 (2016)
3. Bansagi, A., Bes, R., Garama, Z., Gosshalk, L.: The scam filter that fights back. <https://repository.tudelft.nl/islandora/object/uuid%3A6099061a-4ca7-469b-b9c0-86d7bc0f52e3> (2016). Accessed 05 Nov 2020
4. Bapat, R., Mandya, A., Liu, X., Abraham, B., Brown, D.E., Kang, H., Veeraraghavan, M.: Identifying malicious botnet traffic using logistic regression. In: 2018 Systems and Information Engineering Design Symposium (SIEDS), pp. 266–271 (2018)
5. Bercovitch, M., Renford, M., Hasson, L., Shabtai, A., Rokach, L., Elovici, Y.: Honeygen: an automated honeytokens generator. In: Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, pp. 131–136. IEEE (2011)
6. Bradbury, D.: Offensive defence. *Netw. Secur.* **2013**(7), 9–12 (2013)
7. Byrne, D.N.: 419 digilantes and the frontier of radical justice online. *Radical Hist. Rev.* **2013**(117), 70–82 (2013)
8. Cormack, G.V.: Email Spam Filtering: A Systematic Review. Now Publishers Inc. (2008)
9. Crawford, M., Khoshgoftaar, T.M., Prusa, J.D., Richter, A.N., Al Najada, H.: Survey of review spam detection using machine learning techniques. *J. Big Data* **2**(1), 23 (2015)
10. Egele, M., Kolbitsch, C., Platzer, C.: Removing web spam links from search engine results. *J. Comput. Virol.* **7**(1), 51–62 (2011)
11. Farrell, N.: Cisco says that 90 percent of email is spam. <https://www.theinquirer.net/inquirer/news/1050056/cisco-90-percent-email-spam> (2008). Accessed 17 Mar 2019
12. Graham, P.: FFB FAQ. <http://www.paulgraham.com/ffbfaq.html> (2003). Accessed 13 Sep 2020
13. Graham, P.: Filters that fight back. <http://www.paulgraham.com/ffb.html> (2003). Accessed 13 Sep 2020
14. Hines, M.: Lycos Europe: ‘Make love not spam’. <https://www.cnet.com/news/lycos-europe-make-love-not-spam/> (2004). Accessed 16 Sep 2020
15. Vergelis, N., Demidova, M.T.S.: Spam and phishing in Q3 2018. <https://securelist.com/spam-and-phishing-in-q3-2018/88686/> (2018). Accessed 13 Sep 2020

16. Papalitsas, J., Rauti, S., Tammi, J., Leppänen, V.: A honeypot proxy framework for deceiving attackers with fabricated content. In: *Cyber Threat Intelligence*, pp. 239–258. Springer (2018)
17. Papalitsas, J., Tammi, J., Rauti, S., Leppänen, V.: Recognizing dynamic fields in network traffic with a manually assisted solution. In: *World Conference on Information Systems and Technologies*, pp. 208–217. Springer (2018)
18. Rauti, S.: Towards cyber attribution by deception. In: *International Conference on Hybrid Intelligent Systems*, pp. 419–428. Springer (2019)
19. Rusland, N.F., Wahid, N., Kasim, S., Hafit, H.: Analysis of Naïve Bayes algorithm for email spam filtering across multiple datasets. In: *Proceedings of the IOP Conference Series: Materials Science and Engineering* (2017)
20. Siponen, M., Stucke, C.: Effective anti-spam strategies in companies: an international study. In: *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS 2006)*, vol. 6, pp. 127c–127c. IEEE (2006)
21. Spitzner, L.: *Honeypots: Tracking Hackers*. Addison-Wesley Longman Publishing Co. Inc., Boston (2002)
22. Spitzner, L.: Honeytokens: The other honeypot. <http://www.symantec.com/connect/articles/honeytokens-other-honeypot> (2003)
23. Statista: Spam: share of global email traffic 2014–2018. <https://www.statista.com/statistics/420391/spam-email-traffic-share/> (2018). Accessed 13 Sep 2020
24. Zdziarski, J.A.: *Ending spam: Bayesian content filtering and the art of statistical language classification*. No starch press (2005)
25. Zingerle, A.: Scambaiters, human flesh search engine, perverted justice, and internet haganah: Villains, avengers, or saviors on the internet? In: *ISEA Conference* (2015)



Malware Security Evasion Techniques: An Original Keylogger Implementation

Álvaro Arribas Royo¹, Manuel Sánchez Rubio¹, Walter Fuertes²(✉) ,
Mauro Callejas Cuervo³ , Carlos Andrés Estrada² ,
and Theofilos Toulkeridis²

¹ Faculty of Engineering, Universidad Internacional de la Rioja,
Logroño, La Rioja, Spain

{alvaro.arribasroyo,manuel.sanchezrubio}@unir.net

² Department of Computer Sciences, Universidad de las Fuerzas Armadas ESPE,
Sangolquí, Ecuador

{wmfuertes,caestrada4,ttoulkeridis}@espe.edu.ec

³ Faculty of Engineering, Universidad Pedagógica Tecnológica de Colombia,
Tunja, Boyacá, Colombia
mauro.callejas@uptc.edu.co

Abstract. The current study evaluates the malware life cycle and develops a keylogger that can avoid Windows 10 security systems. Therefore, we considered the requirements of the malware in order to create a keylogger. Afterward, we developed a customized and unpublished malware, on which we added as many features as necessary using the Python programming language. At the end of this process, the resulting executable program will execute three main threads responsible for collecting the screenshots, keystrokes, and creating the backdoor in the infected system. Furthermore, we added the required methods to avoid the leading security tools used in Windows environments. Finally, we tested the executable file resulting on different websites as proof of concept in a real scenario. As a result, the keylogger has avoided Windows 10 firewalls, user account control, and the antivirus. Moreover, it gathered a significant amount of confidential information about user behavior, including even the credentials of the users, without noticing them.

Keywords: Keylogger · Malware · Evasion techniques

1 Introduction

In recent years, information and communication technologies have significantly impacted the economic, social, and political sectors. Due to this high impact, the so-called *cybercriminals* have identified these sectors as the most prosperous to commit their cyber-crimes [23]. In this way, during a period when malware evolved over the years and hereby becoming attractive for cybercriminals to obtain benefits from, the preventive measures implemented in the different operating systems advanced as well, offering more protection against these threats to

their users. However, these users often lack technical training or sufficient knowledge to improve their systems' security. A typical example would be represented by any antivirus program, which continuously offers a false sense of security to the end-user, who believes to be protected against all kinds of threats [22].

Within this context, cybercriminals have focused their efforts on developing new malware specimens, which are capable of evading the most common security protections. In order to counterattack such criminal activities, some protective measures have been developed. Among many others, there are (1) Antivirus programs, which detect and prevent malicious actions of malware; (2) Firewalls, which block network connections and prevent attackers from executing remote commands; (3) Systems for privileged control, such as Windows user account control (UAC), which tries to maintain system integrity by controlling the actions that users may perform. Therefore, in the current study, the knowledge and evaluation of these security protection operations have been deepened. Predominantly we detailed the evasion techniques which are often implemented in the malware in order to avoid them. Herewith, it is possible to demonstrate the misleading belief of security that such tools pretend to provide [21].

Therefore, we focus firstly, to simulate the malware cycle phases, which are represented by design, development, propagation, infection, and performance. Secondly, different existing evasion methods will be combined. Specifically, the malware will evade all the mentioned security protections and run on the computer persistently with administrator privileges. The main contribution of this study is the creation of a personalized and previously unpublished software keylogger, which has been added, like many other functionalities, as to be necessary without using existing software.

The remainder of this article is organized as follows: Sect. 2 reports the related work. Section 3 introduces the theoretical framework in regards to the keyloggers. Section 4 explains the design and implementation of the malware. Then in Sect. 5, we evaluated the keylogger using a set of malware evasion techniques. In Sect. 6, we present the obtained results. Finally, in Sect. 7, we clarify the conclusions of this study.

2 Related Work

The market for data theft and privileged information of computer users through keyloggers continues to increase. As reported in [7], a work which indicates such attacks have been generated based on information from more than 70 illicit data collection points anonymously. The authors stated that they encountered more than 33 GB of keylogger data with information stolen from more than 173,000 victims. In [19], the authors presented the concept and types of keyloggers. Also, they describe the techniques to recognize and isolate them. Some of these techniques are Anti-Hook [2], HoneyID [6], Black-Box [15], Dendritic Cell Algorithm, which is an immune-inspired Machine Learning technique [9], and Bots detection [1], among others. Most are activated based on memory-resident software with lines of code entered in the operating system start-up files. In [4],

they developed a technique that consists of simulating a sequence of keystrokes used as tallow at the entrance. In [10], the development of an application called SMMDecoy has been presented as a cheat-based technique to detect GPU keyloggers. Furthermore, in [14], it is mentioned that the approaches or techniques previously described aim to identify, prevent, and/or eradicate the presence of keyloggers or other malicious software, sometimes without any success.

In another context, the evolution in technology to authenticate income through security pins and passwords is no longer as secure as previously assumed. Instead, new development approaches have been considered to have external control of mobile devices [13] in which the users surf the Internet. Such is the case reported in [11], where a gestural recognition system has been proposed. In [20], the aspects related to Social Engineering have been described. Therefore, it applies techniques where users are tricked into revealing personal information or other data they use in their private applications. Finally, the given literature also reports some other techniques about detecting keyloggers in different environments, including virtual environments [3, 12].

It is word notice that all these research efforts severely lack to coincide with the one presented in the current study. Furthermore, we have identified the interest in developing new keylogger proposals to publicize and promote users' safety and information, which is the predominant reason to propose the present work. In this way, our proposal has a more precise technological application, given that users lack to be aware of the new methods of virus propagation. The efforts made allowed the developed keylogger to have a new approach in the development of security that can prevent it.

3 Keyloggers

A keylogger is a software or hardware that can intercept and save the keystrokes performed on the infected computer's keyboard. This malware is placed between the keyboard and the operating system. It is used to intercept and record the information without noticing the user. Also, the keylogger stores the data locally on the infected computer.

Although there is a wide range of keyloggers, the two main categories are software and hardware. The most used one is the software keylogger, which is part of other malware such as Trojans or rootkits. It appears that this is the easiest to install inside a computer as it does not need to access the machine physically. Another type of software keylogger has the function of imitating an API of the infected computer's operating system, allowing the keylogger to save every press that is performed. Keyloggers with the software are usually part of older malware. Computers are usually infected through a malicious website, which attacks the vulnerable computer by installing malware.

In this context, our study developed a software keylogger to infect Windows-based systems, which occupy the largest market share. Here, three primary protection tools are installed by default on all computers and offer the highest use and the manufacturer's best defensive line. These tools are Firewall, User Account Control (UAC), and the Antivirus system.

The firewall is a mechanism capable of controlling the network traffic that the system receives. It is highly configurable by the user and filters the reception of data to the outside equipment. It is based on filtering rules to control everything the computer crosses. The attackers take this as an advantage. Cybercriminals do not try to deceive it but take advantage of the circumstance that the user often performs many Internet services, located at different IP addresses.

The UAC represents a privilege control mechanism, which checks the user's permissions who have an active computer session. It also limits the actions that endanger the system's integrity by an unauthorized user. The means used to avoid this protection are very diverse and depend primarily on the operating system version. However, the best known and used are those based on DLLs' hijacking, which aims to replace it with another created by the attacker to achieve its execution [24].

As a final point, as for antivirus, all of them follow a systematic and similar analysis process, which often consists of three stages of analysis being static signature, static heuristic, and dynamic. It is intended to identify recognizable patterns in other known malware specimens, such as hashes, instructions, or text strings, by analyzing signatures. This represents a simple but effective technique as the first method of detection. However, it is impossible to recognize new malware specimens since they have not yet been recognized, and their information could not be extracted. Later, heuristic analysis is used, representing a static code analysis, based on heuristic algorithms that try to identify the analyzed software's actions through disassemblers; this determines whether the activities are legitimate or, conversely, represent a risk to the system [18].

4 Development of the Malware

We created a customized and previously unpublished malware. Specifically, we developed our keylogger using the Python programming language. Subsequently, we created an executable for Windows-based systems. Therefore, we performed all implementation phases of the malware life cycle, including the conduction to the specification of requirements and functionalities. This will determine the malware's offensive capabilities, which are detailed in Table 1, along with the fundamental security principle in which they impact.

Table 1. Malware offensive capabilities list.

Funcionality	Principle
Pulsation capture	Confidentiality
Screenshot	Confidentiality
Remote execution of commands	Integrity
Geolocation and capture of hardware capabilities	Confidentiality
Persistence	Integrity

Later, once the malware has been implemented, it is only necessary to generate its corresponding executable file. Then, it allows the program to run on Windows systems without the need for them to have a Python interpreter installed. Therefore, the tool selected for this purpose has been *Pyinstaller*, as it offers encryption functionalities of the executable file generated by a robust cryptographic algorithm. At the end of this process, the resulting executable will execute three main threads responsible for collecting the screenshots, keystrokes, and creating the .backdoor in the infected system. Its operation will be described below by using its flowcharts. The thread is responsible for capturing images on the target screen, which obtains images every seven seconds. Then it sends them to the attacker once he has collected twenty, repeating the process mentioned above continuously (see Fig. 1a).

Afterward, the thread responsible for creating a connection with the attacking server in the form of a backdoor aims to expect orders and execute them at the reception time (see Fig. 1b). Finally, Fig. 1c illustrates how the thread with most of the keylogger functionalities mainly tries to collect the keystrokes, classifying them according to the place they have obtained for a more significant user trace's behavior. Later it sends them to the attacker through electronic mail, in which it encloses a compressed file for the images and a plain text document for the pulsations.

5 Keylogger Validation versus Evasion Techniques

Once there is an active malware based on its specifications and successfully tested in Windows environments, it is necessary to add evasion techniques. It is intended that they complement each other to avoid the security measures discussed previously and run malware without restrictions. Therefore, we may describe below the evaded security measure, the used technique, and its implementation in malware.

5.1 Firewall

The firewall typically notifies and blocks incoming connections to the machine, which have not been previously established. The used evasion technique consists of introducing a client-server program, where the client part is executed as a thread of the keylogger. In this way, a backdoor is created in the system since the attacker keeps listening on an IP address and an invariable port. The client tries to connect continuously, and therefore, the attacker is enabled to execute a reverse shell, through which commands may be sent remotely. Therefore, while the compromised team sends the initial connection to the attacker, the firewall cannot control whether it requests a legitimate service. Even worse, if contrarily it is connected to the attacker's server, it will not be blocked by the firewall when sending its requests once the connection has been established.

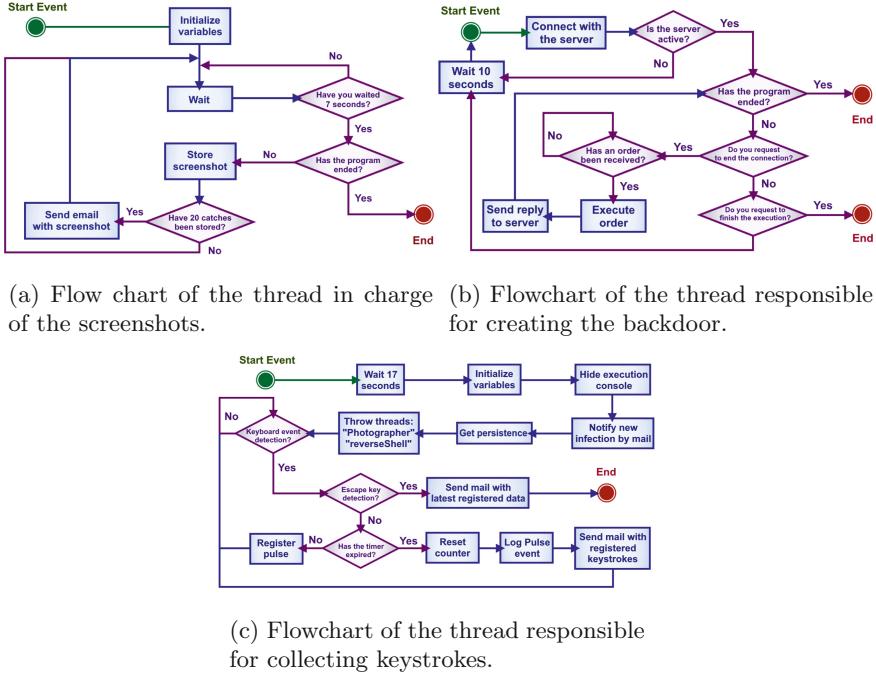


Fig. 1. Flowchart diagrams for creating the keylogger

5.2 User Account Control

This technique is based on how the program *fodhelper.exe* signed by Windows, fulfills the condition *AutoElevate* with the value *True* as explained in [5]. Subsequently, this allows the program to be executed with high integrity. Therefore, all the generated processes are also executed with these permissions. In this way, it has verified that the Windows binary executes a registry key that may be modified by an attacker. Also, it does so with full privileges and without showing any confirmation on the screen. Therefore, to exploit such vulnerability, new functionalities have been included in the malware source code, aiming to modify the system registry keys without special privileges. Furthermore, it can run the program included in Windows *fodhelper.exe* to execute these keys with elevated privileges [16]. Through this technique, malware is transferred to the *C://Programs* directory, which camouflages it below the name of another legitimate program. Later, a new registry key is added at the system start-up to become persistent as the entire process has been executed through a Shell with administrator permissions. After this execution, the directory *C://Programs/SystemUtilitiesC*, where the malware executable file is stored, will be registered. In the same way, a new entry would be registered in the corresponding path. Hereby, due to this new entry, the malware becomes persistent, and it starts on each computer boot indefinitely in time.

5.3 Antivirus

The implementation of evasive measures partially or entirely disables each of the analysis phases performed by antivirus programs on the executable file. In this way, to avoid the first signature analysis, it is only necessary to identify the conflicting code parts. Hereby, a certain antivirus may be considered malicious and modify them. Thus, the heuristic analysis should be avoided. Furthermore, the use has been conducted of compression and encryption techniques provided by the *pyinstaller* tool. On this occasion, the use of this module developed in Python allows it to create executable files compressed using the UPX compressor [17]. It also allows the use of keys to perform encryption using the secure cryptographic algorithm AES-256 [8]. Thus, upon completion of creating the executable, the final file has reduced its size and has obfuscated its executable through encryption. Finally, as the last part of this process, we need to avoid dynamic analysis, which tries to understand malware's actions at runtime, leaving a trace of its operation. After completing this entire evasion process and having included all the described techniques, the resulting executable file is then ready to run and infect computers.

6 Evaluation of Results

We tested the keylogger using Windows and executed it several times in the same VMware virtualized environment. We used its Performance Analyzer tool, of which results were exported to a CSV format dataset to evaluate results. Then, we used R software in order to perform the statistical analysis of them. The parameters evaluated were CPU, RAM, and GPU, consumption during the virus's execution.

The computational device was working correctly. However, when the keylogger was activated, the processing started to increase CPU consumption. Figure 2a illustrates the power of this malware. We can observe that the processing increases steadily, resulting in the collapse of the device. The color diagrams can observe the different states of the processor. The different variants highlight the cycles they obtain in each thread that the processor receives. Concerning RAM, we can observe that the size is continually changing. However, when the malware is activated, RAM begins to collapse. Fig. 2b indicates when the keylogger is on, the RAM is at the highest point. Regarding the GPU, the malware affected the graphical interface, causing a slowdown. Figure 2c illustrates a tremendous change when the keylogger is activated.

In addition, we use the Range diagram to validate the results, which demonstrates the performance when the keylogger is activated. Figure 3a indicates that the throughput, i.e., the device, is not working correctly. In this way, we can observe that the performance objectively grows while remaining constant due to the device's collapse. Likewise, Fig. 3b illustrates the device's performance variant after the activation of the keylogger, which causes its effectiveness. With this, we demonstrated the collapse of a system. Also, we observe the number of critical points in the limits beyond when the device begins to fail.

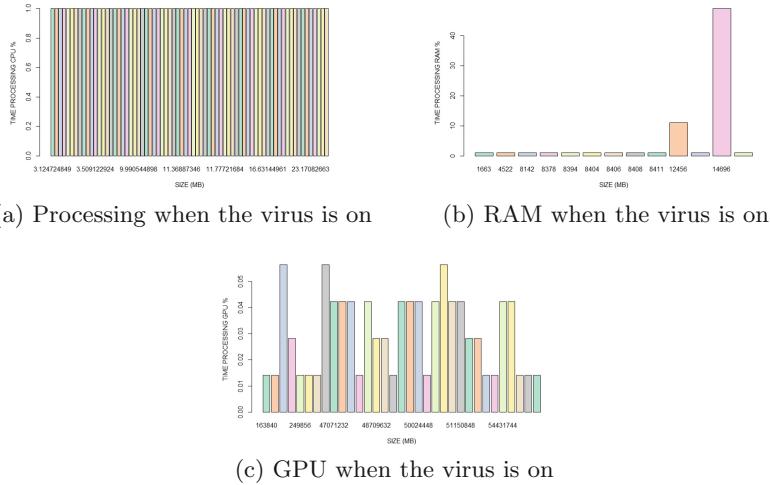


Fig. 2. Statistical processing at the moment of executing the keylogger

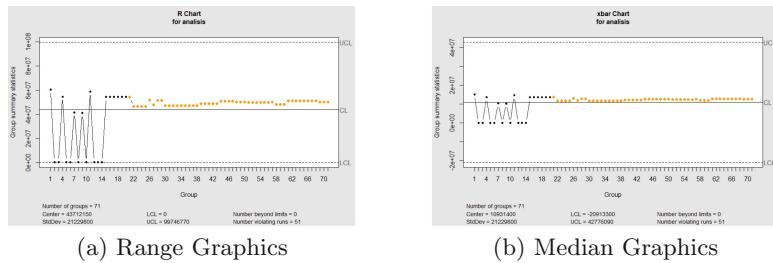


Fig. 3. Performance of the device when the keylogger is activated

From another proof of concept, once all the needed methods have been added in order to avoid the leading security tools used in Windows environments, the executable file resulting from malware was tested on different platforms such as (1) www.virustotal.com; (2) virusscan.jotti.org; and (3) metadefender.opswat.com. Therefore, we may conclude that the antivirus evasion process has been successful. A 100% antivirus evasion rate has been obtained after being analyzed by up to 95 different systems. Besides, it has been verified in a real scenario that the evasion techniques implemented have been sufficient to infect a target computer equipped with a firewall, user account control, and an AVAST antivirus. In this way, we may gather that active and undetectable malware has been developed. This malware can be installed on Windows without the user's consent. It has also demonstrated that malware is hidden among legitimate systems due to its migration to privileged directories by evading user account control, which allows the malware to run with the highest privileges. Therefore, based on the actions mentioned above, conventional keyloggers' benefits

have been improved by including new functionalities focused on monitoring user behavior and trace. However, it has also been possible to avoid security measures installed in the system with administrator permissions and without restrictions, transparent to any user by having a 100% evasion against the antivirus present in the current market.

7 Conclusions

In this study, we demonstrate the effectiveness of the evasion techniques currently implemented in malware. Due to its correct combination during the false life cycle, we managed to generate an executable file undetectable by the antivirus. Also, we managed to bypass the protection systems installed in Windows. The results show the limited security levels of the users and the collapse of the equipment when its performance is affected. Likewise, they demonstrate the need for more significant defensive measures since cybercriminals improve their offensive techniques day by day. In addition to this, we expose the security weaknesses present by default in Windows.

Acknowledgment. We want to thank the resources granted for developing the research project entitled “Detection and Mitigation of Social Engineering attacks applying Cognitive Security, Code: PIC-ESPE-2020-Social-Engineering.” The authors would also like to thank the financial support of the Ecuadorian Corporation for the Development of Research and the Academy (RED CEDIA) in the development of this study within the Project Grant GT-Cybersecurity.

References

1. Al-Hammadi, Y., Aickelin, U.: Detecting bots based on keylogging activities. In: Proceeding of 3rd International Conference on Availability, Reliability and Security, pp. 896–902 (2008)
2. Aslam, M., Idrees, R., Baig, M., Arshad, M.: Anti-hook shield against the software keyloggers. In: Proceeding of the National Conference on Emerging Technologies, pp. 189–191 (2004)
3. Estrada, Z.J., Spraberry, R., Yan, L., Yu, Z., Campbell, R., Kalbarczyk, Z., Iyer, R.K.: Using OS design patterns to provide reliability and security as-a-service for VM-based clouds. In: Proceedings of the 2017 ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, pp. 157–170 (2017)
4. Fangzhou, G., et al.: A novel detection technique for keyloggers. Lecture Notes in Computer Science, vol. 6307, pp. 198–217 (2010)
5. GithubGist: Bypass UAC via fodhelper binary in windows 10 systems. Technical Report. <https://gist.github.com/netbiosX/a114f8822eb20b115e33db55deee6692>
6. Han, J., Kwon, J., Lee, H.: Unveiling hidden spywares by generating bogus events. In: The Proceeding of IFIP 23rd International Information Security Conference, pp. 669–673 (2008)
7. Holz, T., Engelberth, M., Freiling, F.: Learning more about the underground economy: a case-study of keyloggers and dropzones. Lecture Notes in Computer Science, vol. 5789, pp. 1–18 (2009)

8. Jha, A., Sharma, S.: Quantitative interpretation of cryptographic algorithms. In: Emerging Technology in Modelling and Graphics, pp. 459–469. Springer, Singapore. (2020)
9. Jun, F., Yiwen, L., Chengyu, T., Xiaofei, X.: Detecting software keyloggers with dendritic cell algorithm. In: Proceeding of the International Conference on Communications and Mobile Computing, pp. 111–115 (2010)
10. Loutfi, I.: Smmdecoy: detecting GPU keyloggers using security by deception techniques. In: Proceedings of the 5th International Conference on Information Systems Security and Privacy, pp. 580–587 (2019)
11. Luzbashev, A.V., Filippov, A.I., Kogos, K.G.: Continuous user authentication in mobile phone browser based on gesture characteristics. In: Proceedings of the 2nd World Conference on Smart Trends in Systems, Security and Sustainability, WorldS, vol. 8611589, pp. 313–316 (2019)
12. Mallikarjunan, K., Preethi, S.R., Selvalakshmi, S. Nithish, N.: Detection of spyware in software using virtual environment. In: Proceedings of the 3rd International Conference on Trends in Electronics and Informatics, pp. 1138–1142 (2019)
13. Mohsen, F., Bello-Ogunu, E. Shehab, M.: Investigating the keylogging threat in android – user perspective. In: Proceedings of the Second International Conference on Mobile and Secure Services (MobiSecServ), Gainesville, pp. 1–5 (2016)
14. Ortolani, S., Crispo, B.: Noisykey: tolerating keyloggers via keystrokes hiding. In: Proceedings of the 7th USENIX Conference on Hot Topics in Security (HotSec 2012). USENIX Association (2012)
15. Ortolani, S., Giuffrida, C., Crispo, B.: Unprivileged black-box detection of user-space keyloggers, pp. 40–52 (2013)
16. Provecho, E.F.: Testing user account control (UAC) on windows 10
17. PyInstaller: Using pyinstaller – pyinstaller 3.5 documentation. Technical Report
18. Sahay, S.K., Sharma, A., Rathore, H.: Evolution of malware and its detection techniques. In: Information and Communication Technology for Sustainable Development, pp. 139–150 (2020)
19. Solairaj, A., Prabanand, S.C., Mathalairaj, J., Prathap, C., Vignesh, L.S.: Keyloggers software detection techniques. In: Proceedings of the 10th International Conference on Intelligent Systems and Control, ISCO (2016)
20. Tekawade, N., Kshirsagar, S., Sukate, S., Raut, L., Vairagar, S.: Social engineering solutions for document generation using key-logger security mechanism and QR code. In: Proceedings 4th International Conference on Computing, Communication Control and Automation, ICCUBEA, vol. 8697420 (2018)
21. Ucci, D., Aniello, L., Baldoni, R.: Survey of machine learning techniques for malware analysis. *Comput. Secur.* **81**, 123–147 (2019)
22. Willems, E.: Tips for companies: surviving on the internet. In: Cyberdanger, pp. 145–159 (2019)
23. Zambrano, P., Torres, J., Tello-Oquendo, L., Jácome, R., Benalcázar, M.E., Andrade, R., Fuertes, W.: Technical mapping of the grooming anatomy using machine learning paradigms: an information security approach. *IEEE Access* **7**, 142,129–142,146 (2019)
24. Zheng, Y., Liu, F., Hsieh, H.P.: Security risks from vulnerabilities and backdoors. In: Cyberspace Mimic Defense, pp. 3–38. Springer, Cham (2020)



Legal Ethical Implications in the Exercise of Communication and Information Technologies (ICT) in Telemedicine and e-Law in Medellín - Colombia

José Antonio García Pereáñez^(✉) and David Alberto García Arango

Corporación Universitaria Americana, 050012 Medellín, Colombia

jgarcia@coruniamericana.edu.co

Abstract. The categorical assumption of this text is framed in the fact that the practice of Telemedicine in Colombia has allowed an ostensible decrease in human and economic resources in patient care without greater complexity. Likewise, the incipient task with e-Law has agreed to resolve the constant asymmetries between citizens deprived of liberty in the Medellín prisons and the judicial authorities, to invoke rights that are largely unknown. Both Telemedicine and e-Law do not have a guarantee of legal ethical security with the data, thereby generating the presence of an imminent risk that the authorities have not yet intervened. This writing sets out the results of a study carried out at Corporación Universitaria Americana University Institution of Medellín - Colombia, which had three phases: a brief literature review, a consultation analysis by Telemedicine and e-Law and considerations from the ethics and legality of the Technologies of Communication and Information (ICT).

Keywords: Telemedicine · E-Law · Ethics · Legality

1 Introduction

This paper expresses the result of a research project that investigates the social and ethical responsibility of telemedicine and e-Law. For this, a review of the literature was used, and an analysis of the teleconsultation, to then propose the discussion on the legal and the legitimate based on these communication and information technologies (ICT).

The above is based on a first principle that the bioethics discipline proposes, as long as any action that is carried out with the intervention of the Medical Act and the Legal Act are oriented to “do no harm”, which as moral support is located in the order to the ancient Hippocratic precept. The media context for both telemedicine and e-Law is exposed in an unclear regulatory environment and dissipated in its regulation, which inevitably impacts the moral conscience and ethical expertise of the doctor and the lawyer.

According to the World Health Organization (WHO), telemedicine is the provision of health care services, in which distance is a critical factor, by professionals who use information and communication technologies in order to exchange data to make diagnoses, advocate treatments and prevent diseases and injuries, as well as for the permanent

training of health care professionals and in research and evaluation activities, in order to improve the health of people and the communities in which they live [1].

According to [2], the purpose of e-Law is to generate and promote a teaching-learning environment through Communication and Information Technologies (ICT) for those who are in the prison and penitentiary system of the city of Medellín Since the application of Restorative Justice in Colombia is intended to achieve an administration of justice for those deprived of liberty, for the victims, for the officials who ensure their safety and for society in general. E-Law is a bet between the exercise of Colombian Law and its access, through cyberspace. It is an interactive avant-garde communication sustained through the interspatial web, between the legal operator and those deprived of liberty in prisons in the city of Medellín. Through the interactive exercise of e-Law, the benefits of the Penitentiary and Prison Code in Colombia are invoked for those deprived of liberty, according to Law 65 of 1993, especially what is enshrined in article 144.

Telemedicine and e-Law are two bets that, using Communication and Information Technologies (ICT), serve society. The first in health services and e-Law, providing permanent advice to those deprived of liberty. Both are a reality in the Colombian epicenter, each time expanding its coverage, both in health, telemedicine, and e-Law in the legal field. In this way, both doctors and lawyers carry out a social work that transcends the philanthropic scene, attending to an unprecedented moral responsibility in the Colombian deontological context.

Maintaining the life, health and freedom of the population is in the duty of the Medical Act and the Legal Act and in reciprocity with the legislation and Human Rights. But with this new telemedicine and e-Law models, the national norms for their reproduction and function are very limited to determine responsibilities, whether civil, criminal, or administrative, in the provision of the service. The code of medical ethics in Colombia is 39 years behind schedule. Now there is a project that has only passed a debate in the Congress of the Republic, but so far nothing has been resolved in this regard. This is also the case with medical research according to Resolution 8430 of 1993, it is delayed in terms of events in telemedicine and advances in telecare. This is also the case with the deontological codes of the professional practice of the lawyer in Colombia, they are not up to date if they pretend that the Legal Act could be attended by cyberspace.

Therefore, with the exercise of telemedicine and e-Law, they are questioned and without an ethical personality, who is responsible for health situations and legal matters involved in telecare. In these situations, equity, and compliance with the old principles of bioethics are questioned: beneficence, non-maleficence, autonomy, and justice in favor of society. The context of telemedicine and e-Law appears contrary to “anonymous subjects” who remain in cyberspace, which questions a goodwill procedure.

2 Context

The ethical and legal implications between telemedicine and e-law are necessary and pressing in the Colombian context. First, due to the precariousness of health services and judicial procedures. Second, because ICT information and communication technologies, in this case, are at the service of the poorest and most vulnerable, in the understanding - that until now - the Medical Act and the Legal Act are face-to-face, personal and

asymmetric. This study demonstrated another alternative and another implementation model.

Since the 1970s, the term telemedicine appeared as “Distance healing” and included the use of information technologies in health actions that allow improving access to health services [3]. After reviewing the definitions of this concept, it was proposed that telemedicine is “The provision of health care services (where distance is a critical factor) by all health professionals with the use of information and technology technologies. Communication for the exchange of valid information for the diagnosis, treatment and prevention of diseases and injuries, research and evaluation, and for the continuing education of health professionals, all in order to advance the health of individuals and their communities” [4]. This definition is in constant advance and changes due to the constant advancement of technology and its application to health care. For some, tele-medicine and e-health are different concepts and propose e-health as the broadest concept that implies health promotion and disease prevention activities and the continuing education of professionals. In other hand, e-health is defined as “an emerging field at the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies” [5].

The WHO identifies four pertinent elements for telemedicine: 1) Support to the usual activities of the clinic, 2) Reduction of geographical barriers, 3) use of different information technologies and 4) Focus on improving health outcomes.

In Colombia there are several e-health developments such as the telemedicine center of the National University of Colombia. This was created 15 years ago and has provided this service in different regions of the country such as Vichada, Amazonas, Guaviare, Caquetá, Cesar and Sucre, among others. It offers a response service to consultations on topics such as internal medicine, pediatrics, psychiatry, dermatology, gynecology, orthopedics, cardiology, infectiology, urology, otorhinolaryngology, neurology, nutrition and dietetics. Similarly, the Cardiovascular Foundation in Bucaramanga has the National Telemedicine Center in which the interaction between general practitioners and specialties internal medicine, pediatrics, cardiology, dermatology, pain clinic, vascular surgery is carried out. Peripheral, among others.

In other hand, there is a telemedicine plan, Law 1419 of 2010. In this, congress of Colombia established the guidelines for Telehealth in Colombia. In this law, telehealth was defined as the set of activities related to health, services and methods, which are carried out remotely, with the help of information technologies and telecommunications. Tele-medicine and tele-education in health are included [6].

The University of Antioquia, in cooperation with the Government of Antioquia and institutions such as the Red Cross, UbiCuo, UPB and Universidad CES, have developed a strategy to promote telehealth in the department of Antioquia more than two years ago. In its phase II, this strategy aims to carry out actions of telemedicine, pre-hospital telecare, home telecare and tele-education that impact the health of the department of Antioquia. This project is being developed in 50 municipalities of the department and it is intended to expand to the rest of the department in the next two years.

One of the challenges for Telehealth in the world and in our country is to carry out follow-up and evaluation of the impact of these interventions that allow us to know the effectiveness of the intervention, costs and create strategies for constant improvement

of the implementation processes of these programs [7]. In addition, these evaluations make it possible to outline the business model that generates greater feasibility to insert Telehealth programs in the models of health service provision in the countries and regions.

The evaluation of telehealth and e-law projects and programs has been fundamental during the last two years. Through systematic evaluations and reviews, telehealth and e-law have been found to be effective [8]. Systematic reviews have also been found that show an increase in the degree of satisfaction of patients and judicial users with care through ICT [9]. Some of these carried out in developing countries show systematic reviews carried out in Africa and in some developing countries, turning telemedicine and e-law into a promising strategy [10]. After a review of the literature, no impact evaluations of telemedicine and e-law were found in Colombia.

The study of the ethical and legal implications between telemedicine and e-law is necessary, because it undertakes an evaluation process of impact and implementation, which allows actors and decision-makers to recognize the progress and difficulties so that they can make decisions in favor of the improvement of Telehealth and e-law, and the expansion of investments for the department of Antioquia and other municipalities in the country.

This document contains the general evaluation proposal of e-health and e-law strategy for each of the components. It is a design according to the needs of the actors, the moment of development of each component, and the future projections in the department of Antioquia, Colombia. The evaluation proposal that is presented has a comprehensive approach that measures different aspects of telemedicine and e-law such as effectiveness, costs, satisfaction and processes, among others.

The ethical and legal evaluation proposal is based on three systematic reviews of information on telemedicine and e-law issues, in order to identify the results of the evaluations and methodological strategies carried out a priori by other authors. The systematic reviews were: 1) systematic review of ethical studies, 2) systematic review of legal studies, and 3) systematic review of evaluations of both projects regarding their projection with communication and information technologies. For the design of the evaluation methods, a documentary review of the macro project, technical reports of the components, presentations and in-depth interviews carried out with the leaders of the components or with participants was carried out.

3 Methodology

The techniques proposed for the collection of information on the Telemedicine and e-law components were group activities such as focus groups and semi-structured interviews. These techniques made it possible to retrieve information and data derived from the practice and reflections of the subjects. It should be clarified that given the nature of this evaluation, the information collection plan is emerging and changing based on the findings made during the progress of the evaluation process [11, 12].

Semi-structured interviews: For this technique, an interview guide was used mainly with open questions, this allowed collecting detailed information about the perceptions, attitudes, barriers and facilitators identified by participants of the Telemedicine and e-law programs.

In-depth individual interview: It is a special type of individual interview. This interview option was conducted with “key informants”, that is, people with extensive experience and knowledge about the implementation of Telemedicine and e-law. It was implemented primarily to retrieve information about the study context.

Focus group: this technique has a collective character and it was worked with a number of 6 to 8 people. It is important to highlight that this technique is getting richer and reoriented as the field work progresses. For this evaluation, it was used either as a basic data source or as a means to deepen the analysis. Aspects such as can be seen in Table 1 and Table 2.

Table 1. Qualitative techniques used - Tele-medicine and e-law component

Patients	
Objective	Technique
Implementation context	Focus groups: Patients with arterial hypertension or Diabetes Mellitus who make use of Telemedicine
Acceptability of the intervention	Focus groups: Subjects deprived of liberty in prisons of the city of Medellín
Benefits obtained	Focus groups: Medical Act and Legal Act

Table 2. Qualitative techniques to be used in health professionals - Tele-medicine component.

Health and law professionals	
Objective	Technique
Implementation context	Semi-structured interviews: Medical personnel from referral or referral centers or personnel in training who work under the telemedicine modality In-depth interviews: “key informants”, people who have extensive experience and knowledge about the implementation of e-law in prisons in Medellín - Colombia
Perception of use	Semi-structured interviews: Medical staff and practicing lawyers
Acceptance of the intervention	Semi-structured interviews: Medical personnel from referral or reference centers or personnel in training who work under the modality of telemedicine and e-Law
Benefits obtained	Semi-structured interviews: Medical and e-law personnel from referral or reference centers or personnel in training who work under the telemedicine and e.-law modality

This evaluation bases its analysis on the founded theory, which implies an approach focused on the construction of explanatory models that are supported by the data. The analysis process is based on the constant comparison between the data that, if possible,

should be ordered and examined the closest in time, in order to achieve certain precision and avoid forgetting or interference with other information found during field work [12].

The data obtained during the field work were captured, compared with each other and analyzed, to later arrive at the final reflections. Although the collection of information and the analysis are presented in two different steps, in reality, both will be carried out at the same time and concurrently. Under this method, the investigative process was dynamic, recursive and orderly, this because the data will be systematically categorized and the final results will depend on the operation of these categories.

This research followed the principles of the Declaration of Helsinki, especially what is mentioned in Article 6, which stipulates that the well-being of the person participating in the research must always take precedence over all other interests. Additionally, it responds to article 23, taking all kinds of precautions to protect the privacy of the person participating in the investigation and the confidentiality of their personal information and to minimize the consequences of the investigation on their physical, mental and social integrity [13].

This is an investigation that had no risks according to resolution 8430/1993 of the Ministry of Health [14], since no diagnostic tests, laboratory or medical treatments were performed. In the same way, although the patients intervened are monitored through telemedicine and e-law, they continue with their usual treatments for the control of the disease for which they consult without putting their health at risk, as well as on the legal procedures of e-law.

This study is relevant for socializing at the Ninth World Conference on Information Systems and Technologies | Terceira Island, Azores, Portugal, because the results of this research are related to two of the conference topics: Ethics, Computing and Security (ECS) and Human-Computer Interaction (HCI).

4 Telemedicine

Telemedicine is a tool that is imperatively providing a solution during the health crisis and has made its way in the context of the coronavirus pandemic. Previously criticized and with many other edges for its interpretation, telemedicine now becomes an estimable instrument with much social acceptance for representing a feasible solution for the care of patients in the context of isolation due to a disease of public interest.

There is a conjunctural dilemma of communication in the medical act and in scenarios of violation of human rights, which by rigor are immersed in virtual care. Telemedicine brings solution-focused medical attention to the patient with health problems, but at the same time, it blurs the value of the physical, of human contact, since it dispenses with highly valued aspects of the medical act such as: inspection physical, palpation and auscultation of the patient. The individual is digitized, it becomes a difficult scenario and not a convalescent human being, who needs to subordinate his will and corporality to the expertise of the doctor. García, José and others [11].

Telemedicine, understood in its original state, responds to the requirements of a non-face-to-face medical consultation, and that one of its advantages is the decrease in the use of health resources, thus it also greatly reduces unnecessary travel to hospitals and treatment centers. The implementation of telemedicine is useful as a complementary

virtual tool in the follow-up of patients, especially the chronic patient, who on many occasions is far from the medical center and mediated by a rugged geography with difficult communication routes. Despite all this, telemedicine cannot claim to generalize all healthcare processes into a single virtual component [12].

It is also true that the practice of telemedicine creates uncertainty among patients and among physicians, which is expressed in multiple expressions of dissatisfaction on the part of users. Doctors with some experience assure that it is a risky exercise and that it focuses its usefulness on unique and specific cases. However, young, inexperienced doctors see telemedicine as a very useful tool and incorporate it very easily into their customary practice. This is evidenced by several studies, among which it stands out [13].

Although medical teleconsultation is in the genesis of telemedicine, this does not stop generating skepticism, first in the patient and in the doctor. It is evident that the lack of physical contact compromises the autonomy of both actors: the patient and the doctor, this implies doubts with the management. All this can have an influence on patient safety, especially because in Colombia there are no regulations regarding legality and legitimacy for telemedicine and the legal risks for this practice do not exist. In this context, the Medical Record, the Informed Consent, and the Medical Sigil are exposed to a free interpretation, exempt from any guarantee by the media standardization of teleconsultation. Garcia, Jose [14].

Law 1419 of 2010 [15] establishes the government guidelines that must be followed in Colombia in Telehealth, with the purpose of supporting the General System of Social Security in Health, invoking the principles of: quality, unity, integrality, solidarity, universality, and efficiency. The Law proposes provisions and definitions related to the implementation of communication and information technologies that are being adopted in Colombia with the Health service, especially with the practice of telemedicine. Additionally it defines telemedicine as the provision of distance health services in the components of promotion, prevention, diagnosis, treatment and rehabilitation, by health professionals who use information and communication technologies, which allow them to exchange data with the purpose of facilitating access and opportunity in the provision of services to the population with limited supply, access to services or both in their geographic area [16].

Although the law in Colombia proposes guidelines for telehealth and telemedicine, it does not provide effective protection for patients who use these services. Rather, it appears as a proposal for the creation of an advisory committee to program virtual health assistance services, with implications for several ministries, including: Ministry of Health and Social Protection, Education, Finance, Housing and Environment. With this, it is intended a business line that produces money for the State and not, an alternative that improves the health processes of Colombians and protects them from possible abuses and depersonalized management of the virtual interactive web in telecommunications.

The legal ethical implications with the exercise of informatics in telemedicine are not mentioned in Law 1419 of 2010 [17], on the contrary, some possible functions are proposed for an advisory committee on education for telehealth, in order to identify connectivity management and the use of information technology and telecommunication as a “business idea” and for financing the State’s own programs. With this, resources are produced for the Ministries involved in the project. The presence of the legal entity

as the controller of this practice and of bioethics as the legitimate interlocutor in the telehealth and telemedicine procedure are totally absent, they are not even named. Law 1419 of 2010 identifies itself as a "Knowledge Management" of Telehealth learning and for this it involves entities in higher education so that they include this training in the area of health, as well as in the studies of systems engineering and telecommunications. All this presents the telehealth and telemedicine programs as a state program to produce profits.

5 e-Law

E-Law as a neologism and as a practice with the exercise of the right in a virtual way - through synchronous and asynchronous actions proposed by the internet - is an idea and a practice of the Corporación Universitaria Americana University Institution, of Medellín - Colombia. This experience has been implemented since 2018 and consists on the integration of technological resources with legal procedures. It is an innovative, genuine, and unique proposal of its kind, which has been carried out with success especially with the prison population. For this and through cyberspace, the rights of those deprived of liberty are invoked. Their cases are analyzed by eminent jurists, who from different geographical points access the interactive web platform or by an app configured in remote law for the legal defense of life -Biolaw- [13].

E-Law is made visible through an interactive web platform, through which a permanent Legal Council is maintained, via chat, made up of prestigious jurists from Colombia, Spain, and Peru, for now. But an attempt is being made to extend the legal ties and professional solidarity with other Latin American jurists who give their vote and assent to this proposal to analyze the case of persons deprived of liberty, which have not yet received a ruling. E-Law as an exercise in legal informatics is inserted in the process of reintegration and re-socialization of the social life of inmates in the city of Medellín and in the Department of Antioquia - Colombia. It is intended that by means of information technology and telecommunication tools, inmates invoke their rights before the competent authorities in the context of Restorative Justice.

The purpose of e-Law is to generate and promote a teaching and learning environment for those who are in the prison and penitentiary system, since the application of restorative justice is intended to achieve an administration of justice for those deprived of liberty, to victims, for officials who ensure their safety and for society in general [12].

The Colombian penitentiary system during the coronavirus pandemic is collapsed, the levels of overcrowding in the prisons of the main cities of the country present high levels of contagion and a large number of those deprived of liberty await due process, which has been ostensibly delayed by the declaration of public calamity. Therefore, the administration of justice represented in the Judicial Branch made an unprecedented authorization: judicial processes could be audited by teleconferences, guardianships could be established by email, hearings and trials could be proposed virtually. All of this set up a new context for e-Law, which was unthinkable a few months ago.

E-Law is a proposal of restorative justice in Colombia, this will allow in a timely manner the re-socialization of those deprived of liberty and their social inclusion. Non-Governmental Organizations, national and international NGOs whose objective is to

defend Human Rights, as well as the International Red Cross and Cáritas, have denounced the overcrowding in Colombian prisons, unjust detentions, serious deficiencies and increasing slowness in the judicial processes that they keep men and women behind bars unfairly. E-Law is proposed as an interactive avant-garde virtual communication, sustained through the interspatial web, between the legal operator and those deprived of liberty in prisons in Antioquia-Colombia. Through the interactive exercise of e-Law, the benefits of the Colombian Penitentiary and Prison Code are invoked for those deprived of liberty, according to Law 65 of 1993 [18], especially what is enshrined in article 144.

6 Discussion

Although telemedicine is a virtual practice that achieves access to health services, it is also possible to affirm that we are facing another situation of the Medical Act mediated in cyberspace by a computer connected to the internet. The achievements of this practice are widely known: patient protocols are identified and socialized with specialists in different geographical locations, having opinions on diagnostic images in a very short time, documentation on clinical cases is managed, remote controls are carried out, communication is carried out. Synchronous and asynchronous with patients, among others.

Telemedicine and e-Law are two computing bets in communication and information technologies, which fulfill an important social service in two different and pressing contexts for the Colombian population, but in this country none of these practices has a clear control on the part from the authorities [19, 20]. They are two activities that could remain in the purposes of indeterminate, of anonymous characters, who instead of doing good, could cause great damage. Therefore, the construction of public policies is urgently needed to propose ethical-legal controls to these practices [21, 22], which, although they are marked in the sphere of the noblest beneficence, could also remain in the space of the most malicious slander and for this, for now, there would be no responsible.

7 Conclusions

The benefits of telemedicine exercise are not in doubt, but both for this practice and for e-Law in Colombia, an efficient and effective control is urgently needed to account for decision-making. Therefore, its regulation cannot be postponed. These two practices cannot leave their duty to be during the uncertainty and skepticism of the population. In ordinary legislative life, both the doctor and the lawyer are objects of multiple lawsuits and their exercise is scrutinized by medical ethics tribunals and disciplinary chambers under the discretion of the Superior Council of Adjudication; the first for doctors and the second for lawyers. Colombia is a Social State of Law, which in its principles demands truth and neatness in procedures, especially when they mediate the rights of the most vulnerable.

Despite the advances in information and telecommunication technologies, our context is still very little aware of these practices, especially when they are the virtual vehicle for situations as worthy and pressing as they are, the implications of the Medical Act

of the Legal Act. Health and freedom for the citizen are determined in their fundamental rights. For this reason, it is necessary that the Congress of the Republic legislate on this matter. To the date this article is written, there are so many citizens who feel their rights violated due to malpractice mediated by information and telecommunication technologies (ICT).

References

- Burg, G., Hasse, U., Cipolat, C., Kropf, R., Djamel, V., Soyer, H.P., et al.: Teledermatology: just cool or a real tool Dermatology. **210**, 169–173 (2005)
- García, J.: Telederecho: una alternativa de justicia restaurativa a través del uso de las tecnologías de información y comunicaciones, en las cárceles de la ciudad de Medellín. Revista Ibérica de Sistemas e Tecnologías de Informação; Lousada N.º E28, April 2020, pp. 1042–1050. RISTI (2020)
- World Health Organization: WHO. TELEMEDICINE. Opportunities and developments in Member States. Report on the second global survey on eHealth. Global Observatory for eHealth series, vol. 2, p. 94 (2010). https://www.who.int/goe/publications/goe_telemicine_2010.pdf
- World Health Organization, & WHO Group Consultation on Health Telematics: A health telematics policy in support of WHO's Health for all strategy for global health development (1997)
- Eysenbach, G.: What is e-health? J. Med. Internet Res. **3**(2) (2001). <https://doi.org/10.2196/jmir.3.2.e20>
- Congreso de Colombia: Ley 1419, “por la cual se establecen los lineamientos para el desarrollo de la telesalud en Colombia”. Congreso de Colombia, Bogotá (2010). <https://wsp.presidencia.gov.co/Normativa/Leyes/Documents/ley141913122010.pdf>
- Khoja, S., Durrani, H., Scott, R.E., Sajwani, A., Piryani, U.: Conceptual framework for development of comprehensive e-health evaluation tool. Telemed. J. E Health **19**(1), 48–53 (2013). <https://doi.org/10.1089/tmj.2012.0073>
- Ekeland, A.G., Bowes, A., Flottorp, S. (desde 736 hasta 771 páginas): Efectividad de la telemedicina: Una revisión sistemática de las revisiones, p. 79. <https://doi.org/10.1016/j.ijmedinf.2010.08.006>
- Mair, F., Whitten, P.: Systematic review of studies of patient satisfaction with telemedicine (2000). <https://doi.org/10.1136/bmjj.320.7248.1517>
- Blaya, J.A., Fraser, H.S., Holt, B.: E-health technologies show promise in developing countries. Health Aff (Millwood) **29**(2), 244–251 (2010). <https://doi.org/10.1377/hlthaff.2009.0894>
- Aubel, J.: Manual de Evaluación Participativa del Programa. Involucrando a los participantes del programa en el proceso de evaluación. USAID (2000). <https://evalparticipativa.net/wp-content/uploads/2019/05/11.-manual-de-evaluacion-participativa-del-programa.pdf>
- Sandoval, C.: Especialización en teoría, métodos y técnicas de investigación social: Investigación cualitativa. [Sitio en Internet] (2002). <https://docs.google.com/viewer?a=v&pid=sites&srid=ZGVmYXVsdGRvbWFpbnxjdWFsaXRhdGl2YXVuWNvcmR8Z3g6MWZIYTk4MWNjOGU4ODUwNw>
- Asociación Médica Mundial: Declaración de Helsinki de la Asociación Médica Mundial. Principios éticos para las investigaciones médicas en seres humanos. Helsinki, Finlandia (1964)
- República de Colombia, & Ministerio de Salud: Resolución 8430 de 1993, “Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud” (1993)

15. García, J., Otros: Situaciones sobre telemedicina en Colombia: entre lo legal y lo legítimo. In: 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), Coimbra, Portugal, 19–22 June 2019, ISBN: 978-989-98434-9-3 (2019)
16. García, J., Lince, A.: Telederecho: un avance en la administración de justicia. Derecho, Sociedad y Justicia para el Desarrollo. Corporación Universitaria Americana. Sello Editorial. En (2020). <https://americana.edu.co/medellin/wp-content/uploads/2020/09/Derecho-Sociedad-y-Justicia-para-el-desarrollo.pdf>. Consultado el 2 de noviembre de 2020
17. García, J., Lince, A., Ledezma, J.C.: Telederecho: una alternativa de justicia restaurativa a través del uso de las tecnologías de información en el Consultorio Jurídico de la Corporación Universitaria Americana. Corporación Universitaria Americana. Sello Editorial. En (2020). <https://americana.edu.co/medellin/wp-content/uploads/2020/11/Realidades-transversales-al-derecho.pdf>. Consultado el 27 de noviembre de 2020
18. García, J.: Ethical and legal arguments about telemedicine in Colombia. J. Comput. Commun. Scientific Research Publishing. ISSN Online: 2327-5227 En (2017). https://www.scirp.org/pdf/JCC_2017042617172349.pdf. Consultado: Octubre 12 de 2020
19. República De Colombia, Ley 1419 de 2010, artículo 2. En. https://www.funcionpublica.gov.co/eva/gestornormativo/norma_pdf.php?i=40937. Consultado el 2 de noviembre de 2020
20. República De Colombia, Ley 65 de 1993 En. <https://wp.presidencia.gov.co/sitios/normativa/leyes/Documents/Juridica/Ley%2065%20de%201993.pdf8>
21. República De Colombia, Constitución Política de 1991. En. <https://pdbs.georgetown.edu/Constitutions/Colombia/colombia91.pdf>. Consultado el 2 de noviembre de 2020
22. República De Colombia, Código de Procedimiento Penal, en sus artículos 518 a 521. En. https://perso.unifr.ch/derechopenal/assets/files/legislacion/l_20190708_03.pdf. Consultado el 1 de octubre de 2020



A Proposal for Artificial Moral Pedagogical Agents

Paulo Roberto Córdova¹(✉) , Rosa Maria Vicari¹(✉), Carlos Brusius²(✉), and Helder Coelho³(✉)

¹ Federal University of Rio Grande do Sul, Porto Alegre, Brazil
paulo.cordova@ifsc.edu.br, rosa.inf@ufrgs.br

² Hospital Moinhos de Ventos, Porto Alegre, Brazil
cbrusius@uol.com.br

³ Faculty of Sciences, University of Lisbon, Lisbon, Portugal
hcoelho@di.fc.ul.pt

Abstract. While Artificial intelligence technologies continue to proliferate in all areas of contemporary life, researchers are looking for ways to make them safe for users. In the teaching-learning context, this is a trickier problem because it must be clear which principles or ethical frameworks are guiding processes supported by artificial intelligence. After all, people education are at stake. This inquiry presents an approach to value alignment, in educational contexts using artificial pedagogical moral agents (AMPA) adopting the classic BDI model. Besides, we propose a top-down approach explaining why the bottom-up or the hybrid one may would not be advisable in educational grounds.

Keywords: Value alignment · Artificial Intelligence · Education

1 Introduction

The growing and fast advancement in Artificial Intelligence (AI) field have raised different concerns regarding, among other things, the impacts of its increasingly widespread application in a wide variety of areas. One possible reason for this quick expansion of AI, according to Cervantes et al. (2019), consists in aiming to delegate part of their decision-making power to artificial agents. Bots (algos) are becoming popular and dangerous.

As AI continues to proliferate into several areas of life, it becomes evident that society needs to think about the potential impact that it will have. AI needs to present reliability, so that people can trust that it is safe to coexist and interact with it. In order to fully benefit from the potential of AI, it needs to make sure that these technologies are aligned, with our moral values and ethical principles (Dignum et al. 2018).

Thus, researchers in many different areas have examined how to implement moral intelligence, facing a wide range of challenges as: how to translate ethical principles into computational models, how to avoid data bias that imply in the replication of human

prejudices, how to turn intelligent systems accountable for its decision and choice-making, etc. Such research efforts are organized under the broader term so-called Value Alignment (VA) (Kim et al. 2019).

In this context, VA has becoming an important issue in many areas because there are no consensual answers capable to definitely solve all questions regarded to it. In educational context, it is specially a problem as educational systems, in most cases, are connected to the socio-ethnic contexts of their users.

This means that we must consult not only AI specialists, but also experts in other areas related to educational context to be able to establish the right priorities regarding moral values. Considering the need for principles and values to guide people's behavior in this context, we propose a model for value alignments in educational context using a top-down approach.

Therefore, to justify our proposal, we first present some definitions and challenges concerned the value alignment problem. Then, we present our proposal justifying why a top-down approach is better than a bottom-up or hybrid one, in educational contexts. To conclude, we expose some benefits and constraints of our proposal in a real environment.

2 The Value Alignment Problem and Its Challenges

Value Alignment in AI can be defined as the set of efforts to build systems adhering to human ethical values (Aliman and Kester 2019). It is an area to which the need for researches and solutions is clear, as we will show along this paper. This need, in turn, comes from the fact that, being increasingly involved in social relations and interacting with humans, it is almost inevitable that will be situations in which artificial agents will need to deal with ethical dilemmas in their decisions and choices-making. Beyond that, ethics has becoming a crucial issue in the technological realm (Costa and Coelho 2019).

However, despite the notable efforts to deal with the so-called VA problem, we are still far from a consensus on the best solutions. To demonstrate this, we will describe some approaches for VA, categorizing the artificial moral agents (AMA) according to the classification proposed by Allen et al. (2005), namely: top-down approaches, that are based on logical representations of ethical theories such as deontological and utilitarianism ethics; bottom-up approaches, that make use of learning mechanisms to guide their behaviors and; hybrid approaches, in which there are the use of both, top-down and bottom up approaches simultaneously. In the latter, agents are able to show an evolving capability moral judgments.

Regarding top-down approaches, the use of deontological, utilitarian structures, the double-effect doctrine and variants of these frameworks, as exemplarism and augmented utilitarianism, have been proposed over the last years. Logical representations, pure and structured utility functions to support multi-objective approaches, have been observed frequently to implement them. Besides, the top-down approaches still face many tricky challenges, namely: perverse instantiation, temporal complexity and context changing in decision and choice-making (Aliman et al. 2019; Thornton et al. 2019; Vamplew et al. 2018; Dehghani et al. 2008; Anderson and Anderson 2008; Cervantes et al. 2019).

Agents based on bottom-up approaches, as aforementioned, make use of learning mechanisms to improve its decision-making and choice-making process, and hence, its

ethical behavior. In this sense, reinforcement learning and inverse reinforcement learning have been utilized more frequently to implement this approach. As for its challenges, one can highlight the data bias, problems regarding generalization, avoiding naturalistic fallacy and complicated norms representations (Arnold et al. 2017; Kim et al. 2019; Cervantes et al. 2019).

Finally, in hybrid approaches, commonly one can see proposals presenting learning mechanisms being either guided by rules or constrained by them in its learning processes (Arnold et al. 2017; Wallach et al. 2010). Furthermore, it is also possible to find proposals to validate ethical principles by using empirical observation aiming to determine whether values previously defined into the systems are applicable in the real world or not (Kim et al. 2019).

Concerning its challenges, as hybrid approaches make use of both ethical frameworks and learning mechanisms, it faces the same set of challenges that top-down and bottom-up approaches. Besides, identifying and combining different ethical frameworks, machine learning methods and even neuropsychological approaches are among some of challenging tasks tackled by researchers (Thornton et al. 2019).

As one can see, there are many questions and issues when it comes to the VA problem. To answer them seems to be crucial to take the next step towards the future of reliable AI.

3 A Proposal for Value Alignment for Educational Context

Ethical issues are always complex, mainly due to their temporal, cultural and context dependence, in addition to the subjectivity of judgments and multiple points of view. The lack of agreement among moral philosophers, on which theory of ethics should be followed, also may be considered an obstacle to the development of machine ethics (Bostrom 2014; Brundage 2014).

On the other hand, identify ethical frameworks to guide human behavior is still important and has been one of the primary themes of philosophical thought (Vamplew et al. 2018). When it comes to educational context, it must to be clear which principles and ethical frameworks drive people behavior as there are serious concerns regarding behaviorism in the classroom.

Yet, it is not the intention of this paper to define which principles or ethical frameworks should be implemented in educational contexts, but to defend the need to have them. This need, in turn, is as true for human agents as it is for artificial agents.

Intelligent systems have becoming more and more present in our lives and it is not different in classrooms. Pedagogical agents, for instance, due to its properties of autonomy, social ability, persistence, capability to learn and be represented by characters, are capable to support teaching-learning processes in many different ways (Giraffa et al. 1999). Hence, ethical concerns in AI must be taken account in classroom likewise that in any other environment.

Thus, we propose a model to Artificial Moral Pedagogical Agent (AMPA), similar to artificial moral agents (AMA), but focused on pedagogical issues, supporting teaching-learning processes by means of different possible strategies. Such pedagogical agents

should be structured in a top-down approach, so that it can be guided for some ethical framework, such as deontological or utilitarianism, turning it more predictable and controllable.

In this sense, we adopted a mental states approach, more specifically the BDI model, since our team have an expertise on this, having developed different solutions using this model from affective computing to intelligent tutoring systems (Giraffa and Viccari 1998; Viccari and Giraffa 2002; Jaques and Viccari 2004). As for its application for VA, there are some attempts to extend BDI architecture implementing AMAs whereby a top-down approach (Honarvar and Aghaei 2009; Wiegel 2015). Besides, BDI approaches have been applied to implement AMA using bottom-up and hybrid approach (Dennis et al. 2016). However, what we are proposing is the use of the classic BDI model as showed below (Fig. 1).

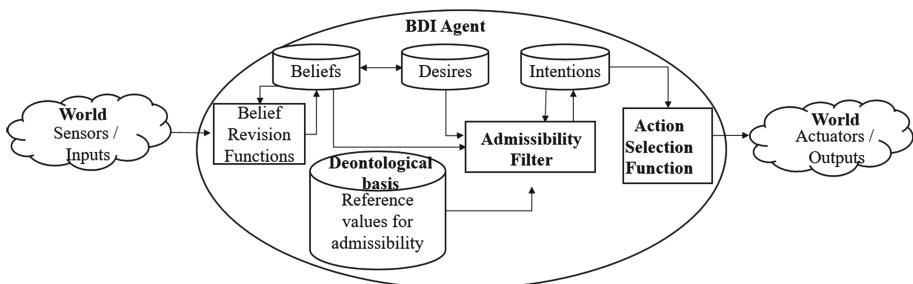


Fig. 1. BDI model for AMPA

In this model, we delegate responsibility for the ethical selection of intentions on the Admissibility Filter (AF). For this, the AF will be linked to a set of ethical rules implemented in an ontology, giving the model a deontological basis for decision making able to judge when a given action is against an ethical principle. In addition, the AF will be endowed of ethical reasoning capability whereby the Hedonistic Act Utilitarianism (HAU), based on the Jeremy's theory (Anderson and Anderson 2008), giving the model a utilitarian basis for decision to deal with ethical dilemmas.

There are other architectures for agent's implementation as reactive architectures, logic-based architectures and more recently, Agent_Zero architecture (Wooldridge 2001; Epstein 2013). However, as aforementioned, our team has been researching and developing solutions using the BDI architecture for educational software development for several years. For these reasons, we choose BDI architectures to propose our solution. The BDI architecture manages to achieve our goals for this application.

For the time being, we discourage the bottom-up or hybrid approach to the educational context, especially when applied to the teaching-learning processes. That's because there are problems like data bias and naturalist fallacy that could pose unnecessary risks to other actors of the process. Besides, different from adaptive systems that use machine learning (ML) to customize teaching processes (Daniel et al. 2015), moral agents using this resource could learn unethical, reprehensible behavior while observing student's behavior (Arnold et al. 2017).

4 Discussion and Final Considerations

Artificial moral pedagogical agents are pedagogical agents capable to deal with ethical dilemmas when it is needed. The main function of AMPAs remains being to support teaching-learning processes, acting autonomously and making explanations and decisions in order to accomplish its major objective.

However, sometimes an AMPA may come across an ethical dilemma, that is, a situation where there is no satisfying decision, and hence, one decision making will override one moral principle (Aroskar 1980). According to Cervantes et al. (2019), there are two non-exclusive situations where ethical conflicts may occur: within an agent, when its ethical norms are in conflict; and, between two agents, when they diverge on what the appropriate ethical decision. The latter, may involve both an interaction between two artificial agents as well as an interaction between an artificial agent and a human.

In a teaching-learning process one could face a situation where a pedagogical agent ought to decide whether interfere or not in a student's actions, considering the autonomy principle. Another situation could involve a student that, for justified reasons, needs more time to conclude a task whose time has expired. In this case, the agent will have to decide if ought to follow the rule about the timing of the task or violate it for the sake of the student.

In this work we propose to endow with ethical reasoning capability, agents involved in interactions of collaborative learning that occurs into the collaborative Learning Management Systems (LMS). In such scenario, due to the implicit nature of the interactions among the involved agents, either human or artificial ones, many ethical dilemmas may appears, as afore-mentioned.

Thus, AMPAs should be able to deal with moral decisions in order to prevent system's undesirable, unethical behavior to improve its reliability or to guide teachers and students regarding moral values and ethical behavior or both. Such characteristics can be useful in many teaching approaches, like cooperative learning, serious games, problem-based learning, Intelligent Tutors Systems, etc.

Up to now, we do not find similar studies aiming to solve the kind of problems and challenges we have described in teaching-learning processes supported by artificial intelligence. There are works on the VA problem in many areas and contexts, but not much in Education area, which makes this proposal very relevant.

To conclude, a BDI model is a good way to implement such kind of artificial moral pedagogical agents in a top-down approach. Considering the state of the art of the AMA technologies and the current solutions proposed for value alignment in AI, we believe in the feasibility of this proposal.

Acknowledgements. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

- Anderson, M., Anderson, S.L.: Ethical healthcare agents. In: Sordo, M., Vaidya, S., Jain, L.C. (eds.) Advanced Computational Intelligence Paradigms in Healthcare-3, pp. 233–257. Springer, Heidelberg (2008)

- Aliman, N.M., Kester, L.: Requisite variety in ethical utility functions for AI value alignment. In: Workshop on Artificial Intelligence Safety 2019, vol. 2419. CEUR-WS, Macao (2019)
- Aliman, N.M., Kester, L., Werkhoven, P.: XR for augmented utilitarianism. In: IEEE International Conference on Artificial Intelligence and Virtual Reality 2019, pp. 283–285. IEEE, San Diego (2019)
- Allen, C., Smit, I., Wallach, W.: Artificial morality: top-down, bottom-up and hybrid approaches. *Ethics Inf. Technol.* **7**(3), 149–155 (2005)
- Arnold, T., Kasenberg, D., Scheutz, M.: Value alignment or misalignment—what will keep systems accountable? In: AAAI Workshop on AI, Ethics, and Society 2017. AAAI Press, Palo Alto (2017)
- Bostrom, N.: Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Oxford (2014)
- Brundage, M.: Limitations and risks of machine ethics. *J. Exp. Theor. Artif. Intell.* **26**(3), 355–372 (2014)
- Cervantes, J.A., et al.: Artificial moral agents: a survey of the current status. *Sci. Eng. Ethics* **26**(2), 501–532 (2019)
- Daniel, J., Cano, V., Cervera, M.G.: The future of MOOCs: adaptive learning or business model? *Revista de Universidad y Sociedad del Conocimiento* **12**(1), 64–73 (2015)
- Costa, A.R., Coelho, H.: Interactional moral systems: a model of social mechanisms for the moral regulation of exchange processes in agent societies. *IEEE Trans. Comput. Soc. Syst.* **6**(4), 778–796 (2019)
- Dennis, L.A., Fisher, M., Lincoln, N.K., Lisitsa, A., Veres, S.M.: Practical verification of decision-making in agent-based autonomous systems. *Autom. Softw. Eng.* **23**(3), 305–359 (2016a)
- Dignum, V., et al.: Ethics by design: necessity or curse? In: AAAI/ACM Conference on AI, Ethics, and Society 2018, vol. 18, pp. 60–66. ACM, New York (2018)
- Kim, T.W., Donaldson, T., Hooker, J.: Grounding value alignment with ethical principles. arXiv preprint (2019)
- Epstein, J.M.: Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science. Princeton University Press, Princeton (2013)
- Giraffa, L.M.M., Viccari, R.M.: The use of agents techniques on intelligent tutoring systems. In: Proceedings SCCC 1998, 18th International Conference of the Chilean Society of Computer Science 1998. IEEE, Antofogasta (1998)
- Giraffa, L., Móra, M., Viccari, R.: Modelling an interactive ITS using a MAS approach: from design to pedagogical evaluation. In: IEEE Third International Conference on Computational Intelligence and Multimedia Applications 1999, vol. 3. IEEE, New Delhi (1999)
- Honarvar, A.R., Ghasem-Aghaei, N.: Casuist BDI-agent: a new extended BDI architecture with the capability of ethical reasoning. In: International Conference on Artificial Intelligence and Computational Intelligence, pp. 86–95. Springer, Heidelberg (2009)
- Jacques, P.A., Viccari, R.M.: A BDI approach to infer student's emotions. In: Ibero-American Conference on Artificial Intelligence (IBERAMIA). Advances in Artificial Intelligence, Puebla, vol. 3315, pp. 901–911. Springer, Heidelberg (2004)
- Thornton, S.M., et al.: Incorporating ethical considerations into automated vehicle control. *IEEE Trans. Intell. Trans. Syst.* **18**, 1429–1439 (2019)
- Vamplew, P.: Human-aligned artificial intelligence is a multi-objective problem. *Ethics Inf. Technol.* **20**, 27–40 (2018)
- Wallach, W.: Robot minds and human ethics: the need for a comprehensive model of moral decision making. *Ethics Inf. Technol.* **12**(3), 243–250 (2010)
- Wooldridge, M.: Intelligent agents: the key concepts. In: ECCAI Advanced Course on Artificial Intelligence, vol. 2322, pp. 3–43. Springer, Berlin (2001)

Human-Computer Interaction



State of the Art of Human-Computer Interaction (HCI) Master's Programs 2020

Gabriel M. Ramirez V.¹ , Yenny A. Méndez² , Antoni Granollers³ , Andrés F. Millán¹ , Claudio C. Gonzalez¹, and Fernando Moreira⁴

¹ Escuela de Ciencias Básicas Tecnología e Ingeniería, Universidad Nacional Abierta y a Distancia, Av. Roosevelt #36 - 60, Cali, Colombia

{gabriel.ramirez, andres.millan, claudio.gonzalez}@unad.edu.co

² Centro de Investigación en Ciberseguridad, Universidad Mayor, Santiago, Chile

yenny.mendez@umayor.cl

³ Escola Politècnica Superior, Universitat de Lleida, Lleida, España

toni.granollers@udl.cat

⁴ REMIT. IJP, Universidade Portucalense, Porto & IEETA, Universidade de Aveiro, Aveiro, Portugal

fmoreira@upt.pt

Abstract. This paper presents the wide review of the state of the university masters or graduate programs in the discipline of Human-Computer Interaction (HCI) and related areas up to 2020. This work has been done as part of the design and development process of a new interuniversity and international master's program graduate developed between the Universidad Nacional Abierta y a Distancia (UNAD) of Colombia and the Universidad de Lleida (UdL) in Spain. The review was necessary to know what other institutions did, the needs of the market and organizations, contrast with our own experiences and ideas to propose of graduate program. It was done by searching for information using educational databases, listings that group masters or graduate program at the level of master in these areas, search engines for master's degrees, articles, documents of organizations that work on HCI. The objective of achieving the state of the masters in HCI is to know what universities are teaching worldwide compared to the needs of companies or organizations and if there is a congruence between what is being taught at the level world in the academy and the requirements of organizations.

Keywords: Human interaction computer · User experience · Graduate education · Higher education

1 Introduction

The paper presents the review of the state of the art of university graduate programs offered in the topics related to Human-Computer Interaction discipline (and related areas) up to 2020. The study does not include non-official programs, or programs with other denominations that include some topics related to HCI.

This work has been done with the propose of building an interuniversity and international master's between Universidad Nacional Abierta y a Distancia, UNAD, (Colombia), and Universidad de Lleida, UdL, (Spain). The review was done by searching information using educational databases, websites that group together graduate programs in these areas, search engines, articles, documents from organizations that work in HCI field (including known topics such as usability or UX), university websites, and interviews with experts in the area.

Contextualized in the field of HCI, the objective of the study is to acquire a deep knowledge about what is taught in the academia around the world and which are the needs of the industry, in terms of which are the professional skills and competences of the professionals to be contracted [1]. Our main interest rises on knowing the concordance between what is currently being taught in the universities at graduate level and the real needs of organizations [2]. This review is one of the initial steps to build the curriculum, the educational processes, the themes, and the required competences in the academy [3].

The paper is structured as follows: the first part shows the need for review; second, the process of finding information and the tools used; third part presents the needs of the organizations; fourth, the comparison between what is taught in academia and the needs of the organizations; and finally, the conclusions and future work are presented.

2 Need for Review

The development of technology generated an important space to study and to research in the Human-Computer Interaction (HCI) discipline. In the design, development, testing, and use of hardware and software products that are related to the needs and successful experiences of users with the use and interaction of technology, and thus generate new aspects, such as forms, methods, media to provide comfortable and exciting user interactive experiences with new technologies applied in different areas, experiences that will also be as much accessible and usable as possible [4].

Moreover, the evolution of the needs of information and communication technologies, users, organizations, contexts, and education generated the need to review and verify if what is taught in the academic programs of universities at different levels of academic degrees corresponds with the needs and abilities of related companies [5]. It is necessary to review the consistency between organizational requirements and academic offerings, as well as to review research trends at universities and whether they are applied in industry [6].

3 Search Process

Based on Kitchenham systematic review methodology [7], a search for information was proposed. The processes of searching the primary data sources were defined, and the review of each of the pages of the master's programs found was developed. With the information obtained, a classification was made according to geographical location and an analysis of the curricula programs to define areas, and topics work is in the programs. In addition, some studies about the state of the art in HCI in different regions were taken into account, such as What is Being Taught on Computing Courses in the UK [8],

HCI Education in Brazil: Challenges and Opportunities [9], Human-Computer Interaction in Ibero-America: Academic, Research, and Professional Issues [10], Analysis of formal and informal degrees in Ibero-America of UX: challenges for online training [11], Human-Computer Interaction in the Curriculum of Higher Education Institutions in Colombia [12] and The State of HCI in Ibero-American countries [13].

The review process follows next steps:

- 1) Research question, 2) Concepts and definitions, 3) Information media consulted,
- 4) Inclusion and exclusion criteria, 5) Search query, 6) Review process and 7) Search results.

3.1 Research Question

Bearing in mind that our main goal is to know the state of HCI's graduate programs up to the year 2020, the following question emerged:

RQ: What are the worldwide graduate programs in Human-Computer Interaction filed?

3.2 Concepts and Definitions

The concepts definitions used in the current search for information are as follows:

- **Master degree program:** a master's program as a purpose to deepen or investigate in an area of knowledge and the development of competences that allow the solution of problems or the analysis of particular situations of a disciplinary, interdisciplinary or professional nature, through the assimilation or appropriation of knowledge, methodologies and scientific, technological or artistic developments or to research a specific area [9, 25]. In this article, the term graduate program will be used to refer to a Master degree program.
- **Human-Computer Interaction (HCI):** is a research area within computer science that aims to understand and improve all the elements that are present in the process of interaction between a human and a computer [16, 17].
- **User eXperience Design (UX):** relates to the process of “creating products that provide meaningful and relevant experiences for users,” including the process of product acquisition and integration, branding, design, usability, and functionality [17].

3.3 Information Media Consulted

Different concepts were defined to carry out the information search process. The initial reviews were carried out in educational databases, lists of graduate programs websites, scientific articles, organization documents, and interviews with experts in the area, among others:

- **Interviews:** different media conducted interviews with experts in HCI and UX.

- **Databases:** information systems from different countries were searched for masters in different countries such as the European Union, Spain, United Kingdom, United States, Mexico, Colombia, Chile, Argentina, Brazil, Australia.
- **Search engines and meta-serachers:** Google was the search engine with more information available for the work.
- **Scientific Papers:** scientific databases that had bodies of knowledge related to our areas of study: ACM, IEEE, Science Direct, and SCOPUS.

3.4 Inclusion and Exclusion Criteria

As it has been mentioned, our focus is on graduate programs, then the inclusion and exclusion criteria of the search were focused on it.

- Inclusion criteria were: 1) Articles published between 2010 and 2020 2) Articles published in conferences, journals, and book chapters, 3) Articles written in Spanish, English, and Portuguese, 4) Articles related to HCI and related areas 5) Searches in educational databases, 6) Listings of graduate programs in HCI and areas, 7) Interviews with experts, 8) Graduate programs official level and 9) the search for postgraduate was conducted until 2020.
- The exclusion criteria were: 1) Articles not available to be downloaded, 2) Articles written in other languages than Spanish, English, and Portuguese, 3) Articles not found in the indicated databases, 4) Information on undergraduate programs, specialization programs, free courses, diplomas, non-official or unrecognized graduate programs, 5) Nonofficial degrees and 6) Inactive graduate programs.

3.5 Search Query

The search query used was the following:

English: ((“Master” OR “Magister”) OR (“Human-Computer Interaction” OR “HCI”) OR (“User Experience” OR “UX”)).

3.6 Review Process

Following the previously mentioned criteria and procedure, the lists of HCI-related graduate programs were found. After that, the interviews (consisting of a set of short questions about the knowledge of master’s programs related to the study) with different experts in HCI, UX design, and related areas enabled us with the data to analyse. It was registered in data arrays with a code for each program, including the name and URL, the country, the number of credits, the emphases, and areas of work. Once cleared and classified, the revision process raised the information that we need to continue.

3.7 Search Results

Table 1 summarizes the programs found worldwide, it has been classified by continent, country, its number of programs, the modality of education and, its duration.

In summary, we found the next 34 graduate programs worldwide (map of Fig. 1):

Table 1. Master's programs in HCI and related programs worldwide with the topics, educational methodology, and average time to complete the programs.

Continent	Country (# programs)	Topics Master's programs	Modality of education		Average time (months)
			Presence	Virtual	
Europe	Spain (6) Netherlands (3) Sweden (2) Portugal (1) France (1) Finland (1) United Kingdom (1) Germany (1) Norway (1)	HCI User-Centred Design UX Interaction	15	2	12
Asia and Oceania	Hong Kong (1) India (1) Singapore (1) Australia (2)	HCI User-Centred Design UX	4	1	16
North America	United States (12)	HCI, UCD, UX	8	4	12
Latin America	0				

- 17 in Europe (Spain, Netherlands, Sweden, Germany, Finland, Norway, France, the United Kingdom, and Portugal).
- 5 in Asia and Oceania (India, Singapore, Australia, and Hong Kong).
- 12 in North America (United States of America).

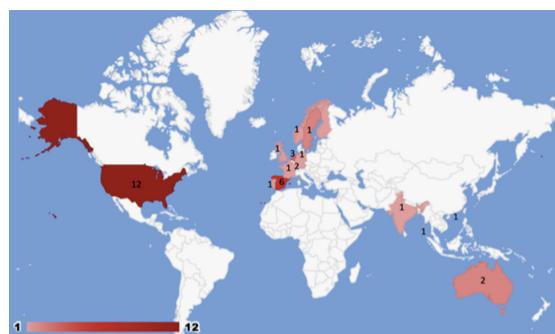


Fig. 1. World map of HCI master's programs located in each country

As no high education programs were found in Latin America, we enlarged the search in other “close” areas (such as, Computer Science or Design), exploring for masters that include HCI related topics within their curriculum or academic program. Then, we found 3 courses in Mexican master programs, 2 courses in Colombia and 3 in Chile.

4 The Need for HCI Knowledge in Companies and Organizations

During the review we found different articles specifically focused on education in HCI, showing us a wide and general teaching overview [17]. When analysing, we observe an important consensus about what should be taught in the academia, that is the same to say which are the skills and knowledge that future graduates need to acquire and demonstrate to work for private companies [18].

According to gathered information, interaction design is an essential issue in the development of the academic processes of HCI related topics, to know: user experience, experience design, accessibility, information architecture, social computing, ubiquitous computing, universal design, usability, prototyping, usability testing, interactivity, mobile technology development, social media, pervasive computing, data mining, machine learning, social network analysis, and product and service development [19].

In the competitive context of the global IT industry, one of the essential premises is the agile and effective application of the knowledge, technical abilities, and creativity in the development of the technological products because “time is money” [20, 21]. In addition, for industry experts, users’ experiences with different technologies are more important than the products and services themselves. Companies started to change the way they create, deliver, and measure the value of their products and services [22], although they still have a long way to go.

For those organizations and companies, providers of useful and friendlies contents, methods, and products, as mediation between humans and electronic devices, now and in the future need human talent trained in the field of HCI -or currently in UX-, to generate satisfactory user experiences that become positive recommendations of products and services, to increase revenue and profits [23].

The analysis led to the emergence of a list of common topics to be addressed. To better understand, we organized in three categories (see Table 2): first, topics found in masters, needs of the companies and, third, the collection of common topics [23–26].

While it is true that graduate education is working with some of the topics that companies currently need, it should also be noted that there is a majority of topics that are not observed in the company’s needs. At the same time there are uncovered topics that are needed in companies and other demanded by companies that are not enough studied/worked in the current programs (to know, from the 34 master’s programs found, very few works on topics such as ubiquitous computing, machine learning, pervasive computing, social computing, and interaction). Most of the programs deeply work on topics related to User Centred Design (UCD), User eXperience (UX), Human-Computer Interaction (HCI), and Interaction Design (ID). Table 3 shows the analysis, in number of programs and the percentage that it is done, of the common topics of the programs [27–29].

Once the analysis has been carried out according to the data obtained and the comparisons made, a more in-depth analysis should be proposed within the master’s programs to

Table 2. Academy-Organization comparison common topics

Topics graduate programs		Organizational needs	Common Topics
User-Centred Design, User Experience, UX Evaluation, Human-Computer Interaction, HCI Evaluation, Usability, Accessibility, Prototype Design, Evaluation, Testing, Interactive Environments, Sensors, Interaction Design, Human-Robot Interaction, Human Media Interaction, User Interfaces, Interface Design, Product and Service Design, Innovation, Development and Creativity, Multimedia, Digital Media, Visual Communication Design	Immersive Environments, Virtual and Augmented Reality, Mobile Applications, Web Design, Mobile Design, Information, Tangible Interaction Design, Human-Human Interaction Design, Video Game Design, Animation, Simulation, Machine Learning, Artificial Intelligence, Ubiquitous Computing, Affective Computing, Real World Computing, Social Computing, Mobile Computing, Thought Design, Context, Collaboration, Information Architecture	Experience design, Accessibility, Information architecture, Social computing, Ubiquitous computing, Universal design, Usability, Prototyping, Usability testing, Interactivity, Mobile technology development, Social media, Pervasive computing, Data mining, Machine learning, Social network analysis, Probability computing, Product and service Development	Experience design, Accessibility, Information Architecture, Social computing, Ubiquitous Computing, Usability, Usability testing, Prototypes, Accessibility, Machine Learning, Product and service design, Social Media

verify and validate whether the curricula are up to date to define some correspondence with current needs. Based on the comparison, we can conclude that a new graduate program should include new topics to be adapted to current needs.

5 Conclusions and Future Work

We have conducted a deep and serious analysis that validated or refuted our suspicions and provided us some important new findings such as:

- 34 HCI related graduate programs were found worldwide, according to the inclusion and exclusion criteria defined for the search. These programs are geographically divided as follows: 17 programs are in European context, 5 in Asia and Oceania, meanwhile North America owns 12 programs.
- From these, only 7 are completely virtual (2 in Europe, 1 in Asia and Oceania and 4 in North America), that is, 20.5% of the total, while remaining 79.5% are face-to-face.

Table 3. Analysis of common topics

Common topics	Number of masters working on the subject	Masters working on the subject (%)
Experience Design	22	11%
Accessibility	1	1%
Information Architecture	4	2,5%
Social Computing	1	1%
Ubiquitous Computing	1	1%
Usability	12	3%
Usability testing	2	1.5%
Prototypes	14	5%
Accessibility	1	1%
Machine Learning	2	1,5%
Product and service design	4	2,5%
Social Media	2	1,5%

- Although some programmes include HCI topics in their curricula (or in some of its courses), there not exist any HCI related graduated program in all Latin America.

These insights confirm our suspicion that there exists a real opportunity/need to contribute to Latin America with HCI-related graduate programs.

Another important finding is that UCD and UX are important topics worldwide meanwhile ID is also important in Asia and Oceania and in North America but in a lower level in Europe [30, 31]. However, it is also necessary to strengthen the programs to include themes related to interaction paradigms since it is a current need by organizations.

In this sense, and due to the importance, that this type of studies is acquiring, we believe that, in a near future, HCI graduate programs in other languages such as Chinese, Japanese, Arabic or Russian will be needed.

The analysis and the experiences of some HCI authors discovered two knowledge problems to design a new curriculum: a) Discovery of the needs, characteristics, and contexts of people, from the experiences mediated by technology; 2) Development of technological solutions focused on the design of user experience and interaction through appropriate paradigms and methods.

Another important insight is that there exists a gap between that existing graduate programs are working and some of the topics that companies currently need. We identified some but more research must be done. In this sense, one important future work is too deep for a better understanding of what company's needs, and, at the same time, to teach them which are the advantages and richness about what HCI can bring to them.

In accordance with the high-level training proposals for graduates, in the field of HCI, we consider it important to incorporate, in new programs, features such as: i) Virtual mode to take advantage of the benefits of information and communication technologies

to provide access to education for professionals anywhere in the world and eliminate mobility restrictions for multiple reasons; ii) Articulate academic networks to leverage capacities of researchers from multiple universities; iii) Spanish language to enable inclusion of the Spanish-speaking population of all Latin America.

Considering the results of this study, both universities, UNAD and UdL, see themselves as having the strength to meet the challenge to design and to offer the first online interuniversity and international master's program in the field of HCI in Latin America. It answers to the academic and organizational needs found. We must include all the findings and considerations discovered to include courses that support the discovery of the user needs, characteristics, and contexts of people from experiences mediated by technology and the development of technological solutions focused on the design of UX and the interaction through appropriate paradigms and methods.

Funding. This work was supported by the FCT – Fundação para a Ciência e a Tecnologia, I.P. [Project UIDB/05105/2020].

References

1. ACM e IEEE CS: Curriculum Guidelines for Baccalaureate Degree Programs in Information Technology – IT 2017. ACM, Nueva York (2017)
2. Baćíková, M.: User experience design: contrasting academic with practice. In: 13th International Conference on Emerging eLearning Technologies and Applications (ICETA), Stary Smokovec, pp. 1–6 (2015)
3. Ernst & Young: The Upside Disruption: Megatrends shaping 2016 and beyond (2017). <https://www.ey.com/gl/en/issues/business-environment/ey-megatrends>
4. BBC: Las 10 habilidades más demandadas por las empresas, según LinkedIn (2019). <https://www.bbc.com/mundo/noticias-46913563>
5. Deloitte: Más allá de la frontera digital (2019). <https://www.iasplus.com/en/publications/colombia/other/deloitte-insights/tech-trends>
6. Foro Económico Mundial: The Future of Jobs 2018 (2018). <https://reports.weforum.org/future-of-jobs-2018/>
7. Kitchenham, B.A.: Systematic review in software engineering. In: Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies - EAST 2012, p. 1 (2012)
8. Kirby, M., Life, A., Istance, H., Hole, L., Crombie, A.: HCI Curricula: what is being taught on computing courses in the UK. In: Nordby, K., Helmersen, P., Gilmore, D.J., Arnesen, S.A. (eds.) Human—Computer Interaction. IFIP Advances in Information and Communication Technology. Springer, Boston (1995)
9. Boscaroli, C., Bim, S.A., Silveira, M.S., Prates, R.O., Barbosa, S.D.J.: HCI education in Brazil: challenges and opportunities. In: International Conference on Human-Computer Interaction, pp. 3–12. Springer, Berlin (2013)
10. Collazos, C.A., Ortega, M., Granollers, A., Rusu, C., Gutierrez, F.L.: Human-computer interaction in Ibero-America: academic, research, and professional issues. *IT Prof.* **18**(2), 8–11 (2016)
11. González, C., Gil, R., Collazos, C., Gonzalez, J.: Análisis de las titulaciones formales e informales en Iberoamérica de UX: desafíos para la formación online (2020). <https://doi.org/10.13140/RG.2.2.35474.27842>

12. Riascos-Pareja, C., Loaiza-Duque, Á., Estrada-Esponda, R.: The human computer interaction in the curricula of Colombian higher educational institutions. *Revista de Investigación Desarrollo e Innovación* **9**(1), 147–162 (2018)
13. Granollers, T., Collazos, C., González, M.: The state of HCI in Ibero-American countries. *J. UCS.* **14**, 2599–2613 (2008)
14. Ramirez, G.M., Collazos, C.A., Moreira, F.: All-Learning: the state of the art of the models and the methodologies educational with ICT. *Telematics Inform.* (2017). <https://doi.org/10.1016/j.tele.2017.10.004>
15. Ministerio de Educación Nacional y El Ministerio de Tecnologías de Información y Comunicaciones. Marco Nacional de Cualificaciones Sector TIC. MINTIC, Bogotá, D.C. (2017)
16. Soegaard, M., Dam, R.F.: The Encyclopedia of Human-Computer Interaction, 2nd edn. The Interaction Design Foundation (2013)
17. IDF. (s.f.) Interaction Design Foundations. The Encyclopedia of Human-Computer Interaction, 2nd edn. <https://www.interaction-design.org/literature>
18. Granollers, I.S.T., Lorés, V.J., Cañas, D.J.J.: Diseño de sistemas interactivos centrados en el usuario. Editorial UOC, Barcelona, ES (2005)
19. Hewett, B., Card, C., Gasen, M., Perlman, S.: ACM SIGCHI Curricula for Human-Computer Interaction (2009). <https://www.worldcat.org/title/acm-sigchi-curricula-for-human-computer-interaction/oclc/25902619>
20. ISO: ISO/IEC 25010 Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—System and software quality models (2011)
21. Lamprecht, E.: The Difference between UX and UI Design: A Layman's Guide (2015). <https://careerfoundry.com/en/blog/ux-design/the-difference-between-ux-and-ui-design-a-laymans-guide/>
22. Magdy, H., Hassan, G.: From usability to user experience. In: Iciibms, pp. 216–222 (2017)
23. Nielsen: The State of The Media: The Social Media Report 2012 (2012). <https://www.nielsen.com/in/en/insights/reports/2012/state-of-the-media-the-social-media-report-2012.html>
24. Nielsen Norman Group: A 100-Year View of User Experience (2017). <https://www.nngroup.com/articles/100-years-ux/>
25. Nielsen, J., Farrell, S.: User Experience Career Advice: How to Learn UX and Get a Job (2014). <https://www.nngroup.com/articles/ux-career-advice/>
26. Norman, D.A.: The design of everyday things. *Hum. Factors Ergon. Manuf.* (2013). <https://doi.org/10.1002/hfm.20127>
27. Pascual, A., Ribera, M., Granollers, T., Coiduras, J.L.: Impact of accessibility barriers on the mood of blind, low-vision, and sighted users. *Procedia Comput. Sci.* **27**(Dsai 2013), 431–440 (2013). <https://doi.org/10.1016/j.procs.2014.02.047>
28. Saffer, D.: The Disciplines of User Experience (2008). <https://www.kickerstudio.com/2008/12/thedisciplines-of-user-experience/>
29. UXPA. (s.f.) User Experience Association. <https://uxpa.org/publications>



On Bridging the Gap Between Far Eastern Cultures and the User Interface

Antoine Bossard^(✉)

Graduate School of Science, Kanagawa University,
2946 Tsuchiya, Hiratsuka, Kanagawa 259-1293, Japan
abossard@kanagawa-u.ac.jp

Abstract. Chinese characters cement Far Eastern cultures: they can be found for example in the Chinese, Japanese, Korean and Vietnamese writing systems. Digital computer systems originate from the Occident, during or shortly after the second world war. Their usage has thus been centred on horizontal, left-to-right layouts, typesetting and I/O, which has created a gap between these systems and, for instance, Far Eastern cultures whose writing systems are traditionally vertical right-to-left (RTL). Although some improvements have been made in this field over time, there are still several major issues. In this paper, we conduct an experiment, which is a proof of concept, to investigate the feasibility and usability of a user interface that abides by the vertical RTL typesetting rules. The results obtained show that while some of these gaps could be filled, there remain several technological challenges.

Keywords: HCI · GUI · UX · Asia · Chinese character · Japanese · Software · Design

1 Introduction

It is common for cultures of the Far East to rely on Chinese characters: not only China, but also Korea with *hanja*, Japan with *kanji*, Vietnam with *chữ nôm* characters – this is the CJKV family of scripts [1]. Traditionally, these are vertical right-to-left (RTL) writing systems, as it can be seen for instance in Japanese novels and newspapers. In addition, the Mongolian script, based on its own alphabet, is also a vertical writing system.

Digital computer systems originate from the occidental world (e.g. the ENIAC and EDSAC machines), thus with horizontal left-to-right writing a *de facto* standard for text input and output, including user interfaces (refer for instance to the ASCII [2] and ANSI [3] standards); the cultures relying on writing systems that do not abide by these principles were for years neglected mostly for computer hardware (e.g. input devices), performance (e.g. RAM memory utilization) and software (e.g. text processing, fonts) reasons [4,5]. This situation has improved, with now input methods for instance for logographic scripts [6]

and software support for right-to-left text, but there is still a wide gap between Far Eastern cultures and systems as explained next.

It is arguably time, and legitimate, for Far Eastern cultures to reclaim their heritage: while technical limitations might have been argued up to now, they do not hold any more. Yet, most people of these countries have got accustomed to using an “occidental” horizontal left-to-right (LTR) layout. So, various efforts are likely to be required in order to make the switch. Nevertheless, sticking to the horizontal LTR layout because it is convenient enough is not satisfactory from a cultural point of view: not using the localisation capabilities of modern computer systems because the conventional approach is just fine is harmful to localisation (multiculturalism) itself.

Although the input and output of computer systems have seen significant improvements with respect to these issues, user interfaces in most software – desktop and mobile – seem to remain stuck to a classic horizontal LTR approach. In this paper, we investigate this matter by conducting an experiment which is at the same time a constructive proof of the usability of a graphical user interface that is based on a vertical RTL writing system such as those of the Far East. The obtained results are discussed especially to document the remaining issues of such a, rather unexpectedly novel, user interface approach.

This research relates to cross-cultural issues as discussed, for instance, in [7], although not restricted to Web matters. Culture-centred human-computer interaction (HCI) design is widely reviewed in [8], with the author notably addressing the sometimes confusing terminology used in the field. It is interesting to note that while in recent years user interface (UI) cultural issues are also considered for Web-based interfaces [9], although having been addressed for a longer time conventional software graphical user interface (GUI) design has never completely adapted to Far Eastern cultures as explained previously. This paper applies to both contexts.

The rest of this paper is organised as follows. The principles and implementation techniques of the experimentation are detailed in Sect. 2. Then, the obtained results are given in Sect. 3 and subsequently discussed in Sect. 4. Finally, this paper is concluded in Sect. 5.

2 Methodology

The principles for the proposed design approach are described before giving some technical details for their implementation.

2.1 Design Principles

We describe here several principles of a vertical RTL user interface with respect to a conventional graphical user interface as provided by a window manager such as GNOME [10] and that of Microsoft Windows (it is handled by the `user32.dll` library). This discussion is also applicable to portable devices such as smartphones and tablets.

The window chrome is adjusted as follows. The menu bar (and, for instance, tool bars) that is typically below the title bar of a window (or at the top of the screen in fullscreen mode applications) is moved to the right of the window, and becomes a vertical GUI component. In the case of a window with a title bar and possibly other non-client area elements on the same bar, the title bar is similarly moved and placed vertically on the rightmost part of the window. Text labels, buttons and other GUI components (a.k.a. widgets, controls), such as those included in the window chrome but not only, become vertical RTL items. A mock-up of a vertical RTL window chrome is given in Fig. 1b; it is to be compared with the corresponding illustration shown in Fig. 1a which implements the conventional approach. It should be noted that this mock-up is given mostly with rotated English labels for the sake of readability, but makes obviously more sense when using a vertical script.

The close button is retained in the top right-hand corner to flatten the learning curve of this novel interface. Contrary to the conventional layout, the window content is to be (primarily) scrolled horizontally. It is interesting to note that on a mainstream, wide display, the proposed vertical RTL layout is more convenient with respect to scrolling: the amount of information displayed is greater than with the conventional horizontal LTR approach. For example, a page of vertical RTL text can more easily fit the display with the proposal than can a page of horizontal LTR text with the conventional horizontal LTR approach, which is somewhat paradoxical, even though computer systems' applications are obviously not to be restricted to such text processing.



Fig. 1. A mock-up of a vertical right-to-left window chrome (a) matching a conventional window layout (b). The usage of a vertical script should be envisioned in place of the rotated English labels.

As another principle, we enforce a shrinking factor on the text so that the interface space (e.g. in menus, title bar, tool bars) is used more efficiently and can be read more rapidly. This method is for instance used in Japanese newspapers [1]. A shrink factor of 0.8 gave satisfactory results. So, whereas Chinese characters conventionally have a width:height ratio of 1:1, they are transformed here to respect a 1:0.8 ratio.

2.2 Technical Details of the Implementation

We have realised a proof of concept that implements a vertical RTL graphical user interface in a Web context: the program is a standard HTML application with extended features such as styling and scripting provided by standard CSS and standard JavaScript. The selection of a Web context is highly relevant considering the software engineering trend that consists in relying on 1) thin (Web) clients [11], such as Google Chromebooks, and 2) Web technologies to build desktop applications, and more generally the popularity of Web technologies for the development of smart devices' applications. Concrete examples include desktop applications with Microsoft's UWP, Electron and the precursor HTML applications (HTA) [12], and iOS applications with Apache Cordova and Adobe PhoneGap for mobile devices, amongst others. Nevertheless, the implemented interface elements are also common for conventional software GUI (they implement, amongst others, the design principles given in Sect. 2.1), this proposal and discussion thus applying to UI in general.

First and foremost, vertical RTL typesetting in Web technologies is still very unevenly supported [13]. As of November 2020, it is not yet part of any current standard, although the CSS property `writing-mode` on which our implementation relies is part of the CSS Writing Modes Level 3 proposed recommendation [14], which is in other words still in a draft state. The value of this property is set to `vertical-rl` for our work.

The character shrinking described previously, especially for menus and title, tool bars, is realised with the CSS property `transform` whose value is set to `scale(1, 0.8)` to enforce the desired character ratio; the CSS property `transform-origin` is besides set to `top`. Once again, it should be noted that this property has not made it to a standard yet, and is still a working draft (CSS Transforms Module Level 1) [15].

The previously described techniques all deal with information output (display). We have relied on the `contenteditable` attribute of, say, a `div` element in order to achieve the input of vertical RTL text. The form elements such as radio buttons and check boxes are less critical in that in a group they can each be positioned separately to achieve the desired layout.

3 Experimental Results

We show in this section the obtained results depending on various scenarios. These experiments were conducted with the Blink browser engine, a (the) major mainstream engine for both desktop and mobile devices. Considering the purpose of this experiment, we have tried to rely as much as possible on standard GUI components, which is hereinafter the meaning of “native”: system-defined in contrast to user-defined.

First, a sample view of the main menu of a window as per our vertical RTL user interface implementation is shown in Fig. 2. The main menu includes conventional menu items written in Japanese. Menu labels (text strings) are shrunk with a 0.8 factor as explained previously, unlike the window content (greyed filler

text). Since a vertical RTL layout, the (main) scrolling bar is the horizontal one: it appears as the content of the window extends beyond the viewport.

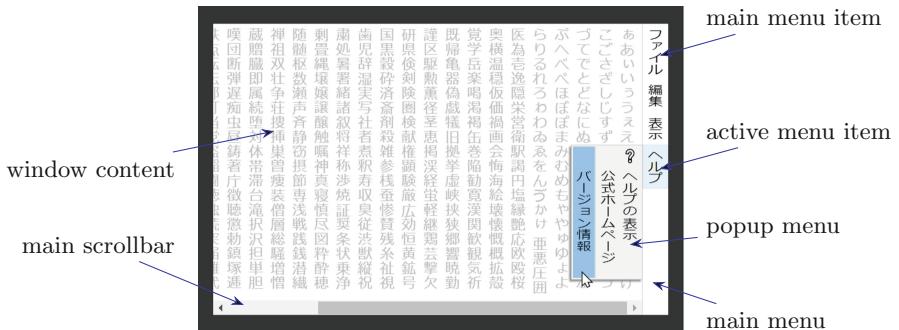


Fig. 2. A sample view of our implementation of a window's main menu in a vertical RTL user interface.

Next, a view of a sample modal dialog box is given in Fig. 3. This dialog box is shown when the user clicks the corresponding menu item of the window's main menu; it is here a replica of a program's "About" dialog box. It can be noticed that the content of the dialog box consists of Japanese text mingled with Latin text. Also, the hyperlink underlining is automatically adjusted for vertical typesetting.

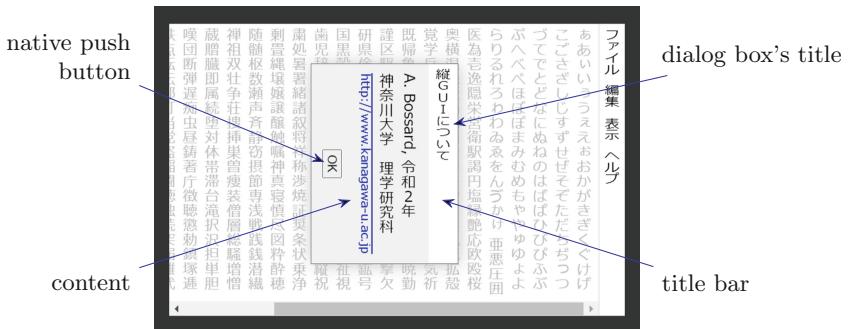


Fig. 3. A view of our implementation of a sample modal dialog box in a vertical RTL user interface.

Third, we illustrate in Fig. 4 a real-world scenario with a questionnaire that could be handed out to the students of a computer architecture lecture. Various form elements are used, with most notably a native text field (an editable div as explained, not an `input`) in order to demonstrate what has been discussed in the previous section.

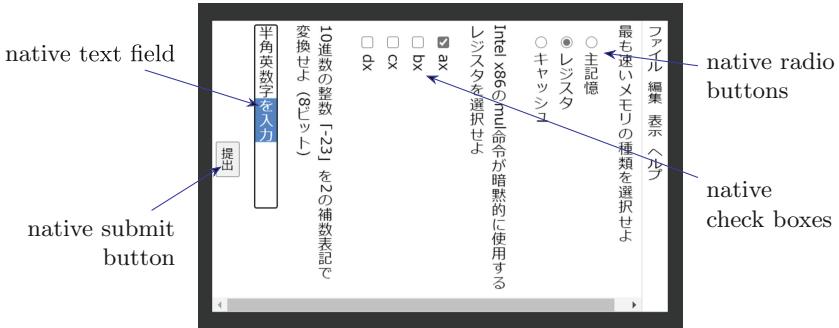


Fig. 4. An illustration of a real-world use case: a possible questionnaire for the students of a computer architecture lecture.



Fig. 5. The case of (simplified) Chinese: a shrink factor of 0.9 (b) seems more suitable than 0.8 (a) for the interface controls (here the main menu).

Finally, whereas the previous sample views are based on a Japanese language interface, we show in Fig. 5 the results obtained in the case of a Chinese language interface (simplified Chinese). A shrink factor of 0.9 (Fig. 5b) seems more suitable than 0.8 (Fig. 5a) in the case of Chinese for the interface controls (here the main menu).

4 Discussion

We distinguish three cases with respect to the implementation of a vertical RTL layout, and especially in the case of Web technologies; in short: what is currently possible, difficult and impossible.

4.1 Possible with a Satisfactory Practicability

First and foremost, relying on a vertical layout with a wide screen has as said advantages (less scrolling), but also disadvantages: for instance, the space avail-

able for the title bar, menu bar, tool bars and similar is reduced. But since as explained we rely on shrunk text for these GUI components, this issue is minor, especially since the applications that use full-width bars are not that frequent (mostly highly-specialised programs, such as the ribbon of Microsoft Excel, which by the way automatically adjusts itself to the space available, so less space is not an issue). For instance, on a wide screen with a 1920×1080 display resolution, we can raise the usable space from $1080/1920 \approx 56\%$ compared to a conventional horizontal layout to $(1080/0.8)/1920 \approx 70\%$ when relying on a 0.8 shrink factor as demonstrated previously. On a tablet-like device such as Microsoft's Surface Pro 7 of display resolution 2736×1824 , the usable space is raised from $1824/2736 \approx 67\%$ compared to a conventional horizontal layout to $(1824/0.8)/2736 \approx 83\%$ when relying on the same shrink factor. There is obviously no problem at all with devices that support rotation of their display, such as most modern mobile devices and some conventional graphical chipsets (e.g. NVIDIA's NVRotate technology).

Besides, the shrinking difference obtained between the Japanese case and the simplified Chinese case (a suitable shrink factor is 0.8 for Japanese but 0.9 for Chinese) originates from font difference: the character bounding box of the Chinese font used in this experiment (Microsoft YaHei) is "tighter" than that of the Japanese font used in this experiment (Meiryo). This is illustrated in Fig. 6: for the same character and at the same font size, the character in the Chinese font is about 7% taller compared to the character in the Japanese font.

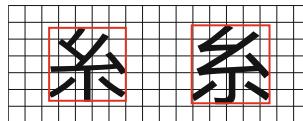


Fig. 6. At the same font size, the height of a same character inside its bounding box may differ between the Japanese (left) and Chinese (right) font, thus the different shrink factor.

The vertical RTL typesetting of basic HTML elements, including text content, poses no problem although this is made possible by Web technologies features that are still not standard (they are at the draft level). In addition, thanks to these modern features, the scrolling of vertical RTL content, thus with the horizontal scrollbar, and the text flow in general is automatically setup without any further adjustments.

Regarding form elements, a push button created with the `button` element (and not with the `input type = button` element) supports as child element one with the CSS property `writing-mode`, so does not induce any particular issue when minding this capability.

4.2 Possible but with a Low Practicability

The scale transformation (with the CSS property `transform` as described earlier) applied to a character string retains the original (i.e. before transformation) size of the HTML element. Hence, in order to effectively and practically rely on shrunk characters, the parent element of the character string is usually resized to fit the actual length of the shrunk characters. This is a non-trivial process, which involves word-breaking issues, that we have realised with JavaScript: the number of characters in the string is multiplied by the scaling ratio and font size to obtain the final (i.e. after transformation) string length, for instance in pixels. This again shows the still lacking support for practical vertical RTL typesetting.

Regarding the radio buttons (`input type = radio`) and check boxes (`input type = checkbox`) form elements, their labels are set with the `label` HTML element which supports vertical RTL typesetting. Yet, the positioning of the label in this case is by default suboptimal and thus requires additional adjustment, for example with a CSS tweak such as `position: relative; right: -4px`. While the range control (`input type = range`) has a non-standard, rarely supported `orient` CSS property to make the control vertical, a similar effect can be achieved with a CSS transformation: `transform: rotate(90deg)`. But again, the positioning requires additional adjustment.

4.3 Not Currently Possible

We have confirmed that the HTML form elements such as the text field (`input type = text`), push button (`input type = button`), multiline text field (`text area`) do not support vertical RTL typesetting. However, as a mitigation and as explained previously, we were able to realise the input of vertical RTL text thanks to the `contenteditable` attribute of an HTML element such as `div`. Nonetheless, it should be noted that this is more a trick than a proper solution, especially considering the accessibility issue [16]. For instance, setting the `contenteditable` attribute shifts, and thus breaks, the page layout if not specifying in addition the CSS property `overflow-y` (or `overflow-x`) to either of `hidden`, `scroll` or `auto`.

Although in general other controls, such as the date control (`input type = date`), can be rotated as described for the range control previously, and with the same positioning limitations, this is without vertical RTL support for their content.

5 Conclusion

For decades, computer systems have tried to adjust their human interfaces to users of various cultural backgrounds, starting with the notorious character encoding issue. Although significant improvements have been witnessed in this field, some sort of complacency has taken place within some cultures: the conventional, foreign approach (e.g. horizontal LTR) being convenient enough, too

little efforts were made to adjust interfaces to cultural principles, such as the vertical RTL writing system of several Far Eastern countries. We have given in this paper a constructive proof of the feasibility of vertical RTL graphical user interfaces. Furthermore, we have shown that several UI principles can be reused to flatten the learning curve for users making the switch between the conventional horizontal LTR layout and the proposed vertical RTL one. The obtained results have been discussed from a technical point of view: what is currently possible, difficult and impossible.

Regarding future works, a meaningful task would be to improve vertical typography: for instance, the handling of Latin words mingled inside vertical text. This is however a complex issue [17]. In addition, considering the case of out-of-browser applications could be another future work: native GUI controls and capabilities seem not as advanced as what a modern Web browser supports with respect to vertical RTL layout and typesetting. Finally, defining and trialling patterns through additional experiments, involving several users, to further investigate the cultural needs, applicability and to estimate the learning curve of the proposal's approach would also be very relevant.

Acknowledgement. The author sincerely thanks the reviewers for their helpful comments and suggestions.

References

1. Lunde, K.: CJKV Information Processing, 2nd edn. O'Reilly Media, Sebastopol (2009)
2. Robinson, G.S., Cargill, C.: History and impact of computer standards. Computer **29**(10), 79–85 (1996). <https://doi.org/10.1109/2.539725>
3. ISO/IEC JTC 1/SC 2 Coded character sets: ISO/IEC 6429:1992 Information technology—Control functions for coded character sets, third edn. (1992)
4. Sakamura, K.: Japanese-language processing as conceived in the TRON project. TRONWARE, vol. 50, pp. 36–44 (1998). In: Japanese; translation. <http://tronweb.super-nova.co.jp/jpnlangpro.html>. Accessed November 2020
5. Bossard, A., Kaneko, K.: Unrestricted character encoding for Japanese. In: Lupekiene, A., Vasilecas, O., Dzemyda, G. (eds.) Databases and Information Systems X - Selected Papers from the Thirteenth International Baltic Conference, Trakai, Lithuania, 1–4 July, Frontiers in Artificial Intelligence and Applications, vol. 315, pp. 161–175. IOS Press (2018). <https://doi.org/10.3233/978-1-61499-941-6-161>
6. Bossard, A.: aIME: a new input method based on Chinese characters algebra. In: Lee, R. (ed.) Computer and Information Science - Papers from the 15th IEEE/ACIS International Conference on Computer and Information Science, Okayama, Japan, 26–29 June, Studies in Computational Intelligence, vol. 656, pp. 167–179. Springer International Publishing (2016). https://doi.org/10.1007/978-3-319-40171-3_12
7. Collazos, C.A., Gil, R.: Using cross-cultural features in web design patterns. In: Proceedings of the Eighth International Conference on Information Technology: New Generations, Las Vegas, NV, USA, 11–13 April, pp. 514–519. IEEE (2011). <https://doi.org/10.1109/ITNG.2011.95>

8. Heimgärtner, R.: Intercultural user interface design – culture-centered HCI design – cross-cultural user interface design: different terminology or different approaches? In: Marcus, A. (ed.) Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience, Las Vegas, NV, USA, 21–26 July, pp. 62–71. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39241-2_8
9. Alexander, R., Murray, D., Thompson, N.: Cross-cultural web design guidelines. In: Proceedings of the 14th Web for All Conference on The Future of Accessible Work, Perth, Western Australia, Australia, April. Association for Computing Machinery (2017). <https://doi.org/10.1145/3058555.3058574>
10. Warkus, M.: The Official GNOME 2 Developer's Guide. No Starch Press, San Francisco (2004)
11. Fanning, K.: Thin client: new cost savings? J. Corporate Acc. Financ. **25**(3), 7–12 (2014). <https://doi.org/10.1002/jcaf.21933>
12. Cooper, P.R., Kohnfelder, L.M., Chavez, R.A.: Method and apparatus for writing a Windows application in HTML. United States Patent no. 6,662,341 (2003)
13. MDN: writing-mode (2020). <https://developer.mozilla.org/en-US/docs/Web/CSS/writing-mode>
14. W3C: CSS Writing Modes Level 3 (2020). <https://drafts.csswg.org/css-writing-modes-3/>
15. W3C: CSS Transforms Module Level 1 (2020). <https://drafts.csswg.org/css-transforms/>
16. ISO/IEC JTC 1 Information technology: ISO/IEC 40500:2012 Information technology—W3C Web Content Accessibility Guidelines (WCAG) 2.0 (2012)
17. Mukai, H.: An introduction to Japanese typography based on structure and algorithm, Seibundo-Shinkosha, Tokyo, Japan (2018). In Japanese



Promotion of Social Participation in Smart City Developments: Six Technologies for Potential Use in Living Labs

Marciele Bernardes¹✉, Francisco Andrade¹, Paulo Novais², Herbert Kimura³, and Jorge Fernandes³

¹ JusGov/University of Minho Braga, Braga, Portugal

² University of Minho Braga, Braga, Portugal

³ University of Brasília, Brasília, Brazil

Abstract. With the advancement of Information and Communication Technologies (ICTs) in providing new tools for interaction between citizens and government agents, public participation has been improved, allowing access to a more significant number of participants in the public policy process, and allowing faster and better monitoring of public servants' initiatives.

These specific ICTs may be seen as generators of information systems, which promote processes and social values among citizens and public agents acting as users. This research is a descriptive one, that presents six cases of technologies for improvement of social participation, and which can be potentially useful for smart city development through institutions termed living labs. The paper describes real cases analyzed under the perspective of information systems, presenting challenges and results aligned to six models of use of ICTs for improvement of social participation, as proposed by [1]. The research collaborates to interpreting the design of smart cities under the perspective of information systems development.

Keywords: Social participation · Smart cities · Living labs

1 Introduction

Participation and public deliberation are essential elements in democratic processes [2, 3], and may contribute to more effective, efficient, and legitimate decision making [1, 4]. Besides that, the lack of popular participation may impose obstacles on implementation of public policies. Subsequent paragraphs, however, are indented.

Considering public participation as a redistribution of power that allows a deliberate inclusion of individuals previously excluded from political and economic processes, [5] and [6] suggested that participation mechanisms may vary in three dimensions: i) who does participate; ii) how participants do communicate and make decisions altogether, and iii) how decisions are connected with public actions or policies.

With the advancement of Information and Communication Technologies (ICTs), public participation in the process of public policies has been enhanced, giving access to a more significant number and diversity of people. It also makes available new ways

of interaction between citizens and governments, and entails a more efficient follow-up of government's results and initiatives.

This paper presents six case studies aligned with six different social participation models, describing real cases, putting in evidence results and challenges. The research seeks to understand smart cities's design under the perspective of information systems, considering the participation of citizens and public servants in the role of users of information systems engaged in "business" processes, which produce added value through the identification of problems and proposal of solutions for urban areas.

The paper is structured as follows: the next section presents a brief consideration of public participation. Despite the fact that there are several typologies that help in the categorization of technologies to promote greater interaction between citizens and government, this paper focuses on discussing the models proposed by [1]. Some cases will be debated, exemplifying the use of technologies for promoting public participation, specific processes and values. Finally, the paper debates on the use of said technologies in allowing public participation in living labs and smart cities, regarding possible results and challenges for implementation.

2 Relevance of Information System Technologies for Public Participation

The impact of information system technologies in public policies and thus in the well-being of society may be influenced by citizens' capacity to participate as users in the processes of decision making. However, regardless of the general perception that the community participation in the government's decision might induce important benefits, several real efforts result in costly and inefficient solutions [7].

According to [8], with the increase in access to the Internet and the advancement of technologies for user location detection, several barriers to citizen participation have diminished, and a massive amount of information may be generated. However, there are other factors that still inhibits greater public participation: (i) logistic, due to difficulties of access to places with an adequate technological infrastructure or short time available for participation; (ii) dynamic of power reflecting different relevance and influence of the various stakeholders; and (iii) communication, arising out of different styles and communication capabilities [9]. Besides that, cultural and territorial aspects may cause civic alienation [9, 10].

Considering the context of smart cities, in which living labs [11, 12] and [13] may contribute to the development of solutions that bring together different stakeholders to find answers to urban challenges, the warranty of public participation becomes essential. Considering the economic, social, and educational disparities in cities, a fair use of technologies and performance of citizens in specific roles as information system users may contribute to a higher degree of inclusion. It also may enhance the population's influence as a whole in the process of political decision making. The relationships and influences between technology and public participation may be analyzed under different dimensions. Under the perspective of the influence of digital technologies in the components of the political process, involving different levels and interests, [1] presents models that identify information flows among agents which participate in the political

decision. From a basic model, [1] propose six models: (i) Muscular Public Sphere, (ii) Here Comes Everybody, (iii) Direct Digital Democracy, (iv) Truth-Based Advocacy, (v) Constituent Mobilization, and (vi) Social Monitoring.

Our hypothesis is that if higher social participation can be addressed by a special set of models of digital technologies, these technologies may indeed foster information systems aimed for use in the “business” of smart cities. In this paper we depart from the definition of information systems proposed by [14, p. 38], that declares an information system as “a cohesive combination of processes, concerning the collection, transformation, storing and retrieval of (output) data which contain news for the user, regardless of the technical means applied”. Furthermore, it must be stressed that the user and also the ICTs (technical means) are integral parts of an information system. Put in simple ways, an information system may be described as a specialized assembly of (i) users, (ii) processes and (iii) ICTs, producing a (iv) new and distinctive values or beneficial outcomes to an organization. The users in an information system are the people that develop specific roles when using the technological service interfaces presented inside the system. The processes in an information system are the internal flows of tasks executed inside the information system that produce specific values for the users and the organization.

The generic value of the information systems to be identified further in this paper is the improvement of social participation in the development of public policies for a city. The processes are those flows and transformations of information that help to create such values. The users in general are citizens and public servants tied to the city in context, developing specific roles. The framework in which the studied technologies are classified are briefly presented in the next section.

2.1 Models of Digital Technology for Public Policy Development

In the work “Six models for the internet + politics”, [1] presents a basic model of the relation between citizens and policies. This basic model and its refinements suggest that citizens originally joined interest groups and social organizations. Such groups, as they grow, recruit and mobilize other citizens, strengthening the defense of their interests.

In the model Muscular Public Sphere, ICTs promotes citizens’ higher participation in the definition of public opinion and a higher degree of communication between citizens, public agents, and politicians. The use of technologies by citizens, in this model, may partly withdraw the direct influence of traditional organizations over public agents [1].

Another model established by [1], Here Comes Everybody, emphasizes a direct involvement of citizens in production of public policies, in detriment of communication and information. In this context, ICTs are used to aggregate individuals aiming at obtaining conquests for the public. Citizens develop through ICTs a direct influence on public actions, with a corresponding reduction of the role of traditional organizations, politicians, public agents or specific policies and norms.

In the model Digital Direct Democracy [1] ICTs are employed to exert a more direct influence of citizens in politics, reducing the role that traditional organizations have on the public sphere as a whole [1].

In the model Truth-Based Advocacy, digital platforms become mechanisms through which organized groups bring facts and concerns that give a direction to public opinion, which, by its turn, may influence public agents [1]. In this context, new media platforms

have a fundamental role in leading traditional organizations to transmit information to form public opinion. This shall cause a stronger pressure on public agents and politicians.

The model Constituent Mobilization [1] focuses on using ICTs to strengthen the interaction between citizens and traditional organizations, bringing greater attention to other citizens and the general public. This model is based on crowd-sourcing with the production and solving of distributed problems.

The research of [1] brings the argument that, due to political incentives and institutional restrictions, the models (i) Muscular Public Sphere, (ii) Here Comes Everybody, and (iii) Direct Digital Democracy, although more revolutionary and transformative, are less likely to be implemented than the models (iv) Truth-Based Advocacy, (v) Constituent Mobilization, and (vi) Social Monitoring, which have an incremental contribution to the use of technology in policies.

3 Case Studies

This section presents six cases of information system technologies, mostly in the brazilian context, for participative governance applied to citizens, which implement most aspects of the theoretical constructions developed by [1]. This section is split into six parts, corresponding to examples of implementation of the six models [1]: (i) Muscular Public Sphere (MPS), (ii) Here Comes Everybody (HCE), (iii) Direct Digital Democracy (DDD), (iv) Truth-Based Advocacy (TBA), (v) Constituent Mobilization (CM), and (vi) Social Monitoring (SM).

Muscular Public Sphere (MPS). To exemplify the study of a technology aligned to the MPS model [1], we refer to Operação Serenata do Amor – OSA [15] as a way of public participation and contribution to social control. OSA's main goal was to present how Open Government Data - OGD can be employed to oversee reimbursements made by a Quota to Exercise Parliamentary Activity - QEPA. Departing from the supposition that the mere availability of Open Government Data in XML format does not necessarily comply with the principles of Open Government, OSA searched for alternatives for citizens to use these data to oversee the ways members of the brazilian parliament use QEPA. To test the hypothesis, OSA built two internet-based robots, Rosie and Jarbas. Rosie was built to detect eventual frauds, and Jarbas to facilitate the visualization of data generated by Rosie. Among the results of OSA, it was verified that more than 80% of denunciations of suspicious expenses did not get any attention from politicians, who just ignored the messages sent by these robots; thus, to present the results of eventual suspicious expenses, Rosie got an interface on Twitter [15]. While studying the referred project, [16] verified that the use of big data in fighting against corruption was quite efficient with the project, and found a considerable volume of irregularities in public expenses. Concerning the publicity of results, the authors explain that the suspicious expenses' presentation is done through the Twitter platform. Departing from the hypothesis that OSA is a ICT assembly that fostered a new information system, we may state that: (i) OSA's users are, in more specific terms, citizens overseeing public spending, and politicians using quotas"; (ii) OSA's main processes are detection of fraudulent behavior using machine learning; data visualization; and robot-based tweeting. Finally, the (iii) values added by the system are: social engagement; social control; public spending

control; transparency; and accountability. These information system elements of OSA are declared in a corresponding line in Table 1.

Here Comes Everybody (HCE). To exemplify the model HCE [1] we studied Fab Lab Livre SP, from the city of São Paulo, Brazil. Inspired on the philosophy of "Cultura Maker – Faça Você Mesmo" (maker culture – do it yourself), Fab Labs are collaborative and creative spaces, totally free of cost, that give to citizens access to high technology maker spaces in support to development and implementation of ideas and projects [17]. The methodology adopted by Fab Lab follows a teaching process stimulating the sharing of information and collective construction of ideas. It is based on four dimensions: (i) knowledge, science, technology; (ii) participation, citizenship, and democracy; (iii) education; (iv) social relevance. With the use of social technology, Fab Lab allows the democratization of access to new digital technologies, making available to the population a set of last generation ICT tools and the possibility of living in an innovative and collaborative ambiance. The specific characterization of information system's users, processes and added values generated by Fab Lab Livre SP are declared in Table 1.

Digital Direct Democracy (DDD). One of the easiest ways for public participation in city issues involves public consultations. This was the case of Digital Participative Budget in the city of Belo Horizonte, Brazil (OP Digital BH), which is mentioned in the study of [1], as an example of Digital Direct Democracy.

In Belo Horizonte there were two ways for citizens to engage in the Participative Budget: the In Person Participative Budget (PPB), implemented since 1993, and the Digital Participative Budget (DPB), implemented in 2006. DPB was created so that citizens could use the Internet (<https://opdigital.pbh.gov.br>) to choose which construction projects are to be built. To do so, it is sufficient for the citizen an access to the Internet and to be registered as an elector in the city [18, 19]. Its pioneer introduction and the high population's adherence made the case "Digital Public Budget in Belo Horizonte" to be recognized as an innovative experience in the field of participative democracy by the International Observatory of Participatory Democracy – IOPD [20]. The specific characterization of information system's users, processes and added values generated by OP Digital BH are declared in Table 1.

Truth-Based Advocacy (TBA). In the fourth model of technologies for promotion of public participation for public policy development, "Truth-Based Advocacy" [1], digital platforms become mechanisms through which citizens may bring facts and concerns directing the public opinion, which, by its turn, may influence public agents. For this case, it is worth the experience implemented in Lisbon, Portugal, termed LisBOAIDEIA [21]. In this social participation technological model, citizens may, by accessing an online portal, submit an idea on a general subject of interest (examples of published ideas: parking in Lisbon, urban art in electric power poles). These ideas stay available for online voting for sixty days and, if they get more than 100 votes, They will be analyzed by the Municipality, which shall have the last word [17]. The specific characterization of information system's users, processes and added values generated by LisBOAIDEIA are declared in Table 1.

Constituent Mobilization (CM). Aligned to the fifth model, "Constituent Mobilization" of [1], it is worth referring to the experience of participative budget developed

on the platform “Decide Madrid”, from Spain. This one was awarded the prize of public service 2018 for the United Nations - UN, in the category “Turning institutions inclusive and ensuring participation in decision making”. The decision was precisely based on transcending the mere logic of participative budgets functioning, in which the administration selects a set of projects, then citizens just vote. Based on the use of Consul free software, this experience advances in terms of digital participation, paving the way for democracy in a bottom-up mode through direct and binding mechanisms [22]. The specific characterization of information system’s users, processes and added values generated by “Decide Madrid” are declared in Table 1.

Social Monitoring (SM). Finally, the last model proposed by [1] is “Social Monitoring,” in which public agencies and civil organizations use digital techniques of a survey on the information of individuals to identify problems, bringing attention to other citizens and the public. For this model, it must be referred to the strategy of popular participation, comprising cases in which citizens may digitally interact in monitoring and evaluating the impacts of a public policy. This is an experience of the app Promise Tracker [23], developed by MIT’s Media Lab, an open code tool designed to help communities, individuals, and civil society organizations to monitor the commitment of public authorities and, this way, to require greater civic responsibility of managers and politicians. An example of real use of this technology aligned with the “Social Monitoring” model is the project Belém’s Scholar Snacks (“merenda escolar”), were civil society organizations, agencies of government control, a university and a scholar network adopted the app Promise Tracker for monitoring the quality of the Scholar Snack offered by the public schools of the city of Belém, Brazil. The vast diversity of actors in the implementation of this initiative became one of the key elements for the monitoring, whose lessons learnt were observed in different categories: for the snacks (improvement in preparing and storing and in the quality of the ingredients); for the participants (understanding of the lunch as a right); the commitment of the students (citizen monitoring as a tool for learning); apprenticeship on technology and mobilization (technology helping to scale, streamline and make visible the use of multiple technological platforms and relevance of a campaign organization); apprenticeship on network organization (roles and division of work, the power of inter-sectoral partnerships for monitoring, the value of collaboration with governmental control agencies, the development and consolidation of partnerships) [23, 24]. The specific characterization of information system’s users, processes and added values generated by “Belem’s Scholar Snack” are declared in Table 1.

Next, Table 1 is presented with a synthesis of all the information system elements related to each case studied. For each ICT solution, are presented its users), processes and added values.

Analysis of data from Table 1 shows that each case studied presents distinctive processes characteristic of the models to which they adhere: The “musculature” of the public sphere was improved by machine learning processes working on open data; FabLab’s harmony occurs through the sharing of ideas, information, tools and physical resources, projects and so on; The Digital Participative Budget in Belo Horizonte city was made possible by easy online access, use of strong user authentication mechanisms, and online voting processes. Advocacy in finding the “truth” in LisBOAIDEIA occurred because of the system’s capacity to organize separate proposals and debates in thematic

Table 1. Information system's elements present in case investigated.

Case@Model	Users (roles developed by people)	Processes (internal flows of tasks)	Added value (beneficial outcome produced)
Operação Serenata de Amor-OSA @MPS	Citizens overseeing public spending; Politicians using quotas	Detection of suspicious behavior using machine learning; data visualization; robot-based tweeting	Social engagement; social control; public spending control; transparency; accountability
Fab Lab Livre SP @HCE	Citizens engaging in the maker culture	Capacity building; sharing of information and high tech artifacts; collective development of ideas and projects	Democratization of knowledge, science and technology; collaborative participation; promotion of maker culture; social engagement
Digital Participative Budget (DPB) @DDD	Citizens prioritizing public spending in his territory	User identification and authentication; online and telephone voting; online debate; georeferencing	Social engagement; social control; public spending control; transparency; accessibility
LisBOAIDEIA @TBA	Citizens and activists engaged in debates	Authentication via social media; web pages and forms; online voting; organization of proposals by thematic area	Social inclusion; deliberative quality; multichannel participation; protection of personal data; transparency
Decide Madrid @CM	Citizens selecting budget application proposals of popular origin	Authentication via social media, online discussion forum, georeferencing	Participatory deliberation, proposal and refinement of projects and laws
Belem's Scholar Snacks @SM	Citizens and civil society organizations, and control bodies that oversee public policies	Crowdsourcing; georeferencing; dashboard of indicators	Understanding and monitoring the implementation of public policies, data collection to facilitate engagement and civic action

Source: the authors

areas and multichannel participation. Mobilization in Madrid depended on the use of social media-based authentication. Finally, social monitoring relied on crowdsourcing processes, production of indicators and dashboards.

In view of this, the question that should be investigated in a deeper study of this work is whether the selected items of each information system could be employed as building blocks of information systems, as in a “Lego” approach. These “Lego” blocks could be made available within living labs for promotion of collaborative emergent designs of information systems to be built by citizens. Thus, it might be possible to use these guidelines to identify which processes and values that are of most interest to citizens who experience these living labs, in search of the improvement of social participation practices in smart cities.

4 Final Remarks

The study aimed to conduct a survey of technologies for social participation, presenting a description of six real cases, evidencing challenges and results. For this, the “Six Models for the Internet + Politics” proposed by [1] was used as a framework for identifying models of social participation technologies. The paper provided a description of the six models and case studies related to each model.

The case studies are just one small and documented sample, considering the adoption of social participation technologies and political decision-making concerning their positive aspects and the challenges that managers may find while adopting such technologies.

While in a position of neither advocating the use of the technologies nor showing a total skepticism on that, we tried to unveil how technologies are seen as tools/means to promote public participation through the deployment of information systems, but not as a remedy to solve all the democracy deficits.

This approach is in line with [23] and [24] when clarifying that a technique is produced within a specific culture, and the society is conditioned by the methods it possesses.

Although not exhaustive, we believe that this initial analysis allows evidence of strategies of validation for social participation products and services as criteria for improving urban efficiency. Future research must not forget that other citizen participation developments, such as the public sector’s opening of data allowing new services and solutions, must be created by private initiative, third sector, or civil society.

Besides that, and specifically, concerning the fight against pandemics coronavirus disease (Covid-19), transparent management of data is more than necessary because it helped managers follow up on the virus situation in different urban contexts and make effective control decisions about the disease.

From this perspective, technologies will contribute to increasing the dialogue and allowing a better sense of commitment and efficiency. It also will allow greater collaboration and responsibility on the part of the involved actors, in the exact measure of an existing openness and transparency of data, as a crunch strategy for passing from a traditional government to an open government, in which citizens trust and know.

Finally, it is believed that this work can contribute to future studies focusing on information systems and social participation, to identify which processes and values are more efficient for citizens who experience living laboratories for innovation in urban settings.

Acknowledgments. Our thanks to the CNPQ National Council for Scientific and Technological Development of Brazil (by CNPQ project n. 400278/2020-0 e 350341/2020-6), and also to the JUSGOV- Research Centre for Justice and Governance (supported by National Funds through the FCT -Portuguese Foundation for Science and Technology; Project UIDB/05749/2020), and to the Algoritimi Centre; both at University of Minho.

References

1. Fung, A., Russin Gilman, H., Shkabatur, J.: Six models for the internet + politics. *Int. Stud. Rev.* **15**(1), 30–47 (2013)
2. Ingram, H., Rathgeb-Smith, S.: *Public Policy for Democracy*. Brookings Institution Press, Washington, DC (1993)
3. Pahl-Wostl, C., et al.: From applying panaceas to mastering complexity: Toward adaptive water governance in river basins. *Environ. Sci. Policy* **23**, 24–34 (2012)
4. Mukhtarov, F., Dieperink, C., Driessen, P.: The influence of information and communication technologies on public participation in urban water governance: a review of place-based research. *Environ. Sci. Policy* **89**(April), 430–438 (2018)
5. Arnstein, S.R.: A ladder of citizen participation. *J. Am. Plan. Assoc.* **35**(4), 216–224 (1969)
6. Fung, A.: Varieties of participation in complex governance. *Public Adm. Rev.* **66**(SUPPL. 1), 66–75 (2006)
7. Irvin, R.A., Stansbury, J.: Citizen participation in decision making: is it worth the effort? *Public Adm. Rev.* **64**(1), 55–65 (2004)
8. Yu, J., et al.: Towards a service-dominant platform for public value co-creation in a smart city: evidence from two metropolitan cities in China. *Technol. Forecast. Soc. Change* **142**(September 2018), 168–182 (2019)
9. Pflughoeft, B.R., Schneider, I.E.: Social media as E-participation: can a multiple hierarchy stratification perspective predict public interest? *Gov. Inf. Q.* **37**(1), 101422 (2020)
10. Liu, W., et al.: Connecting with hyperlocal news website: cause or effect of civic participation? *Am. Behav. Sci.* **62**(8), 1022–1041 (2018)
11. Eriksson, M., Niitamo, V.P., Kulkki, S.: State-of-the-art in utilizing Living Labs approach to user-centric ICT innovation-a European approach. Lulea: Center for Distance-spanning Technology. Lulea University of Technology Sweden: Lulea. Center for Distance-spanning Technology, vol. 1, no. 13, p. 13 (2005)
12. Westerlund, M., Leminen, S.: Managing the challenges of becoming an open innovation company: experiences from living labs. *Technol. Innov. Manag. Rev.* **1**(1), 19–25 (2011)
13. Schuurman, D.: Bridging the gap between open and user innovation? Exploring the value of Living Labs as a means to structure user contribution and manage distributed innovation. Dissertation in order to obtain the title of Doctor in the Communication Sciences (2015). <https://biblio.ugent.be/publication/5931264/file/5931265.pdf>. Accessed mai 2020
14. Prakken, B.: *Information, Organization and Information Systems Design: An Integrated Approach to Information Problems*. Springer (2000)
15. Lima, C.: Dados Abertos Governamentais no Contexto Da Ciéncia Cidadã: O Caso Da “Operação Serenata De Amor”. *e-prints in library & information science*, p. 15 (2019)

16. Rodrigues, J., Fontes, C.: “Operação Serenata de Amor”: a análise de Big Data no combate à festa dos gastos públicos. In: XIV Congreso de la Asociación Latinoamericana de Investigadores de la Comunicación (ALAIC), San Jose, Costa Rica, July 2018
17. São Paulo. Fab lab livre. O que é. Informação. <https://fablablivresp.art.br/o-que-e>. Accessed July 2020
18. Bernardes, M.B.: Democracia na Sociedade Informacional. Saraiva, São Paulo (2013)
19. Mathias, I.R.M.: Orçamento participativo: uma nova abordagem para Belo Horizonte – MG. Universidade Federal Fluminense, Brazil (2017)
20. International Observatory on Participatory Democracy. Best practices in citizen participation, Barcelona (2007). Access: <doc298.pdf (oidp.net)>. Accessed Dec 2020
21. Bernardes, M.B.: Proposta de Modelagem Regulatória para a Governança Participativa, no Contexto Lusobrasileiro. Universidade do Minho, Portugal (2019)
22. Decide Madrid. Presupuestos Participativos 2019 (2020). <https://decide.madrid.es/presupuestos>. Accessed July 2020
23. Promise Tracker. Monitorando a Cidade. MIT Center for Civic Media. <https://play.google.com/store/apps/details?id=com.ionicframework.monitorandoacidade&hl=pt&showAllReviews=true>. Accessed July 2020
24. Martano, A., Reiser, E., Craveiro, G., et al.: Fomentando o Monitoramento Cidadão no Pará Aprendizados de Três Estudos de Caso. Relatório, Maio 2017. <https://drive.google.com/file/d/0B65ni2sQNdUFbEpWc3RpU19wMzQ/view>. Accessed set. 2020



Electroencephalography as an Alternative for Evaluating User eXperience in Interactive Systems

Sandra Cano¹ , Jonathan Soto² , Laura Acosta² , Victor Peñeyor² , and Fernando Moreira³

¹ Escuela de Ingeniería Informática, Pontifícia Universidad Católica de Valparaíso, Valparaíso, Chile

Sandra.cano@pucv.cl

² Facultad de Ingeniería, Universidad de San Buenaventura de Cali, Cali, Colombia
vmpeneno@usbcali.edu.co

³ REMIT, IJP, Universidade Portucalense, Porto & IEETA, Universidade de Aveiro, Aveiro, Portugal
fmoreira@upt.pt

Abstract. Electroencephalography is proposed as an alternative for evaluating user experience with two interactive systems. Traditionally, evaluation methods are applied either during the interaction, which can disturb the user, or at the end where the user does not usually remember all of their interactions. Therefore, using a BCI (Brain Computer Interface) device as OPEN-BCI as an alternative to evaluate the user experience of a subject while interacts with two interactive systems. In this evaluation were analyzed the sub-bands: alpha, theta and beta. The results show differences in workload and emotions. In addition, consistent analysis of the EEG data, were applied questionnaires, as: SUS, NASA-TLX y SAM, where data showed high consistency.

Keywords: EEG signals · User eXperience · Brain-computer interfaces · Ubiquitous computing

1 Introduction

The focus of technology today lies in integrating itself within different types of users and contexts, as new forms of interaction are created. The research line of HCI (Human-Computer Interaction) thus enjoys an important role in recognizing user behavior, detecting it, processing it and responding according to this behavior and its interest has centered on how to evaluate these types of interaction. Traditionally, evaluation methods are applied either during the interaction, which can disturb the user, or at the end where the user does not usually remember all of their interactions, especially unsatisfactory when the aim is to evaluate emotions during the interaction [1]. Researchers have therefore been interested in evaluating user interactions through implicit methods, one of these implicit methods is electroencephalography, a method by which to monitor a person's

brain activity [2] by means of electrodes that adhere to the scalp, although different types of electrodes allow it to be non-invasive. Therefore, researchers have been interested in analyzing patterns related to emotions, through brain activity [3]. These types of EEG signals are measured through electrical impulses produced by the brain cells of a person and, through electrodes that are placed on the subject's scalp, it is possible to capture these signals [4, 5]. Studies have used EEG signals for different purposes, such as analyzing the effect of frequency bands [6–9], recognizing emotions supported by machine learning techniques [10, 11] and rehabilitation [12]. Elsewhere, user experience is defined by ISO 9241-210 as a person's perceptions and responses that result from the use or anticipated use of a product, system or service [13, 14]. Subjective experience has received much attention in evaluating the experience when a user interacts with technology. However, from the field of neurosciences, people's emotions and cognition are affected by an event [15]. Meanwhile, the development of BCI devices has led researchers to take more interest in EEG signals. The research proposal therefore arises: How can electroencephalography be an alternative to evaluate the user experience of interactive systems?

2 Background

Brain signals involve cognitive and emotional states that, in order to capture a subject's brain activity, can be captured by electrodes placed in certain positions on the head. These positions follow an international standard 10–20 system [17], which is an internationally recognized method for describing and applying electrode placement. A well-known method to capture a person's brain activity is the electroencephalogram (EEG), which allows a neurophysiological exploration of a person. BCI's equipment makes use of the EEG method. EEG measure voltage fluctuations result from ionic current with the neurons of the brain [18]. The oscillations of the EEG signals are thus represented in different frequency bands or ranges: delta (δ , <4 Hz), theta (θ , 4–8 Hz), alpha (α , 8–13 Hz), beta (β , 13–35 Hz), and gamma (γ , >35 Hz) [7]. EEG signals, each band corresponds to specific characteristics. The theta band (4–8 Hz) is related to emotional information [19]; the alpha band (8–13 Hz) is related to cognitive processing [20]; the beta band (13–30 Hz) to logical thinking and to stimulation effects [21]; while the gamma band (over 30 Hz) is related to memory, linguistic processing, cognition and attention [22]. Electrodes are labelled with a letter and a number. The letter refers to the area of brain underlying the electrode, such as: F - Frontal lobe, T - Temporal lobe, C - Central lobe, P-Parietal lobe, O - Occipital lobe. The numbers denote the right/left side of the head. The electrodes are then placed at points that are 10% and 20% of these distances. Therefore, the selection of the channels and band types are related to the purpose of the study, such as cognitive load and emotions. Some works as Calcagno et al. [23] have used 29 electrodes to analyze the programmer's experience. They analyzed the delta, theta, alpha and beta bands for each channel, where they found that delta and theta increased in the frontal and parietal-occipital regions. Meanwhile, Shivsevak et al. [24] analyze cognitive load and affective, where they have found variations in the alpha and theta bands, where there is a decrease in the spectral potential of alpha, while theta there is an increase. In addition, channel selection to analyze cognitive load can change for each

subject. Tian et al. [25] found that the selected channels vary from subject to subject. Also, the feature and lead selection is highly dependent on the task being performed as they directly relate with the activated lobes of the brain [26]. A study conducted by Peng et al. [27] found an increase in stress at Fp1 and Fp2. Another study were used F3 and F4 position to measure the motor imagery tasks involve imagination of movement [28]. The O1 and O2 positions are located above the primary visual cortex [29], and C4 sensory and motor functions. Therefore, to measure cognitive load have been reported as sensitive to task difficulty manipulations alpha and theta bands [30, 31].

3 Materials and Methods

An experiment was designed that consisted of capturing the brain activity of a group of people as they interacted with a classic game called Snake [32]. Interaction in this game was performed in two different ways. The first interaction, System 1, involved a traditional interaction with computer and keyboard. The second, System 2, involved the use of NFC cards and a mobile device such as a smartphone. NFC cards were designed for interaction with system 2. Each card featured a representation of the movement symbol (left, right, up and down) also associated with a color. The movements associated with each color were: Up (brown), Down (green), Right (purple), Left (blue).

3.1 Participants

Brain activity data was captured from 6 subjects (3 women and 3 men) between the ages of 18 to 25. Each subject provided their signed, informed consent to participate in the study, at which point the conditions and protocol of the experiment regarding publication of the data were explained. In turn, the procedures followed met the human experimentation ethical standards according to the Helsinki declaration.

3.2 Instruments

The OPENBCI device was used to measure the electrical activity of the brain using electrodes placed on the skin of the scalp. The type of electrode used, therefore, was Gold Cup, applying conductive gel to the scalp of the subject. When too much gel is applied, however, it excessively presses on the scalp. In the skin preparation, the subject was required to wash his or her hair with coconut soap, with the aim of establishing a better contact area between skin and electrode and to reduce the impedance (between 10 kOhm to 5 kOhm).

The questionnaires applied for each subject were as follows: SAM (Self-Assessment Manikin) [33], NASA-TLX [34] and SUS (System Usability Scale) [35]. SAM is a picture-oriented questionnaire developed to measure an emotional response. The questionnaire measures three characteristics of an emotional response according to the theory of emotion proposed by Lang et al. [36], in which he defines an emotion represented by three dimensions. These are: Affective valence, representing the image with a smile and indicating a maximum score of 9 to a sad face representing a minimum score of 1; Arousal, represented in the figure with eyes open (maximum score 9) or another very

relaxed figure (score 1). Lastly, Dominance would vary from a very small figure to a very large one. The NASA TLX (Task Load Index) questionnaire is a subjective, multi-dimensional assessment tool used to evaluate the perceived workload and effectiveness of a task. The SUS (System Usability Scale) questionnaire relates to measuring user experience, since it measures user satisfaction on interacting with a product. This questionnaire is composed of 10 questions in one, with five response options, ranging from strongly agree (5) to strongly disagree (1). When using SUS, participants are asked to rate the 10 questions on a Likert scale of 1 to 5. The scores that each participant assigns to each answer are added together and then multiplied by 2.5 to convert the scores on a scale from 0 to 100. As such, a SUS score above 68 is considered above average, and below 68 below average.

3.3 Experimental Procedures

For data capture, an OPEN-BCI device was used using the Cyton card (+reference, +ground) and 8 channels (Fp1, Fp2, F3, F4, C4, P4, O1 and O2) using as reference and ground those located in positions A2 and A1. A sampling frequency of 250 Hz was used to capture the EEG signals, that is to say that every 4 ms approximately 256 samples were captured, with a total of 5,160,960 data being collected through the BCI device. Electro-gel sensors were used, since these facilitate longer recordings and improve patient care.

The process consisted of two phases. In the first phase, subject was asked to close their eyes for 90 s to neutralize their emotions. They were then required to interact for 120 s with the first interactive system. Once the interaction was over, the SAM, NASA-TLX and SUS questionnaires were applied. Interaction with the second interactive system featured the same procedure, with subjects first being required to close their eyes for 90 s, followed by an interaction for 120 s, before finally responding to the SAM, NASA-TLX and SUS questionnaires.

Since the captured EEG data is large (8 channels) and contains redundant and noisy information, the raw data was pre-processed, eliminating artifacts and reducing the sampling of the raw data to reduce computational overhead in feature extraction. A second order Butterworth filter was first applied, to filter from 0 to 64 Hz. From the information thereby obtained, Wavelet transform [37] was applied. The EEG signals were thus sampled at 250 Hz. Six levels of decomposition were selected using Daubechies Wavelet, in particular Wavelet db4. This selection was based on previous work related to smoothing, to analyze EEG signals [38, 39]. The approximation coefficients at level 6 (A6: 0–4 Hz) and the detailed coefficients at level 6 (D6: 4–8 Hz), level 5 (D5: 8–16 Hz) and level 4 (D4: 16–32 Hz) produce the delta, theta, alpha and beta sub-bands of the EEG.

3.4 Results

Figure 1 shows the relative energies for the theta, alpha and beta bands for systems 1 & 2. In the theta band the relative energies for each channel are greater for system 1 than system 2, while for the alpha and beta bands, the energies for system 2 are greater than system 1. When comparing the alpha and beta frequencies for each channel, it is

observed that the greatest changes occur for channels C4, F3, Fp1, Fp2, O1, O2, P4. However, for the Fp2, O1, O2 and P4 channels it behaves very similar in both channels (beta and alpha), while for the theta band the Fp2 and O1 channels behave the same, and F3 and Fp1 likewise. It is also observed that in the beta band a higher value is observed for the Fp1 channel. The beta band is associated with active thinking, active attention, focusing on the outside world or solving concrete problems [40].

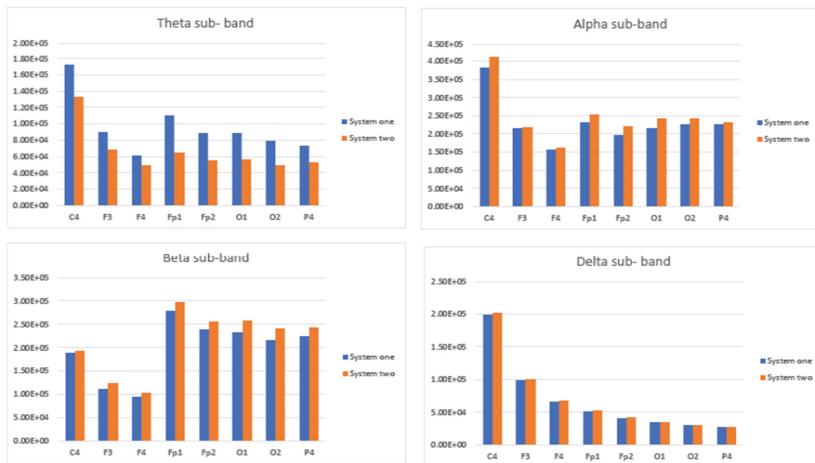


Fig. 1. Energy for each sub-band of beta, alpha, theta and delta frequencies for the both interactive systems.

Therefore, cognitive tasks related to decision-making are related to the beta band, as a maximum value is observed for the Fp1 and Fp2 channels compared to the other bands. The beta band is associated with the state of waking and active thinking, attention and problem solving in an adult. It is found in the frontal region [41]. It is also observed that theta, alpha and delta have a higher value of C4, while for alpha and theta F3 and F4 are higher. Emotions meanwhile are more related to the prefrontal region, such as Fp1 and Fp2 and theta waves are related to states of relaxation.

In Fig. 2, the alpha, beta and theta sub-bands of each system by gender are observed. The C4 channel corresponding to the central region is higher for women in the alpha sub-band for both systems (one and two), while the C4 channel for the beta sub-band is much smaller compared to the other sub-bands. A higher value is also observed for the Fp1 and Fp2 channel for system one in the male gender for the theta and alpha sub-bands. However, the beta sub-band shows a greater activation in women for system two.

The results of the SAM tool are: system 1 for men (Valence: 7; Arousal: 3; Dominance: 3.66), system 1 for women (Valence: 8.33; Arousal: 6.33; Dominance: 6), system 2 for men (Valence: 7.33; Arousal: 6.33; Dominance: 6), system 2 for women (Valence: 5.33; Arousal: 7.66; Dominance: 7.66). It is observed that men and women presented greater activation (arousal) for system 2 compared to system 1. It is also observed that the activation values between men and women are different, as is the valence value. In women they tend to have higher values compared to men. The arousal is related to the activation of emotion, and valence is the motivational component of emotion (pleasure vs. displeasure). Dominance, meanwhile, is related to the degree of control the person perceives over their emotional response, in which men obtained higher scores. These differences found between men and women in values have already been found in the work carried out by [7]. Table 1 and 2 show the results of the NASA-TLX tool, where it is observed that the highest results are obtained in women compared to men. In turn, it is observed that the highest values are obtained with system two. However, effort for men in system one and two does not make a difference, while for women it does. In the same way, frustration is presented at a higher level for system two. It is also related to the values obtained in the theta and alpha sub-bands, the channels Fp1, Fp2, O1 and O2, decreased in system two, while in system one they increased.

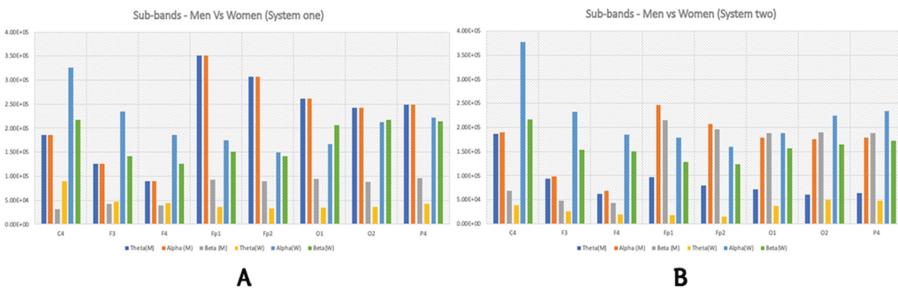


Fig. 2. Energy for each sub-band of beta, alpha, theta and delta frequencies for the two interactive systems separated by gender (A: female, B: male).

The SUS questionnaire is focused on evaluating the usability of a system: System 1 (Men: 64.16, Women: 85), System 2 (Men: 58.33, Women: 30), that system one had a higher value for women. This indicates that system one felt more familiar to interact with compared to system 2. Likewise, this is evidenced in the results obtained from the beta sub-band for system two, where the Fp1 and Fp2 channels had a greater activation in women, as did the alpha sub-band with the C4 channel. System one also had a higher rating for men, but it was not as differentiating compared to the female gender.

Table 1. Results of the NASA-TLX for system 1 (a) and system 2 (b) for the male gender.

General score	30.42	General score	64.58
Mental	53.33	Mental	73.33
Physical	13.33	Physical	61.67
Temporal	28.33	Temporal	75.83
Performance	23.33	Performance	78.33
Effort	55.00	Effort	55.00
Frustration	28.33	Frustration	55.00
a		b	

Table 2. Results of the NASA-TLX for system 1 (a) and system 2 (b) for the female gender.

General score	43.75	General score	68.75
Mental	40.00	Mental	75.00
Physical	28.33	Physical	63.33
Temporal	45.00	Temporal	57.50
Performance	30.00	Performance	46.67
Effort	51.67	Effort	78.33
Frustration	26.67	Frustration	61.67
a		b	

4 Conclusions and Future Work

Electroencephalography has been used to capture brain activity of a subject. The results have shown that different aspects of the user on interacting with a system can be analyzed, such as the workload and emotions that can influence the interaction. For this study we have relied on subjective tools such as SUS, NASA-TLX and SAM that have served to validate the results obtained when analyzing the characteristics of the EEG signals. This electroencephalography technique is an alternative that allows evaluation of the interaction of non-traditional systems. Today, the growth of technology has led to the creation of new interaction paradigms, which can become a challenge to evaluate these types of interface, which not only involve a digital but also a physical interface. The evaluation methods that exist from the HCI field usually ask the user directly and indirectly, before, during or after the interaction, and can be uncomfortable or the user very often cannot remember their emotional states or interaction events. Using encephalography, signals are captured as the user interacts with the system, and are not interruptions from questions about the interaction or how they feel with the interface. In the EEG signal, the frontal activity in the theta sub-band increases and the activity in the alpha sub-band decreases with an increase in load. In turn, the increase in load is more evident

in the female gender for system two compared to the male gender. These types of results serve to support EEG-based assessment methods to monitor cognitive load when a user interacts with a system. The rise of technology has made BCI systems more accessible due to their price and use, which allows these types of system to be used not only as interaction, but also as an evaluation tool in the field of HCI. As future work we want to make use of pattern recognition techniques, to recognize emotions and estimate the workload.

References

1. Frey, J., Mühl, C., Lotte, F., Hachet, M.: Review of the use of electroencephalography as an evaluation method for human-computer interaction, arXiv (2013)
2. Subha, D.P., Joseph, P.K., Acharya, U.R., et al.: EEG signal analysis: a survey. *J. Med. Syst.* **34**, 195–212 (2010). <https://doi.org/10.1007/s10916-008-9231-z>
3. Kragel, P.A., LaBar, K.S.: Decoding the nature of emotion in the brain. *Trends Cogn. Sci.* **20**(6), 444–455 (2016). <https://doi.org/10.1016/j.tics.2016.03.011>
4. Herrmann, C.S., Strüber, D., Helfrich, R.F., Engel, A.K.: EEG oscillations: from correlation to causality. *Int. J. Psychophysiol.* **103**, 12–21 (2016)
5. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Res. Rev.* **29**(2–3), 169–195 (1999)
6. Sulthan, N., Mohan, N., Khan, K.A., Sofiya, S., Shanir, P.P.M.: Emotion recognition using brain signals. In: 2018 International Conference on Intelligent Circuits and Systems (ICICS), Phagwara, pp. 315–319 (2018). <https://doi.org/10.1109/ICICS.2018.00071>
7. Cano, S., Araujo, N., Guzman, C., Rusu, C., Albiol-Pérez, S.: Low-cost assessment of user eXperience through EEG signals. *IEEE Access* **8**, 158475–158487 (2020). <https://doi.org/10.1109/ACCESS.2020.3017685>
8. Figueira, J.S.B., David, I., Lobo, I., et al.: Effects of load and emotional state on EEG alpha-band power and inter-site synchrony during a visual working memory task. *Cogn. Affect Behav. Neurosci.* **20**, 1122–1132 (2020). <https://doi.org/10.3758/s13415-020-00823-3>
9. Xiao, D., Zhang, W.: Electroencephalogram based brain concentration and its human computer interface application. In: 2015 IEEE International Conference on Computer and Communications (ICCC), Chengdu, pp. 21–24 (2015). <https://doi.org/10.1109/CompComm.2015.7387533>
10. Lahane, P., Sangaiah, A.K.: An approach to EEG based emotion recognition and classification using kernal density estimation. *Procedia Comput. Sci.* **48**, 574–581 (2015)
11. Sohaib, A.T., Qureshi, S., Hagelbäck, J., Hilborn, O., Jerčić, P.: Evaluating classifiers for emotion recognition using EEG. In: Schmorow, D.D., Fidopiastis, C.M. (eds.) Foundations of Augmented Cognition. AC 2013. Lecture Notes in Computer Science, vol. 8027. Springer, Berlin (2013). https://doi.org/10.1007/978-3-642-39454-6_53
12. Fok, S., Schwartz, R., Wronkiewicz, M., Holmes, C., Zhang, J., Somers, T., Bundy, D., Leuthardt, E.: An EEG-based brain computer interface for rehabilitation and restoration of hand control following stroke using ipsilateral cortical physiology. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, pp. 6277–6280 (2011). <https://doi.org/10.1109/IEMBS.2011.6091549>
13. Ergonomics of Human-System Interaction_Part 11: Usability: Definitions and Concepts, ISO Standard 9241-11, International Standardization Organization (ISO), Geneva, Switzerland (2018)

14. Nielsen, J.: Usability inspection methods. In: Conference Companion on Human Factors in Computing Systems, pp. 413–414. ACM (1994)
15. Beaudouin-Lafon, M.: Interaction instrumentale: de la manipulation directe à la réalité augmentée. In: Actes des Neuvièmes Journées sur l'Interaction Homme-Machine, IHM 1997 (1997)
16. Abiri, R., Borhani, S., Kilmarx, J., Esterwood, C., Jiang, Y., Zhao, X.: A usability study of low-cost wireless brain-computer interface for cursor control using online linear model. *IEEE Trans. Hum.-Mach. Syst.* **50**(4), 287–297 (2020). <https://doi.org/10.1109/THMS.2020.2983848>
17. Herwig, U., Satrapi, P., Schönenfeldt-Lecuona, C.: Using the international 10–20 EEG system for positioning of transcranial magnetic stimulation. *Brain Topogr.* **16**, 95–99 (2003)
18. Niedermeyer, E., da Silva, F.L.: Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins (2005). A New EEG Acquisition Protocol for Biometric Identification Using Eye Blinking Signals. https://www.researchgate.net/publication/275830679_A_New_EEG_Acquisition_Protoocol_for_Biometric_Identification_Using_Eye_Blinking_Signals. Accessed 03 Jan 2021
19. Saby, J.N., Marshall, P.J.: The utility of EEG band power analysis in the study of infancy and early childhood. *Dev. Neuropsychol.* **37**(3), 253–273 (2012)
20. Lin, Y., Liu, Z., Gao, X.: Alpha-band oscillation during speech recognition under different sensory conditions. In: 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), pp. 153–157. IEEE (2015)
21. Gola, M., Magnuski, M., Szumska, I., Wrobel, A.: EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects. *Int. J. Psychophysiol.* **89**(3), 334–341 (2013)
22. Miltner, W.H., Braun, C., Arnold, M., Witte, H., Taub, E.: Coherence of gamma-band EEG activity as a basis for associative learning. *Nature* **397**(6718), 434 (1999)
23. Calcagno, A., et al.: EEG monitoring during software development. In: 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), Palermo, Italy, pp. 325–329 (2020). <https://doi.org/10.1109/MELECON48756.2020.9140717>
24. Negi, S., Mitra, R.: EEG metrics to determine cognitive load and affective states: a pilot study. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp 2018), pp. 182–185. Association for Computing Machinery, New York (2018). <https://doi.org/10.1145/3267305.3267618>
25. Lan, T., Erdogmus, D., Adamo, A., Mathan, S., Pavel, M.: Channel selection and feature projection for cognitive load estimation using ambulatory EEG. *Comput. Intell. Neurosci.* **2007** (2007). <https://doi.org/10.1155/2007/74895>. Article no. 74895. PMID: 18364990, PMCID: PMC2267884
26. Roland, P.E.: Brain activation. Wiley-Liss, New York (1993)
27. Peng, H., Hu, B., Zheng, F., et al.: A method of identifying chronic stress by EEG. *Pers. Ubiquit. Comput.* **17**, 1341–1347 (2013). <https://doi.org/10.1007/s00779-012-0593-3>
28. Kawala-Janik, A., Pelc, M., Podpora, M.: Method for EEG signals pattern recognition in embedded systems. *Elektronika Ir Elektrotehnika* **21**(3), 3–9 (2015). <https://doi.org/10.5755/j01.eee.21.3.9918>
29. Herrmann, M.J., Huter, T., Plichta, M.M., Ehliis, A.-C., Alpers, G.W., Mühlberger, A., Fallgatter, A.J.: Enhancement of activity of the primary visual cortex during processing of emotional stimuli as measured with event-related functional near-infrared spectroscopy and event-related potentials. *Hum. Brain Mapp.* **29**, 28–35 (2008). <https://doi.org/10.1002/hbm.20368>
30. Gevins, A., Smith, M.E.: Neurophysiological measures of cognitive workload during human-computer interactions. *Theoret. Issues Ergon. Sci.* **4**, 113–131 (2003)

31. Klimesch, W., Schack, B., Sauseng, P.: The functional significance of theta and upper alpha oscillations for working memory: a review. *Exp. Psychol.* **52**, 99–108 (2005)
32. Stelios, X., Aristea, T.: Studying student's attitudes on using examples of game source code for learning programming. *Inform. Educ. Int. J.* **2**, 265–277 (2014)
33. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **25**(I), 49–59 (1994)
34. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*. North Holland Press, Amsterdam (1988)
35. Brooke, J.: SUS – a quick and dirty usability scale. In: *Usability Evaluation in Industry*, pp. 189, 194 (1996)
36. Lang, P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O.: Looking at pictures: evaluative, facial, visceral, and behavioral responses. *Psychophysiology* **30**(3), 261–273 (1993)
37. Dhiman, R., Priyanka, Saini, J.S.: Wavelet analysis of electrical signals from brain: the electroencephalogram. In: Singh, K., Awasthi, A.K. (eds.) *Quality, Reliability, Security and Robustness in Heterogeneous Networks*. QShine 2013. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 115. Springer, Berlin (2013). https://doi.org/10.1007/978-3-642-37949-9_24
38. Omerhodzic, I., Avdakovic, S., Nuhanovic, A., Dizdarevic, K.: Energy distribution of EEG signals: EEG signal wavelet-neural network classifier. arxiv.org/abs/1307.7897 (2013)
39. Jacob, J.E., Nair, G.K., Iype, T., Cherian, A.: Diagnosis of encephalopathy based on energies of EEG subbands using discrete wavelet transform and support vector machine. *Neurol. Res. Int.* (2018). 1613456. <https://doi.org/10.1155/2018/1613456>
40. Klimesch, W., Schimke, H., Pfurtscheller, G.: Alpha frequency, cognitive load and memory performance. *Brain Topogr.* **5**(3), 241–251 (1993)
41. Sanei, S., Chambers, J.A.: *EEG Signal Processing*. Wiley, Chichester (2007)



Heuristic Evaluation Method Applied to the Usability Assessment of Smart Homes Applications

Ana Isabel Martins¹ Ana Carolina Oliveira Lima²,
and Nelson Pacheco Rocha³

¹ Institute of Electronics and Informatics Engineering of Aveiro, Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal
anaisabelmartins@ua.pt

² Institute of Telecommunications, Department of Electronic, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal
ana.carolina.lima@ua.pt

³ Institute of Electronics and Informatics Engineering of Aveiro, Department of Medical Sciences, University of Aveiro, Aveiro, Portugal
npr@ua.pt

Abstract. The usability of smart homes applications is crucial to ensure their correct and pleasant use, thus contributing to the safety, comfort, and entertainment of the inhabitants of a smart home. Since heuristic evaluation is a widespread method for usability assessment, the objective of the study reported by the present paper was to verify the applicability of this method to assess the usability of smart homes applications. Therefore, an experience was conducted to apply the heuristic evaluation method to assess the usability of three different smart homes applications. This evaluation involved three evaluators and was based on the Heuristic Evaluation System Checklist (HESC). The results reveal that the application of the heuristic evaluation method using the HESC is a demanding and time consuming task, since 292 sub-heuristic should be considered. In turn, the results also show that heuristic evaluation is effective in the exhaustive identification of the concrete flaws of the interfaces of smart homes applications and generates a large quantity of objective information valuable for the improvement of the applications.

Keywords: Usability · Heuristic evaluation · Smart Homes

1 Introduction

Smart homes are intended for the safety, comfort and entertainment of the respective inhabitants, thus facilitating their daily activities [1]. The interaction with the devices of a smart home must be intuitive and adaptable to the characteristics of the inhabitants and their contexts [2]. In order to make smart homes applications user-friendly and usable by all types of inhabitants in different contexts, a large investment in usability design

and assessment is necessary. One way to ensure a high level of usability is to involve experts in heuristic evaluations of smart homes applications [3].

In the literature, several heuristic evaluations of smart homes applications are described, namely for assessing home control applications [4–6], smart homes devices [7], assistive robotic furniture [8], and pervasive games for smart homes [9]. However, these studies solely address the results of the evaluations themselves, and do not analyse the applicability of the heuristic evaluation method, nor the respective advantages and disadvantages. In this sense, the objective of the study reported by this article was to verify the applicability of the heuristic evaluation method to assess the usability of smart homes applications.

2 Background

Usability assessment is a demanding task especially when it comes to assess complex systems such as smart homes. For instance, considering the complexity of integrated control systems for residential environments, it is evident that the assessment of their usability is a challenging task, since several interactive home applications (e.g., to control heating systems) are commonly misused or under used because they are difficult to understand and operate [1].

According to ISO 9241-11, usability is a measure of how well a specific user in a specific context can use a system or service to achieve specific objectives with effectiveness, efficiency, and satisfaction [10, 11]. Usability can affect the acceptance of a particular system and applies to all the aspects of the interaction, including the procedures of installation and maintenance [12]. It is not derived from the aesthetic, the latest generation interaction mechanisms or the intelligence integrated in the interfaces, but it is achieved when the design of the interfaces attends to the real needs of the users [13]. Moreover, usability is dependent on the context of use, meaning that the level of usability obtained depends on the specific circumstances in which the system or service is used [10], including tasks, equipment (i.e., hardware and software), and the physical and social environment, since all these factors can influence the usability of a system or service [14].

Usability assessment can be empirical (based on data from real users) or analytical (based on the analysis by experts). Empirical models include test and inquiry methods, that are based in the observation of the users' performance while they are using the system or service to perform a task, while analytical models involve the inspection the attributes of the interface design and are typically conducted by experts, without involving the participation of users [15, 16]. This paper focuses on heuristic evaluation, which is an inspection method.

According to Nielsen, heuristic evaluation involves the systematic inspection of a user interface to analyse its usability [3]. The method comprises the judgment of one or more usability experts (e.g., ergonomists or computer engineers) about the suitability of the interface of a system or service according a set of criteria, recommendations standards, and usability principles (i.e., the heuristics) [17]. Heuristics are established set of principles of interface design organized on written sentences [18]. Heuristics have a dual-use as they can be used both for creating an interface (typically used by designers

and developers) and to assess its compliance in terms of usability (typically performed by usability evaluators) [19].

In a heuristic evaluation, the participation of three to five evaluators is recommended since a single evaluator may not be able to identify all the problems of an interface. It is also advised that the assessment should be carried out in two stages to ensure independent and uninfluenced assessments. In the first stage, each evaluator should inspect the interface individually at least twice, considering the heuristics, on order to identify usability problems. In the second stage, known as a debriefing, a consensus meeting should take place so that evaluators are able to discuss the usability problems they encountered and to generate a unique list of problems [12].

3 Methods

3.1 Study Design

Considering the aforementioned objective, the study reported by this article was informed by the following research questions:

- RQ1: Is heuristic evaluation suited to assess the usability of smart homes applications?
- RQ2: What are the advantages and disadvantages from using heuristic evaluation to assess the usability of smart homes applications?

An experience was conducted to apply the heuristic evaluation method to assess the usability of three different smart homes applications.

In this usability assessment three evaluators were involved. They had a vast experience in usability, user experience and human factors, and have been involved in research in human-computer interaction for over ten years. To perform the inspection, two evaluators analysed all the screens of the applications' interfaces individually at least twice, based on the heuristics, to identify and note possible problems.

The heuristic evaluation was based on the Heuristic Evaluation System Checklist (HESC) [20]. The HESC covers the ten original Nielsen heuristics [11] and also includes three heuristics related to user skills, user experience and privacy. In total, the HESC consists of 292 sub-heuristics that help the evaluators to review thoroughly the applications.

The assessment was carried out in two stages to ensure independent and uninfluenced assessments. In the first stage, each evaluator individually inspected the all the screens of the applications' interfaces twice considering the HESC, aiming to identify usability problems and their locations. In the second stage, the debriefing, there was a discussion in which the evaluators clarified the problems uncovered and generated a single list of usability problems. The third evaluator participated in this debriefing and established an agreement whenever there were doubts.

The procedures and work sessions held to carry out the heuristic evaluation of the three smart homes applications were the following:

- Training session on the applications - A presentation session of the applications and its functionalities was held on June 2020. In this session information about the applications was provided to the evaluators. None of the evaluators had any previous contact with the applications or were involved in their development.
- First heuristic evaluation session - The evaluators performed the heuristic evaluation based on the HESC [20], independently, indicating the non-applicable heuristics, those verified without problems and those with problems. This first review took place on the middle of June 2020.
- Second heuristic evaluation session - The evaluators reviewed the applications for the second time, adjusting whenever necessary the evaluation made in the first session. This second review took place on the end of June 2020.
- Consensus session - The evaluators compared the results of their individual evaluations and established a consensus regarding the sub-heuristics in which there was disagreement. A third evaluator participated in this session that was held on the beginning of July 2020.

Once the assessment of the smart homes applications was concluded, the three evaluators conduct a brain storming discussion to systematize the main difficulties related to the application of the heuristic evaluation method, to identify the respective advantages and disadvantages, as well as to analyse the respective adequacy in terms of usability assessment of smart homes applications.

3.2 The Assessed Applications

The three smart homes applications selected for the experimental work (i.e., HomeCom, HomeComPro and EMMA) were developed by Bosch Termotecnologia (Fig. 1).

HomeCom is an application aiming to control the heating system over the internet using a smartphone, tablet or computer. In turn, HomeCom Pro is a version of HomeCom that allows not only the interaction with inhabitants but also the interaction with installers. It allows the technicians to access remotely the heating systems of their customers for providing help desk functions. Therefore, the application allows managing small repairs in a more accessible way.

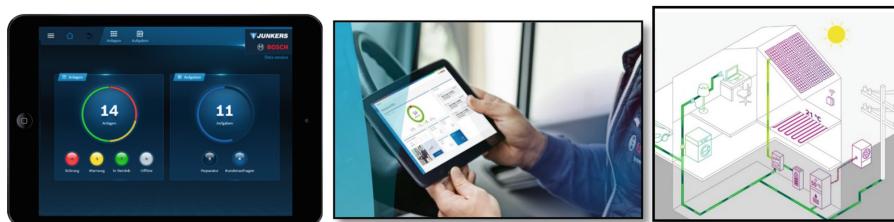


Fig. 1. HomeCom, HomeCom Pro and EMMA.

Finally, EMMA is an energy management application for smart homes. It is part of Bosch Smart Home Controller that is connected to a photo voltaic system via a

communication network and its objective is to distribute solar energy in an intelligent and automatic way in a smart home. Household appliances have priority and after they are supplied, energy flows to the heat pump and the battery storage. Superfluous solar energy only flows into the public grid when home devices and batteries are already supplied. Through the management application, the inhabitants have access to the energy distribution in the building at any time.

4 Results

The application of the heuristic evaluation method based on the HESC [20] is not a simple process, even for experts, as it presupposes the verification of a large number of heuristics and sub-heuristics.

Each heuristic is divided into sub-heuristics and each sub-heuristic is presented as a question. For example, the first two sub-heuristics of the heuristic 1 (*Visibility of the System Status*) are sub-heuristic 1.1 (*Does every display begin with a title or header that describes screen contents?*) and 1.2 (*Is there a consistent icon design scheme and stylistic treatment across the system?*). The evaluators are able to identify usability problems by answering the sub-heuristics questions. However, before answering each sub-heuristic question, the evaluators must decide if it is applicable to the system or service being evaluated or not.

The HESC includes 292 sub-heuristics as it is presented in Table 1. Moreover, the number of sub-heuristics per heuristic is not constant. The heuristic with more sub-heuristics is the heuristic 4 (*Consistency and Standards*) with 51 sub-heuristics, and the heuristic with less sub-heuristics is the heuristic 13 (*Privacy*) with only three sub-heuristics. The fact that there are so many sets of sub-heuristics, with different number of elements, makes the method difficult to apply.

In the HomeCom and HomeCom Pro evaluations, of the 292 sub-heuristics included in the HESC, 114 were not applicable, while in EMMA, 124 sub-heuristics were not applicable. The sub-heuristics that were not applicable are those that address features that do not make sense for the applications under analysis. Examples of sub-heuristics that were not applicable when assessing HomeCom and HomeCom Pro include sub-heuristic 13.1 (*Are protected areas completely inaccessible?*), because there are no protected areas in both applications, or sub-heuristic 4.48 (*If the system has multipage data entry screens, do all pages have the same title?*) since both applications do not have multipage data entry screens. In the case of EMMA, examples of sub-heuristics that were not applicable are sub-heuristic 7.8 (*Is there an obvious visual distinction made between “choose one” menu and “choose many” menus?*), since the application does not have chosen menus, or sub-heuristic 12.15 (*Is the numeric keypad located to the right of the alpha key area?*) because the application do not support the use of keypads.

Although a significant number sub-heuristics were not applicable for the evaluation of HomeCom, HomeCom Pro and EMMA, there was no quick way to select which ones were applicable and non-applicable. In fact, within the same heuristic, some sub-heuristics may be applicable, and others may not, which forced the evaluators to always review all of them.

Another aspect that was evident in the completion of the heuristic evaluations was the fact that it took a significant amount of time to carry out the evaluations. Since the

Table 1. The HESC and corresponding number of sub-heuristics.

Heuristics	Number of sub-heuristics
1. Visibility of the System Status	29
2. Match between System and the Real World	24
3. User Control and Freedom	23
4. Consistency and Standards	51
5. Help Users Recognize, Diagnose, and Recover from Errors	21
6. Error Prevention	15
7. Recognition Rather Than Recall	40
8. Flexibility and Minimalist Design	16
9. Aesthetic and Minimalist Design	12
10. Help and Documentation	23
11. Skills	21
12. Pleasurable and Respectful Interaction with the User	14
13. Privacy	3
Total	292

HESC is so exhaustive and contains so much detail, the evaluators took a long time to go through all the screens of the applications' interfaces and test all the interaction paths. On average, for each application, each evaluator took 428 min (i.e., 7 h and 15 min) in the first review, and 298 min (i.e., around five hours) in the second review (Table 2). The duration of the reviews included checking the applicable heuristics, when they are being complied, and, in the case of failures, the detailed record of the failure. The second review was considerably faster than the first review, because in the first one the evaluators were interacting with the applications for the first time, while in the second review the evaluators already knew the applications and were just confirming the flaws and checking if any other failure went unnoticed.

Considering the number of screens of the interfaces of the three assessed applications, on average, each evaluator took from seven to nine minutes for the first review of each screen and respective connections of the application' interfaces, and from five to six for the second review of each screen and respective connections (Table 2).

A contribution for the long time consumed to perform each evaluation is the fact that, at times, sub-heuristics are difficult to interpret. Many are very similar to each other and it is difficult for the evaluators to select among them. What happens is that the evaluators go back and forth in the HESC to decide which heuristic best suits the identified flaw. For example, just considering the labels there are 12 different sub-heuristics, some of which are similar and require careful reading by the evaluators. Some examples are sub-heuristic 2.24. (*Are function keys labelled clearly and distinctively, even if this means breaking consistency rules?*), sub-heuristic 4.5 (*Are icons labelled?*), sub-heuristic 4.18 (*Are field labels and fields distinguished typographically?*), sub-heuristic 4.19 (*Are field labels*

Table 2. Summary of the heuristic evaluation process – average time spend by each evaluator during the two revisions and for each screen of the applications' interfaces.

	Time 1 st review (min)	Time 2 nd review (min)	Number of screens	Average 1 st review time per screen (min)	Average 2 nd review time per screen (min)
Home Com	445	305	49	9	6
Home Com Pro	360	280	51	7	5
EMMA	480	310	63	8	5
Average	428	298	54	8	5

consistent from one data entry screen to another?), sub-heuristic 4.20 (Are fields and labels left-justified for alpha lists and right-justified for numeric lists?), sub-heuristic 4.21 (Do field labels appear to the left of single fields and above list fields?), sub-heuristic 7.15 (Are field labels close to fields, but separated by at least one space?) and sub-heuristic 9.8 (Are field labels brief, familiar, and descriptive?). Moreover, it should be noted that these sub-heuristics are part of different heuristics (i.e., heuristics 2, 4, 7, and 9).

After identifying a problem in a label, the evaluators reviewed the different sub-heuristics and select the one that best describes the failure. The fact that the sub-heuristics are similar also led to a greater disagreement between the evaluators. For instance, frequently, both evaluators identified the same usability flaw in the application but attributed different sub-heuristics. This difficulty in applying the method was overcome in the debriefing session, in which the evaluators, together with the third evaluator, were able to reach a consensus on the failed sub-heuristics.

All flaws identified in the applications were recorded, along with the description of the problem and the proposal for improvement to correct it. This evaluation resulted in a detailed report with the results of the heuristic evaluation. Table 3 describes the number of detailed problems, prints and suggestions for improvement for each application evaluated. The number of problems' descriptions, prints and suggestions for improvement diverge, as, sometimes, a problem description is applicable to different failures, in the same print there are various problems, or the same improvement suggestion is applicable to different problems. The opposite also happens, when for the same problem, different solutions are suggested depending on the problem specific context.

As real examples, some flaws and corresponding suggestion for improvement are presented below (prints were omitted due to commercial property confidentiality):

- In HomeCom, the sub-heuristic 10.8 (*Is the help function visible; for example, a key labelled HELP or a special menu?*) failed because the help is located in a footer in the contact part. As it is too hidden, the suggestion for improvement was to create a menu option for help.

Table 3. Usability report results.

	Heuristics with flaws	Problems descriptions	Prints	Suggestion for improvements
Home Com	28	25	21	26
Home Com Pro	29	26	22	27
EMMA	21	28	36	23

- Still in HomeCom, the sub-heuristic 6.11 (*Does the system prevent users from making errors whenever possible?*) failed because the application does not advise users how to adjust the desired temperature on the panel instigating the occurrence of errors. The user must deduce what is the mechanism used to adjust the temperature. The suggestion for improvement was to clarify how to set the desired temperature (information in mouse-over).
- In HomeCom Pro, the sub-heuristic 6.14 (*Do data entry screens and dialog boxes indicate the number of character spaces available in a field?*) failed because in the forms there is no information of the total number of characters that the field supports. The suggestion for improvement was to add at the end of the text box the number of characters that the field supports, and this number must be updated as the user writes.
- In EMMA, the sub-heuristic 7.36 (*Do GUI menus offer affordance: that is, make obvious where selection is possible?*) failed because the configuration operations look like a menu, but have no functionality. It is not clear whether it is a menu or a sequence of information. It looks like menu options but are not selectable, and for that reason the suggestion for improvement was to change the appearance of the configuration.
- Still in Emma, the sub-heuristic 3.15 (*If the system has multiple menu levels, is there a mechanism that allows users to go back to previous menus?*) failed because in two second level menus there is no return option and the suggestion for improvement was to place a back button as there is in other menus.

Therefore, the heuristic assessment method, despite the difficulties of its application, generates a large amount of concrete data, which is essential for the improvement of smart homes applications.

5 Discussion

The application of the heuristic assessment method to evaluate smart homes applications was effective in the exhaustive identification of failures of the applications' interfaces, and allowed to list the concrete aspects of the interface problems and served as a basis for creating improvement suggestions adapted to each particular situation.

The great advantage of this method is that it generates a lot of objective information on how to improve the assessed applications. Even robust applications, with good levels of usability, such as HomeCom, HomeCom Pro and EMMA, can benefit from this type of inspection because it is possible to collect clear and detailed information with a high

degree of precision. This is a challenge in terms of usability assessment, as most of the existing test and inquiry methods provide a global usability score but do not deliver clear indications about application failures or suggestions on the aspects that must be improved. In addition, the fact that sub-heuristics are questions posed in the positive makes it easier to establish the reasoning to resolve the failure, since the sub-heuristic itself describes the final objective.

Despite the success in applying the method and its positive results, it also presented some challenges and difficulties.

One of the main difficulties is that when identifying a usability problem, it is difficult to match with the corresponding sub-heuristic because the HESC is exhaustive and has many similar sub-heuristics, which makes it difficult to know exactly which one best fits.

The fact that the sub-heuristics were very similar led to a greater disagreement between the evaluators (i.e., on several occasions both identified the same problem in the application but attributed different sub-heuristics). This difficulty was overcome in the debriefing session, in which the two evaluators, together with the third evaluator, were able to reach a consensus on the failed sub-heuristics.

Moreover, the application of the heuristic evaluation method is a demanding and time-consuming task, not only because the HESC is extensive, but also because each evaluator makes two rounds of verification. Each verification round took almost an entire working day to be completed.

It should be noted that, ideally, this type of assessment should be done before the application is tested with end users. The fact that most flaws are so exhaustively identified and corrected before smart homes applications are presented to users will help them to have a better user experience with a smoother use and without severe usability problems. In this case, users will be able to detect more discreet flaws that go unnoticed and that are only identified with real use continued over time.

6 Conclusion

Considering the first research question (i.e., *Is heuristic evaluation suited to assess the usability of smart homes applications?*), it can be concluded that it is confirmed because, as it was clear in the results, this method allows to collect a large amount of important information to improve the application towards a better usability.

Regarding the second research question (i.e., *What are the advantages and disadvantages from using heuristic evaluation to assess the usability of smart homes applications?*), as main advantages it is possible to list the fact that it generates a large quantity of objective information, including concrete aspects of the interface usability problems and serve as a basis for improvement suggestions adapted to each particular application. As disadvantages, it is difficult to match the problem identified with the corresponding sub-heuristic because they are often similar and hard to interpret, and it is a time-consuming task.

Acknowledgments. The present study was developed in the scope of the Smart Green Homes Project [POCI-01-0247 FEDER-007678]. The publication was financially supported by National

Funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the project UI IEETA: UID/CEC/00127/2020.

References

1. Alshammari, A., Alhadeaf, H., Alotaibi, N., Alrasheedi, S., Saudi, A.: The usability of HCI in smart home. *Int. Res. J. Eng. Technol.* **2**, 1–7 (2008)
2. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: a survey of principles, models and frameworks, Berlin, pp. 3–26. Springer, Heidelberg (2009)
3. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: *Proceedings of the Conference on Human Factors in Computing Systems*, pp. 249–256 (1990)
4. Kartakis, S., Antona, M., Stephanidis, C.: Control smart homes easily with simple touch. In: *Proceedings of the 2011 ACM Multimedia Conference and co-located Corkshops - UBI-MUI 2011*, pp. 1–6 (2011)
5. Marques da Silva, A., Ayanoglu, H., Silva, B.: An age-friendly system design for smart home: findings from heuristic evaluation. In: *International Conference on Human-Computer Interaction*, pp. 643–659 (2020)
6. Lopez, J., Textor, C., Hicks, Pryor, W.B. McLaughlin, A.C., Pak, R.: An aging-focused heuristic evaluation of home automation controls. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 6–10 (2019)
7. Demiris, G., Skubic, M., Rantz, M., Keller, J., Aud, M., Hensel, B., He, Z.: Smart home sensors for the elderly: A model for participatory formative evaluation. *Hum.-Comput. Interact.* **6**, 7 (2020)
8. Threatt, A., Merino, J., Keith, S., Green, E., Walker, I., Brooks J.O.: An Assistive Robotic Table for Older and Post-Stroke Adults: Results from Participatory Design and Evaluation Activities with Clinical Staff (2014)
9. Röcker, C., Haar, M.: Exploring the usability of video game heuristics for pervasive game development in smart home environments In: *Proceedings of the Third International Workshop on Pervasive Gaming Applications–Pergames*, pp. 199–206 (2006)
10. International Organization for Standardization, ISO 9241-11:2018 Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts (2018)
11. Nielsen, J., Mark, R.: Usability Inspection Methods. *Heuristic Evaluation* (1994)
12. Nielsen, J.: Usability Engineering. Academic Press, Cambridge (1993)
13. Constantine, L., Lockwood, L.: *Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design*. Pearson Education, London (1999)
14. Bernsen, N.O., Dybkjær, L.: *Multimodal Usability*. Springer, London (2010)
15. Martin, B., Hanington, B.: *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers, Beverly (2012)
16. Dix, A., Finlay, J., Abowd, G., Beale, R.: *Human-Computer Interaction*, 3rd edn. Prentice Hall, Upper Saddle River (2004)
17. Rogers, Y., Sharp, H., Preece, J.: *Design de interação : além da interação humano-computador*. Bookman (2013)
18. Costa, R., Canedo, E., Sousa, R., Albuquerque, R., Villalba, L.: Set of usability heuristics for quality assessment of mobile applications on smartphones. *IEEE Access* **7**, 116145–116161 (2019)
19. Sauro, J.: MeasuringU: Understanding Expert Reviews and Inspection Methods (2019). <https://measuringu.com/inspection-methods/>. Accessed 21 May 2020
20. Pierotti, D.: Heuristic Evaluation - A System Checklist (2004). <ftp://cs.uregina.ca/pub/class/305/lab2/example-he.html>. Accessed 05 Sept 2020



Smart Glasses User Experience in STEM Students: A Systematic Mapping Study

Ronny Santana^{1,3,4(✉)}, Gustavo Rossi^{1,2}, Gonzalo Gabriel Méndez⁴, Andrés Rodríguez¹, and Viviana Cajas^{1,5}

¹ LIFIA, Fac. de Informática, Universidad Nacional de La Plata, La Plata, Argentina
{gustavo.rossi.andres.rodriguez}@lifia.info.unlp.edu.ar

² CONICET, Buenos Aires, Argentina

³ Universidad de Guayaquil, Guayaquil, Ecuador
ronny.santanae@ug.edu.ec

⁴ Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador
{roensant,gmendez}@espol.edu.ec

⁵ Universidad Tecnológica Indoamérica, Quito, Ecuador
vivianacajas@uti.edu.ec

Abstract. User experience (UX) is related to the feelings and emotions that people undergo when interacting with technology. This concept also applies to wearable devices, such as smart glasses, which have been widely adopted in a myriad of contexts in recent years. This paper analyzes the literature on user experience with smart glasses, with a particular focus on STEM educational settings. Our goal is to identify gaps and opportunities within this area and contribute to inform future research. To this end, we conducted a systematic mapping study of papers published between 2014 and 2020 indexed by four scientific databases and repositories: the ACM digital library, the IEEE Xplore, Scopus, and Web of Science. Our selection and systematic classification processes considered studies conducted in educational settings or with educational purposes. A total of 485 studies were initially identified and mapped. After revising and analyzing this set of publications, 51 studies were selected and further classified according to their research and contribution, and the educational setting in which they were conducted. This mapping study offers the first systematic exploration of the state of the art on user experience with smart glasses within the context of STEM education.

Keywords: Human-computer interaction · Smart glasses · STEM education · STEM students · User experience · Systematic mapping

1 Introduction

In a broad sense, wearable computing could be defined as technology that is worn on the body, like clothing [1]. This paradigm assumes the existence of a device, with which humans somehow interact and that is able to perform diverse types of computational tasks [2]. In educational settings, wearable technologies can be

particularly useful, as they can become active tools in the classroom, improving the students' instruction and supporting new ways of participation [3]. Smart glasses are a type of wearable devices with capabilities that augment or superimpose real-world objects [4] on top of what a person can see. Recent technological advancements, have made smart glasses increasingly more innovative and powerful. Ultimately, this has led to their use in a wider range of scenarios. A notable example of this is the Augmented Reality (AR) capabilities featured by most current smart glasses [5]. AR brings numerous advantages when smart glasses are used in educational contexts and several research efforts have exploited AR-based solutions to facilitate and enhance learning. In medicine education, for example, smart glasses enable access to virtual representations of the human body that can be used in surgery training. This allows students to get significant experience before engaging in real-world practical scenarios [6].

The widespread and increasing application of wearable technologies and, in particular, of smart glasses has led to a myriad of research efforts in this area. In turn, this has contributed to the generation of knowledge collected in a large variety of scientific publications. We present a systematic mapping of this literature. More specifically, we focus on user experience of smart glasses in STEM educational settings (science, technology, engineering, mathematics). We seek to contribute to the understanding of the current state of the art in this area, and to identify opportunities that motivate and inform future research efforts.

2 Related Work

Most of the work that has reviewed the literature on smart glasses in educational contexts has focused on areas related to medicine and medical training. Dougherty et al., for example, explored the use of Google Glass in nonsurgical medical settings in a systematic review that covered literature published between 2013 and 2017 [7]. Along the same lines, Yoo et al. focused on works that involved augmented reality and wearable head up displays in surgical use [8]. Mitrasinovic et al. [9] also review salient uses of smart glasses in healthcare with a particular focus on practical capabilities and patient confidentiality. More recently, Badiali et al. explored the adoption of AR guidance in surgical practice in oral and crano-xillofacial surgery [10]. Given that these works are centered on medical settings, they review databases of publications on life sciences and biomedical topics (e.g., PubMed MEDLINE¹, Embase², EBSCO³), as well as repositories oriented to the publication of technological advancements (e.g., IEEE Xplore).

Closer to our area of interest, the work by Kumar et al. [11] investigated the typical applications of smart glasses in the education sector and identified several ways in which wearable technology supports teaching and learning processes (e.g., documentation of lectures, capturing lectures' essential points, telementoring, trainee's evaluation, on-site report preparation). A similar investigation was

¹ <https://pubmed.ncbi.nlm.nih.gov>.

² <https://www.embase.com>.

³ <https://www.ebsco.com>.

conducted by Sapargaliyev [12] but only for scenarios that used Google Glass as a teaching and learning tool.

The body of knowledge referred above, explores specific facets of the use of smart glasses, both from a medical perspective and from an educational point of view. We share with these research efforts the goal of reviewing the literature on the use of smart glasses to identify gaps and research opportunities. We, however, are interested in how the use of smart glasses in educational contexts relates to the concept of user experience. This is a central principle of our work. Thus, our systematic mapping focuses on studies that involved the use of smart glasses with STEM students and that somehow characterized aspects related to user experience (such as gathering of usability metrics, or identifying important human factors). To this end, we follow a structured searching protocol that focuses only on the publications produced in the last five years—as most commercial smart glasses started to become available more widely since 2015 [1]. This allows us to draw conclusions in the light of the most recent research.

3 The Systematic Mapping Study

We conduct our systematic mapping following the strategies and frameworks proposed by Kitchenham [13], Brereton [14] and Petersen [15]. In the development of our literature search, we followed the PICO process [16], that defines four conceptual elements to drive the publications search: a **P**opulation (or area of interest), an **I**ntervention, a **C**omparison (also called control or comparator), and an **O**utcome. These elements are defined as follows:

Population: The scientific studies with Smart Glasses that have been conducted with a focus on user experience.

Intervention: Smart glasses and related technologies.

Comparison: Where are smart glasses used.

Outcome: Identification of UX related aspects when STEM students used smart glasses in learning environments.

3.1 Mapping Questions

We aim at conducting a systematic mapping study on the user experience with smart glasses in STEM educational settings. To this end, we consider aspects related to usability, interaction, and human factors that hint at the possible benefits and limitations of this type of technology. More specifically, our mapping questions are:

MQ1: What are the uses of smart glasses in STEM educational contexts?

MQ2: What types of smart glasses are used to evaluate UX and/or usability in these studies?

RQ1 seeks to identify the application areas of smart glasses in STEM educational contexts in STEM educational contexts. On the other hand, RQ2 inquires on the types of smart glasses that are most frequently used to explore user experience and related concepts.

3.2 Search Strategy

Our search strategy involved the definition of search strings to identify relevant primary studies, according to what Kitchenham et al. [13] suggest. On top of these initial results, we used the PICO criteria to structure and further refine our search results. The PICO criteria we used are detailed in Table 1.

Table 1. Terms included in the search

Criteria	Main terms	Alternative terms
Population	User experience	UX, student, pupil, trainee, undergraduate, undergrad, apprentice, disciple, learner, learners, scholar, teacher, lecturer, professor, tutor, instructor, trainer, educator experience
Intervention	Wearable	Smart Glasses, smart glass, smart eyewear, google glass, augmented reality glasses, ar glasses
Comparison	STEM	Stem education, stem subjects, stem majors, stem learning, learning for stem, stem teaching, knowledge in stem, stem careers, stem disciplines, stem areas, higher education, student learning, education, learning, teaching, knowledge, training, study, learnedness
Outcome	Application	Use, usage, utilization, utility, usefulness, application, implementation, manipulation

We identified our primary studies by using search strings on the selected scientific databases. These strings contained the main and alternative terms listed above joined through conjunction and disjunction logical operators.

3.3 Databases and Inclusion & Exclusion Criteria

We decided to query four digital repositories of scientific literature: the ACM Digital Library⁴, the IEEE Xplore⁵, Scopus⁶, and Web of Science (WoS)⁷.

⁴ <https://dl.acm.org>.

⁵ <https://ieeexplore.ieee.org>.

⁶ <https://www.scopus.com>.

⁷ <https://www.webofknowledge.com>.

We chose these digital libraries because they are amongst the most worldwide recognized repositories of research results in the areas of engineering, computing, and informatics. Moreover, they have excellent bibliographic indicators and metrics for journals, conference papers, book chapters, and magazines [17].

Based on our mapping questions, we defined the following inclusion and exclusion criteria:

Inclusion criteria: Papers published from 2014 to 2020 that report research with smart glasses in STEM educational settings and focus on UX aspects. This includes publications from journals and conferences that are written in English and appear indexed by any of the databases mentioned above.

Exclusion criteria: Papers whose full text was not available for download, short papers (e.g., position papers, extended abstracts), other systematic revisions or literature surveys, duplicated entries (e.g., papers that are simultaneously indexed in more than one database), papers that do not focus on smart glasses and UX or that do not involve studies with STEM Students, grey literature (e.g., reports, working papers, government documents, white papers and evaluations), and studies outside the period [2014 – 2020].

From each paper we considered, we extracted the following data:

General information: Title, author, and publication date.

Document type: Conference paper, journal, symposium, and tech report.

Application scope: Educational setting, industry.

Smart glasses type: E.g., Google Glass, Epson Moverio, Microsoft Hololens.

3.4 Studies Selection Process

The studies we selected resulted from a two-stage procedure—as done in other systematic mappings (e.g., [18]). In the first stage, a researcher reviewed the title and abstract of the papers of our initial search. In this step, irrelevant documents (e.g., those that involved wearable devices but not smart glasses) were discarded. The list of resulting papers was then revised by another researcher, who conducted a verification step. This consisted on reading the papers' title and abstract to assess their relevance. When the second researcher disagreed with the opinion of the first one, the study was discussed further until a unanimous decision was reached. In the second stage, we obtained the full text version of the documents selected in the previous step. On these, we applied the inclusion and exclusion criteria defined earlier. This process was done by two researchers and disagreements were resolved with the opinion of a third person.

4 Results

Our initial search resulted in 485 studies (see first column of Table 2). Duplicated papers were then removed. This step reduced the publication pool to a

total of 382 papers. We then applied the inclusion and exclusion criteria defined earlier. The full text of all the papers that complied with these criteria was then reviewed. Our final set of studies was composed by 51 papers. A detailed breakdown of this process (per database) is show in Table 2.

Table 2. Processing phases of the search results.

Database	Results of initial search	After removing duplicates	Records screened by abstract & other metadata	After applying inclusion & exclusion criteria	Selected papers after full-text analysis
Scopus	125	85	38	15	15
IEEE Xplore	45	24	11	8	8
ACM	214	214	58	26	26
WoS	101	59	16	2	2
Total	485	382	123	51	51

Our systematic mapping resulted in a collection of 51 relevant publications⁸. Our results show that a few studies conducted with smart glasses in STEM educational settings and that focused on user experience, were published in 2014 ($n = 3$). 2015 is the year with more research activity in this area ($n = 14$), followed by 2016 ($n = 10$). The following years have a more or less oscillating number of publications: 2017 ($n = 4$), 2018 ($n = 9$), 2019 ($n = 4$), and 2020 ($n = 7$). Figure 1 shows this evolution together with the distribution of papers per education level (Fig. 1a) and the publications' type of venue (Fig. 1b).

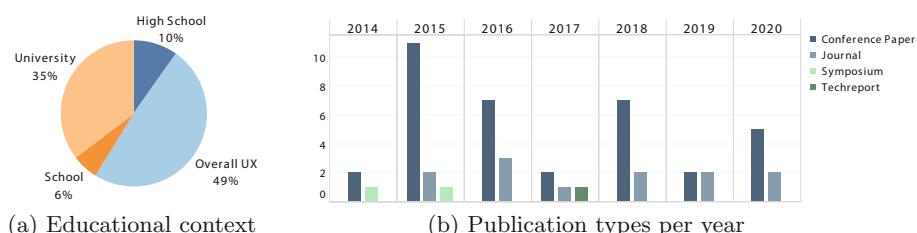


Fig. 1. Contexts in which smart glasses are used and evolution of publications.

Our review revealed that smart glasses have been widely used and studied at different educational levels (**MQ1**). In primary schools, for example, Silva et al. used Glassist [19], an application designed to help teachers in management

⁸ The full list of papers analyzed is available at <http://bit.ly/2WMha83>.

tasks, on Google Glass. The results of a preliminary evaluation of Glassist seemed promising. At the high school level, Kuhn et al. [20] described the development of an application to perform physical experiments. Students who evaluated this self-reported higher levels of cognitive load when working with Google Glass versus other devices.

A myriad of additional studies have explored the use of smart glasses in higher education activities (e.g., [21–24]). Among many others, this includes applications oriented to support learning in science (e.g., [25–28]), as well as technology and engineering (e.g., [29–37]). In the latter category, Cao et al. [38] used a pair of Epson Moverio BT-350 smart glasses for an AR guidance system for experimental teaching. The system supports learning of basic hardware information together with training of a programming environment.

Less conventional educational settings in which smart glasses have been used include museums (e.g., [39]), public speaking (e.g., [40]), procedural knowledge training (e.g., [41]) and fire safety (e.g., [42]). Our systematic mapping did not surface uses of smart glasses in mathematical teaching or learning activities.

Regarding the most frequent types of smart glasses used in STEM educational settings (**MQ2**), we found that the Google Glasses have been, by far, the most popular wearable device used for scientific research in this space between 2014 and 2020. 50 % ($n = 26$) of the studies we reviewed used them. The Google Glasses were followed by the Epson Moverio ones and the Microsoft Hololens—each used in 4 studies. Other types of devices were used in either one or two of the studies we reviewed. These results are summarized in Fig. 2.

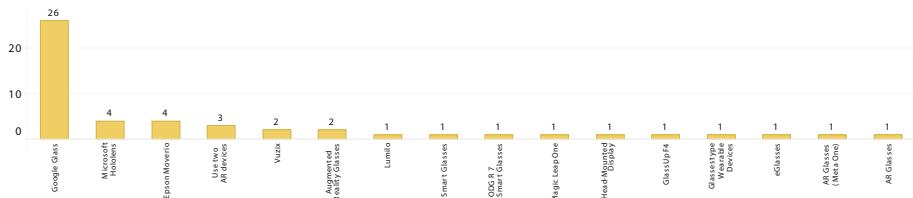


Fig. 2. Number of smart glasses used in the studies we mapped.

5 Discussion

Our mapping study shows that smart glasses have been steadily used in educational settings between 2014 and 2020. This is consistent with the increasing adoption of wearable devices promoted by communities from other academic disciplines—such as Human-Computer Interaction and Information Visualization. Our work also highlights that virtual and augmented reality are important supporting technologies in the use of smart glasses in teaching and learning environments. In fact, none of the studies we reviewed prescinded from these

technologies. Thus, VR and AR play a crucial role in advancing the adoption of smart glasses in a wider range of scenarios.

Our literature exploration also confirmed that the vast majority of work with smart glasses—both within and outside educational settings—has focused on medical and clinical applications. This suggests that there is potential for exploiting smart glasses and related technologies in a larger variety of educational settings. For example, we believe these devices could be beneficial to train students on concepts that involve physical manipulation of objects (such as tools or specialized equipment). We also believe this is especially relevant for the current world context, in which face-to-face instruction has shifted greatly to remote learning due to the sanitary crisis around the COVID-19 pandemic. An interesting counterpart of the many applications of smart glasses in medical contexts is the absence of these devices in the teaching of Mathematics. This may be explained due to the abstract nature of many math concepts. Overcoming abstraction would require representing such concepts through animations or virtual models whose production can be time consuming and often requires specialized knowledge. This type of applications, however, could pave the way for new and interesting research questions in the area of wearable devices.

We also note that just a comparatively small number of the studies we reviewed focused on user experience. Most smart glasses applications are built and designed oriented to answer specific research questions. This could explain why researchers do not often investigate user experience as a main aspect of their studies. Nevertheless, there are notable exceptions (e.g., [43–56]). This body of work also includes examples in which usability and user experience have been investigated from specific perspectives: with users with different levels of computer skills [57], in gaming [58] and public spaces [59], with palm-based text entry [60], in the automotive industry [61], in manufacturing [62, 63], and in farming [64]. Finally, other studies also show that there exists human factors that can also affect the acceptability and technological adoption of smart glasses [65, 66]. This highlights the importance of considering the “humans in the loop” when designing and studying studies that use smart glasses and related technology.

6 Conclusions and Future Work

This paper presented a systematic mapping study on user experience with smart glasses in STEM educational settings. We focused on educational scenarios, extending previous work that explored smart glasses and UX in medicine and medical training contexts. We considered scientific publications generated between 2014 and 2020 and indexed in four digital repositories (the ACM digital library, the IEEE Xplore, Scopus, and Web of Science). From an initial set of 485 papers, we selected 51 papers that reported studies with a focus on UX. Based on these publications, we identified the most common uses of smart glasses in educational settings and the most common types of devices. We also discussed potential opportunities that lay at the intersection of the research with smart glasses and the evaluation of usability.

The main limitation of our mapping study—also evidenced by similar works (e.g., [67])—is that the studies we considered were obtained using tools integrated into the indexing systems and the digital libraries we consulted. Additionally, the results output by these tools were retrieved based on handcrafted chains of logical operators and keywords that we constructed. Moreover, our results are dependent on the keywords we used to characterize the analyzed studies, which might not have captured information reflected by synonyms or other types of semantic variations. Because of this, some sources could have been omitted and, for this reason, our mapping should not be considered comprehensive.

Our results, nevertheless, can be used to motivate research on new uses of smart glasses in educational contexts. One of the future perspectives of this work is to extend this mapping study into a systematic revision of the literature. This would allow to deepen our understanding of the results described in the papers we considered. Ultimately, this would support a more focused identification of research opportunities beyond educational contexts.

References

1. Mann, S.: Wearable computing: a first step toward personal imaging. *Computer* **30**(2), 25–32 (1997)
2. Toh, P.K.: The new age of consumer wearables: internet of smart things (wearable computers) (2013)
3. Borthwick, A.C., Anderson, C.L., Finsness, E.S., Foulger, T.S.: Special article personal wearable technologies in education: value or villain? *J. Digit. Learn. Teach. Educ.* **31**(3), 85–92 (2015)
4. Rzayev, R., Hartl, S., Wittmann, V., Schwind, V., Henze, N.: Effects of position of real-time translation on AR glasses. In: Proceedings of the Conference on Mensch und Computer, pp. 251–257 (2020)
5. Rauschnabel, P.A., Ro, Y.K.: Augmented reality smart glasses: an investigation of technology acceptance drivers. *Int. J. Technol. Mark.* **11**(2), 123–148 (2016)
6. Hafsa, S., Majid, M.A.: Learnability factors for investigating the effectiveness of augmented reality smart glasses in smart campus. In: IOP Conference Series: Materials Science and Engineering, vol. 958, p. 012005. IOP Publishing (2020)
7. Dougherty, B., Badawy, S.M.: Using google glass in nonsurgical medical settings: systematic review. *JMIR mHealth uHealth* **5**(10), e159 (2017)
8. Yoon, J.W., Chen, R.E., Kim, E.J., Akinduro, O.O., Kerezoudis, P., Han, P.K., Si, P., Freeman, W.D., Diaz, R.J., Komotar, R.J., et al.: Augmented reality for the surgeon: systematic review. *Int. J. Med. Robot. Comput. Assist. Surg.* **14**(4), e1914 (2018)
9. Mitrasinovic, S., Camacho, E., Trivedi, N., Logan, J., Campbell, C., Zilinyi, R., Lieber, B., Bruce, E., Taylor, B., Martineau, D., et al.: Clinical and surgical applications of smart glasses. *Technol. Health Care* **23**(4), 381–401 (2015)
10. Badiali, G., Cercenelli, L., Battaglia, S., Marcelli, E., Marchetti, C., Ferrari, V., Cutolo, F.: Review on augmented reality in oral and crano-maxillofacial surgery: toward “surgery-specific” head-up displays. *IEEE Access* **8**, 59015–59028 (2020)

11. Kumar, N.M., Krishna, P.R., Pagadala, P.K., Kumar, N.S.: Use of smart glasses in education-a study. In: 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on, pp. 56–59. IEEE (2018)
12. Sapargaliyev, D.: Learning with wearable technologies: a case of google glass. In: International Conference on Mobile and Contextual Learning, pp. 343–350. Springer (2015)
13. 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)
14. Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., Khalil, M.: Lessons from applying the systematic literature review process within the software engineering domain. *J. Syst. Softw.* **80**(4), 571–583 (2007)
15. Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M.: Systematic mapping studies in software engineering. In: 12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12, pp. 1–10 (2008)
16. Petersen, K., Vakkalanka, S., Kuzniarz, L.: Guidelines for conducting systematic mapping studies in software engineering: an update. *Inf. Softw. Technol.* **64**, 1–18 (2015)
17. Aghaei Chadegani, A., Salehi, H., Yunus, M., Farhadi, H., Fooladi, M., Farhadi, M., Ale Ebrahim, N.: A comparison between two main academic literature collections: web of science and scopus databases. *Asian Soc. Sci.* **9**(5), 18–26 (2013)
18. Cañas, V., Urbieta, M., Rossi, G., Domínguez Mayo, F.: Challenges of migrating legacies web to mobile: a systematic literature review. *IEEE Lat. Am. Trans.* **18**(05), 861–873 (2020)
19. Silva, M., Freitas, D., Neto, E., Lins, C., Teichrieb, V., Teixeira, J.M.: Glassist: using augmented reality on google glass as an aid to classroom management. In: 2014 XVI Symposium on Virtual and Augmented Reality, pp. 37–44. IEEE (2014)
20. Kuhn, J., Lukowicz, P., Hirth, M., Poxrucker, A., Weppner, J., Younas, J.: gPhysics-using smart glasses for head-centered, context-aware learning in physics experiments. *IEEE Trans. Learn. Technol.* **9**(4), 304–317 (2016)
21. Weppner, J., Hirth, M., Kuhn, J., Lukowicz, P.: Physics education with google glass gPhysics experiment app. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 279–282 (2014)
22. Lukowicz, P., Poxrucker, A., Weppner, J., Bischke, B., Kuhn, J., Hirth, M.: Glass-physics: using google glass to support high school physics experiments. In: Proceedings of the 2015 ACM International Symposium on Wearable Computers, pp. 151–154 (2015)
23. Fun Man, F.: Exploring technology-enhanced learning using google glass to offer students a unique instructor's point of view live laboratory demonstration. *J. Chem. Educ.* **93**(12), 2117–2122 (2016)
24. Holstein, K., Hong, G., Tegene, M., McLaren, B.M., Aleven, V.: The classroom as a dashboard: co-designing wearable cognitive augmentation for k-12 teachers. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 79–88 (2018)
25. Hu, G., Chen, L., Okerlund, J., Shaer, O.: Exploring the use of google glass in wet laboratories. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2103–2108 (2015)

26. Oh, S., Park, K., Kwon, S., So, H.-J.: Designing a multi-user interactive simulation using AR glasses. In: Proceedings of the TEI 2016: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 539–544 (2016)
27. Scholl, P.M., Wille, M., Van Laerhoven, K.: Wearables in the wet lab: a laboratory system for capturing and guiding experiments. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 589–599 (2015)
28. Suarez, A., Ternier, S., Kalz, M., Specht, M.: Supporting inquiry-based learning with google glass (GPIM). *Interact. Des. Archit. J.-IxD&A* **24**, 100–110 (2015)
29. Zarraonandia, T., Díaz, P., Montero, Á., Aedo, I., Onorati, T.: Using a google glass-based classroom feedback system to improve students to teacher communication. *IEEE Access* **7**, 16837–16846 (2019)
30. Bazarov, S., Kholodilin, I.Y., Nesterov, A., Sokhina, A.: Applying augmented reality in practical classes for engineering students. In: IOP Conference Series: Earth and Environmental Science, vol. 87, p. 032004. IOP Publishing (2017)
31. Kommera, N., Kaleem, F., Harooni, S.M.S.: Smart augmented reality glasses in cybersecurity and forensic education. In: 2016 IEEE Conference on Intelligence and Security Informatics (ISI), pp. 279–281. IEEE (2016)
32. Dafoulas, G.A., Maia, C., Loomes, M.: Using optical head-mounted devices (OHMD) for provision of feedback in education. In: 2016 12th International Conference on Intelligent Environments (IE), pp. 159–162. IEEE (2016)
33. Berque, D.A., Newman, J.T.: Glassclass: exploring the design, implementation, and acceptance of google glass in the classroom. In: International Conference on Virtual, Augmented and Mixed Reality, pp. 243–250. Springer (2015)
34. Benninger, B.: Google glass, ultrasound and palpation: the anatomy teacher of the future? *Clin. Anat.* **28**(2), 152–155 (2015)
35. Sidiya, K., Alzanbagi, N., Bensenouci, A.: Google glass and apple watch will they become our learning tools?. In: 2015 12th Learning and Technology Conference, pp. 6–8. IEEE (2015)
36. Kocejko, T., Ruminski, J., Bujnowski, A., Wtorek, J.: The evaluation of eGlasses eye tracking module as an extension for scratch. In: 2016 9th International Conference on Human System Interactions (HSI), pp. 465–471. IEEE (2016)
37. Bermejo, C., Braud, T., Yang, J., Mirjafari, S., Shi, B., Xiao, Y., Hui, P.: Vimes: a wearable memory assistance system for automatic information retrieval. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 3191–3200 (2020)
38. Cao, Y., Tang, Y., Xie, Y.: A novel augmented reality guidance system for future informatization experimental teaching. In: 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 900–905. IEEE (2018)
39. Mason, M.: The MIT museum glassware prototype: visitor experience exploration for designing smart glasses. *J. Comput. Cult. Heritage (JOCCH)* **9**(3), 1–28 (2016)
40. Tanveer, M.I., Lin, E., Hoque, M.: Rhema: a real-time in-situ intelligent interface to help people with public speaking. In: Proceedings of the 20th International Conference on Intelligent User Interfaces, pp. 286–295 (2015)
41. Hobert, S., Schumann, M.: LearningGlasses app: a smart-glasses-based learning system for training procedural knowledge. In: European Conference on e-Learning, pp. 185–194. Academic Conferences International Limited (2017)

42. Somerkoski, B., Oliva, D., Tarkkanen, K., Luimula, M.: Digital learning environments-constructing augmented and virtual reality in fire safety. In: Proceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning, pp. 103–108 (2020)
43. Bai, H., Lee, G., Billinghurst, M.: Free-hand gesture interfaces for an augmented exhibition podium. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, pp. 182–186 (2015)
44. Wichrowski, M., Koržinek, D., Szklanny, K.: Google glass development in practice: Ux design sprint workshops. In: Proceedings of the Multimediala, Interaction, Design and Innovation, pp. 1–12 (2015)
45. Seok, A., Choi, Y.: A study on user experience evaluation of glasses-type wearable device with built-in bone conduction speaker: focus on the zungle panther. In: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video, pp. 203–208 (2018)
46. Hernandez, J., McDuff, D., Infante, C., Maes, P., Quigley, K., Picard, R.: Wearable ESM: differences in the experience sampling method across wearable devices. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 195–205 (2016)
47. Häkkilä, J., Vahabpour, F., Colley, A., Väyrynen, J., Koskela, T.: Design probes study on user perceptions of a smart glasses concept. In: Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia, pp. 223–233 (2015)
48. Rzayev, R., Woźniak, P.W., Dingler, T., Henze, N.: Reading on smart glasses: the effect of text position, presentation type and walking. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–9 (2018)
49. Koskela, T., Mazouzi, M., Alavesa, P., Pakanen, M., Minyaev, I., Paavola, E., Tulin-İemi, J.: Avatarex: telexistence system based on virtual avatars. In: Proceedings of the 9th Augmented Human International Conference, pp. 1–8 (2018)
50. Damian, I., Tan, C.S., Baur, T., Schöning, J., Luyten, K., André, E.: Augmenting social interactions: realtime behavioural feedback using social signal processing techniques. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 565–574 (2015)
51. Kosmalla, F., Daiber, F., Wiehr, F., Krüger, A.: ClimbVis: investigating in-situ visualizations for understanding climbing movements by demonstration. In: Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, pp. 270–279 (2017)
52. Hsieh, Y.-T., Jylhä, A., Orso, V., Gamberini, L., Jacucci, G.: Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4203–4215 (2016)
53. Al-Maroof, R.S., Alfaisal, A.M., Salloum, S.A.: Google glass adoption in the educational environment: a case study in the gulf area. Educ. Inf. Technol. 1–24 (2020)
54. Vlahovic, S., Mandurov, M., Suznjevic, M., Skorin-Kapov, L.: Usability assessment of a wearable video-communication system. In: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–6. IEEE (2020)
55. Vrellis, I., Delimitros, M., Chalki, P., Gaintatzis, P., Bellou, I., Mikropoulos, T.A.: Seeing the unseen: user experience and technology acceptance in augmented reality science literacy. In: 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT), pp. 333–337. IEEE (2020)

56. Rao, N., Zhang, L., Chu, S.L., Jurczyk, K., Candelora, C., Samantha, S., Kozlin, C.: Investigating the necessity of meaningful context anchoring in AR smart glasses interaction for everyday learning. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 427–432. IEEE (2020)
57. Xue, H., Sharma, P., Wild, F.: User satisfaction in augmented reality-based training using Microsoft HoloLens. *Computers* **8**(1), 9 (2019)
58. Hsu, C.-Y., Tung, Y.-C., Wang, H.-Y., Chyou, S., Lin, J.-W., Chen, M.Y.: Glass shooter: exploring first-person shooter game control with google glass. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 70–71 (2014)
59. Tung, Y.-C., Hsu, C.-Y., Wang, H.-Y., Chyou, S., Lin, J.-W., Wu, P.-J., Valstar, A., Chen, M.Y.: User-defined game input for smart glasses in public space. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3327–3336 (2015)
60. Wang, C.-Y., Chu, W.-C., Chiu, P.-T., Hsiu, M.-C., Chiang, Y.-H., Chen, M.Y.: PalmType: using palms as keyboards for smart glasses. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 153–160 (2015)
61. Pismag, J.K.V., Alawneh, H., Adam, C., Rawashdeh, S.A., Mitra, P., Chen, Y., Strumolo, G.: Augmented reality for improved dealership user experience. Technical report, SAE Technical Paper (2017)
62. Neumann, A., Strenge, B., Uhlich, J.C., Schlicher, K.D., Maier, G.W., Schalkwijk, L., Waßmuth, J., Essig, K., Schack, T.: Avikom: towards a mobile audiovisual cognitive assistance system for modern manufacturing and logistics. In: Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1–8 (2020)
63. Liu, C.-F., Chiang, P.-Y.: Smart glasses based intelligent trainer for factory new recruits. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 395–399 (2018)
64. Caria, M., Todde, G., Sara, G., Piras, M., Pazzona, A.: Performance and usability of smartglasses for augmented reality in precision livestock farming operations. *Appl. Sci.* **10**(7), 2318 (2020)
65. Mentler, T., Berndt, H., Herczeg, M.: Optical head-mounted displays for medical professionals: cognition-supporting human-computer interaction design. In: Proceedings of the European Conference on Cognitive Ergonomics, pp. 1–8 (2016)
66. Adenuga, K.I., Adenuga, R.O., Ziraba, A., Mbuh, P.E.: Healthcare augmentation: social adoption of augmented reality glasses in medicine. In: Proceedings of the 2019 8th International Conference on Software and Information Engineering, pp. 71–74 (2019)
67. Cajas, V., Urbieta, M., Rybarczyk, Y., Rossi, G., Guevara, C.: Portability approaches for business web applications to mobile devices: a systematic mapping. In: International Conference on Technology Trends, pp. 148–164. Springer (2018)



Multimodal Assistive Technology for the Support of Students with Multiple Disabilities

Valentim Realinho^{1,2}(✉) , Luís Baptista² , Rafael Dias², Daniel Marmelo²,
Paulo Páscoa³, and João Mourato³

¹ VALORIZA - Research Center for Endogenous Resources Valorization, Portalegre, Portugal
vrealinho@ipportalegre.pt

² Polytechnic Institute of Portalegre, Portalegre, Portugal

³ Escola Secundária Mouzinho da Silveira, Portalegre, Portugal

Abstract. This paper describes a multimodal assistive technology, called NJOY, to help students with multiple disabilities or with some type of special needs in terms of mobility, communication, and learning impairments. The basis of the entire system is an App that runs on a Tablet with a multimodal interaction system that includes a joystick that was developed for this purpose, a symbol-based communication system that aims to increase or compensate for communication, language, and learning difficulties with audio feedback. The system also integrates with a home automation system, which can be controlled using the same App and the same interaction mechanisms, and an SMS module that allows the read, through text-to-speech technology, of the received SMS, and send predefined SMS messages to predefined users. The system was tested by a sixteen-year-old student with degenerative disease of the central nervous system including severe motor, articulation, visual impairment, and limitations at a cognitive level, which greatly compromises its functionality, depending on adults to carry out the basic tasks of daily life. The results were very promising with the user being able to use the system without any major difficulties.

Keywords: Assistive technology · Multimodal interaction · Audio-visual speech

1 Introduction

The definition of multiple disabilities that has the greatest acceptance today is that of Orelove and Sobsey and the term applies to individuals with mental disabilities, with various associated motor and/or sensory disabilities, requiring specific health care [1]. The integration of individuals with multiple disabilities, especially among children or young people, becomes a challenge for those around them, namely, for parents, teachers, or friends, as it requires extra work in terms of their social interaction, learning, communication, among other problems. In this way, the interaction of the difficulties and needs of people with multiple disabilities currently represents one of the great challenges in terms of education.

The education of students with multiple disabilities requires highly specialized and permanent support to help them to satisfy the uniqueness of their needs, as well as to participate in daily activities and to learn [2]. The uniqueness of their needs requires different and distinct approaches to the concept of learning, levels of participation, and success. It is necessary to create environments that encourage their development and provide opportunities to be able to interact with the physical and social contexts in which they find themselves. They tend to learn better when they are actively involved in the learning process and are given tactile, visual, object clues (real or part of an object) and clear behavioral models associated with verbal clues [3]. Adaptive Learning is a process that provides an individualized learning experience with technologies designed to determine the learner's strengths and weaknesses [4]. The goal of an adaptive learning system is to personalize instructions to improve or accelerate the student's performance gain.

Students with multiple disabilities, often present serious difficulties in terms of language, being unable to efficiently use speech in the communicative process, such as to initiate and maintain interactions [5]. The use of Augmentative and Alternative Communication (AAC) systems facilitate interactions and guarantee them better social inclusion. AAC encompasses methods of communication for those with impairments or restrictions on the production or comprehension of spoken or written language [6]. AAC systems are extremely diverse and depend on the capabilities of the user. They may be as basic as pictures on a board that is used to request food, drink, or other care; or they can be advanced speech generating devices, based on speech synthesis, that are capable of storing hundreds of phrases and words. The use of text-to-speech and speech-to-text software can improve students' sight-reading and decoding abilities and allow the student to bypass the demands of typing or handwriting [15, 16]. Alexandros Pino [9] highlight the key features of an AAC system and give some practical hints for choosing a system. His study includes some of the today used AAC systems like the Aragonese Center of Augmentative and Alternative Communication (ARASAAC) [10], Blissymbolics [11], Makaton [12], Mulberry [13], Picture Communication Symbols [14] or Sclera NPO [15].

Computers and so-called Assistive Technology (AT) have shown an important horizon of new possibilities for autonomy and social inclusion for students with some type of disability or limitation. AT is any equipment or system that is used to increase, maintain, or improve the functional capabilities of a person with a disability. Karpov and Ronzhin [16] define the term "assistive information technology", which is special software and/or hardware that improves information accessibility and communication means for people with disabilities and special needs. Nowadays, there are devices and software such as trackballs, joysticks, gaze, and head tracking, screen readers, speech and haptic interfaces that facilitate access to computers and allows students with special needs to communicate and interact. It is important to note that a device that fits one particular person or problem may not fit another. For example, a device that works for one person with muscular dystrophy might not work for another person with muscular dystrophy. We may also highlight that these technologies provide possibilities of access to situations that most human beings take for granted, but students with more severe multiple disabilities do not.

The form of home automation called assistive domotics focuses on making it possible for elderly and disabled people to live independently. This field uses much of the same technology and equipment as home automation for security, entertainment, and energy conservation but tailors it towards elderly and disabled users. Bissoli et al. [17] present an assistive system called SMAD (System for Multimodal Assistive Domotics), through which a user with motor disabilities can control home devices from a wheelchair through biological signals captured on muscles and eyes.

Multimodal human-computer interaction refers to the interaction with the virtual and physical environment through natural modes of communication such as speech, body gestures, handwriting, graphics, or gaze [18]. Specifically, multimodal systems can offer a flexible, efficient, and usable environment allowing users to interact through input modalities, such as speech, handwriting, hand gesture, and gaze, and to receive information by the system through output modalities, such as speech synthesis, smart graphics, and other modalities, opportunely combined. Multimodal user interfaces provide several alternative ways of human-computer interaction at the same time, which the user can choose the most appropriate to communicate with information systems [16]. The decision for using a multimodal solution should be guided by usability criteria, such as adaptation to different environments due to user behaviours, shorter the learning curve and be more intuitive [19, 20]. Also, for users with some types of limitations, multimodal interfaces offer alternative and more natural ways of user interactions [21]. This is central in our system, which uses multimodal interaction to overcome multiple disabilities users' known difficulties. Due to user limitations on using traditional interfaces, like the mouse and the keyboard, our system has a new channel, an input joystick with separate input buttons, that overcomes those limitations. A multimodal system should replicate the process of human natural communication using proper hardware [22]. In this sense, our system improves communication with the user with multiple disabilities, which is a way of making the communication more natural, from the user's perspective.

This project aims to develop a multimodal interaction system that aims to support the teaching and learning process as well as the basic task of the daily life of students with special educational needs. Besides this introductory section, the rest of this paper is structured as follows. In Sect. 2, we present the NJOY Architecture. Section 3 presents the evaluation performed and Sect. 4 concludes the paper and outlines some future directions.

2 NJOY Architecture

We have developed a multimodal assistive technology, called NJOY, with a special focus for the support of students with multiple disabilities ranged from the motor to cognitive impairments that can compromise the basic tasks of daily life and/or learnability and communication. We start using a joystick as the primary input device for students with mobility impairments, although the system may support any other device, and have chosen ARASSAC [10] as the AAC system. Based on the review of the literature we have defined a set of requirements that have been conducted to the architecture of NJOY (Fig. 1) which consists of three interconnected systems, described next.

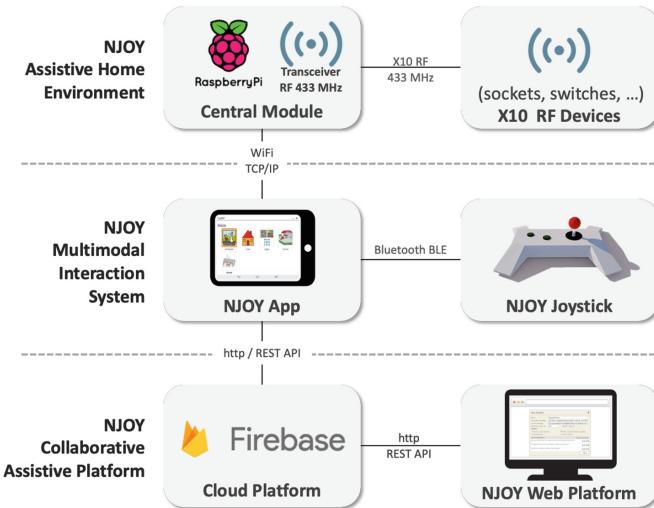


Fig. 1. NJOY architecture.

2.1 NJOY Multimodal Interaction System

The main component of the NJOY Multimodal Interaction System is the Android Tablet application (NJOY App) which can be controlled by a joystick (NJOY Joystick). The communication between these two modules is carried out via Bluetooth BLE using a communication protocol that was created for that purpose. It uses an augmentative and alternative communication system based on ARASAAC [10] with voice synthesis for audio feedback using text-to-speech technology. The NJOY App is a runtime capable of running applications built with the NJOY Collaborative Assistive Platform (see Sect. 2.2).

NJOY Joystick. The NJOY Joystick is the main interaction device used in the system, particularly for people with some motor impairment. It was developed using a Bluno Beetle [23] which is an Arduino [24] compatible Bluetooth 4.0 (BLE) hardware solution designed for smart App controlling, a small arcade joystick with four snap microswitches for directional control, and two 35 mm concave momentary push button similar to the ones used on arcade games. The push buttons are used for “Confirm” and “Back” actions. A small lithium battery was used to power all the hardware, which was assembled in a box made in a 3D printer. Figure 2a shows the 3D project of the box, while Fig. 2b shows the first prototype.

The Bluno Beetle was programmed to send to the NJOY App the information concerned with the state of the switches (joystick snap microswitches and buttons) according to the defined communication protocol, and it also implements some basic filtering to handle the switches bouncing issue.

NJOY App. The application can be controlled like a normal App with a touch-screen interface or through the NJOY Joystick. The basic layout of an NJOY application screen

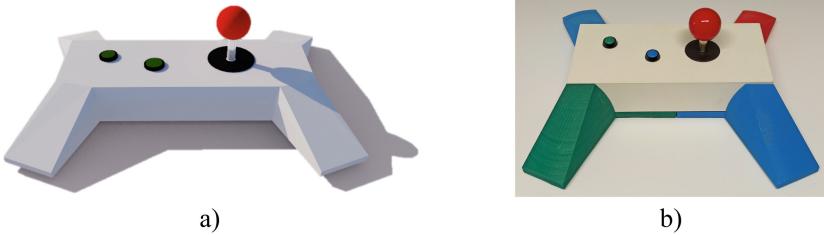


Fig. 2. The NJOY Joystick: (a) the 3D project, and (b) the first prototype.

is based on a 4×2 grid system (Fig. 3a). Each cell of the grid can contain an ARASAAC pictogram or any other image, and a text that describes the option. When using the joystick, options are highlighted while a box moves over items on the screen one after the other, until the user presses the “Confirm” button to select and execute the corresponding action. Every time the selection moves, the corresponding text is spoken using text-to-speech technology.

The behavior of the joystick can be adjusted or adapted to the limitations that the user presents (Fig. 3b). For a user with a small degree of limitation, a full direction movement may apply, but, on the other hand, for a user with more severe limitations like tremor or spasticity of the upper limbs and athetotic movements, the option to always go to the next option, may be more appropriate. To deal with false movements of the joystick and button clicks, we implemented an algorithm that uses timeouts, that can also be adapted to best fit the user limitations. We also have made experiments with machine learning algorithms to predict the movement of the user.

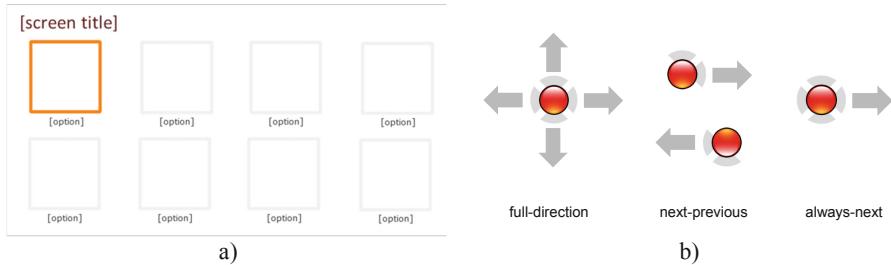


Fig. 3. (a) Layout grid of an application and, (b) transformation of joystick direction to screen scanning movements according to the defined strategies.

Each option can be assigned with an action that corresponds to a defined behavior of the application. We started with a relatively short set of actions (speech, navigate to another screen, run a quiz, view a YouTube video, control a particular home automation device or use SMS to communicate), but the system can be extended with more actions to create more expressive applications.

2.2 NJOY Collaborative Assistive Platform

The NJOY Collaborative Assistive Platform provides a visual environment that allows end-users, like teachers, educators, or parents, to create applications that can run on the NJOY App described above. These applications can be made available through the platform to other users. The idea is the construction of a community that creates and shares applications for several purposes. Examples are creating thematic content for learning (history, mathematics, or geometry, for example) and the support for more specific features that can facilitate some basic tasks of daily life with the integration of the assistive home environment or the SMS module. It uses the Firebase Authentication service as an authentication platform, Firebase Storage to store images, and Firebase Realtime Database as a NoSQL database to store all the information needed by the applications. Each application is stored as a JSON object and can be shared in the platform with other users.

2.3 NJOY Assistive Home Environment

The NJOY Assistive Home Environment is a home automation system specially adapted to be used with the modal interface described above. It provides the ability to control electrical devices and consists of a Raspberry Pi to which a 433 MHz RF transceiver has been connected. It allows wireless control of devices compatible with the X10 RF home automation protocol and is based on a previous work described in detail in [25].

3 Evaluation

The system was tested by a sixteen-year-old student with Pelizaeus-Merzbacher disease [15, 16] that presents tremor of the upper limbs, spasticity of the limbs, athetotic movements, and cognitive impairment, which greatly compromise its learnability and functionality, depending on adults to carry out the basic tasks of daily life. The student presents a 93.9% degree of disability and moves on an electric wheelchair controlled with a joystick. He's unable to use a mouse or keyboard on computers but uses a joystick daily in the wheelchair. For that reason, our first approach was to make a trial test that had the goal to understand more particularly the limitations caused by motor disabilities when using the NJOY Joystick. We have built an application that collects the joystick data. We ask the student to perform each of the four directional movements of the joystick and collected the data with the states of the four snap microswitches, while the user performed the movement.

Table 1 shows the confusion matrix which summarizes the real movement of the joystick when compared with the intended movement. The corresponding accuracy (the true movements divided by all movements) is 39,8%, which is a very poor result that can be explained by the existence of athetotic movements that compromises the possibility to correctly define the direction of the movement.

The second stage of our evaluation was conducted with a full prototype running a base application. We have installed the application on a tablet Samsung Galaxy Tab A and configure the domotic environment with three devices that can be turned on and off

Table 1. Confusion matrix of the trial test

		Real movement				
		down	up	right	left	
Intention		down	52%	0%	0%	48%
		up	6%	56%	1%	38%
		right	28%	0%	32%	40%
		left	19%	27%	16%	39%

(a television, a light, and a radio). As a result of the trial test, we have decided to use the joystick with an “always-next” strategy.

Figure 4 shows some screens of the application: Fig. 4a illustrates the initial screen of the application, Fig. 4b a quiz about Portugal history, Fig. 4c the SMS interface and, Fig. 4d the assistive home environment control.

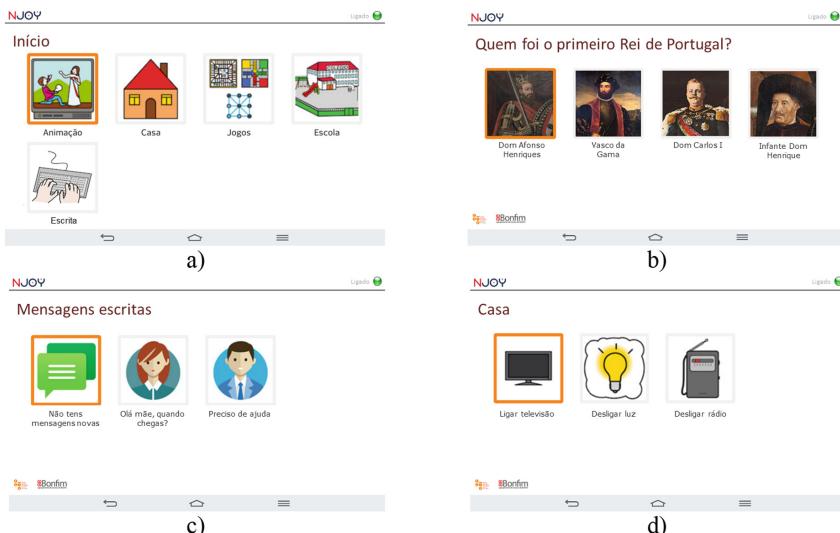


Fig. 4. (a) The main screen of the application, (b) a quiz about Portugal history, (c) the interface for SMS messages, and (d), the assistive home environment control.

This second test was conducted by the researchers with the precious help of the educators who normally support the student while he is at school, including special needs teachers. The student was briefed about the goals of the test and after a small training using the NJOY Joystick, the student was asked to perform some tasks while the researchers play the role of observer. During the tests, the researchers took notes on the student’s behavior while performing the tasks and made some adjustments to the configuration of the algorithms that deal with false movements of the joystick and button clicks.

Overall, the results were very promising with the student being able to use the system without any major difficulties. After this successful test, the student used the system several times in class context with the educators. Figure 5 shows the student using the system during the test and also in a class.

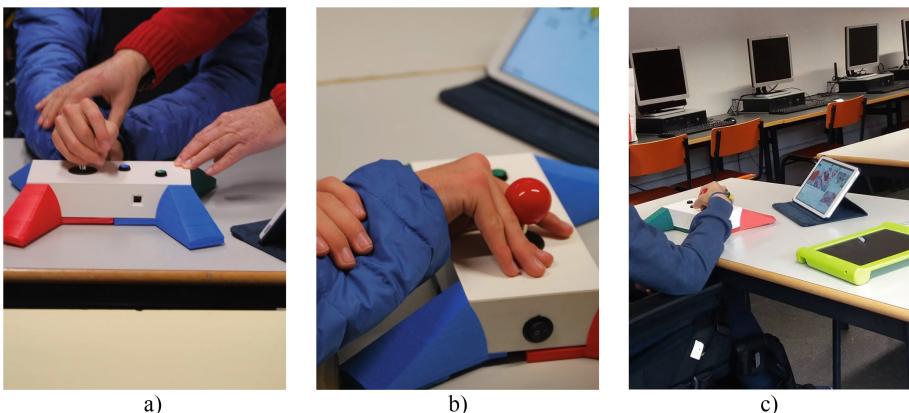


Fig. 5. The system being tested by a student with Pelizaeus-Merzbacher disease (a) and (b), and (c) the student using the system in a class.

4 Conclusions and Future Work

This paper describes a multimodal interaction system, called NJOY, designed to assist and support students with multiple disabilities or with some type of special needs in terms of mobility, communication, and learning. The system comprises three interconnected components: (i) the NJOY Multimodal Interaction System, (ii) the NJOY Collaborative Assistive Platform, and (iii) the NJOY Assistive Home Environment. The system uses an augmentative and alternative communication system (ARASAAC) that aims to increase or compensate for communication, language, and learning difficulties with audio feedback. The tests conducted with a student with Pelizaeus-Merzbacher disease had shown very promising results with the user being able to use the system without any major difficulties. As future work, we intend to test the system with students with other kinds of disabilities, and study/adapt the system to fit their particular needs.

Acknowledgements. This work was funded in part by Fundação Ilídio Pinho through the award Ciência na Escola, in which it won the 3rd national place. The authors would like to thank the valuable collaboration of students and teachers of the Polytechnic Institute of Portalegre and Escola Secundária Mouzinho da Silveira involved in this project.

References

- Orelove, F.P., Sobsey, D., Silberman, R.: *Educating Children with Multiple Disabilities: A Collaborative Approach*, 4th edn. Brookes Publishing Company, Baltimore (2004)

2. Mansell, J.: Raising our sights: services for adults with profound intellectual and multiple disabilities. *Tizard Learn. Disabil. Rev.* (2010). <https://doi.org/10.5042/tldr.2010.0399>
3. Downing, J.E., Eichinger, J.: Educating students with diverse strengths and needs together: rationale for inclusion. In: *Including Students with Severe and Multiple Disabilities in Typical Classrooms: Practical Strategies for Teachers*, p. 19. Paul H. Brookes Publishing Co., P.O. Box 10624, Baltimore, MD, pp. 21285–0624 (2008)
4. Sharma, N., Doherty, I., Dong, C.: Adaptive learning in medical education: the final piece of technology enhanced learning? *Ulster Med. J.* **86**, 198–200 (2017)
5. Harding, C., Lindsay, G., O'Brien, A., Dipper, L., Wright, J.: Implementing AAC with children with profound and multiple learning disabilities: a study in rationale underpinning intervention. *J. Res. Spec. Educ. Needs.* (2011). <https://doi.org/10.1111/j.1471-3802.2010.01184.x>
6. Fossett, B., Mirenda, P.: Augmentative and alternative communication. In: *Handbook of Developmental Disabilities*. The Guilford Press, New York (2007). <https://doi.org/10.1007/s12453-010-0015-0>
7. Strangman, N., Dalton, B.: Using technology to support struggling readers: a review of the research. In: Edyburn, D., Higgins, K., Boone, R. (eds.) *Handbook of Special Education: Research and Practice*, pp. 325–334 (2005). Whitefish Bay, Wis.: Knowledge by Design
8. Graham, S.: The role of text production skills in writing development: a special issue. *Learn. Disabil. Q.* (1999). <https://doi.org/10.2307/1511267>
9. Pino, A.: Augmentative and alternative communication systems for the motor disabled. In: *Disability Informatics and Web Accessibility for Motor Limitations*, pp. 105–152. IGI Global (2013). <https://doi.org/10.4018/978-1-4666-4442-7.ch004>.
10. Aragonese Center of Augmentative and Alternative Communication (ARASAAC). <https://arasaac.org/>. Accessed 5 Nov 2020
11. Bliss, C.K.: Semantography - Blissymbolics, 3rd enlarged edn. Semantography-Blissymbolics Publications, Sydney (1978)
12. Grove, N., Walker, M.: The makaton vocabulary: using manual signs and graphic symbols to develop interpersonal communication. *Augment. Altern. Commun.* **6**, 15–28 (1990). <https://doi.org/10.1080/07434619012331275284>
13. Mulberry Symbols. <https://mulberrysymbols.org/>. Accessed 18 Nov 2020
14. Mayer-Johnson: PCSTM Symbols - Picture Communication Symbols. <https://goboardmaker.com/>. Accessed 18 Nov 2020
15. Sclera NPO. <https://sclera.be>. Accessed 18 Nov 2020
16. Karpov, A., Ronzhin, A.: A universal assistive technology with multimodal input and multi-media output interfaces. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 369–378. Springer (2014). https://doi.org/10.1007/978-3-319-07437-5_35
17. Bissoli, A.L.C., Coelho, Y.L., Bastos-Filho, T.F.: A system for multimodal assistive domotics and augmentative and alternative communication. In: *ACM International Conference Proceeding Series*. Association for Computing Machinery (2016). <https://doi.org/10.1145/2910674.2910733>
18. Bourguet, M.-L.: Designing and prototyping multimodal commands. In: *Human-Computer Interaction (INTERACT 2003)*, pp. 717–720 (2003)
19. Martin, J.C., Veldman, R., Béroule, D.: Developing multimodal interfaces: a theoretical framework and guided propagation networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (1998). <https://doi.org/10.1007/bfb0052318>
20. Ferri, F., Paolozzi, S.: Analyzing multimodal interaction. In: Grifoni, P. (ed.) *Multimodal Human Computer Interaction and Pervasive Services*, pp. 19–33. IGI Global (2009). <https://doi.org/10.4018/978-1-60566-386-9.ch002>

21. Silva, S., Almeida, N., Pereira, C., Martins, A.I., Rosa, A.F., e Silva, M.O., Teixeira, A.: Design and development of multimodal applications: a vision on key issues and methods. In: Antona, M., Stephanidis, C. (eds.) Universal Access in Human-Computer Interaction. Access to Today's Technologies. UAHCI 2015. Lecture Notes in Computer Science, pp. 109–120. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20678-3_11
22. Rafael, S.: Multimodality, naturalness and transparency in affective computing for HCI. In: Marcus, A., Rosenzweig, E. (eds.) Design, User Experience, and Usability. Interaction Design. HCII 2020. Lecture Notes in Computer Science, vol. 12200, pp. 521–531. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49713-2_36
23. DFRobot: Bluno Series. <https://www.dfrobot.com/bluno.html>. Accessed 09 Nov 2020
24. Arduino. <https://www.arduino.cc>. Accessed 09 Nov 2020
25. Realinho, V.: Low cost domotic system based on open hardware and software. In: Proceeding of The Eighth International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services-oriented and Personalized Mechanisms, Technologies, and Services (CENTRIC 2015), pp. 13–16. IARIA XPS Press, Barcelona (2015). <https://doi.org/10.13140/RG.2.1.2061.7687>
26. Pelizaeus, F.: Ueber eine eigenthümliche Form spastischer Lähmung mit Cerebralerscheinungen auf hereditärer Grundlage. (Multiple Sklerose). Arch. Psychiatr. Nervenkr. **16**, 698–710 (1885). <https://doi.org/10.1007/BF02057569>
27. Merzbacher, L.: Eine eigenartige familiär-hereditäre erkrankungsform (Aplasia axialis extracorticalis congenita) (1910). <https://doi.org/10.1007/BF02893591>



Hunter-Gatherer Approach to Math Education - Everyday Mathematics in a San Community and Implications on Technology Design

Samuli Laato^{1(✉)}, Shemunyenge T. Hamukwaya², Laszlo Major¹,
and Shindume L. Hamukwaya³

¹ Department of Computing, University of Turku, Turku, Finland
{sadala,laszlo.major}@utu.fi

² Department of Mathematics and Statistics, University of Turku, Turku, Finland
shtaha@utu.fi

³ Department of Mining and Metallurgical Engineering, University of Namibia,
Ongwediva, Namibia
shamukwaya@unam.na

Abstract. K-12 math education is struggling as despite obvious job market benefits, several students choose to discontinue math education when given the possibility. At the same time, advances in learning technologies now enable modes of learning that were impossible up until ten years ago. In this study we analyse math education from scratch by returning to the hunter-gatherer time period and empirically observing how mathematics is present in the lives of a San tribe in Southern Namibia with ethnographic analysis. With this work, we propose two high level design considerations for integrating math learning technologies with the hunter-gatherer way of living: (1) integration of learning technologies and real world objects; and (2) the introduction of physical activity and social communication to math education. We discuss how these two design considerations could boost students' motivation to continue learning math also beyond the formal school environment.

Keywords: Mathematics · Education · Hunter-gatherer · Evolutionary psychology · San people

1 Introduction

Mathematics has been identified as one of the most important if not the most important school subject at primary and secondary schools in terms of students' future prospects in the labour market [1, 2]. It is the core subject in the STEM fields which governments have acknowledged creates jobs and which citizens need in their daily lives in industrialized societies [3]. At the same time, mathematics is perceived as boring or mundane by a significant proportion of students, and

when given the choice, many decide to discontinue studying it despite obvious benefits [2,4].

As mathematics is a subject that heavily builds on top of prior knowledge, perceived difficulty and lack of confidence in abilities have been shown to be the main reasons for students to discontinue their learning [5]. To counter this, educators could invoke strategies such as teaching learning to learn [6], which in the case of math can mean tolerating challenge and not being discouraged when facing difficult problems [5]. Formal mathematics education historically advances at a certain rate and students who fall behind have trouble catching up. Educational math games and technologies have been proposed as solutions to these problems. Here the pedagogical quality [7] and technical quality [8] are important, but more holistic design and ideas are needed to create effective learning solutions [9].

In this study, we take the evolutionary psychology approach to educational math technology design in that we focus on the hunter-gatherer way of living which the homo sapiens and our primate ancestors lived off for up to several millions years before the agricultural revolution some 10k years ago [10]. We gather empirical evidence from one of the last cultures on earth which followed the hunter-gatherer lifestyle: the San people in Northern Namibia. For this analysis, we formulate the following two research questions that guide our work:

RQ1: What mathematics can we observe in the daily lives of the hunter gatherers?

RQ2: What design considerations can we draw from the hunter gatherers for mathematics learning technologies?

2 Theoretical Background

2.1 Educational Math Technologies

The types of educational math technologies are numerous. There exists games, learning assistant tools, study diaries, technologies with visual, audio and force feedback, technologies with various sorts of input and so on. A single category of educational math games contains thousands or even tens of thousands of apps on popular market places [7]. Recent work has emphasized that it is important for math learning technologies to support students' deliberate practise [7,11] i.e. instead of routine drill-and-practise tasks, students should be pushed to deliberately improve their skills with tasks that require problem-solving, conceptualization, reflection and deliberate pushing for further development of skill [11].

As primary and secondary level math teachers must adapt their teaching to match the slowest learners, this inhibits the most talented students from advancing their mathematical skills as fast as otherwise possible. Educational technologies can assist in this situation by introducing personalized learning, allowing students to proceed each at their own pace. Another advantage of educational technology is that it can make learning fun and motivating [8]. Math learning technology can also provide modes of learning that were previously not possible.

2.2 Evolutionary Psychology and the Hunter-Gatherers

Evolutionary psychology has been suggested as a meta-theory for answering questions such as why humans play [12] and what kinds of activities appeal to humans [13]. For example, scholars have discussed how dormant territorial control instincts may be invoked by games [14,15] and why children universally like to practise fighting, climbing trees and taking care of babies through playing [13]. This way, evolutionary psychology can act as a theory that guides the design of educational math games.

One of the pioneers of observing the mathematics of hunter-gatherers is Peter Denny [16,17], who conducted research in particular among the Inuit, Cree and Ojibway hunters. Denny argues that mathematics has little use in hunter/gatherer societies [17]. The reasoning is that hunters adapt to the surrounding environment as they are dependent on wildlife for food. While agricultural and industrial societies depend on making changes to their environment to live, hunter-gatherers need specific knowledge about their environment but have no need or desire to change it. They know that any action that throws their environment off balance can lead to the loss of food. Basic mathematics such as counting can become useless in a way of living where every place and item is known by their name and characteristics [17]. It follows from here, that math is not an inherent skill in the same way as pattern recognition is, and instead, it needs to be learned [11]. To do this, education is needed, and here we need to investigate how educational technologies could serve the hunter-gathering mind of humans to make mathematics engaging, natural and fun. Previous work has investigated integrating math with almost all school subjects such as music [18] and computation [19], and findings from these studies invite research into looking at what new areas of life mathematical thinking and learning could potentially be combined with.

3 Empirical Study

3.1 Study Process and Participants

In this study as a primary method for uncovering the mathematics and opportunities for learning mathematics in the selected San community was ethnographic analysis [20]. We harnessed the knowledge of two informants who had lived in a San community in northern Namibia their entire childhood, as well as conducted specific on-sight observations in autumn 2018. We received permission from the locals to participate in the research and to publish photographs where they, or items they created appear. Interviews about the items and daily lives were conducted in Oshiwambo, the participants' native language. In our reporting we refer to the locals with pseudonyms.

Three San individuals who work as art producers were interviewed: designers Tom and Mika, and a musician Olavi. All three were male aged between 40 and 70. Their education levels were as follows. Tom finished only Grade 1; he cannot read nor write. By contrast, Mika and Olavi can read and write. We

interviewed the three of them about mathematics. None of the three perform any mathematical measurements when producing their artwork and do not see any need for it. Tom did not have knowledge about geometric shapes, while Mika was able to identify some letters and triangles on his artwork. They only make different shapes and designs for their products to make them look good and beautiful. Their artwork is a way for them to generate income to support their families. They sell products, such as knives, omaholo (cups), bowls, arrows, and spears, to local people at different prices depending on the size. Olavi only participates in music and dancing, and he sees an opportunity to make an income out of it for him in the future.



Fig. 1. Omaholo traditional cups (left) and an air blowing fire chamber (right)

3.2 What Mathematics We Observe in the Daily Lives of the San People

Figure 1 (left) depicts Omaholo traditional cups that are made from the Omupopo or Omupalala trees on the left. These omaholo are used for drinking Oshiwambo alcoholic beverage (e.g. omalunga, omagongo, ombike). We notice that the cup design features various geometric shapes. The omaholo themselves can be used for mathematical measurements. With regards to the inner part, it can be used to measure the volume of a liquid once we know the diameter and height of the cup. With regards to the outer part, the artistic decor enables the learning of geometric shapes and related mathematics. Despite this potential for learning mathematics, Tom did not identify any geometric shapes nor saw need for it. Mika could identify triangles, but did not develop this understanding further.

On the right side of Fig. 1 we see an air blowing chamber that the San people in the observed village use to light a fire. In the past, people used Oryx horns

instead of tubes (i.e. metal iron). At present, people use metal tubes, especially bicycle parts, because Oryx is now highly protected by the government. The chambers are made from the local tree Omukanga. Here, we can deduce mathematical shapes from the design, such as cubes and cylinders. Subsequently, we can find and calculate the volume and the length of the blowing tubes and the channels. In addition to mathematics, this fire creation chamber in particular serves as a construal [21] to teach physics. For example, the following can be observed: (1) The bigger the tubes, the higher the resulting air pressure will be; (2) If the tubes (metal iron) are long, the pressure will be low; (3) The handle stickers, which are circular in shape, are used for air volume control. The more accelerated the pumping or blowing of air, the more charcoal oxidizes in the iron casting; and (4) The length of the tubes mostly depends on the size of the blower's chamber.

Music instruments of the San people are displayed in Fig. 2. The longest instrument here is the bass and the shorter ones produce higher notes. Here we identify several types of mathematics as well. Again, we can observe cylinders and ovals, find out the volume of the musical instruments, observe that the diameter and length of the instruments determine the tones and sound that is produced and so on. In addition, the music that is produced with the instruments can be used to teach mathematics [1, 18]. For example, rhythm can be written down with decimal or fractional numbers, and several numerical representations also exist for pitch.



Fig. 2. These instruments are used to produce musical sound through a blowing technique. They are made from the roots of Onyu, a local shrub. Roots are dinged from the soil, and this process usually takes about three hours.

3.3 What Mathematics We Observe in the Hunting and Gathering Traditions of the San People

Knives play an important part in the culture of the San people. Knives are made by the people themselves, and are always in the shape of a leaf, as shown in Fig. 3.

This shape makes it easy to cut and penetrate into something. To ensure safety, the knife must be in its full compartment and always kept with the head up. The wooden part is made from local trees Omupopo or Omupalala, while the metal part is made from either an old panga knife or a knife razor, regarded as the best and strongest material. With regards to mathematics, the knife's body has an oval form. It has a 3-dimensional figure, with a head and a tail. The tail is almost diagonal in shape, while the head is circular. Here we observe that there are diagonals and other geometric shapes (e.g. triangles) in the way the knives are decorated.



Fig. 3. Traditional knives made by the San.

As the San people are among the final hunting-gathering cultures left on earth, we wanted to specifically focus on this aspect of their lives. Unfortunately, they have been forced to stop living this way due to restrictions on movement, laws forbidding hunting and territory claimed by landlords. Because of this, we had to interview the village elders specifically to uncover this lost knowledge. In Fig. 4 we see a hunting pouch and arrows that up until 1990's were used by the San to hunt wildlife. Today, only a single such pouch remains in the village. On the right of Fig. 4 we see a 3-level basket that was used to store gathered fruit and other goods and carried on the back. Today, these types of baskets are mostly used as decorations or sold to tourists.

Exemplar mathematics embedded into the baskets in Fig. 4 are as follows. Knowing the amount of raw material used for one of the baskets, we can estimate how much dry palm leaves was needed to make the other two baskets (assuming that the thickness of the different baskets is the same). This estimate is based on the fact that the area depends quadratically on the size: the area of a shape x times larger is x^2 times greater. This also holds for areas in three dimensions, namely the area of a spherical segment (the shape of our baskets) is proportional to the square of its diameter. Knowing the storage capacity of one of the basket



Fig. 4. A hunting pouch and arrows (left) and a traditional 3-level basket used for gathering (right)

we can estimate the capacity of the other two baskets. This estimate is based on that spatial volume depends cubically on the size: the volume of a shape x times larger is x^3 times greater. That is, in our case, the volume of a spherical segment is proportional to the cube of its diameter. After the calculations, we check our estimate experimentally: we measure the weight and the storage capacity of each basket.

An essential element of the traditional hunting strategy of the San people is the observation of animal tracks. Tracking is a complex and challenging task and it is not limited to following an animal from footprint to footprint, but its goal is to understand the movement and behavior of the animals by interpreting their tracks. As pointed out by Liebenberg [22], tracking was the earliest manifestation of scientific thinking in human history. The tracks can be considered as symbols that contain implicit information about animals. Reading and understanding these symbols, as practiced by hunters of the San communities, is indeed a science that requires the similar intellectual abilities as modern mathematics. It is therefore reasonable to link the learning of mathematics with tracking experiences. Understanding the similarity between reading the tracks in the sand and reading the symbolic language of mathematics can help students develop a positive attitude towards mathematics.

4 Discussion

4.1 Design Consideration for Math Technologies

Based on our observations as well as previous work [17], mathematics is not something that hunters-gatherer societies naturally develop. The question arises that as all mathematics beyond simple counting is learned, how can we motivate students to learn mathematics and think about the world in a more mathematical

way. Via embedding mathematics into the everyday lives, it can become a natural part of thinking that enables further development of mathematical skills [23]. However, it is arguably not only the hunter-gatherers that benefit from embedding math into their daily lives. As hunter-gatherer societies manifest many of the natural tendencies that make humans happy [10], these societies can be seen as inspiration for the development of educational math technologies. Here we specify two fruitful design considerations that arise from our observations.

Use of Everyday Items as Construals for Learning Mathematics. We observed that several opportunities exist within the everyday lives of the San people for teaching and learning mathematics, both through everyday items but also through culture. Yet, this learning process is not automatic, and teaching is required to draw out the mathematics that is associated with the many observed items and habits. For technology designers this is a major challenge. Designers need to break free from the constraints of creating software and applications and move to create construals, tools to learn mathematics with.

Include More Exertion and Social Communication into the Learning of Math. Exertion and social communication are central to human behavior and something that modern societies lack. While doing mathematics is mostly working with pen and paper, there is no reason why mathematical thinking would not be introduced to exercise and social situations. In fact, dancing is a good example of a social and physical activity that also includes mathematics in the form of counting steps and keeping up with the rhythm. In our empirical work we showed that the musical instruments have a lot of music embedded into them, not to mention the music itself that is played [1]. Other technologies such as location-based apps [15] could also be utilized to combine exercise, social interaction and mathematics.

4.2 Design Considerations for the Namibian School System

The findings suggest possibilities of promoting the use of mathematical knowledge of hunter-gatherers in the Namibian school curriculum, especially for cultural confidence [24], and contributes to the understanding of African culture in math learning technologies. The integration of local artwork in school subjects may create confidence, and meaningful learning of mathematics among students, especially those who perceive mathematics as a difficult subject. Furthermore, integration of an important school subject, mathematics, to the local culture also serves to preserve the culture, as it gains more value through its association to mathematics.

4.3 Limitations and Future Work

Due to the large scope of our approach and the study topic, the resulting design considerations remain at a general level. Our findings show promising research

directions which future work could explore further. Especially embedded math learning technologies and technologies supporting everyday mathematical thinking are worth investigating further. Furthermore, this work opens the research avenue on how the hunter-gatherer tendencies of humans could be utilized to boost the learning of math and learning in general. One of the criticisms towards our approach is that it might not be as cost or time-efficient as currently favored modes of teaching. Also while mathematical thinking outside the classroom setting is important, it remains unclear what is the best way to achieve this goal.

5 Conclusions

In this work we showed that geometry, symmetry, combinatorics and other kinds of mathematics are present ubiquitously in the San community and other human societies, but without additional teaching and support, humans do not learn to think about their everyday objects and activities in mathematical terms. Here we proposed two ways in which we can re-imagine mathematics education from the perspective of observations in the hunter-gatherer society: (1) the use of props and real world objects as construals through which math can be learned; and (2) introducing physical activity and social communication more prominently into math education. The main benefit of this kind of an approach is that mathematics will be integrated into the daily lives more holistically. While pen and paper are still useful tools for learning and doing mathematics, seeing mathematics in the everyday world can guide thinking and lead to further development of mathematical knowledge.

References

1. Laato, S., Shivoro, R., Pope, N., Gideon, F., Sutinen, E.: Identifying factors for integrating math and music education at primary schools in Namibia. In: 2019 IEEE International Conference on Engineering, Technology and Education (TALE), pp. 1–8. IEEE (2019)
2. Pursiainen, J., Rusanen, J., Partanen, S.: Lukion tärkein aineevalinta. *Dimensio* **4**(2016), 21–24 (2016)
3. Reyna, V.F., Brainerd, C.J.: The importance of mathematics in health and human judgment: numeracy, risk communication, and medical decision making. *Learn. Individ. Differ.* **17**(2), 147–159 (2007)
4. Kislenko, K., Grevholm, B., Lepik, M.: Mathematics is important but boring: students' beliefs and attitudes towards mathematics. In: Nordic Conference on Mathematics Education: 02/09/2005–06/09/2005, pp. 349–360. Tapir Academic Press (2007)
5. Brown, M., Brown, P., Bibby, T.: “i would rather die”: reasons given by 16-year-olds for not continuing their study of mathematics. *Res. Math. Educ.* **10**(1), 3–18 (2008)
6. Novak, E., Tassell, J.: Video games that improve ‘learning to learn’: focus on action video game play elements. In: 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), pp. 142–144. IEEE (2017)

7. Laato, S., Lindberg, R., Laine, T.H., Bui, P., Brezovszky, B., Koivunen, L., De Troyer, O., Lehtinen, E.: Evaluation of the pedagogical quality of mobile math games in app marketplaces. In: 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–8. IEEE (2020)
8. Bui, P., Rodríguez-Aflecht, G., Brezovszky, B., Hannula-Sormunen, M.M., Laato, S., Lehtinen, E.: Understanding students' game experiences throughout the developmental process of the number navigation game. *Educ. Technol. Res. Dev.* **68**, 2395–2421 (2020)
9. Kiili, K.: Digital game-based learning: towards an experiential gaming model. *Internet High. Educ.* **8**(1), 13–24 (2005)
10. Harari, Y.N.: *Sapiens: A Brief History of Humankind*. Random House, New York (2014)
11. Lehtinen, E., Hannula-Sormunen, M., McMullen, J., Gruber, H.: Cultivating mathematical skills: from drill-and-practice to deliberate practice. *ZDM* **49**(4), 625–636 (2017)
12. Mendenhall, Z., Saad, G., Nepomuceno, M.V.: *Homo virtualensis: evolutionary psychology as a tool for studying video games*. In: *Evolutionary Psychology and Information Systems Research*, pp. 305–328. Springer (2010)
13. Pellegrini, A.D., Smith, P.K.: *The Nature of Play: Great Apes and Humans*. Guilford Press, New York (2005)
14. Laato, S., Kordyaka, B., Najmul Islam, A.K.M., Papangelis, K.: Landlords of the digital world: how territoriality and social identity predict playing intensity in location-based games. In: *Proceedings of the 54th Hawaii International Conference on System Sciences*, pp. 744 (2021)
15. Papangelis, K., Chamberlain, A., Lykourentzou, I., Khan, V.-J., Saker, M., Liang, H.-N., Sadien, I., Cao, T.: Performing the digital self: Understanding location-based social networking, territory, space, and identity in the city. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **27**(1), 1–26 (2020)
16. Peter Denny, J.: Context in the assessment of mathematical concepts from hunting societies. In: *Human Assessment and Cultural Factors*, pp. 155–161. Springer (1983)
17. Peter Denny, J.: Cultural ecology of mathematics: ojibway and inuit hunters. In: *Native American Mathematics*, pp. 129–180 (1986)
18. Laato, S., Laine, T., Sutinen, E.: Affordances of music composing software for learning mathematics at primary schools. *Res. Learn. Technol.* **27** (2019)
19. Laato, S., Rauti, S., Sutinen, E.: The role of music in 21st century education—comparing programming and music composing. In: 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT), pp. 269–273. IEEE (2020)
20. Hammersley, M.: Ethnography. In: *The Blackwell Encyclopedia of Sociology* (2007)
21. Harfield, A., Alimisi, R., Tomcsányi, P., Pope, N., Beynon, M.: Constructionism as making construals: first steps with JS-Eden in the classroom. *Proc. Constructionism* **2016**, 42–52 (2016)
22. Liebenberg, L.: *A Field Guide to the Animal Tracks of Southern Africa*. New Africa Books, Cape Town (1990)
23. D'ambrosio, U.: Multiculturalism and mathematics education. *Int. J. Math. Educ. Sci. Technol.* **26**(3), 337–346 (1995)
24. Gerdes, P.: On mathematics in the history of Sub-Saharan Africa. *Historia Mathematica* **21**(3), 345–376 (1994)



Emotions and Intelligent Tutors

Rámon Toala^{1,2} , Dalila Durães¹ , and Paulo Novais¹

¹ Algoritmi Research Centre, Department of Informatics, University of Minho, Braga, Portugal
id7410@alunos.uminho.pt, dalila.duraes@algoritmi.uminho.pt,
pjon@di.uminho.pt

² Technical University of Manabí, Portoviejo, Manabí, Ecuador

Abstract. In the last year, schools, universities, teachers, and students had to adapt to distance learning because of the pandemic situation. The e-learning situation is very different from a face-to-face situation. One of the problems of the e-learning is that the students should be more responsible to not distracted. Another problem is that the type of computer for each student varies significantly. Finally, the traditional interaction between teacher, student and content is made more complicated by introducing technology. When new tools are applied, and there is an improvement in e-learning education, student, teacher, and educational institutions benefit from it. Emotion plays an essential role in the knowledge, acquisition, and decision process of an individual. There is also significant evidence that rational learning in humans is dependent on emotions. In this paper, we presented a solution with an Intelligent Tutor Application, that analyzed emotions in a non-intrusive and non-invasive way.

Keywords: Intelligent Tutors · User emotions · E-Learning

1 Introduction

Due to the pandemic situation, teaching has become mixed, that is, face-to-face and online. If the student stays at home, schools need to have online classes. However, the traditional class where the teacher gives in a face-to-face situation does not work in an online situation [1]. A system is needed to engage the student in a better way of learning [2]. Hence, several technologies have been applied to maintain and promote learning [2, 3]. One example is the Intelligent Tutors, which are training software systems that use intelligent technologies to offer personalized systems, ambient learning, and content learning to students, depending on their characteristics and behaviour [4].

An Intelligent Tutor's goal is to make these technologies adaptable to users, based on their characteristics and needs [4]. Thus, an Intelligent Tutor provides individual benefits of automatic and autonomous tutoring, making each user progress at their own pace. To these systems, it is necessary to apply the concept of adaptive and interactive learning to make it a powerful learning tool [5].

For students, better learning, one factor to be considered is emotion [6]. So, emotion plays an essential role in the process of knowledge acquisition and an individual's decision. In this way, emotion directly influences perception, learning process and way

people communicate. Consequently, several theories attempted to specify the interrelationships of all the components involving emotion and the causes, the reasons, and the function of an emotional response. Some research intended to relate emotion and computer, so for Ortony, Clore and Collins [7] emotion identification is generally used in cognitive science and connected to affective computing enabling computers to recognize and express emotions.

In this paper, we conducted an experience with Intelligent Tutors, applying a type of emotion in the students. This paper is organized as follows. Firstly, Sect. 2 introduces the concepts with state of the art, namely student behaviour and emotions. Then, Sect. 3 presents a proposal architecture. Next, Sect. 4 presents result and discussion. Finally, Sect. 5 concludes by performing a global conclusion and some future work.

2 State of Art

Currently, there are several types of Intelligent Tutors. However, these tutors have not entirely achieved the desired objectives, since they are either autonomous or adaptable, but not both. Besides, they do not consider an essential element that affects users' learning in real-time: their emotional state. Some of these tutors assess the user's emotional state only at the end of the work sessions, which is not enough to improve the learning environment [4, 8].

One of ITS's main applications involves personalized guidance, adaptation to learning materials, analysis of students' learning styles, and applying various techniques to support a harmonious teaching process [8].

A definition of ITS is “a new generation of a learning system that offers individualized instruction” one to one “stimulating the activities of a human teacher, like a teacher for a student” [9]. There are different architectures proposed for IT'S. However, according to [10–12], the traditional architecture of an ITS is illustrated in Fig. 1.

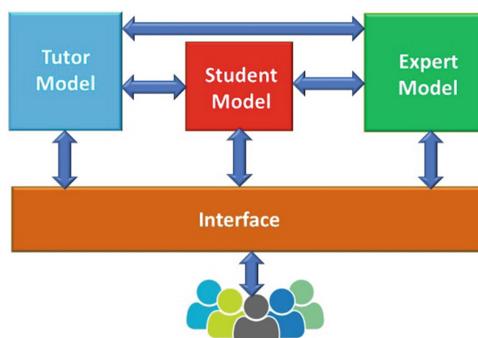


Fig. 1. The general structure of an Intelligent Tutor.

- Expert module: the expert module is defined as the function of harnessing knowledge about a specific topic that must be taught or learned. The expert module is responsible for generating and storing knowledge on a given subject.

- Student module: The student module consolidates students' fundamental data on their learning progress, conduct, and mental attributes. It is additionally responsible for processing and storing accumulated data about students.
- Tutor module: The tutor module, also known as the pedagogical module, determines the learning and tutoring strategies. In addition to updating the exhibition procedure, this module is responsible for keeping pedagogical knowledge.
- Interface module: This module provides interaction between the system and students through various input/output devices.

2.1 Previous Work

Part of the framework presented in this paper was implemented in previous work. This first version focused on the general framework for an intelligent tutor, which includes behaviour biometrics like mouse velocity or acceleration, click duration, etc. For a complete list of features and the process of their acquisition and extraction, please see [4].

While this early work focused on the monitored attention [4, 13] and student model [4] from the analysis of Human-Computer Interaction, we also found out that people tend to interact differently with different applications, and different contexts. For example, although both tasks involve typing, people tend to type differently if they are in a messaging application and a word processing application [14].

The present work adds a new feature, and a new framework is present. It provides a completed framework of intelligent tutors, analyzed the user emotions states, the behaviour biometrics, and the user student style. It thus constitutes a much more precise and reliable mechanism for intelligent tutors, while maintaining all the advantages of the existing system: non-intrusive, lightweight, and transparent.

2.2 Affective Computing

According to [15], affective computing's evolution is related to the need to put computers interacting, thinking, receiving, and transmitting people's personalities. Picard and Hassin [15, 16], highlight affective computing as a research area, which explores how computer systems can identify, classify, and prove human personality.

Affective computing can increase conflict management capabilities with the customer and increase the efficiency of recommendation systems. Affective computing is about: (a) understanding how emotions play vital roles in persons; (b) regulating our intention; (c) helping people make right decisions; and (d) changing the way we emphasize and prioritize things. Consequently, it is possible to build a personalized computer system to perceive and interpret the human being's feelings, providing intelligent, sensitive and adapted responses to situations [17].

3 Architecture

Based on the state-of-the-art section, the idea is to create an Intelligent Tutor adapted to each student. In this first phase, an Intelligent Tutor's general structure was developed, which is shown in Fig. 2.

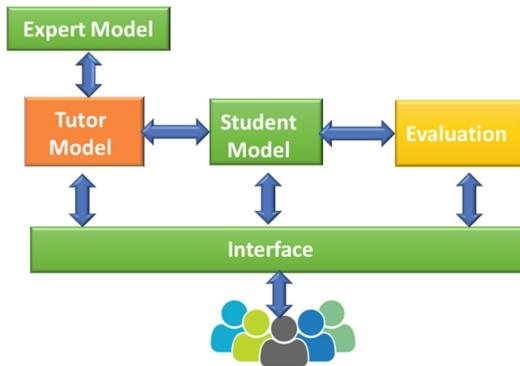


Fig. 2. The general structure of our Intelligent Tutor.

The idea is to have an interface that communicates directly with users and captures the data necessary to create a student profile. The system, based on the content it must address, and the student profile apply the student's tools to acquire the necessary knowledge. Another function of the interface module is to show the reports to the students or teachers and the students' formatted presentation. Making a complete description of the system, Fig. 3 presents an Intelligent Tutor's framework in more detail.

The interface module is the front-end interaction of the Intelligent Tutor. This system integrates all types of information necessary to interact with the user, through graphics, text, multimedia, video, menus. The interface module is the Intelligent Tutor's communication component that controls the user and the system's interaction. He captures data from the user's interaction with the Intelligent Tutor. Data capture is done using a non-invasive and non-intrusive approach. A log application runs in the background, saving the user's necessary events with Intelligent Tutor. This application has a device that generates raw data that describes the user's interaction with the system: mouse, keyboard, and activity. Flexible sensors use the information available from other measurements and process parameters to calculate and estimate the amount of raw data. The raw data generated is stored locally until it is synchronized with the web server in the cloud at regular intervals, usually every 5 min. The interface module relates to the tutor module to receive the exercises. Then the interface module sent the exercise and the solution to the tutor module. Finally, the tutor module gives feedback if the answer is correct or give suggestions and the solutions if the answer is incorrect. The interface module is also connected with the student module for exchange the login information of the student. The interface module and the evaluation module's connection is to exchange the usage reports and the student learning outcomes.

The tutor module accepts information from the student module, the interface module, and the expert module. The tutor module has the definitions of the problem solver strategies, the exercises corrector, the definitions of the pedagogical approach, the learning instructor, and the predictions knowledge. The tutor module is connected to the expert model, and they exchange the declarative, procedural, and conditional knowledge. It is closely linked to the student module since it uses knowledge about student learning

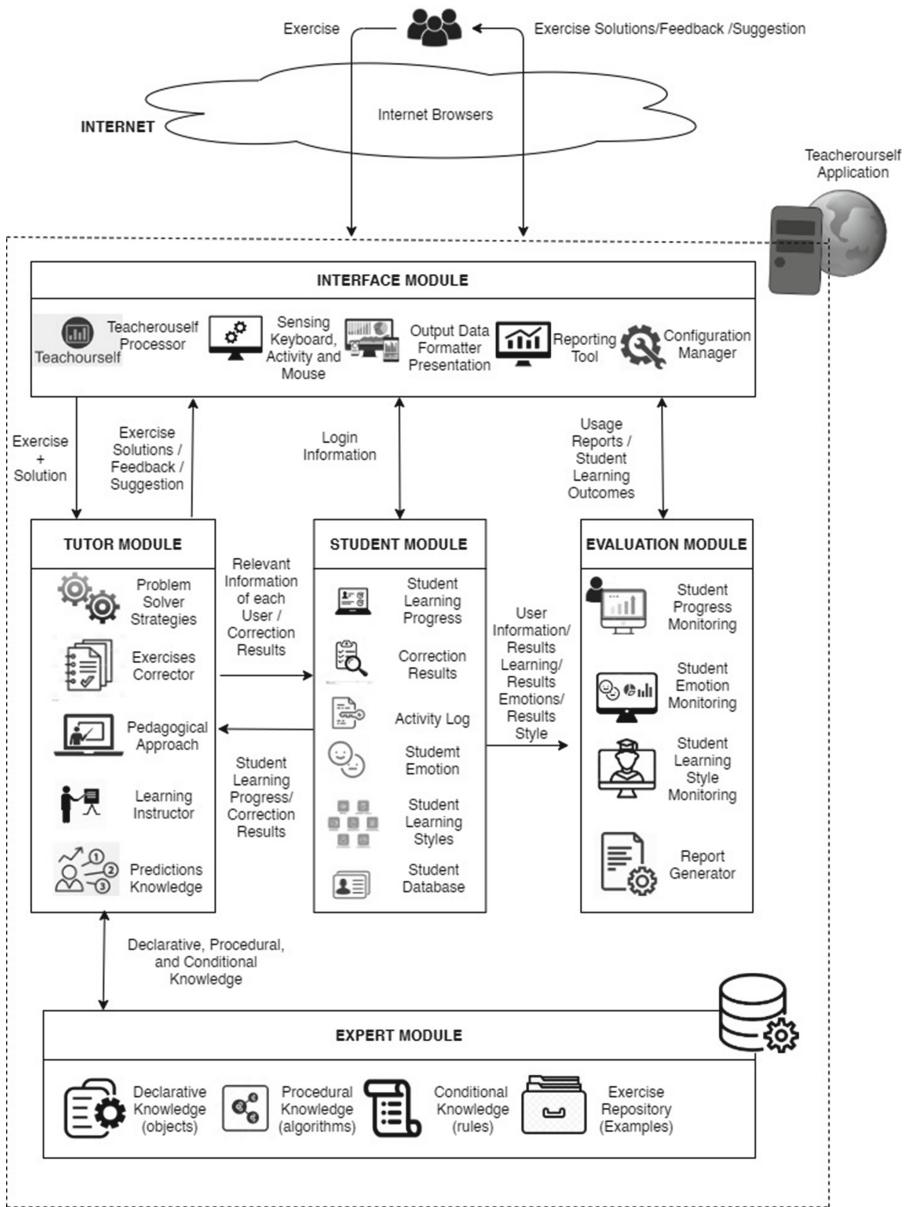


Fig. 3. Framework of an Intelligent Tutor.

progress and correction results. The tutor module gives The student's relevant information and the correction results at the student module. It also monitors user progress, creating a profile of strengths and weaknesses concerning production rules.

The student module has the student learning progress, the corrections results, the activity log, the student emotion, and the student learning style stored in the student

database. The student module also gives information at the evaluation module related to the user information, learning results, emotions results, and learning style results.

The evaluation module has the student progress monitoring, the student emotion monitoring, the student learning style monitoring, and generator the necessary reports.

The expert module contains the declarative knowledge (objects), procedural knowledge (algorithms), the conditional knowledge (rules), and the exercises repository (with examples).

4 Methodology and Methods

We have created several contents for a Statical Subject at the Technical University of Manabí, Portoviejo, Manabí Ecuador to implement the proposed system.

4.1 Population

Based on the framework presented in Sect. 3, we have created an Intelligent Tutor Application presented in Fig. 4. This application will be applied to 160 students in two phases with the same conditions. The first phase is a regular assessment with several levels, where the student did not have time count for each question. The second phase has a time count for each question.



Fig. 4. Intelligent Tutor application.

When the student opens the application for the first time, we need to register. It is necessary to indicate the following data: date of birth, gender; course, and address. The system creates a student profile from this data. When the student makes a test, we can indicate the test's difficulty level to be performed. The data capture is based on the mouse's dynamics and the keyboard to propose an entirely non-intrusive method for

evaluating student-computer interaction. Each computer has a keyboard, a mouse, and a monitor. The assessment activity starts simultaneously for all students, they log in to the standard software using their credentials, and the activity begins.

5 Results and Discussion

Figure 5 present a comparison of the real answers for each assessment. The first assessment without time count and the second assessment with time count. We can observe that in the phase without time count, the correct answer is 34% and in the phase with time count, it is 35%. It seems that the stress caused by time will not affect the results.

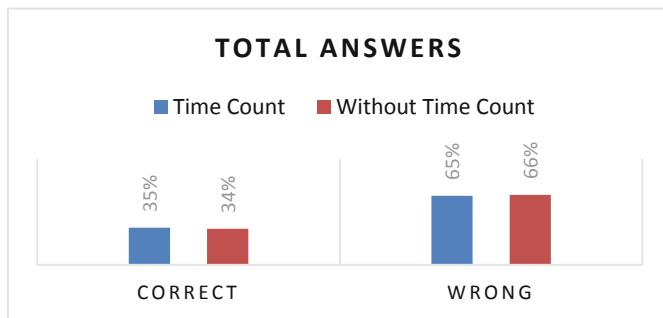


Fig. 5. Total student answers.

However, the number which the backspace keyword is press in total is 162 times when the assessment has time count, and zero without time count. Also, the number of keys pressed with time count is 47891, and 12938 without time count. It seems that the students intend to give a complete answer when the assessment has a time count (Fig. 6).

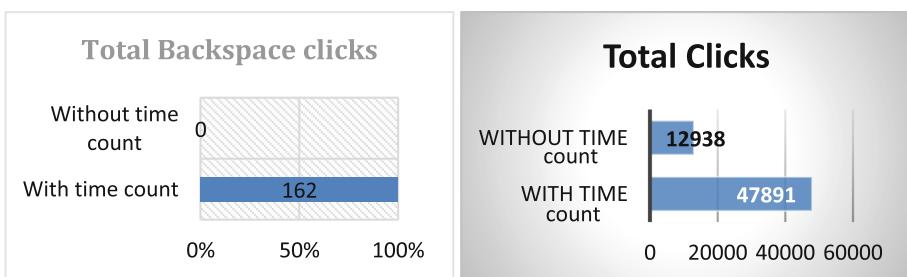


Fig. 6. Total backspace and total clicks.

Emotion is present on the abscissa axis. For level 1, emotion is very negative for level 2, negative emotion, for level 3, neutral emotion, level 4, positive emotion, and level 5, very positive emotion. Figure 7 shows that 26% has neutral emotions, 23% has positive emotion, and 11% is very positive emotion. On the other hand, 20% has negative and very negative emotion, respectively.



Fig. 7. Total backspace and total clicks.

6 Conclusions and Future Work

This paper presents an approach to Intelligent Tutoring. The approach is non-invasive and non-intrusive. It is proposed based on the biometric analysis of work behaviour in different students applying a stress emotion. The system monitors and analyzes the mouse and keyboard dynamics to determine the student's interaction with the computer. These results are crucial to improve learning systems in an e-learning environment and to predict student behaviour based on their interaction with mobile devices or the computer.

This Intelligent Tutor makes possible the enhanced learning/teaching processes. The architecture of an ambient intelligent learning environment is proposed to address these issues, especially to monitor the students' emotion students in distance learning. With this architecture, it is possible to detect those factors dynamically and non-intrusively, making it possible to foresee negative situations and mitigate them.

In future work, the door is then open to analyze students' emotion profile, consider their comments and propose new strategies and actions, minimizing issues such as stress, which can influence students' results and are closely related to the abandonment occurrence. Moreover, we intend to inform the teacher about the emotion of each student. Another improvement in this work is making correlations with two types of the questionnaire (before and after the tests).

Acknowledgement. This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

1. Rodrigues, M., Novais, P., Santos, M.F.: Future challenges in intelligent tutoring systems: a framework (2005)

2. Carneiro, D., Pimenta, A., Gonçalves, S., Neves, J., Novais, P.: Monitoring and improving performance in human–computer interaction. *Concurrency Comput.: Pract. Experience* **28**(4), 1291–1309 (2016)
3. Carneiro, D., Novais, P., Durães, D., Pego, J.M., Sousa, N.: Predicting completion time in high-stakes exams. *Futur. Gener. Comput. Syst.* **92**, 549–559 (2019)
4. Durães, D., Toala, R., Gonçalves, F., Novais, P.: Intelligent tutoring system to improve learning outcomes. *AI Commun.* **32**(3), 161–174 (2019)
5. Brusilovsky, P.: From adaptive hypermedia to the adaptive web (invited talk). In: Proceedings of Mensch Computer, pp. 21–24 (2003)
6. Rincon, J.A., Julian, V., Carrascosa, C., Costa, A., Novais, P.: Detecting emotions through non-invasive wearables. *Logic J. IGPL* **26**(6), 605–617 (2018)
7. Ortony, A., Clore, G.L., Collins, A.: *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge (1990)
8. Hasan, M.A., Noor, N.F.M., Rahman, S.S.A., Rahman, M.M.: The transition from intelligent to affective tutoring system: a review and open Issues. *IEEE Access* **8**, 204612–204638 (2020)
9. Hooshyar, D., Ahmad, R.B., Yousefi, M., Fathi, M., Horng, S.J., Lim, H.: Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. *Comput. Educ.* **94**, 18–36 (2016)
10. Petrovica, S., Anohina-Naumeca, A., Ekenel, H.K.: Emotion recognition in affective tutoring systems: collection of ground-truth data. *Proc. Comput. Sci.* **104**, 437–444 (2017)
11. Ramírez-Noriega, A., Juárez-Ramírez, R., Martínez-Ramírez, Y.: Evaluation module based on Bayesian networks to intelligent tutoring systems. *Int. J. Inf. Manage.* **37**(1), 1488–1498 (2017)
12. Victorio-Meza, H., Mejía-Lavalle, M., Ortiz, G.R.: Advances on knowledge representation of intelligent tutoring systems. In: 2014 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE), pp. 212–216. IEEE Computer Society, November 2014
13. Durães, D., Jiménez, A., Bajo, J., Novais, P.: Monitoring level attention approach in learning activities. In: *Methodologies and Intelligent Systems for Technology Enhanced Learning*, pp. 33–40. Springer, Cham (2016)
14. Durães, D., Carneiro, D., Jiménez, A., Novais, P.: Characterizing attentive behavior in intelligent environments. *Neurocomputing* **272**, 46–54 (2018)
15. Hassin, M.H.M., Aziz, A.A., Norwawi, N.M.: Affective computing: knowing how you feel. In: *The National Seminar of Science Technology and Social Science (STSS 2004)*, UiTM Pahang (2004)
16. Picard, R.W.: Affective computing for HCI. In: *HCI*, no. 1, pp. 829–833, August 1999
17. Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Strohecker, C.: Affective learning—a manifesto. *BT Technol. J.* **22**(4), 253–269 (2004)

Health Informatics



Factors affecting the Usage of e-Health Services in Kuwait

Issam A. R. Moghrabi^(✉) and Manal H. A-Farsi

College of Business Administration, Gulf University for Science and Technology, Mubarak Al-Abdullah, Kuwait
moughrabi.i@gust.edu.kw

Abstract. The increasing quality of ICT seems to be paralleled with an escalating complexity and scope of e-Health processes. E-Health, which it is normally perceived as a component of e-government, offers an opportunity for hospitals and medical centers to integrate their services and communicate standardized medical data intended for easy sharing of patient medical records electronically. Such services are expected to deliver cost-effective and high quality health care. This research aims to investigate the factors leading to enhancing and adopting e-health in Kuwait. Such factors range from human to socio-technical issues, access to electronic health resources, awareness of HICT, utilization of HICT, and perceived factors responsible for use or non-use of HICT among professionals to deliver best understanding of e-Health systems and highlight the significant objectives, tools, models, and obstacles of such systems.

Keywords: Health Information and Communication Technologies (HICT) · E-health · Electronic health record system · Computer science · Kuwait

1 Introduction

The escalating development of communication technologies has resulted in equally advancing modes of applying technology in people's daily lives. This applies to both the developed and developing parts of the world. As a result of applying technology in the medical care, notable leaps have been made towards overcoming many diseases and the general improvement of health care systems. The "e" in e-Health stands for a number of concepts [5]. It includes enhancing quality and encouraging trust between the patients and health professionals. It also includes educating physicians by means of online sources, enabling information exchange and communications. So, an electronic health record is a digital store of patient data made accessible to multiple authorized users for continuity and efficiency in an integrated healthcare delivery system [2]. Information and communication technologies (ICTs) are incorporated in the health sector to create e-health solutions which lead to additional benefits such as; minimizing coordination and transaction costs in patients and physicians' relationships, better deliberativeness, and promoting the ability of information-processing in information technology [6]. There is a growing interest in implementing a system of shared electronic health records among organizations as a quality improvement initiative [1].

This research discusses e-Health in Kuwait and addresses factors leading to the adoption and utilization of HICTs. The results are essential to determine the obstacles Kuwait is facing to use e-Health portals [13]. Furthermore, the paper will conclude with suggestions provided by both patients and physicians working in the health sector to face the challenges. End user input in the design and development of e-Health technologies should be considered as a way of overcoming barriers of adaptability [14].

2 The status e-Health services in Kuwait

Following the recent growth in communication technologies, many nations in the world have realized notable improvements in the delivery of public health services. Technology solutions have helped health organizations and governments to combat most common world's fatal diseases and health conditions due to improved flow of health information. Electronic health means the application of information communication technologies in the delivery of health services, professional development in health fields, strategic management of adverse health conditions, and general development of public health sector [18]. Electronic health adoption, therefore, involves efforts of both governments and global health organizations with the main aim of improving health services. In the delivery of public health services, health practitioners and professionals apply technology solutions as tools to aid quality and efficiency in their operations. For instance, collection and management of health information are areas where information communication technology plays a key role. In Kuwait, the technology application process involves collaboration with patients through media services, and online communication platforms such as telephone services, social media, and public websites to collect health information such as patients' needs and health conditions. This is easier, broader in scope, and more effective than empirical field research that requires physical meetings with target patients when conducting medical research.

Technological advancements always pose varied effects on the provisions of health-care. According to Ross et al. [14], e-Health is seen as essential for solving challenges associated with healthcare systems with regard to increasing demand due to an aging population and required improved treatment, all coupled with limited resources. As a result, many world nations have adopted both short-term and long-term strategies aiming at exploiting technology to transform their management of the health sector. For instance, England has invested at least £12.8 billion in a national program for incorporating the latest technology in healthcare while in the United States, the Obama administration committed to a \$34 billion investment in e-Health through the Department of Health Services [7].

Moreover, according to Ross et al. [14], tele-monitoring of diabetic foot ulcer patients is one of the proven e-health applications that have proven to enable better monitoring of healthcare systems. It is a cost-effective method besides being accurate, a factor that enables saving of many lives. In healthcare, ICT plays a critical role of improving efficiency of service delivery, enhancing satisfaction through reduction of costs and improvement of safety in offering healthcare services [9]. Majority in the public population suffer health risks mostly because of financial constraints, this is especially in the developing countries. Therefore, access to expensive health facilities is left to be a privilege of the financially endowed minority. Kuwait may not be classified as a fully developed economy, as it belongs more to the category of underdeveloped nations.

According to Farhan & Sanderson [2], Kuwait has come a long way in terms of information technology development. Primarily, it is a country whose educational system has not traditionally placed great emphasis on the need to engage in IT and ICT courses. Nevertheless, with the increased technological advancements, the country has been among those that have recorded the highest uptake rates of modern technology. The Kuwaiti Ministry of Health has played a key steering role in ensuring that the development of the country's IT sector positively impacts the provisions of healthcare practice advancement. In particular, the current government under the leadership of Sheikh Sabah Ahmad AL Sabah, has ensured that the country maintains an improvement trajectory in matters of e-Health to keep up with other current trends practiced in developed countries [13]. According to Rabaai et al. [11] the government and its leadership has always maintained a futuristic vision for Kuwait. This has been demonstrated by strong embracing of state-of-the-art technology practices in Kuwait. The launching of the Kuwait Central Agency for Information Technology in 2006 is a good example of the national technology road mapping strategies adopted by Kuwait. Mandated by the national government, the Ministry of Health ensures that the country procures up-to-date data storage and analytics systems to ensure good keeping of records and proper interpretation of phenomena related to the health cases data collected. This has enabled private institutions in the health sector to conform to adopting such technology standards by ensuring that they adopt e-Health methods in order to capitalize on the satisfaction of patients, adhere to national government requirements and also capitalize on efficiency.

The increased adoption of technology in healthcare has been attributed to high literacy levels, the economic endowment resulting from oil production, large budget surpluses and high levels of disposable income [11]. Coupled with these national growth factors, technology has improved healthcare, governance, education, industrialization, the private sector, and general human growth in the State of Kuwait [10, 11]. Kuwait is currently focusing on research and development with an aim of building more capacity for exploring new technology-aided methods for improving healthcare [2]. E-Health is one of the indicators of the improved health sector. The country also expects the research and development efforts to yield competitive and affordable means of securing the best healthcare services for all citizens.

The tremendous growth in communication technologies in the State of Kuwait is influenced by a number of factors that affect participation in e-Health. Those factors are shared across various nations but the extent to which their effect is displayed differs. Such factors include, but not limited to, economic development, national productivity, culture and the general attitude towards technology [13]. Literature shows that between the years 2006 and 2014, Kuwait's global position in terms of e-Government Readiness Index has been alternating with those of the neighboring Gulf Corporation Council members. For instance, it was very low in 2012, while it rose in 2014. This is because participation in e-government, e-Health, and other areas of ICT applications is influenced by a number of variables. Some variables affect the index positively while others have negative effects. Such different variables can exist or originate from the same environment. They include users, executives, organizational cultures, external environments, and global forces [10, 11].

However, Rabaai et al. [11], advance that the implementation of health ICT solutions has a number of challenges in the State of Kuwait. Limited awareness of information technology and lack of online information security are good examples on obstacles for adoption. Some Kuwaitis have limited awareness of the importance of communication technologies in health service practices. This is the result of ignorance and insufficient public education on the role of ICT in the health sector. A notable percentage of the citizens have apprehensions of online data insecurity; for instance, sensitive patients' information is subject to cyber-attacks [5]. Medical information security and privacy is a major concern as an online leak causes irreparable damage in terms of reputation and confidence. Cyber security is, therefore, a major limiting factor for e-participation in e-Health.

The culture of Kuwaiti people has a significant effect on how they embrace e-Health information and platforms. Essentially, many people in the country seem to prefer the traditional manual procedures for obtaining medical care that is basically based on calling for appointments, visiting private facilities for treatment and obtaining over-the-counter drugs. The prevailing culture also influences organizational cultures, a factor that largely dictates how healthcare is provided both in private and public organizations [14]. For instance, if a health institution maintains an operation mode in which all employees are given equal opportunities to utilize technology resources and build technical skills the e-Health may gain value.

The culture of a health institution, may it be private or public, defines its mode of operations both internally and externally in terms of the roles played by every stakeholder, how responsibilities are shared, and how various functions coordinate in the implementation of managerial plans and in-service delivery [12]. This partly explains why organizational culture in Kuwait health sector directly influences e-participation in e-Health.

If a health institution maintains an operation mode in which all employees are given equal opportunities to utilize technology resources and build technical skills, the e-Health can gain value. It can be perceived as the easiest way to interacting with patients, collect and manage their information while maintaining information security and privacy. Otherwise, if ICT services are privileges limited to senior stakeholders only, the growth of e-participation in e-Health remains remote. Therefore, the culture of a health institution can aid or, otherwise, undermine effective participation in e-Health.

If a health institution maintains an operation mode in which all employees are given equal opportunities to utilize technology resources and build technical skills, the e-Health can gain value. It can be perceived as the easiest way to interacting with patients, collect and manage their information while maintaining information security and privacy. Otherwise, if ICT services are privileges limited to senior stakeholders only, the growth of e-participation in e-Health remains remote. Therefore, the culture of a health institution can aid or, otherwise, undermine effective participation in e-Health [7, 8].

This research has unveiled a number of facts concerning Kuwait health sector and economy. For example, the private and public sectors have grown in terms of service efficiency since 2006, following the good policies adopted by the leaderships. The state spearheaded the adoption of information technology solutions in health service delivery with the support of United Nations Public Administration Network. In addition, Kuwait's oil industry has boosted the economy, leading to high levels of education and fast growth of telecommunication industry. The country utilizes all telecommunication channels, with almost 80% of the population recognized as internet users. This means that e-participation in e-Health is reasonably high. It has been facilitated by both internal and external stakeholders of health institutions. The private sector is, however, more advanced, compared to the public health sector, because of the competitive initiatives that have led to faster adoption of technology solutions as compared to public institutions. This calls for close monitoring of the management of public health institutions by the government.

As cyber security remains the major concern for e-participation in e-Health [4, 7], the main recommendation here is the adoption of cybersecurity approaches such as sophisticated online data encryption, security, and vulnerability analysis tools that aid in promoting the fact that E-Health is a revolutionary approach to healthcare provision [17, 19].

3 Problem Statement and Methodology

Many barriers and parameters exist when it comes to embracing and implementing e-Health systems [1, 3, 15]. This research aims to deliver a better understanding of e-Health prospects in Kuwait and highlight the significant elements and factors affecting e-Health adoption. We, therefore, propose a model to facilitate a better understanding of the dimensions affecting indulging in HICT/e-Health systems in Kuwait. To do so, it is emphasized here that the recognition of health providers' specialization, their e-Health background, and prior experience in such systems. Also, the study presents health providers' point of view and evaluation of e-Health systems through the use of a carefully prepared questionnaire. The research hypothesis is that the implementation of e-Health

systems in hospitals and medical centers has positive impact on: cost savings, quality control, increased efficiency and effectiveness. The researcher surveyed a various group of employees in the health sector to get their point of view about Kuwait e-Health system, their experiences, and reasons behind the fact that considerable portion of practitioners abstain from adopting and practicing the electronic services.

A cross-sectional design was habituated to accumulate quantitative data utilizing a structured questionnaire among health practitioners working in the both the public and private sectors. Systematic arbitrary sampling was administered to 211 practitioners, out of which 193 consented to participate. Data analysis was done utilizing MatLab. Hypotheses were tested at p value < 0.05 utilizing Chi square and correlation coefficient. The questionnaire consisted of 10 questions, direct and clear for participants to answer. Some questions were of general demographic focus; others were questions about Kuwait e-Health system with multiple choice answers; one question was a short answer to collect suggestions for future services in the system. The last question consisted of multiple statements to be measured on a five-point scale that range between "Strongly disagree" to "Strongly agree". The questionnaire was conducted using a professional survey website (Google Forms) and was distributed online via Twitter, and WhatsApp in order to reach the targeted population of participants efficiently. It was distributed among hospitals and medical centers around Kuwait. The survey was open for a whole month (April 2018 to May 2018) during that time 132 responses were received.

The study of the factors influencing adoption of e-Health technology is inspired by a variant of the Technology Acceptance Model, given in Fig. 1.

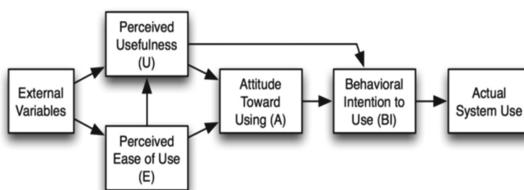


Fig. 1. A variant of the traditional TAM model.

4 Validity and Reliability of the Research Instrument

Face and content validity techniques were employed to establish the validity of the structured questionnaire. The structured questionnaire was critically reviewed for felicitous structuring and opportuneness of the test item to answer the research questions. Pilot study was conducted by sending the questionnaire to 38 arbitrarily selected respondents who are directly involved in HICT usage for evaluation and participation. Internal consistency of data engendered by the instrument or questionnaire was ascertained by evaluating the collected responses. Equivocal questions were reframed for pellucidity and pertinence to the expressed research objectives of the study.

Test-retest method was employed to determine the Cronbach's Alpha value of the Likert scale, which establishes the reliability and consistency of the instrument. The

Cronbach's Alpha was calculated utilizing 16 culled test cases, of which the result was 0.79. Hence, it was concluded that the research instrument possesses an acceptable level of reliability with good internal consistency.

5 The Results

The table below present a summary of the results obtained from the study.

Table 1. Hypotheses testing results

Null hypothesis	Crosstabs (p-value < 0.05)	Chi-Square	Interpretation
(i) There is no significant relationship between Age of Users and Actual Use of HICT	Age range of respondents * Frequency of HICT use at Workplace	0.041	Significant
(ii) There is no significant relationship between the level of Academic Qualification and Actual use of HICT	Level of Academic Qualification * Frequency of HICT use at Workplace	0.135	Not significant
(iii) There is no link between the perceptual usefulness concerning the quality of service of HICT in health practices	Perceptual usefulness of HICT * Frequency of HICT use at Workplace	0.009	Significant
(iv) There is no significant relationship between Training on the use of HICT system and Actual Use	Training on the use of HICT * Frequency of HICT use at Workplace	0.001	Significant

The questionnaire embodies a number of five-scale Likert type questions that range from "strongly agree" to "strongly disagree". Some of those are reported in Table 1. One variable was not included in Table 1 that is the quality criteria. Our results reveal that the significance of the quality determinant is not a major determinant when it comes to users in Kuwait. Another determinant has to do with the significance of the technical factors, such as availability, support and reliability. This criterion was not found to be as a serious determinant, as opposed to others. Usefulness is another criterion that was measured for significance.

For the respondents, the prominent reasons for using e-health systems wavered among saving time and money, facilitating job processes, saving effort and compulsion to use as it was imposed as the only way for conducting health-related functionalities.

6 Conclusions

Although our study is restricted to one country, namely Kuwait, and concentrated on 12 clinics and 5 hospitals, both private and public, the results obtained seem to confirm the significance of three of the proposed hypotheses in Table 1, namely (i), (iii) and (iv).

There is still room for making the scope of the coverage more extensive geographically and hence the results obtained would be progressively solid and more convincing. More health firms need to be included. We propose that future research would fill in whatever yawning holes are seen between our present discoveries and what other missing criteria that need to be revealed. Another confinement is that the investigation is cross-sectional, which is viewed as quick, straightforward, and affordable to perform. Nonetheless, a cross sectional examination could be one-sided, where members who partook in the investigation neglected to present the input of others working in the health domain, influencing subsequently, the generalizability of the findings. This is on the grounds that it depends on a poll review, and members are met just once [16]. Consequently, future work ought to consider conducting a longitudinal methodology, where members participate more than once over a stretch of time.

Regardless of these impediments, our conducted tests had the capacity to explore and recognize the basic factors that influence health domain practitioners' adoption of HICTs in conducting health services. Overall, the benefits of e-Health systems much outweigh any disadvantages perceived by some users, such as releasing the grip of power in the hands of health administrators, especially when it comes to sharing resources such as patient files. This transparency can be disadvantageous to some. The results obtained from the statistical analysis concluded that a large portion of health practitioners perceive the advantages of such systems. The measurement of overall satisfaction was, in general, good as per our results. The main issue that we need to emphasize is the lack of integration among the hospitals in Kuwait, the thing that inhibits a full appreciation of embracing HICT. It is important to have the integration; it will allow the availability of electronic medical records in all hospitals. This will allow each patient having a unified electronic medical record that he will use in all of hospital and medical centers around Kuwait. It is important to mention that the results, in this study, are helpful for decision makers to understand the user's needs and requirements for future evolution of e-Health systems in Kuwait.

Appendix

Factors Affecting e-Participation in Kuwait e-Health

e-Health is using Information and Communication Technologies (ICTs) to deliver health care and services like the electronic medical record. The survey will investigate the importance of installing e-health systems in Kuwait hospitals. It also addresses the users' acceptance for using e-Health systems. We highly appreciate your participation. Privacy will be maintained, please answer questions as they relate to you.

Personal information:

1-Gender:

Male

Female

2- Age:

Less than 20

20–29

30–39

40–49

50–59

60 and more

3-Nationality:

Kuwaiti

Non Kuwaiti

4-Education:

High school

Diploma

Bachelor

Master & above

5-Occupation:

Medical student

I work in the government sector

I work in the private sector

I have my own practice

6-Please choose which category fits your position best:

Patient

Doctor

Nurse or other medical professional

Hospital administrator

Pharmacist

Health Insurance representative

Health care academic

Other (please specify) _____

Factors affecting e-Participation in Kuwait e-Health:

Are you familiar with the term "e- Health systems"

Yes

No

a) Do you use the online information system provided by Kuwait ministries?(ex. <https://www.e.gov.kw>)

Yes

No (if you choose No, answer the following question)

b) Why you didn't use any online information systems?

I don't know how to use it

I didn't hear about any information system

I don't trust the online information system

Other (please specify) _____

How often do you use e-Health systems?

None

Rarely

Often

Always

Indicate your level of agreement with this statement: "e-Health is essential in improving the quality and consistency of health care."

Strongly Agree

Agree

Neutral

Disagree

Strongly Disagree

How useful is the integration of e-health systems in Kuwait Hospitals?

Very useful

Useful

Neutral

Un-useful

very un-useful

How important is e-Health for creating new health care products and services?

Very useful

Useful

Neutral

Un-useful

Very un-useful

Using e-Health will benefit in (you can choose more than one option)

Cost savings

Quality control

- Improvement of health outcomes
- Development of new products and services
- Easier patient access to care
- Other (please specify _____)

What kind of issues do the e-health systems have in Kuwait?(you can choose more than one option)

- None
- Lack of integration.
- Lack of modern electronic devices.
- Lack of availability of IT maintenance team.
- Lack of employees' knowledge about the system.
- Lack of system training courses.
- Employee resistance to change.
- Other (please specify)_____

What services you like e-Health to provide in the future?

References

1. Angel, D., Bjerregaard, J., Oconnor, T., Meguiness, W., Kröger, K., Rasmussen, B.S., Yderstraede, K.B.: Barriers and facilitators for eHealth. *J. Wound Care* **24**, 1 (2015)
2. Farhan, H., Sanderson, M.: User's Satisfaction of Kuwait E-Government Portal; Organization of Information in Particular. University of Sheffield, UK (2012)
3. Gagnon, M.P., Desmartis, M., Labrecque, M., Car, J., Pagliari, C., Pluye, P., et al.: Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. *J. Med. Syst.* **36**(1), 241–277 (2012). <https://doi.org/10.1007/s10916-010-9473-4>
4. Hayrinен, K., Saranto, K., Nykanen, P.: Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int. J. Med. Inform.* **77**(5), 291–304 (2008)
5. Hill, L.: Digital literacy instruction for eHealth and beyond. *ORTESOL J.* **33**, 34–40 (2016)
6. Kearns, M., Kavanagh, A., Curran, M., Collier, D.: ICT supporting clinicians and patients – the key facilitator of integrated care. *Int. J. Integr. Care* **17**(5), 52–61 (2017)
7. Keasberry, J., Scott, I.A., Sullivan, C., Staib, A., Ashby, R.: Going digital: a narrative overview of the clinical and organisational impacts of eHealth technologies in hospital practice. *Aust. Health Rev.* **41**(6), 646–664 (2017). <https://doi.org/10.1071/AH16233>
8. Lium, J.T., Laerum, H., Schultz, T., Faxvaag, A.: From the front line, report from a near paperless hospital: mixed reception among health care professionals. *J. Am. Inform. Assoc.* **13**, 668–675 (2006)
9. Peate, I.: Technology, health and the home: eHealth and the community nurse. *Br. J. Community Nurs.* **18**(5), 222–227 (2013). <https://doi.org/10.12968/bjcn.18.5.222>
10. Qureshi, N.A., et al.: Factors affecting adoption and use of ICTs in hospitals/healthcare. *Mediterr. J. Med. Sci.* **1**(1), 13–20 (2014). https://www.academia.edu/9517442/Factorsaffecting_the_introduction_of_ICTs_for_healthcare_decision-making_in_hospitals_of_developing_countries

11. Rabaai, A.A., et al.: Adoption of e-government in developing countries: the case of the State of Kuwait. *J. Glob. Res. Comput. Sci.* **42**(2), 289–304 (2017)
12. Ravasi, D.L., Schultz, M.: Responding to organizational identity threats: exploring the role of organizational culture. *Acad. Manage. J.* **49**(3), 433–458 (2006)
13. Ridic, G., Gleason, S., Ridic, O.: Comparisons of health care systems in the United States, Germany and Canada. *Mater. Sociomed.* **24**(2), 112–120 (2012). pmid:23678317
14. Ross, J., Stevenson, F., Lau, R., Murray, E.: Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). *Implementation Sci.* **11**(1), 1–12 (2016). <https://doi.org/10.1186/s13012-016-0510-7>
15. Ross, J., Stevenson, F., Lau, R., Murray, E.: Exploring the challenges of implementing e-health: a protocol for an update of a systematic review of reviews. *BMJ Open* **5**(4), e006773 (2015). <https://doi.org/10.1136/bmjopen-2014-006773>
16. Sedgwick, P.: Cross sectional studies: advantages and disadvantages *BMJ* **348**, g 2276 (2014)
17. Singh, B., Muthuswamy, P.: Factors affecting the adoption of electronic health records by nurses. *World Appl. Sci. J.* **28**(11), 1531–1535 (2013)
18. Tundjungsari, V.: E-Participation modeling and developing with trust for decision making supplement purpose. *Int. J. Adv. Comput. Sci. Appl.* **3**(5), 55–62 (2011)
19. Yehualashet, G., Andualem, M., Tilahun, B.: The attitude towards and use of electronic medical record system by health professionals at a referral hospital in Northern Ethiopia: cross-sectional study. *J. Health Inform. Afr.* **3**(1), 19–29 (2015)



Data Mining Approach to Classify Cases of Lung Cancer

Eduarda Vieira¹ , Diana Ferreira² , Cristiana Neto² , António Abelha² , and José Machado²

¹ Department of Information, University of Minho, Braga, Portugal
a83160@alunos.uminho.pt

² Algoritmi Research Center, University of Minho, Campus of Gualtar, Braga, Portugal
{diana.ferreira,cristiana.neto}@algoritmi.uminho.pt,
{abelha,jmac}@di.uminho.pt

Abstract. According to the World Cancer Research Fund, a leading authority on cancer prevention research, lung cancer is the most commonly occurring cancer in men and the third most commonly occurring cancer in women, with the 5-year relative survival percentage being significantly low. Smoking is the major risk factor for lung cancer and the symptoms associated with it include cough, fatigue, shortness of breath, chest pain, weight loss, and loss of appetite. In an attempt to build a model capable of identifying individuals with lung cancer, this study aims to build a data mining classification model to predict whether or not a patient has lung cancer based on crucial features such as the above mentioned symptoms. Through the CRISP-DM methodology and the RapidMiner software, different models were built, using different scenarios, algorithms, sampling methods, and data approaches. The best data mining model achieved an accuracy of 93%, a sensitivity of 96%, a specificity of 90% and a precision of 91%, using the Artificial Neural Network algorithm.

Keywords: Healthcare · Lung cancer · Data mining · Classification · CRISP-DM

1 Introduction

Lung cancer is the most common cancer in men and the third most common cancer in women [1]. The prognosis for this type of cancer is usually alarming, with less than 15% of patients surviving five years after the diagnosis. This prognosis results from the lack of effective early detection diagnostic methods as well as the lack of successful metastatic disease treatment [2].

The 5-year relative survival estimates the percentage of cancer patients who will not have died of cancer five years after diagnosis. The 5-year relative survival rate for lung and bronchus cancer in the United States is 19.1%, this metric varies between age groups [5]. The lung cancer statistics also vary substantially

with sex, ethnicity, socio-economic status, and geography. Lung cancer rates are higher in countries where smoking begins at an early age, specifically countries in North America and Europe. Low and middle income countries account for more than 50% of lung cancer deaths each year [6]. Smoking is then the major risk factor for lung cancer, accounting for about 90% of the incidence of lung cancer [7]. Additional risk factors include: low fruit and vegetable consumption, genetic predisposition, exposure to non-tobacco procarcinogens, carcinogens and tumor promoters, previous lung diseases such as chronic obstructive pulmonary disease, previous tobacco-related cancer, and passive smoking [7].

The recommendations from the National Institute for Health and Care Excellence (NICE) for the screening of Lung Cancer are: people aged 40 and over if they have two or more of the following unexplained symptoms, or if they have ever smoked and have one or more of the following unexplained symptoms: cough, fatigue, shortness of breath, chest pain, weight loss and appetite loss [8].

As mentioned earlier, the lack of effective early detection diagnostic methods for lung cancer is one of the major causes of low chances of survival five years after the diagnosis [2]. Hence, it is valuable to carry out research in this area in order to try to discover new ways of improving the diagnosis of this condition [9]. A common way to identify clinical conditions is to use Data Mining (DM) algorithms since they enable researchers to find new patterns and hidden knowledge in large and complex datasets [4].

Hospitals and other healthcare facilities are known to generate large amounts of data on a daily basis and therefore the increasing interest of healthcare organizations in using DM as it can greatly benefit the efficiency and quality of the services provided [3]. This study in particular focuses on the use of DM techniques to build a classification model capable of predicting whether or not a patient has lung cancer following the well known Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology.

The present paper is organized into five different sections. After the Introduction, Sect. 2 presents an overview of existing studies carried out within the scope of the topic addressed in this study. The entire DM process implemented is thoroughly detailed in Sect. 3. Then, Sect. 4 presents the discussion of the results obtained. Finally, Sect. 5 presents the main conclusions obtained with this work and some prospects for future work.

2 Related Work

Krishnaiah et al. (2013) used DM classification algorithms namely IF-THEN Rules, Decision trees (DT), Bayesian classifiers and Neural Networks (NN) to build a classification model able to predict whether or not a patient has lung cancer. They concluded that the most effective model to predict if patients had lung cancer was Naïve Bayes (NB) followed by IF-THEN rules, DT and NN. NB performed better than DT as it could identify all the significant medical predictors [10].

Nasser et al. (2019) developed an Artificial Neural Network (ANN) to detect lung cancer in patients based on symptoms such as yellow fingers, anxiety,

chronic disease, fatigue, allergy, wheezing, coughing, shortness of breath, swallowing difficulty, and chest pain, using the “Survey Lung Cancer” dataset. The model had a 96.67% accuracy after 1418105 learning cycles and with less than 1% training error rate. They concluded that age, gender and coughing were the most relevant features [11].

Murty et al. (2017) conducted a study to analyze lung cancer prediction using classification algorithms such as NB, Radial Basis Function (RBF) Neural Network, Multilayer Perceptron (MLP), DT and C4.5 (J48) using the Weka DM software. The attributes age, gender, alcohol usage, genetic risk, chronic lung disease, balanced diet, obesity, smoking, passive smoker, chest pain, coughing of blood, weight loss, shortness of breath, wheezing, swallowing difficulty, frequent cold, dry cough, snoring and some more additional symptoms were taken into consideration for predicting the lung cancer. They found that the NB algorithm achieved a better performance in all aspects over the other classification algorithms previously mentioned [12].

3 Data Mining Process

The dataset used in this study consists of medical data related to the diagnosis of lung cancer disease, collected from surveys, namely demographic facts, habits, symptomatology, and clinical history [14]. Five different Data Mining Techniques (DMTs) were explored namely Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), NB, DT and ANN (Artificial Neural Net). These techniques were chosen based on their Receiver Operating Characteristic (ROC) curve analysis and were further implemented using the RapidMiner software. The DM process was guided by the CRISP-DM methodology, which comprehends six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [13].

3.1 Business Understanding

Due to the increasing number of cases of lung cancer in developed countries and the high number of deaths associated with lung cancer, it has become essential to identify and understand what are the main factors of this disease to try to mitigate this worldwide concern. Early diagnosis of cases of lung cancer would be important, as healthcare providers could intervene earlier, thus reducing the likelihood of the incidence of the disease, preventing it from progressing to more advanced stages, or even reversing it.

The business goal of this paper is to study which factors have a greater impact on the diagnosis of lung cancer as well as to build a predictive system for the early detection of this disease. DM provides the means to accurately classify cases of lung cancer. In this paper, DMTs will be used to build models capable of extracting relevant information from the data of inquired patients in order to classify cases of lung cancer based on demographic facts, habits, symptomatology, and clinical history.

3.2 Data Understanding

The dataset used in this work contains 16 features and 309 instances [14]. Each instance corresponds to a patient and contains the patient's medical data. The statistical data of the features can be consulted in Table 1.

Table 1. Statistical data of the features

Feature	Description	Min	Max	Mean	Deviation	Quantity
gender	Patient gender (M = male, F = female)	-	-	-	-	M (162) F (147)
age	Patient's age in years	21	87	62.673	8.210	-
smoking	Whether the patient smokes or not (1 = No, 2 = Yes)	1	2	1.563	0.497	-
yellow_fingers	Whether the patient has yellow fingers or not (1 = No, 2 = Yes)	1	2	1.570	0.496	-
anxiety	Whether the patient has anxiety or not (1 = No, 2 = Yes)	1	2	1.498	0.501	-
peer_pressure	Whether the patient feels peer pressure or not (1 = No, 2 = Yes)	1	2	1.502	0.501	-
chronic_disease	Whether the patient has a chronic disease or not (1 = No, 2 = Yes)	1	2	1.505	0.501	-
fatigue	Whether the patient has a fatigue or not (1 = No, 2 = Yes)	1	2	1.673	0.470	-
allergy	Whether the patient has an allergy or not (1 = No, 2 = Yes)	1	2	1.557	0.498	-
wheezing	Whether the patient has wheezing or not (1 = No, 2 = Yes)	1	2	1.557	0.498	-
alcohol_consuming	Whether the patient consumes alcohol (1 = No, 2 = Yes)	1	2	1.557	0.498	-
coughing	Whether the patient coughs a lot (1 = No, 2 = Yes)	1	2	1.579	0.494	-
shortness_of_breath	Whether the patient feels shortness of breath (1 = No, 2 = Yes)	1	2	1.641	0.481	-
swallowing_difficulty	Whether the patient has difficulty in swallowing (1 = No, 2 = Yes)	1	2	1.469	0.500	-
chest_pain	Whether the patient feels chest pain or not (1 = No, 2 = Yes)	1	2	1.557	0.498	-
lung_cancer	The patient has lung cancer (true, false)	-	-	-	-	YES (270) NO (39)

Regarding the data, two features are demographic, namely *gender* and *age*, and fourteen are binary and related to habits, symptomatology, and clinical history of each patient. The features *gender* and *lung_cancer* are polynomial and the remaining are integer. As observed in Table 1, the youngest patient is 21 years-old and the oldest is 87 years-old, with the average age being quite high: 62.673 years-old. Of the remaining attributes it is possible to establish that the most recurring features of the patients interviewed are: fatigue (1.673 mean value) and *shortness_of_breath* (1.641 mean value). The least recurring features are: *anxiety* (1.498 mean value) and *swallowing_difficulty* (1.469 mean value).

As a classification problem, it was necessary to select a target, in this case, the *lung_cancer* attribute. Table 1 shows that, out of 309 interviewed patients, 270 have lung cancer, meaning an unbalanced distribution of the target classes.

3.3 Data Preparation

This stage of the CRISP-DM process is crucial to improve the performance of the different Data Mining Models (DMM). It typically involves integrating, cleaning, transforming, reducing, and sampling data [15].

First, an analysis of data inconsistencies was carried out. No missing values were found. Regarding outlier values, the *Detect Outlier (LOF)* operator was used, which identifies outliers based on local outlier factors (LOF). The operator has identified one instance as an outlier, the instance that corresponds to the 21-year-old participant. However, it was decided to keep it since this instance was identified as an outlier only because of the age value, given that the average age of the individuals in this dataset is approximately 63 years and the fact that all the remaining attributes are binary. In addition, there were 33 duplicate instances and, therefore, they were removed using the operator *Remove Duplicates*. This operator removes duplicate instances by comparing all instances, based on specific attributes, and only keeping one of all the duplicate instances.

Then, some data type conversions were performed, particularly in the *gender* attribute and in the target class, *lung_cancer*, which were converted from nominal to numeric and from nominal to binominal, respectively.

In addition, a feature selection was made to determine the features that had greater impact on the target attribute prediction and, consequently, to create different scenarios based on that selection. In order to achieve this, four operators were used: *Weight by Information Gain*, *Weight by Chi Squared Statistic*, *Weight by Gini Index*, and *Weight by Gain Ratio*. Unexpectedly, *gender*, *smoking*, and *shortness_of_breath* were considered, by all the criteria mentioned above, the least relevant features for predicting the target attribute. These features were followed by *fatigue* and *yellow_fingers*.

Finally, as there were far more patients diagnosed with lung cancer (270) than those not diagnosed with lung cancer (39), an oversampling method was used in order to achieve a balanced distribution between the two classes. The oversampling was performed using the *SMOTE Upsampling* operator, which creates new instances for the minority class based on the k nearest neighbours.

3.4 Modeling

After the Data Preparation, the data was finally ready to be used in the modeling phase. In this step, several DMMs were developed according to the Eq. 1:

$$DMM = \{A, S, DMT, SM, DA, T\} \quad (1)$$

Hence, each DMM can be described as belonging to an approach (A), being composed by a data mining technique (DMT), a scenario (S), a sampling method (SM), a data approach (DA) and a target (T).

As mentioned, classification was the chosen A and there was only a single T, the *lung_cancer* attribute. In addition, two DA were performed, namely, no oversampling and oversampling using the *SMOTE upsampling* operator. Furthermore, five DMT were performed namely SVM, k-NN, NB, DT, and NN.

In what concerns the SM, two different methods were used, namely, Cross Validation with 10 folds and Split Validation, with 30% of the data used for testing and the remaining 70% used for training. Split Validation differs from Cross Validation because in the latter the dataset is split into k random folds for which the model is tested using each fold throughout the k iterations. The final value of accuracy is the mean value of each of these iterations.

Three different scenarios were created based on the feature weights: S1 - All the attributes; S2 - All the attributes except *smoking*, *shortness_of_breath*, and *gender*; S3 - All the attributes except *smoking*, *shortness_of_breath*, *gender*, *fatigue*, and *yellow_fingers*

In total, 60 DMM were built in this study, as expressed in Eq. 2

$$DMM = \{1(A) \times 3(S) \times 5(DMT) \times 2(SM) \times 2(DA) \times 1(T)\} \quad (2)$$

3.5 Evaluation

This study is a binary classification problem and thus, to evaluate the predictions of each model, the chosen criteria were the confusion matrix and some evaluation metrics derived from it. The confusion matrix is a predictive classification table, which provides the number of True Negatives (TN), False Negatives (FN), False Positives (FP) and True Positives (TP). These values were used to determine the following evaluation metrics: Accuracy, Sensitivity, Specificity, and Precision, which were calculated through Eqs. 3, 4, 5, and 6 respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Accuracy measures the model's capability of correctly distinguishing the patients with lung cancer and the patients without lung cancer. Sensitivity measures the proportion of patients that have lung cancer that were correctly identified as such by the model. On the other hand, Specificity measures the proportion of patients without lung cancer that were correctly identified as such by the model. Finally, Precision measures the proportion of patients who have been labeled by the model has having lung cancer that actually have it.

Tables 2, 3, and 4 contain the best results obtained, with and without SMOTE, for the S1, S2 and S3 scenarios, respectively.

Table 2. Best results for each DMM, with and without SMOTE, for S1

DMT	SM	DA	Accuracy	Sensitivity	Specificity	Precision
SVM	Split Validation	–	0.915	0.958	0.636	0.944
SVM	Split Validation	SMOTE	0.909	0.916	0.901	0.903
NN	Cross Validation	–	0.892	0.938	0.617	0.937
NN	Cross Validation	SMOTE	0.920	0.942	0.899	0.908
NB	Cross Validation	–	0.881	0.938	0.533	0.928
NB	Cross Validation	SMOTE	0.887	0.862	0.912	0.910
DT	Cross Validation	–	0.856	0.895	0.625	0.937
DT	Cross Validation	SMOTE	0.874	0.912	0.836	0.852
k-NN	Cross Validation	–	0.855	0.971	0.133	0.875
k-NN	Cross Validation	SMOTE	0.855	0.874	0.836	0.847

Table 3. Best results for each DMM, with and without SMOTE, for S2

DMT	SM	DA	Accuracy	Sensitivity	Specificity	Precision
SVM	Cross Validation	–	0.899	0.950	0.583	0.936
SVM	Cross Validation	SMOTE	0.918	0.928	0.908	0.910
NN	Split Validation	–	0.842	0.859	0.727	0.953
NN	Split Validation	SMOTE	0.930	0.958	0.901	0.907
NB	Cross Validation	–	0.885	0.942	0.533	0.928
NB	Cross Validation	SMOTE	0.870	0.835	0.903	0.895
DT	Cross Validation	–	0.852	0.900	0.567	0.929
DT	Cross Validation	SMOTE	0.918	0.950	0.886	0.894
k-NN	Split Validation	–	0.878	0.944	0.455	0.918
k-NN	Split Validation	SMOTE	0.901	0.958	0.845	0.861

4 Discussion

Analyzing the results obtained, it can be observed that there are very low Specificity values when oversampling is not applied, although at first sight it may appear that models with good performance have been obtained as the other evaluation metrics have achieved good performance values. This is due to the imbalance in the distribution of the target attribute, since the models created are unable to distinguish between the two classes, and can only classify cases of the majority class (patients with lung cancer), resulting in low specificity values that, as mentioned above, refer to the proportion of patients without lung cancer that were correctly identified as such. Hence, when using oversampling, in particular the SMOTE Upsampling technique, it was possible to obtain Specificity values higher than 90% for some models, which could not be achieved without applying oversampling.

Table 4. Best results for each DMM, with and without SMOTE, for S3

DMT	SM	DA	Accuracy	Sensitivity	Specificity	Precision
SVM	Cross Validation	–	0.885	0.958	0.433	0.913
SVM	Cross Validation	SMOTE	0.857	0.887	0.827	0.841
NN	Cross Validation	–	0.896	0.946	0.592	0.934
NN	Cross Validation	SMOTE	0.914	0.937	0.890	0.902
NB	Cross Validation	–	0.878	0.934	0.533	0.927
NB	Cross Validation	SMOTE	0.849	0.828	0.870	0.870
DT	Cross Validation	–	0.860	0.900	0.617	0.937
DT	Cross Validation	SMOTE	0.885	0.891	0.879	0.893
k-NN	Cross Validation	–	0.826	0.933	0.150	0.874
k-NN	Cross Validation	SMOTE	0.838	0.857	0.819	0.834

As expected, in general, the models that used Cross Validation as SM achieved better results than those that used Split Validation because the former uses all data for training, while the latter uses only a certain percentage, and algorithms are known to learn more effectively when using more data for training.

Regarding the different scenarios, it can be concluded that the worst results were achieved when the S3 scenario was used. Although the results obtained for the S2 and S1 scenarios are very similar, it can be seen that, in general, the DMMs using the S2 scenario had slightly better results.

As far as DMTs are concerned, overall, all algorithms achieved good performance and there are no striking differences between the results obtained for the different evaluation metrics. Nonetheless, it can be observed that the algorithms

with the best results were NN and SVM. In turn, although it has also achieved a satisfactory performance, the technique with the worst results was k-NN.

In order to choose the best model, a threshold has been established. The threshold combines the four evaluation metrics, but since this problem fits into the scope of medical diagnosis and the consequences that could arise from the existence of FN, Sensitivity was the prioritized metric. Accordingly, the threshold was set at Sensitivity values equal or greater than 93% and Accuracy, Precision, and Specificity values equal or greater than 85%. Table 5 exhibits the models that achieved the threshold previously defined by their ranking order.

Table 5. DMM that achieved a performance within the defined threshold

DMT	S	SM	Accuracy	Sensitivity	Specificity	Precision
NN	S2	Split Validation	0.930	0.958	0.901	0.907
NN	S2	Cross Validation	0.926	0.950	0.903	0.910
DT	S2	Cross Validation	0.918	0.950	0.886	0.894
NN	S1	Cross Validation	0.920	0.942	0.899	0.908
NN	S3	Cross Validation	0.914	0.937	0.890	0.902
NN	S1	Split Validation	0.916	0.944	0.887	0.893

By analyzing Table 5, it can be concluded that most DMMs, about 66.67%, use cross validation as sampling method. In addition, 50% of the best-performing models used the S2 scenario, 33.33% used the S1 scenario, and only 16.67% used the S3 scenario. Finally, with respect to the different DMTs used in this study, it is observed that, besides one DMM that used the DT algorithm, the rest of the models, approximately 83.33%, used the NN algorithm. Thus, it is possible to claim that the most suitable model, from all the 60 induced models, is DMM = {Classification, S2, NN, Split Validation, Oversampling (SMOTE), lung_cancer}, achieving approximately 93% of accuracy, 96% of sensitivity, 90% of specificity and 91% of precision.

5 Conclusion

In this study, a DM process was performed with the goal to build an accurate DMM able to predict whether patients have lung cancer or not. Three scenarios were tested, along with five DMT and two SM. All DMM achieved an accuracy greater than 82%. The DMM with the best performance is characterized by a classification approach, the scenario that removes the three features with the least weight on the target attribute prediction according to the Information Gain, Gain Ratio, Gini Index and Chi Squared criteria (S2), the NN algorithm, the Split Validation method and the SMOTE Upsampling technique. The best metric combination achieved was an accuracy of 93.0%, a specificity of 90.1%, a precision of 90.7% and a sensitivity of 95.8%.

Considering the proposed business goal, this study presented promising results. However, for the model to be implemented in a clinical environment and to be able to successfully assist health professionals in their decision-making, some improvements and further testing are required. In the future, a larger number of cases should be studied by integrating new data in the dataset to obtain a balanced distribution between patients diagnosed with lung cancer and patients without lung cancer. New features should be added and the correlation between them could also be analysed. Also, different scenarios and different DMTs, such as Random Forest, could be applied to improve the model's performance and the reliability of the results.

Acknowledgments. This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

1. World Cancer Research Fund. Lung cancer statistics. <https://www.wcrf.org/dietandcancer/cancer-trends/lung-cancer-statistics>. Accessed 10 Nov 2020
2. Hirsch, F.R., Franklin, W.A., Gazdar, A.F., Bunn, P.A.: Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. *Clin. Cancer Res.* **7**(1), 5–22 (2001)
3. Morais, A., Peixoto, H., Coimbra, C., Abelha, A., Machado, J.: Predicting the need of neonatal resuscitation using data mining. *Procedia Comput. Sci.* **113**, 571–576 (2017). <https://doi.org/10.1016/j.procs.2017.08.287>
4. Hand, D.J., Adams, N.M.: Data mining. In: Wiley StatsRef: Statistics Reference Online, pp. 1–7 (2014). <https://doi.org/10.1002/9781118445112.stat06466.pub2>
5. Centers for Disease Control and Prevention. U.S. Cancer Statistics Data Visualizations Tool. <https://www.cdc.gov/cancer/uscs/dataviz/index.htm>. Accessed 10 Nov 2020
6. Torre, L.A., Siegel, R.L., Jemal, A.: Lung cancer statistics. In: Lung Cancer and Personalized Medicine, pp. 1–19. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-24223-1_1
7. Biesalski, H.K., De Mesquita, B.B., Chesson, A., et al.: European consensus statement on lung cancer: risk factors and prevention. Lung cancer panel. *CA Cancer J. Clin.* **48**(3), 167–176 (1998). <https://doi.org/10.3322/canjclin.48.3.167>
8. Bradley, S.H., Kennedy, M.P., Neal, R.D.: Recognising lung cancer in primary care. *Adv. Ther.* **36**(1), 19–30 (2019). <https://doi.org/10.1007/s12325-018-0843-5>
9. Martins, B., Ferreira, D., Neto, C., Abelha, A., Machado, J.: Data mining for cardiovascular disease prediction. *J. Med. Syst.* **45**(1), 1–8 (2021)
10. Krishnaiah, V., Narsimha, G., Chandra, N.S.: Diagnosis of lung cancer prediction system using data mining classification techniques. *Int. J. Comput. Sci. Inf. Technol.* **4**(1), 39–45 (2013)
11. Nasser, I.M., Abu-Naser, S.S.: Lung cancer detection using artificial neural network. *Int. J. Eng. Inf. Syst. (IJE AIS)* **3**(3), 17–23 (2019)
12. Murty, N.R., Babu, M.P.: A critical study of classification algorithms for lung-cancer disease detection and diagnosis. *Int. J. Comput. Intell. Res.* **13**(5), 1041–1048 (2017)

13. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, pp. 29–39. Springer, London (2000)
14. Kaggle – Lung Cancer Dataset By Staceyinrobert. <https://www.kaggle.com/imrkannan/lung-cancer-dataset-by-staceyinrobert>. Accessed 06 Nov 2020
15. Ferreira, D., Silva, S., Abelha, A., Machado, J.: Recommendation system using autoencoders. *Appl. Sci.* **10**(16), 5510 (2020). <https://doi.org/10.3390/app10165510>. MDPI



Mobile Burnout Estimation Device - An Agile Driven Pathway

Raluca Dovleac , Marius Risteiu , Andreea Cristina Ionica ,
and Monica Leba

University of Petrosani, 332006 Petrosani, Romania
andreeaionica@upet.ro

Abstract. The field of health care began to benefit from devices and solutions provided by both small companies and passionate entrepreneurs developing innovative solutions to modern problems. Therefore, we can know witness an increase in the number of engineering solutions that come to the aid of healthcare professionals not only with diagnosing but also providing assistance for medical procedures. In this paper, we look at one such solution – specifically, a project consisting of developing a wearable device that helps identify when a person is experiencing burnout symptoms and providing useful feedback. The wearable device offers a new “sense”, but the person must close the loop and take action with the help of the specialist based on knowledge of coping strategies (cognitive and emotional) in a person-centered approach. In this paper, we looked not only at the device, that is subject of a patent [1], and its capabilities, presented by test results on the developed prototype, but also at its development process in order to help understand what can be improved and how it can become a feasible option for a marketplace, using a model based on a version of the QFD (Quality Function Deployment) method that integrates Agile practices and workflows for monitoring the development process and measuring the outputs.

Keywords: Health · Device · Product development · Agile · QFD

1 Introduction

1.1 Research Context

The research places itself in the context of this Agile framework by analyzing the possibility of implementing a development approach that fits the requirements of a project concerned with the creation of a wearable device for health monitoring and detection of burnout symptoms, which also integrates an alerting component to help the wearer understand what actions need to be taken in order to reduce his/her stress levels and prevent the offset of burnout.

The research is integrated into the unfortunate present trend of increased work-related stress that end up in too many cases of burnout. Concentrated efforts worldwide to include burnout in the category of occupational diseases are evidence of the increased attention paid to this syndrome. The world is “flooded” with wearable devices for health

monitoring, and in this context the proposed prototype comes as a viable solution for a one-time problem that is vital to solve.

For a while, startup companies and small companies began entering the health and medicine field, providing products and solutions to professionals and organizations in the field. These companies typically offer niche products and/or services and specialize in providing solutions to fit specific requirements.

This can prove to be a very valuable contribution since most of the startups enter these markets with innovative products and/or services most of the time. One aspect however that must be taken into account is the reliability of the solutions offered by these companies and even more so, their ability of meeting customer expectations given the constraints regarding the delivery times and the availability of resources.

1.2 Theoretical Background

Since its inception [2], a large number of companies from the software development industry have adopted Agile as an alternative to the traditional development methods [3] due to the increased flexibility of Agile among other added benefits [2, 3].

The Agile approach implies extensive collaboration [4], working software over comprehensive documentation, customer collaboration and adaptability to changes in customer requirements [5]. All these have made Agile suitable to be adopted in most software development processes and, combined with other management tools, in the new innovative dynamic business world of today that deals with many projects which include software but are beyond just software. Recently, many researchers and also practitioners in the field of product development, have explored the necessity of quality management in an agile world.

The first assumption of the research based on the review of literature and the results of our previous research [6] that led to the research questions was related to the usefulness of QFD in an Agile approach to innovative hardware-software product development. Even if there is a temptation to treat the Agile manifesto as a holder of the absolute truth, we must be aware that Agile in practice has evolved. And without entering debates on which came first: Agile or Scrum and whether or not Scrum is Agile, the second assumption of the research is that Scrum is considered an agile framework for developing hardware-software products.

The Scrum as a framework of Agile offers, besides its traditional usage in software development, the simplicity of dealing with unpredictability and solving complex problems, with respect for people and self-organization and also the strictness brought by time-boxed events, or sprints, managed by the Scrum team (Product Owner PO, Scrum Master SM, and Development Team DT).

Scrum is a framework for complex, adaptive problems, that only prescribes what the teams should do, but not how they should do it. Regarding software development teams, it was proved that each team can find its own way to make the Scrum framework work for them in order to reduce the risk of complex work, start delivering value to their stakeholders faster, and become more responsive. Regarding mixed hardware-software development teams, there were signaled problems regarding the synchronization, correlation and integration between the hardware and the software parts at the Scrum team level.

Using the model [6] based on the modified QFD method allows the quantification of the impact of modifications at any point with the help of the computed index (offset), representing the percentage of accomplishment for the product at the current moment. These modifications can have significant implication, especially when some constraints exist (time, fitting into a budget or the availability of the human resource), by introducing some risks that must be known and managed properly in order to not jeopardize obtaining the new product.

1.3 Short Literature Review

This section is dedicated to the description of the tool used to develop the research, that is based on the Quality Function Deployment (QFD) method that was previously applied in many fields.

The role of quality models and methods in today's companies and agile companies has been noted, studies showing that more companies are becoming aware of these aspects and are looking for ways of integrating quality management methods, models and tools for various reasons. Amongst the many quality management resources available, the method known as QFD has been studied and applied in Agile contexts significantly more frequent than the others, with research analyzing how agility and innovation can be achieved through QFD [7, 8], how QFD can complement agile practices [9, 10], QFD as part of a model for e-CRM framework assessment in agile manufacturing [11], QFD as an approach to achieving agility with the help of fuzzy logic [12], and as an enabler for the implementation of Leagile Supply Chain Management [13], for the correlation of Customer-Specific Requirements and System-Inherent Characteristics [14], as a solution for sustainable supply chain management [15], as a modified model for risk management in accordance to the ISO standards [16], as a way of improving requirements management in order to obtain better estimates [17] and applying in the development lifecycle of software products [18].

2 Materials and Methods

The research question that has been asked in the study was:

RQ. Is any Agile framework to be used for developing innovative products that require tight integration of hardware and software components?

In the context of the undertaken research, to answer the research question, the used model based on QFD plays the role of a canvas for structuring the elements specific to Agile, specifically the Scrum framework. Scrum shows what is going on in the development process, but it needs to be enriched with the knowledge provided by the use of other methods and tools in order to deliver the most value in the new product developed. To be sure that the highest possible value product it is delivered, in the Scrum framework the QFD quality management method is embedded, that, as highlighted in the literature review, proved to be a useful method for different types of product development. The role of QFD is to convert the Scrum specificities into a quantifiable form, that is significant and gives a measure of the current state of the product development process, a "snapshot" of the product at any moment.

The research direction (RD) explored is to apply the model, based on the SCRUM framework of the Agile approach and the QFD method of the quality management in the development process for innovative hardware-software products intrinsically mixing the values brought by both Agile and quality management.

The scope of the research is to demonstrate the usefulness of the proposed model for new/innovative products development and illustrate the application on a case study of a mixed hardware-software product, namely a burnout estimation and alarming wearable device (Fig. 1).

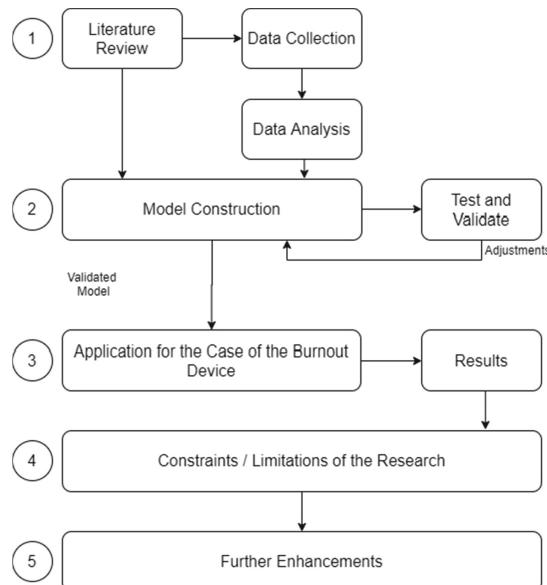


Fig. 1. The research methodology

The model integrates a series of principles, methodologies, concepts and ideas, such as: QFD method, Scrum framework.

The QFD method was mostly applied for iterative product improvement, based on the voice of the customer. But, in case of a new innovative product, it's more difficult to fully articulate the voice of the customer and what is already known about the customer, its needs and preferences can prove to be of great value. This shortcoming is resolved thought the introduction of the Agile manifesto principles based on changing requirements but stable architecture based on a number of well-defined interconnected matrixes, which offers, though the help of an index [16] the possibility of measuring the degree of product accomplishment, from the standpoint of the development team (expressed as how much work has been completed) as well as from the customer's perspective (how much product functionality can be seen).

As a research method we use the case study approach on the burnout estimation and alarming device. This is a device that combines elements of hardware, like sensors and microcontrollers as electronic design, shape and materials as mechanical design with

elements of software, like microcontroller program for physiological data measurement, interpretation and alarming and server program for device setup and data communication [19]. In order to prevent problems that could emerge from hardware-software integration all the sensors data acquisition requires testing in every step, like hardware connection, data communication, data validation, data storage and data interpretation. Another big issue to consider is related to the mechanical design in order to make a good miniaturization of the electronic part but still ensure enough space for the entire software that estimates and alarms regarding the burnout state and also obtain a device that is easy and nice to wear.

The model is implemented starting with understanding and describing the customer requirements or needs. Since the model is based on the Scrum framework, the requirements take the form of User Stories (US). The next step is establishing the degree of importance for each US for the customer, and prioritizing the completion of those requirements that would bring the customer the most satisfaction. Based on this, the PO together with the SM will decide the order and the timeframe for the customer requirements, along with what tasks those requirements imply. From here on, the SM will establish together with the DT how difficult each of the tasks is, and who is responsible for the completion of the tasks, along with the associated timeframe for each task. This information will then be used as inputs for the model in order to determine how much of a US a task is covering and even more, in order to estimate how much work can be achieved and how that will be translated in terms of visible functionality that the customer can experience.

The model can be applied within the innovation departments of companies, as well as in the case of startups, and it is applicable for any kind of hardware-software product.

3 Agile Based Model Application for Product Development

We synthetized in Fig. 2 the model application for the case of the burnout wearable. For the presented project, a number of eight US have been established for the epic which had the goal of building a wearable device that would met the following requirements: monitor and display the user health status (US1); display the status in an intuitive way, by using a color-coded system (US1); assuring the confidentiality of the gathered user data (US2); having an alarming system to inform the user that some action is required regarding his state (US3); having personalization options in order to fit customer preferences (US4); being easy to wear (US5) and resistant to wear and tear (US6); having a long functioning time between charges (US7) and charging rapidly (US8).

Each of the US was graded on a scale from 1 to 10 for its importance in the customer requirements and had one or more tasks that required completion in order to consider it done.

The US degree of importance and the number of tasks required for each US have been established by the development team together with the customer, and can be observed in **US** matrix.

After establishing the US and their degree of importance, the development team had to establish the tasks that were required for each US (US1:8 tasks, US2:3 tasks, US 3:2, US4:3, US5:3, US6:1, US7:2, US8:2) with their degree of difficulty – presented in **T** matrix and the interrelationship between these tasks **TT** matrix.

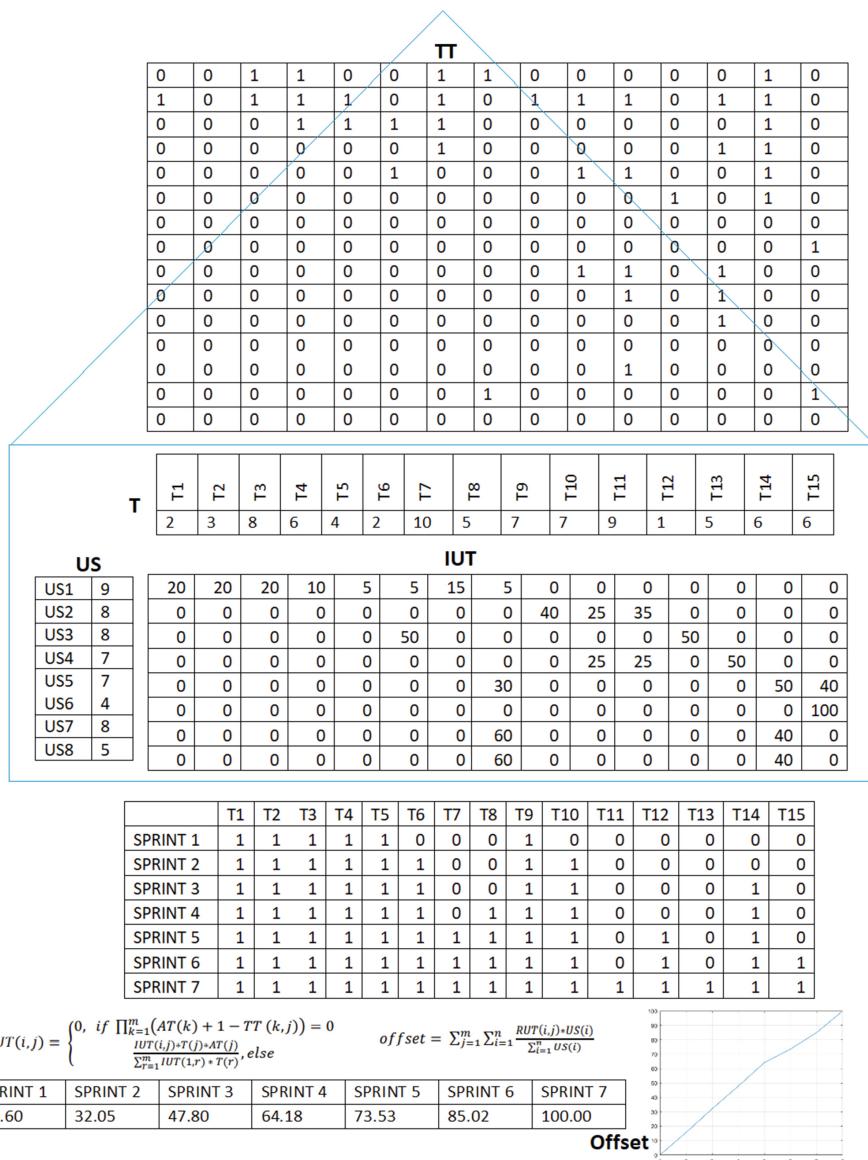


Fig. 2. Results of the Agile based model application for product development

The tasks interrelationship matrix highlights the tasks dependencies, with the role of alerting the development team if one or more tasks cannot be completed unless other tasks are completed first. This information is also useful for calculating the **Offset** indicator value in order to determine which tasks should be considered top priority and be completed so to optimize the development process and to prevent work bottlenecks.

The next step after establishing the US, the tasks required for the completion of each US and the task interdependencies, was the creation of the **IUT** matrix, which highlights how much a task contributes to the completion of a US (this is expressed in percentages).

In the **IUT** matrix for the project can be observed that for most US, except for US6, a number of at least two tasks contribute to their completion. Another thing that can be observed is that US1 is the most complex, and it requires the most tasks for its completion.

Based on this input data, and after establishing the time constraint of the project, a unanimous decision of dividing the work in 7 sprints of 8 days each (covering 56 days for the development of the first epic) has been reached. The decision to divide the tasks and US in short iterative development periods with an output at the end of each iteration has been made in accordance to Agile principles that encourage rapid and iterative releases.

The expected growth rate of the project completion has been established with the help of the offset indicator between 2 consecutive sprints and has been set as at least 12.5% increase per iteration. The results of the expected growth rate are shown in Fig. 2. As it can be observed the project was expected to be half done by the end of the third sprint, and each iteration has a growth rate of at least 12.5% as established.

An algorithm for automatic task division into sprint has been used in order to optimize the decision-making process of selecting the task or tasks that would be best to be completed for each iteration/sprint. The algorithm is subject to a patent registered with the Romanian State Office for Inventions and Trademarks (OSIM) [19] and is using the data about the tasks (in terms of natural order of completion, interdependency, and task difficulty) to assign the tasks that are essential for completion in each sprint.

The results of applying the algorithm with the help of MatLab environment were as follows: First sprint: tasks - 1, 2, 3, 4, 5 and 9; Second sprint: tasks - 6 and 10; Third sprint: task 14; Fourth sprint: task 8; Fifth sprint: tasks 7 and 12; Sixth sprint: task 15; Seventh sprint: tasks 11 and 13.

The average growth rate for each sprint was 14.28% with higher values in the case of the first couple of sprints and lower values towards the end, but cumulating a value of 100% at the end of the last sprint. The growth of the offset indicator value is graphically represented in Fig. 2.

4 Test Results on the Developed Prototype

The Neural Network training [20] was performed using data collected from 30 subjects over the period of one month, data being monitored and interpreted by a psychologist to ensure the consistency of the resulting system. Then, the Neural network integrated in the prototype developed around the Teensy 4.1 system was tested on 5 subjects. Of these, one subject registered a level of burnout higher than 75%, one subject was in the (50–75)% range and the others had a level that was lower than 50% (Fig. 3).

A demonstration of the operation of the system for the possible combinations of the pulse (HR) in (84–98) bps spectra and oximetry (SpO₂) in (96–92)% can be observed in Fig. 4.

As can be seen, there is a directly proportional dependence between pulse and burnout level and inversely proportional dependence between oximetry and burnout. As the

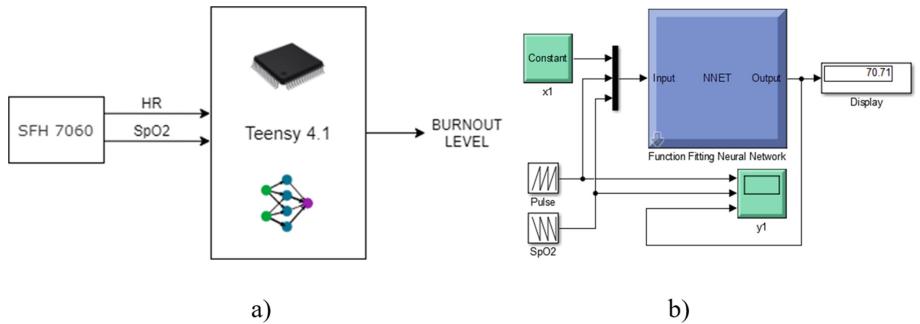


Fig. 3. a) Prototype block diagram; b) System simulation diagram.

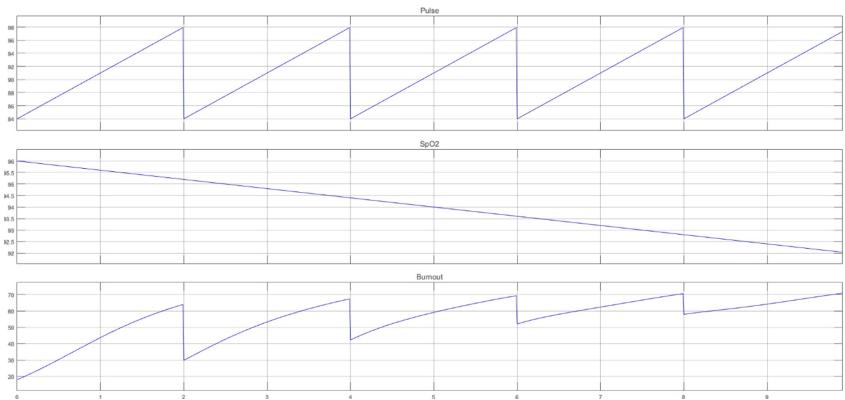


Fig. 4. Simulation results

oximetry decreases, smaller burnout variation is obtained over the same pulse variation interval, but with limit values that indicate high levels of burnout.

5 Conclusions and Discussions

Thus, the proposed model was implemented in the development process of a wearable device that monitors the wearer for signs of burnout and alerts the latter if and when action is required. The model followed the development period of the first epic, which contained a total of eight US and 15 tasks. Applying the model allowed the development team to sketch a clear path to be followed based not only on their previous experience but also taking into account the customer preferences and the complexity of the project and that of the tasks involved in the development of the product.

The paper shows how the project is carried out, highlighting the functionality of the product that can be delivered through the offset values calculated at the level of each sprint until the final product is obtained, when the offset is 100% and all the characteristics necessary for a correct final functionality have been included.

Thus, testing on the 5 subjects showed that the ANN system was well trained based on data from the 30 subjects, so that it provides burnout levels across the spectrum from 0 to 100% based on pulse and oximetry measurements.

The system was developed and tested only on subjects in the field of education, that represents for now a limitation of the research. In order to apply the system for other categories it requires a rigorous testing and possibly a retraining of the network using the same protocol. For this reason, the current and future work is focused on establishing a larger target group and gathering data from this group in order to further customize the network and achieve a general-use burnout estimation device.

References

1. Nassar, Y., Ionica, A., Leba, M.: Burnout Status Identification and Alarming System, 270751/18.11.2019 Israel Patent Office
2. Campanelli, A.S., Parreira, F.S.: Agile methods tailoring – a systematic literature review. *J. Syst. Softw.* **110**, 85–100 (2015)
3. Salah, D., Paige, R.F., Cairns, P.: A systematic literature review for agile development processes and user centred design integration. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London (2014)
4. Inayat, I., Salim, S.S., Marczak, S., Daneva, M., Shamshirband, S.: A systematic literature review on agile requirements engineering practices and challenges. *Comput. Hum. Behav.* **51**, 915–929 (2014)
5. Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M.: Manifesto for Agile Software Development (2001). <https://agilemanifesto.org/>
6. Ionica, A., Leba, M., Dovleac, R.: A QFD based model integration in Agile software development, pp. 1–6 (2017). <https://doi.org/10.23919/CISTI.2017.7975995>
7. Vinodh, S., Sundararaj, G., Devadasan, S., Rajanayagam, D., Murugesh, R.: Agility and innovation through QFD: an endeavour through agile ITQFD technique. *Int. J. Manage. Pract.* **3**(4), 383–404 (2009)
8. Baramichai, M., Zimmers, E.W., Marangos, C.: Agile supply chain transformation matrix: a QFD-based tool for improving enterprise agility. *Int. J. Value Chain Manage.* **3**(2), 282–303 (2007)
9. Tichy, M., Bodden, E., Kuhrmann, M., Wagner, S., Steghöfer, J.-P.: Agile Software Quality Function Deployment (2018)
10. Riesener, M., Rebentisch, E., Doelle, C., Kuhn, M., Brockmann, S.: Methodology for the design of agile product development networks. *Proc. CIRP* **84**, 1029–1034 (2019)
11. Zandi, F., Tavana, M.: A fuzzy group quality function deployment model for e-CRM framework assessment in agile manufacturing. *Comput. Ind. Eng.* **61**, 1–9 (2011)
12. Bottani, E.: A fuzzy QFD approach to achieve agility. *Int. J. Prod. Econ.* **119**, 380–391 (2009)
13. Haq, A.N., Boddu, V.: Analysis of enablers for the implementation of agile supply chain management using an integrated fuzzy QFD approach. *J. Intell. Manuf.* **28**, 1–2 (2017)
14. Schneberger, J.-H., Luedke, T., Vielhaber, M.: Agile transformation and correlation of customer-specific requirements and system-inherent characteristics - an automotive example. *Proc. CIRP* **70**, 78–83 (2018)
15. Büyüközkan, G., Çifci, G.: Extending QFD with pythagorean fuzzy sets for sustainable supply chain management. In: Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making, Istanbul (2019)

16. Carmignani, G., Zammori, F., Cervelli, G.: Modified QFD approach for context analysis and risk management according to ISO standards. In: Twentieth International Working Seminar on Production Economics (2018)
17. Buglione, L., Abran, A., Daneva, M., Herrmann, A.: “Filling in the blanks”: a way to improve requirements management for better estimates. In: Software Quality Assurance, pp. 151–176. Elsevier (2016)
18. Dovleac, R., Ionica, A., Leba, M.: QFD embedded agile approach on IT startups project management. *Cogent Bus. Manage.* **7**(1), 1782658 (2020)
19. Ionica, A.C., Leba, M., Dovleac, R.A.: Method for planning work load, involves determining requests, establishing interdependence between work tasks and prioritizing work tasks using algorithm based on indicator quantifying requests meeting degree of clients. RO133735-A0 Derwent Primary Accession Number: 2019-A0879E (2019)
20. Riurean, S., Leba, M., Ionica, A., Nassar, Y.: Technical solution for burnout, the modern age health issue. In: 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), Palermo, Italy, pp. 350–353 (2020). <https://doi.org/10.1109/MELECON48756.2020.9140516>



Development of Adaptive Software for Individuals with Hearing Loss

Pedro Giuliano Farina¹(✉), Cibelle Albuquerque de la Higuera Amato²,
and Valéria Farinazzo Martins²

¹ Computing and Informatics Department, Mackenzie Presbyterian University,
São Paulo, Brazil

² Developmental Disorders, Mackenzie Presbyterian University, São Paulo, Brazil
{cibelle.amato, valeria.farinazzo}@mackenzie.br

Abstract. This study proposes the creation of a software capable of adapting sounds generated by a smartphone to an intensity range according to the user's needs. Tests were carried out with 11 participants and the results are presented in this work.

Keywords: Hearing loss · Software · Hearing disabilities

1 Introduction

In the area of Physics, sound is defined as a mechanical wave. This wave as far as it is concerned is defined by vibrations propagated in space. Thus, every sound generates a vibration in the medium through which it propagates, but not every vibration produces audible sound for the human being. For example, it is possible to know that when the intensity (dB) of the sound is low, it is difficult to hear [12].

The human ear also perceives the frequency at which the sound vibrates to certain limits. Frequencies 20 Hz and above 20 kHz are inaudible to humans. That is why it is said that the frequency range audible to man is that between 20 and 20,000 Hz [12].

Over time, this capacity deteriorates, a process resulting from physiological degeneration caused by exposure to noise, ototoxic agents and damage caused by disorders and medical treatments [7]. In some cases it is possible to be born with problems or to develop them to the point of losing hearing completely. There are tests that measure this loss accurately and display this information with an audiogram, relating the individual's minimum audible intensity to frequency ranges [1].

This study proposes the development of a software that's able to adapt sounds emitted in a smartphone to an audible intensity range according to the user's needs.

The proposed software for individuals with hearing loss, if eventually developed to a stable point of usability, may, in addition to encouraging greater digital inclusion, also bring a better quality of everyday life for users. Focusing on

meeting the needs of people with hearing loss in mild to severe stages, excluding profound or total loss [2].

Once the prototype was done, usability tests were carried out with 11 people and the results are show in this article. The article is structured as explained below:

- Section 2: Theoretical foundation, necessary for the understanding of the other chapters.
- Section 3: Software development.
- Section 4: Tests Methodology which addresses which public, how it was tested and what risks did it present to the user.
- Section 5 the presentation and discussion of data obtained during the tests.
- Section 6 the conclusion based on the results presented.

2 Hearing Loss

Hearing is one of the pillars of human social interaction, directly impacting how we understand ourselves and how we make ourselves understood. Some of the potential consequences of losing this essential element of the communicative process are: absence from work, resignations, difficulties in acceptance, discrimination and shame [9], which can lead to social exclusion, isolation and even depression [11].

Hearing loss occurs due to countless reasons, and is very present in everyday life. Noise in a work environment is considered the most common harmful physical agent in hearing loss. In a survey with a heterogeneous sample of data, it was found that around 60% of workers exposed to noise and tinnitus had their hearing capacity affected [6].

In another study, it was found that approximately 104 million people in the United States are exposed to noise levels that can damage their hearing, and 1 in 4 adults suffer from hearing loss due to it [3].

Time is also a cause of hearing loss. The percentage of the population with communication difficulties is progressively increased by age. And based on Brazilian studies, it was revealed that around 60% of the elderly population residing in Brazil is affected by hearing loss [7].

In 2018, 28 million elderly people were accounted for in the Brazilian population, representing 13% of the country's population. Percentage which according to the Population Projection tends to double in the coming decades [10], potentially meaning that around 16 million elderly people are affected by hearing problems.

Hearing loss means a lot in the social life of those affected and although in Brazil there are laws that guarantee social inclusion, there is still a lot of prejudice and discriminatory attitudes towards people with a disability. There are personal reports that approach treatment differently from the disabled, causing discomfort and shame of the condition, which in many cases end up leading to social and even professional distance [9].

2.1 Rehabilitation of Hearing Loss

Hearing rehabilitation aims to develop the auditory skills acquired or not after being lost by the hearing impaired. The work is done mainly by the training of detection, discrimination, recognition and understanding of sounds, with the aid of devices that can amplify sounds. This procedure should only start after the diagnosis of hearing loss by a doctor and consultation with a speech therapist, who will analyze the patient's hearing loss and recommend the ideal model for each case [8].

The rehabilitation of hearing capacity is therefore an attempt to bring the impaired individual back to a common routine, allowing and effecting their social insertion. In cases of hearing loss at non-permanent levels, there is even the possibility of reducing the existing loss [5, 11].

The greatest difficulties reported in this process are usually found in the digital environment, such as using the phone, listening to the radio or television, even when the patient is wearing a prosthesis, which may be related to the fact that in the digital environment it is not possible to establish direct contact, making it difficult if not impossible to read lips and expressions [4].

3 Software Development

3.1 Requirement Analysis

- The software must accept any kind of sound file.
- The software must boost the intensity of the sound.
- The user must have access to his audio files.
- The user must be able to control the intensity boost.
- Must be intuitive.
- Must work with the Operational System to assure usability.
- Has no needs of internet connection to work.

3.2 Project

In the software development the integrated development environment chosen was Xcode¹, using Swift Language and AudioKit², an open source framework of sound processing.

3.3 Implementation

The software implementation consists of an application developed for the iOS platform; its execution is completely transparent to the user and is integrated with the operating system. The following minimum requirements for its correct operation were considered: an iPhone with an operating system iOS 14.0 or higher and the installation of the application.

To use the app, all the user has to do is select an audio to start the interaction, which can be done:

¹ <https://developer.apple.com/xcode/>.

² <https://audiokit.io>.

- Forwarding any file through the Operational System (Fig. 1).
- Opening the app and selecting an audio in the list of files (Fig. 2).
- Opening the app and selecting an audio from the recent file (Fig. 3).

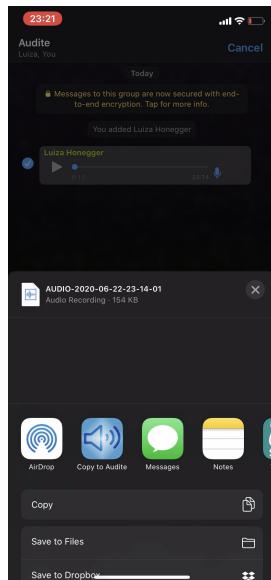


Fig. 1. Forwarding audio

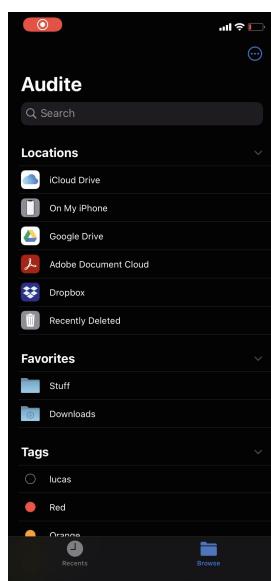


Fig. 2. File browsing

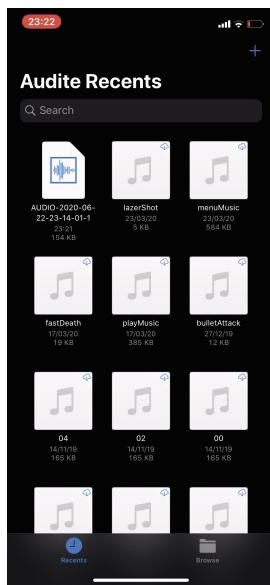


Fig. 3. Recents audios

Once selected, it opens a slider for the user to control the boost (Fig. 4).

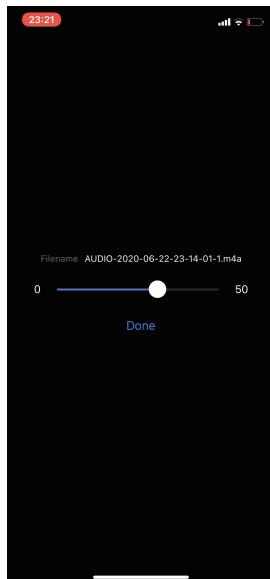


Fig. 4. Audio playing with intensity control slider

4 Usability Tests

4.1 Participants

The study participants were 11 people over the age of 18, who suffer from hearing loss, through the criteria for recruiting participants for convenience. Thus, inclusion criteria were being 18 years of age or older, of both sexes, with hearing loss. Exclusion criteria for the study were to present profound or total deafness, to be illiterate or not to have an iPhone smartphone.

Respecting all ethical precepts, after approval by the institution's ethics committee (approval number CAAE 37127820.6.0000.0084), all participants complied with the informed consent form, through their acceptance via website. Due to the Covid-19 pandemic, the tests were carried out online, in order not to compromise the physical integrity of the participants.

4.2 Location

The research's location was carried out was in the participant's residence or other place of convenience due to the restrictions of social isolation.

4.3 Procedures

Having chosen to participate in the research, in a quiet environment, the participant should, using an iPhone smartphone, click on a hyperlink that was made available to him to automatically download the Audite application.

For about 15 min, the participant performed some proposed tasks, in a single session.

The tasks to be performed by the research participant were:

- 1) Install the software through TestFlight.
- 2) Select an audio in the device to check if the system can identify audio files.
- 3) Forward an audio received in another app directly to Audite.
- 4) Modify the intensity boost of a sequence of prerecorded audios selected to test the efficiency of the software.

In order for the participant to be able to perform these tasks, if he decides to participate in the research, a short tutorial on how to perform the tasks was also sent.

As soon as he finished, he was asked to answer a very simple and quick form with the following questions:

- Age
- Biological sex
- Stage of hearing loss
- App familiarity
- I'd like to use this product in everyday life
- I heard better thanks to the app

- The interaction of the software was clear
- I had to learn new things to use the system
- What you liked the most?
- What you disliked?
- What could've been better?

4.4 Risks and Benefits

The research brings minimal risks to the research participants, which correspond to the inherent fatigue of using a software. Although the estimated time for using and answering the form should not exceed 20 min, if the participant feels tired, the test can be paused. If he wishes to continue at another time, it will be possible to resume; otherwise, the test may be canceled, without prejudice to the participant.

On the other hand, the direct benefits of this research to the participant, are that the experience is expected to be useful to the user so that its use is beneficial in their daily lives. For society, the proposed adaptation software for individuals with hearing loss, if eventually developed to a stable point of usability, may, in addition to encourage greater digital inclusion, also bring a better quality of everyday life for users, focusing on meet the needs of people with hearing loss in mild to severe stages, excluding profound or total loss.

5 Results and Discussions

5.1 Interviews with Speech Therapists

Once the prototype was ready, 3 meetings were held with speech therapists to certify that the application would be beneficial to the user. The following positive points were raised:

- The application cannot be harmful to hearing due to the limitation of the cell phone.
- It should work well specially for people who are losing their hearing but still not enough for a hearing aid.
- Seems easy to use for the everyday user.
- The interaction is simple.
- Works with other apps like WhatsApp

Among the negative points raised by the professionals are:

- Should not exceed 120 dB for users with headphones.
- Whether it was dangerous for the user to keep using the app instead of looking for help.
- Whether the app could reach all the categories of hearing loss.
- Whether it is intuitive enough for the elderly people.

Due to these points, a second version of the prototype had already begun to take shape:

- Implement an audio lock to be sure no audio goes over 120 dB.
- Alert the user when he's using the app too frequently to look for a doctor.
- Rethink the design and user experience and develop an tutorial of how to use the app.

5.2 Results of the Tests

The tests were carried out with a group of 11 participants, composed of men and women, from 22 to 78 years old. At the end of each test, 11 questions were asked users that could be answered by:

- Selecting an option (biological sex and stage of hearing loss).
- Rating from 1 to 5 stars.
- Writing.

Quantitative Data. When asked about the interaction with the software and the need for learning, four showed that they needed assistance to understand and learn how to use the application, as shown in Fig. 5.

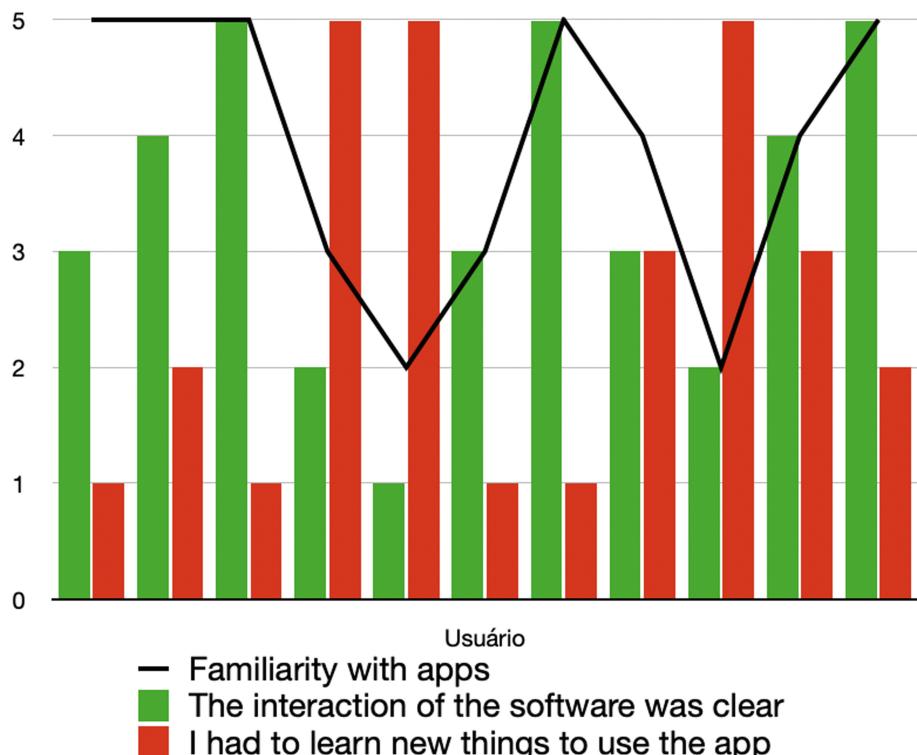


Fig. 5. Usability graph

According to the results found in Fig. 5, it is clear that those people who needed to learn less to use the software considered the interaction with the software more clear. This is quite reasonable, as users do not like having to learn to use an application, especially when it comes to the elderly audience that is less digitally included.

On the other hand, the data about the quality of the experience and intention of use show great acceptance of the product (Fig. 6).

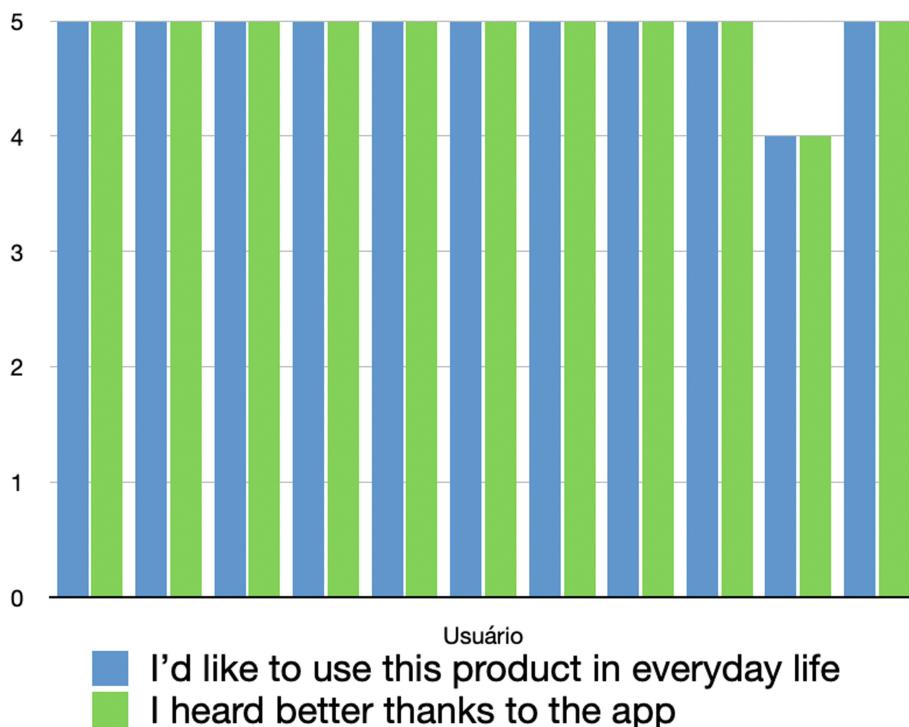


Fig. 6. Opinion graph

Now about the results presented in Fig. 6, it can be seen that the application in question had a high degree of approval from the 11 participants, both in relation to the motivation to use it again and to its effectiveness, that is, they were able to better hear the audio when they used the application. Only 1 of the participants did not give a maximum score for these two items, and even so, 80% of satisfaction.

Qualitative Data. The following are the written responses that users have answered about the application (Table 1).

Table 1. Written answers (these were originally written in portuguese, and translated for this submission)

Age	Degree of hearing loss	What you liked the most?	What you disliked?	What could've been better?
22	Moderate	The clear difference that the audio gets easier to hear with the app. It's a very good and important idea	Not a clear path inside the app	Could've instructions inside the app to help out the user
25	Mild	I was able to forward an audio from WhatsApp	The silent mode got in the way, since I always use my phone on silent	The sound should've worked even in silent mode
28	Moderate	Impossible to understand the male's and female's voice without the app!	I liked everything	The app itself doesn't need to improve, although iOS in general needs to be louder!
74	Mild	Hearing with more clarity	I didn't understand how to use it at firsthand	Should be simpler and easier to access
64	Moderate	It improved my hearing and understanding	It has to be on iPhone	Have a tutorial
67	Harsh	Of the proposition of the app. It helps to hear everything clearer	Nothing	I didn't find it easy to use
48	I don't know/would rather not say	The efficiency of upgrading the quality of the sound	Restrict access to iOS	I can't tell
64	I don't know/would rather not say	The objective of everyday use	Nothing	Open use for Android
71	Moderate	Hearing better	I should be able to boost even more than 50 dB	Increase even more the volume
59	Mild	The innovative idea	The difficulty of understanding the use	Be easier for those who are not used to technology
78	Moderate	The everyday use case and integration to other apps	Small sound sample to be tested	Be available on Android and with no boost limit

Analyzing the results presented in Table 1, it can be said of the 11 research participants, the strongest point of the research is consistent with the main objective, which is the improvement of the perception/understanding of the audio by the participants who have hearing problems. As weaknesses of the application, the participants brought, they are related to the difficulty in using the application, mainly because it is limited to the iPhone. And finally, possible improvements are: being available for Android system and improving usability for those who know little about technology.

6 Conclusion

This work presented the use of amplification of sound intensity in an application for iOS smartphones as assistive technology to assist people with hearing loss.

The limitations of this work were mainly due to the choice of the smartphone system, which does not allow the developer to have control over the audio played on the device at all times. Restricting the scope of the project to work only with audio files, making it impossible to access calls, streaming systems, and among other forms of consumption of sound media. In addition to these, there was also the limitation of the pandemic, which limited the amount and monitoring of tests, restricting the public to users already iOS.

According to the tests presented, it can be concluded, for these participants that there is a market search for this product, the proposal is solid for everyday life and improves the quality and understanding of the sound for users, proving to be beneficial.

As future work, there's the possibility to redo the prototype design, create a tutorial so that the path to the application is clear and easy to use and understand, as well as guarantee its operation outside the iPhone's silent mode, and a version Android to serve all audiences. Ideally, this project serves as a proof of concept so that in the future the operating systems themselves already handle this type of facility for the user, ensuring a more complete and comfortable experience.

Acknowledgment. The work was supported by the Coordenação de Aperfeiçoamento de Pessoal de nível superior - Brazil (CAPES) - Programa de Excelência - Proex 1133/2019.

References

1. Brender, Burke, and Glass: Audiometry | Otolaryngology (2006)
2. Cochlear: Types and causes of hearing loss (2020)
3. Cunningham, L.L., Tucci, D.L.: Hearing loss in adults. *New England J. Med.* **377**, 2465–2473 (2017)
4. de Barros, S., Fernanda, P., Queiroga, M., Arruda, B.: As dificuldades encontradas no processo de adaptação de aparelho de amplificação sonora individual em indivíduos idosos. *Rev. CEFAC* **8**(3), 375–385 (2006)
5. de O. Marques, A.C., Kozlowski, L., Marques, J.M.: Reabilitação auditiva no idoso. *Rev. Brasileira Otorrinolaringol.* **70**(6), 806–811 (2004)
6. Dias, A., Cordeiro, R., Corrente, J.E., de Oliveira Gonçalves, C.G.: Associação entre perda auditiva induzida pelo ruído e zumbidos. *Cad. Saúde Pública* **22**(1), 63–68 (2006)
7. dos Santos Baraldi, G., de Almeida, L.C., de Carvalho Borges, A.C.: Evolução da perda auditiva no decorrer do envelhecimento. *Assoc. Brasileira Otorrinolaringol. Cirurgia Cérvico-Facial* **73**(1), 64–70 (2007)
8. Fonseca, C.: Reabilitação Auditiva: entenda a importância de tratar a perda auditiva (2014)

9. Francelin, M.A.S., Motti, T.F.G., Morita, I.: As implicações sociais da deficiência auditiva adquirida em adultos. *Saúde Soc.* São Paulo **19**(1), 180–192 (2010)
10. IBGE: Idosos indicam caminhos para uma melhor idade (2019)
11. Margall, S.A.C., Honora, M., Carlovich, A.L.A.: A reabilitação do deficiente auditivo visando qualidade de vida e inclusão social. *O Mundo Saúde São Paulo* **1**, 123–128 (2006)
12. Rui, L.R., Steffani, M.H.: Física: som e audição humana. recurso didático (2006)



Ensemble Regression for Blood Glucose Prediction

Mohamed Zaim Wadghiri¹, Ali Idrri^{1,2(✉)}, and Touria El Idrissi¹

¹ SPM Research Team, ENSIAS, Mohammed V University in Rabat, BP 713, Rabat, Agdal, Morocco
ali.idri@um5.ac.ma

² MSDA, Mohammed VI Polytechnic University, Ben Guerir, Morocco

Abstract. Background: Predicting blood glucose present commonly many challenges when the designed models are tested under different contexts. Ensemble methods are a set of learning algorithms that have been successfully used in many medical fields to improve the prediction accuracy. This paper aims to review the typology of ensembles used in literature to predict blood glucose.

Methods: 32 papers published between 2000 and 2020 in 6 digital libraries were selected and reviewed with regard to: years and publication sources, integrated factors and data sources used to collect the data and types of ensembles.

Results: This review results found that this research topic is still recent but is gaining a growing interest in the last years. Ensemble models used often blood glucose, insulin, diet and exercise as input to predict blood glucose. Moreover, both homogeneous and heterogeneous ensembles have been investigated.

Conclusions: An increasing interest have been devoted to blood glucose prediction using ensemble methods during the last decade. Several gaps regarding the design of the reviewed ensembles and the data collection process have been identified and recommendations have been formulated in this direction.

Keywords: Blood glucose prediction · Ensemble methods · Homogeneous ensembles · Heterogeneous ensembles

1 Introduction

Diabetes is a chronic disease caused by a disorder in glucose and insulin metabolism leading to abnormal blood glucose levels (BGL) [1]. When it is not well managed, diabetic patients could face higher risks of complications including cardiovascular diseases, kidney damage, coma, or even death [2, 3].

Diabetic patients need to recurrently measure their BGL to maintain a good control of their glycaemia using either Continuous Glucose Monitoring Sensors (CGMS) or manual sticks [4]. Predicting BGL can assist patients to regulate their glycaemia and avoid hyper- and hypo-glycemic episodes [5].

To predict BGL, Machine Learning (ML) techniques have become widely used in data mining problems [6]. However, due to the complexity of BGL dynamics, using a single technique to predict BGL might not provide accurate values and doesn't capture

in general intra- and inter-patients changes [7]. Ensemble methods are learning algorithms that aggregate multiple ML techniques into one predictive model using specific combination rules. This process helps to reach higher accuracy and to achieve a better variance/bias trade-off [8]. These techniques have been successfully used in many eHealth fields including cardiology, oncology and endocrinology [9–11].

The aim of this paper is to present a systematic mapping study (SMS) to identify and analyze the typology of ensembles adopted in literature to predict BGL. 32 papers published between 2000 and November 2020 were identified by searching in 6 digital databases. The selected studies were discussed through answering 3 Mapping Questions (MQs) with respect to: 1) year and source of publication, 2) incorporated inputs and data sources used to collect the information, 3) adopted ensemble types.

The rest of the paper is organized as follows: Sect. 2 highlights the related work. Section 3 describes the research methodology. Section 4 reports the results which are discussed in Sect. 5. Conclusion and study implications are reported in Sect. 6.

2 Related Work

Several ML techniques have been investigated in the literature to predict BGL in diabetic patients including Artificial Neural Networks (ANN), Support Vector Regression (SVR) and Genetic Programming [7]. Many reviews have also been published to assess data mining techniques and ML approaches in diabetes self-management and BGL prediction. Woldaregay et al. [7] conducted a literature review regarding blood glucose prediction using ML strategies in type 1 diabetes. El Idrissi et al. [12, 13] performed a systematic mapping and review study on the use of predictive techniques in diabetes self-management. Kavakiotis et al. [14] conducted a systematic review of the applications of ML and data mining methods in diabetes research. Moreover, Oviedo et al. [15] presented a methodological review of models for predicting BGL.

3 Research Methodology

This SMS adopted the guidelines suggested by Kitchenham and Charters [16] and Petersen et al. [17]. The systematic map process consists of five steps: (1) mapping questions, (2) search strategy, (3) study selection, (4) data extraction, and (5) data synthesis. All these steps are detailed in the following subsections.

3.1 Mapping Questions

The aim of this paper is to analyze and review studies dealing with the application of ensemble methods to predict BGL. Toward this aim, 3 MQs were identified as shown in Table 1.

Table 1. Mapping questions

Mapping question	Motivation
MQ1- In which sources and years have the studies been published?	To identify where studies can be found and how it evolves over time
MQ2- What information was fed as input to predict BGL using ensembles and which data sources were used to collect the data?	To determine which factors affecting BGL dynamics have been explored and how data has been collected
MQ3- What are the most used ensemble types for BGL prediction?	To determine if the ensembles are homogeneous or heterogeneous

3.2 Search Strategy

The search strategy aims to find primary studies that would answer the MQs as defined in Table 1. A search string was put in place to find relevant papers in literature. We derived the main terms from the identified MQs, searched for alternative spellings and then for alternative synonyms, and finally incorporated the main terms using AND operator and joined synonyms using OR operator. The final search string used in this SMS is the following: (“blood sugar” **OR** “blood glucose” **OR** “glucose concentration”) **AND** (predict* **OR** forecast* **OR** control* **OR** estimat* **OR** regress* **OR** manag* **OR** monitor* **OR** evaluat* **OR** assess*) **AND** (ensemble **OR** taxonomy **OR** committee **OR** resampling **OR** fusion **OR** bagging **OR** “random forest” **OR** boost* **OR** stacking **OR** combin* **OR** cluster* **OR** bootstrap* **OR** meta*).

Six digital libraries have been selected to search for relevant papers: Science Direct, PubMed, IEEE Xplore, SpringerLink, ACM Digital Library and Google Scholar.

3.3 Study Selection

The candidate papers are assessed to include only the relevant studies that pass through a set of inclusion criteria (IC) and exclusion criteria (EC) as presented in Table 2. IC and EC are linked by the OR boolean operator. A paper is retained if it meets at least one IC and no EC and is rejected if it meets at least one EC.

Table 2. Inclusion and exclusion criteria

Inclusion criteria (IC)	Exclusion criteria (EC)
IC1: Papers presenting newly developed ensemble predictive techniques of BGL	EC1: Papers considering medical disciplines other than diabetes
IC2: Papers evaluating existing ensemble predictive techniques of BGL	EC2: Papers considering medical tasks other than blood glucose prediction
IC3: Papers providing comparisons of ensemble predictive techniques of BGL	EC3: Papers published before 2000
IC4: Papers presenting an overview of studies investigating ensemble predictive techniques of BGL	EC4: Abstracts and papers written in a language other than English

3.4 Data Extraction

After selecting the relevant papers to include on this review, we used the data extraction form presented in Table 3 to collect the required information to answer the MQs.

Table 3. Data extraction form

MQ1	Publication year, publication source, publication type (journal, conference, book)
MQ2	Included features to predict BGL, data sources used to collect the patients' data
MQ3	Ensemble type used [18]: Homogeneous ensemble or Heterogeneous ensemble

3.5 Data Synthesis

Once the data is extracted from all the relevant papers, it is summarized with respect to each MQ. The vote counting method was adopted to aggregate the results then a narrative synthesis was used to discuss the obtained results regarding each MQ.

4 Results

This section presents how the search process has been executed to obtain the list of the selected studies. The results of each MQ are presented on a dedicated subsection.

4.1 Overview of Selected Studies

As illustrated in Fig. 1, the first step was to execute an automated search on the 6 digital libraries which gave 2726 papers. 2107 unique papers were identified after removing duplicates (a). Applying the inclusion and exclusion criteria discarded 2076 papers

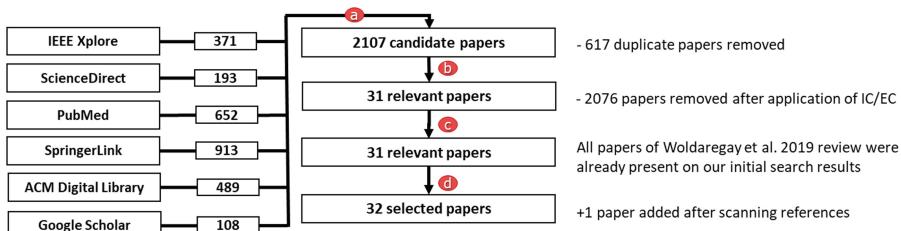


Fig. 1. Search process. The 32 selected along with the required information to answer the MQs are available upon request by email to the authors of this study.

leaving 31 relevant papers (b). Studies from [7] were scanned as well in order to integrate any omitted publications (c). 7 papers dealing with ensemble techniques have been found in [7], all of them were already present on the initial automated search. By scanning the references of each relevant paper, one additional publication was added to the selection (d).

4.2 MQ1: In Which Sources and Years have the Studies been Published?

This MQ aims to identify the research trends over the time and the channels used to publish the selected papers. Figure 2 shows the distribution of the studies according to their publication channel over the years. No paper has been published before 2012 and at least one has been published each year since this date. A higher publication rate is noticed during the last three years as 75% of the papers have been published on 2018 and beyond. Among 32 selected papers, 21 studies were presented at conferences (65.62%) and 11 others were published on journals (34.38%). Many channels have been used but no source comes up more than once except 3 conferences.

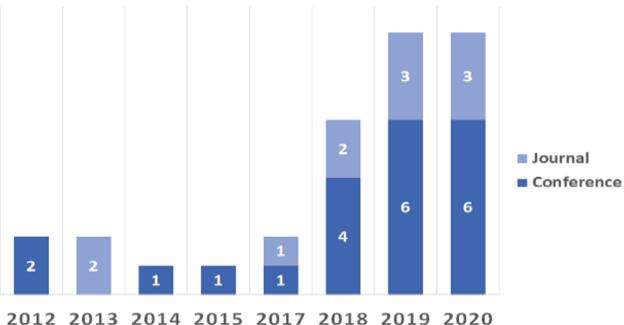


Fig. 2. Distribution of papers over years

4.3 MQ2: What Information was Fed as Input to Predict BGL and Which Data Sources were Used to Collect the Data?

In order to build an accurate model to predict BGL, the most impacting factors should be incorporated to maximize the prediction accuracy. Four main inputs have been identified as follows: Blood Glucose (BG) was used as input in all the papers, insulin was used in 17 papers, meals in 14 papers and physical activity in 10 papers.

The second objective of this MQ is to identify how these four identified main factors have been collected to provide them as inputs to the proposed models.

BG Data Sources: 28 papers used a CGMS to record the blood glucose level (87.5%) while 2 papers used a simulator to generate the data (6.25%). 1 paper performed both experiences using CGMS records and simulators respectively (3.125%) and 1 paper did not provide details about the used technique (3.125%).

31 experiences have been performed by the 30 papers that used CGMS input. Different brands have been used where Medtronic ranked first with 9 experiences (29.03%) followed by Abbott Freestyle with 3 experiences (9.68%), then Guardian Real-Time with 2 experiences (6.45%) and Dexcom with 1 experience (3.23%). The remaining 16 experiences did not provide the used CGMS brand (51.61%). All the papers that simulated the BGL data used the UVA/Padova simulator.

Insulin Data Sources: 17 papers used insulin as input to the proposed models. 11 papers used insulin pumps to record the data (64.71%), 1 paper used a simulator (5.88%), 1 paper used a diary where patients self-report the injected insulin bolus (5.88%) and 4 papers did not provide how insulin data has been captured (23.53%).

Different brands were used by the 11 papers that recorded insulin data from pumps. 1 paper used two brands which makes a total of 12 experiences. 9 experiences used Medtronic (75%) and 3 experiences did not specify the insulin pump brand (25%). All the papers that simulated the insulin bolus used UVA/Padova simulator.

Meal Data Sources: 14 papers used meals as input to the proposed model. 1 paper used 2 methods which makes a total of 15 experiences. 6 experiences used a diary where their food intake is manually reported (40%) and 2 experiences used photos to estimate the food intake (13%). 6 experiences did not specify how data was collected (40%) and 1 experience used UVA/Padova Simulator to simulate meal data (7%).

Physical Activity Data Sources: 10 papers used physical activity data to predict BGL. 4 papers used activity trackers (40%) where Basis Peak was used on 2 papers while Fitbit and SenseWear Armband were used on 1 paper each. On the other hand, 4 papers used a diary allowing patients to report their physical activity habits (40%). Finally, 2 papers did not provide how the physical data have been monitored (20%).

4.4 MQ3: What are the Most Used Ensemble Types for BGL Prediction?

Ensemble methods can be classified into two categories: Homogeneous ensembles where one single technique is used either with 1) different configurations or 2) a meta-algorithm,

and Heterogeneous ensembles where at least two different single techniques are used [15]. The objective of this MQ is to examine the types of ensembles used by the selected papers to predict BGL. For the 32 selected papers, 48 ensemble models have been proposed. 21 papers out of 32 used homogeneous ensembles and proposed 28 ensemble models (58.34%). On the other hand, the remaining 11 papers investigated heterogeneous ensembles and proposed 20 ensembles (41.66%).

5 Discussion

5.1 MQ1: In Which Sources and Years have the Studies been Published?

The publication chronology chart (Fig. 1) shows that BGL prediction using ensemble techniques is a recent research topic as all the studies have been published after 2012 and 75% (24 papers) of the papers were published after 2018. This interest can be attributed to: 1) the positive feedback of those techniques in many fields where models that used ensemble methods outperformed models based on single techniques only [8–10]; 2) the complexity of BG dynamics where a single model shows quickly its limitations to capture inter- and intra-patients changes [7]. Furthermore, the selected studies were published on both journals and conference proceedings. 21 papers out of 32 were presented at conferences and 8 out of 11 journal papers (72.73%) have been published after 2018. This confirms that research has not reached its maturity yet. Moreover, no specialized publication channel was identified as 25 sources have been listed for 32 studies. This diversity can be explained by the multidisciplinary of the topic that deals with artificial intelligence, computer science and medicine.

5.2 MQ2: What Information was Fed as Input to Predict BGL Using Ensembles and Which Data Sources were Used to Collect the Data?

35% of the papers used BGL history data only. This can be explained by the difficulty to incorporate the data of other factors in opposition to BGL which is simple to gather from CGMS and simply fed as a time series to the model [8]. Insulin, food intake and physical activity are the next most investigated inputs as they can be easily quantified and can be gathered through wearable devices.

Most of the papers used CGMS to record BGL. This can be attributed to their wide availability, ease-of-use and cost-effectiveness [19]. This signifies that the models were trained with reliable data, automatically recorded instead of being self-reported.

The administered insulin was captured in general from insulin pumps. This implies again that reliable data was used in most cases to train the proposed models [20].

The food intake information was reported either through a diary or by taking photos of the meals to estimate the corresponding carbohydrate quantity. Both methods are not reliable and a simple omission can weaken the BGL prediction accuracy [21].

The physical activity was collected either through wearable devices or via a diary. The first method is a simple and reliable way to automatically report the patients' activity and record many other physiological data. However, self-reported diaries are prone to errors and hard to represent as input to the model [21].

On the other hand, some papers used a simulator to generate patients' data. UVA/Padova was the only simulator used on all the experiments in this review. In general, simulated data has the advantage to be reproducible and available for comparison but does not always reflect real physiological dynamics [22].

5.3 MQ3: What are the Most Used Ensemble Types for BGL Prediction?

21 out of 32 studies used homogeneous ensembles and proposed 28 ensemble learning models. Most of the homogeneous models used a meta-algorithm to combine multiple instances of a single regressor (19 out of 28 ensembles). Bagging [23–25] and Boosting [26–28] are the two meta-regressors used to construct the ensembles to predict BGL with 11 and 8 models each respectively. The remaining 9 models used a simple combination rule to aggregate results of the base learners. Random Forest regression remains one of the most used homogeneous ensembles as it presents high performance accuracies and are easier to interpret and understand. A Random Forest (RF) of 1,500 decision trees [29] was the largest proposed ensemble model. 2 other papers used RF of 100 decision trees [30, 31] and one paper investigated the combination of 100 Grammatical Evolution models [32]. Boosting meta-algorithms were investigated through XGBoost and LightGBM implementations. They were both published recently but are attracting a growing interest as they demonstrate high accuracy and can be trained in parallel reducing the computation time. In fact, the 6 papers that explored the former meta-algorithms were all published after 2018.

Heterogeneous ensembles were used on 11 papers with 20 proposed ensembles. Stacking is the most investigated meta-algorithm in heterogeneous ensembles with 9 proposed models. Moreover, one paper explored the use of both Bagging and Boosting to build heterogeneous ensemble models [25].

Homogeneous and heterogeneous ensembles were both investigated with a slightly larger number of homogeneous models. In fact, they are easier to understand and simpler to implement. Nevertheless, an interest almost equivalent has also been devoted to heterogeneous ensembles. This can be attributed to the complexity of BGL prediction in various contexts that encouraged researchers to examine heterogeneous ensembles as well since they offer built-in diversity.

Many single techniques were adopted to construct the proposed ensembles where ANN, decision trees, Autoregressive models and SVR were the most investigated. In fact, these techniques have been widely used in literature to predict blood glucose [7]. The investigation of other configuration of these techniques or other exotic and well-known performing methods would help in building more accurate ensembles.

6 Conclusion and Study Implications

In this paper, a SMS has been conducted to analyze the typology of ensemble techniques used in BGL prediction. 32 papers published between 2000 and November 2020 have been reviewed with regard to 3 MQs. The results showed that the use of ensemble methods in predicting BGL is a recent research topic gaining more interest since 2012 and more particularly after 2018. Studies have been published in various channels due to the

multidisciplinary aspect of the topic. Different inputs were incorporated in the prediction models and BGL history, insulin bolus, diet and physical activity were the most investigated features. However, many other information such as stress level or other illnesses should be investigated and integrated in future models to obtain well-performing prediction models. We also noticed that meals information is most of the time self-reported in opposite to BGL, insulin and physical activity that are often collected automatically which might affect the performance of the proposed predictors. 48 ensemble models have been proposed in the 32 identified papers. Both homogeneous and heterogeneous ensembles have been investigated to predict blood glucose. However, larger ensemble sizes should be investigated and a process to select the base learners will help in finding the best performing combinations.

References

1. Williams, G., Pickup, J.C.: *Handbook of Diabetes*. Wiley-Blackwell, Malden, Mass (2004)
2. Leon, B.M., Maddox, T.M.: Diabetes and cardiovascular disease: epidemiology, biological mechanisms, treatment recommendations and future research. *World J. Diab.* **6**, 1246–1258 (2015)
3. Chen, C., Wang, C., Hu, C., Han, Y., Zhao, L., Zhu, X., Xiao, L., Sun, L.: Normoalbuminuric diabetic kidney disease. *Front Med.* **11**, 310–318 (2017)
4. Khadilkar, K.S., Bandgar, T., Shivane, V., Lila, A., Shah, N.: Current concepts in blood glucose monitoring. *Indian J. Endocrinol. Metab.* **17**, 643 (2013)
5. Abraham, S.B., Arunachalam, S., Zhong, A., Agrawal, P., Cohen, O., McMahon, C.M.: Improved real-world glycemic control with continuous glucose monitoring system predictive alerts. *J. Diabetes Sci. Technol.* **15**(1), 91–97 (2019). 1932296819859334
6. Teng, X., Gong, Y.: Research on application of machine learning in data mining. *IOP Conf. Ser.: Mater. Sci. Eng.* **392**, 062202 (2018)
7. Woldaregay, A.Z., Årsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T., Hartwigsen, G.: Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes. *Artif. Intell. Med.* **98**, 109–134 (2019)
8. Seni, G., Elder, J.F.: Ensemble methods in data mining: improving accuracy through combining predictions. *Synth. Lectures Data Mining Knowl. Disc.* **2**, 1–26 (2010)
9. Hosni, M., Carrillo de Gea, J.M., Idri, A., El Bajta, M., Fernández Alemán, J.L., García-Mateos, G., Abnane, I.: A systematic mapping study for ensemble classification methods in cardiovascular disease. *Artif. Intell. Rev.* 1–35 (2020)
10. Hosni, M., Abnane, I., Idri, A., Carrillo de Gea, J.M., Fernández Alemán, J.L.: Reviewing ensemble classification methods in breast cancer. *Comput. Methods Programs Biomed.* **177**, 89–112 (2019)
11. Fernández-Alemán, J.L., Carrillo-de-Gea, J.M., Hosni, M., Idri, A., García-Mateos, G.: Homogeneous and heterogeneous ensemble classification methods in diabetes disease: a review. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2019)
12. El Idrissi, T., Idri, A., Bakkoury, Z.: Systematic map and review of predictive techniques in diabetes self-management. *Int. J. Inf. Manage.* **46**, 263–277 (2019)
13. El Idrissi, T., Idri, A., Bakkoury, Z.: Data mining techniques in diabetes self-management: a systematic map. In: Rocha, Á., Adeli, H., Reis, L.P., Costanzo, S. (eds.) *Trends and Advances in Information Systems and Technologies*. Springer, Cham (2018)

14. Kavakiotis, I., Tsavos, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I.: Machine learning and data mining methods in diabetes research. *Comput. Struct. Biotechnol. J.* **15**, 104–116 (2017)
15. Oviedo, S., Vehí, J., Calm, R., Armengol, J.: A review of personalized blood glucose prediction strategies for T1DM patients. *Int. J. Numer. Methods Biomed. Eng.* **33**, e2833 (2017)
16. Kitchenham, B.A., Budgen, D., Pearl Brereton, O.: Using mapping studies as the basis for further research – a participant-observer case study. *Inf. Softw. Technol.* **53**, 638–651 (2011)
17. Petersen, K., Vakkalanka, S., Kuzniarz, L.: Guidelines for conducting systematic mapping studies in software engineering: an update. *Inf. Softw. Technol.* **64**, 1–8 (2015)
18. Zhou, Z.-H.: Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC, Boca Raton (2012)
19. Fonda, S.J., Graham, C., Munakata, J., Powers, J.M., Price, D., Vigersky, R.A.: The cost-effectiveness of real-time continuous glucose monitoring (RT-CGM) in type 2 diabetes. *J Diabetes Sci. Technol.* **10**, 898–904 (2016)
20. Freckmann, G., Kamecke, U., Waldenmaier, D., Haug, C., Ziegler, R.: Accuracy of bolus and basal rate delivery of different insulin pump systems. *Diabetes Technol. Ther.* **21**, 201–208 (2019)
21. Kerkenbush, N.L.: A comparison of self-documentation in diabetics: electronic versus paper diaries. In: AMIA Annual Symposium Proceedings 2003, vol. 887 (2003)
22. Man, C.D., Micheletto, F., Lv, D., Breton, M., Kovatchev, B., Cobelli, C.: The UVA/PADOVA type 1 diabetes simulator. *J. Diabetes Sci. Technol.* **8**, 26–34 (2014)
23. Rodríguez-Rodríguez, I., Rodríguez, J.V., Chatzigiannakis, I., Zamora Izquierdo, M.Á.: On the possibility of predicting glycaemia ‘On the Fly’ with constrained IoT devices in type 1 diabetes mellitus patients. *Sensors* **19**, 4538 (2019)
24. Liu, J., Wang, L., Zhang, L., Zhang, Z., Zhang, S.: Predictive analytics for blood glucose concentration: an empirical study using the tree-based ensemble approach. Library Hi Tech. ahead-of-print (2020)
25. Saiti, K., Macaš, M., Lhotská, L., Štechová, K., Pithová, P.: Ensemble methods in combination with compartment models for blood glucose level prediction in type 1 diabetes mellitus. *Comput. Methods Programs Biomed.* **196**, 105628 (2020)
26. Midroni, C., Leimbigler, P.J., Baruah, G., Kolla, M., Whitehead, A.J., Fossat, Y.: Predicting glycemia in type 1 diabetes patients: experiments with XGBoost. Presented at the 3rd International workshop on knowledge discovery in healthcare data (2018)
27. Wang, Y., Wang, T.: Application of improved LightGBM model in blood glucose prediction. *Appl. Sci.* **10**, 3227 (2020)
28. Alfian, G., Syafrudin, M., Rhee, J., Anshari, M., Mustakim, M., Fahrurrozi, I.: Blood glucose prediction model for type 1 diabetes based on extreme gradient boosting. *IOP Conf. Ser.: Mater. Sci. Eng.* **803**, 012012 (2020)
29. Xao, W., Shao, F., Ji, J., Sun, R., Xing, C.: Fasting blood glucose change prediction model based on medical examination data and data mining techniques. In: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). IEEE, Chengdu (2015)
30. Georga, E.I., Protopappas, V.C., Polyzos, D., Fotiadis, D.I.: A predictive model of subcutaneous glucose concentration in type 1 diabetes based on Random Forests. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, San Diego (2012)

31. Hidalgo, J.I., Colmenar, J.M., Kronberger, G., Winkler, S.M., Garnica, O., Lanchares, J.: Data based prediction of blood glucose concentrations using evolutionary methods. *J. Med. Syst.* **41**, 142 (2017)
32. Hidalgo, J.I., Botella, M., Velasco, J.M., Garnica, O., Cervigón, C., Martínez, R., Aramendi, A., Maqueda, E., Lanchares, J.: Glucose forecasting combining Markov chain based enrichment of data, random grammatical evolution and bagging. *Appl. Soft Comput.* **88**, 105923 (2020)



Virtual Reality in the Treatment of Acrophobia

Vanessa Maravalhas^(✉) , António Marques, Sara de Sousa, Pedro Monteiro, and Raquel Simões de Almeida

School of Health, Polytechnic of Porto, Porto, Portugal
10150491@ess.ipp.pt

Abstract. Exposure Therapy using Virtual Reality has been indicated as one of the most promising therapeutic approaches in the treatment of Acrophobia, because Virtual Reality generates exposure conditions systematic controlled and customized desensitization of the individual concerning the phobic incitement and the impulsive responses related to it, potentiating the emotional self-regulated competence and the incorporated cognitive restructuration.

This quasi-experimental study had the purpose to analyze the impact of an exposure therapy program using Virtual Reality in the reduction of fear of heights on a sample of the Portuguese population. In this study participated 19 people with and without fear of heights, being distributed by the experimental and control groups. The program was composed of 8 sessions distributed biweekly, corresponding to a total program duration of 4 weeks. The impact of the program was analyzed through the administration of anxiety and acrophobia level assessment instruments and operationalization of psychophysiological procedures, focused on the biofeedback principles, before and after the intervention.

The results from both self-report and psychophysiological data revealed a significant reduction in fear of heights in the participants exposed to the exposure therapy program, showing a positive impact of this protocol on the treatment of Acrophobia.

Keywords: Anxiety disorders · Acrophobia · Exposure therapy · Virtual Reality · Biofeedback

1 Introduction

The World Mental Health Survey Initiative, in 2015, ranked Portugal as the second European Union country with the highest prevalence of neuropsychiatric diseases (22.9%), with anxiety disorders being the most prevalent group of mental illnesses in our country (16.5%) [1–3].

Overall, anxiety disorders are a group of neuropsychiatric disorders characterized by feelings of anxiety and fear. Anxiety does not have a specific driver that defines its onset and usually responds to concern with future events, while fear is a reaction to present events. These feelings can cause physical symptoms, such as rapid heart rate or tremors, which in different combinations and depending on the associated etiology are organized into different types of nosologies [4–6].

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM V), anxiety disorders include social phobia, panic disorder, agoraphobia, separation anxiety, selective mutism, generalized anxiety disorder, substance-induced anxiety disorder, and specific phobia [7].

Specific phobia, the most prevalent of these conditions, is characterized as fear or anxiety in the presence of a particular situation or object, called a “phobic stimulus” [7]. Acrophobia is a type of specific phobia and is defined according to DSM V as an extreme fear of some specific object or situation, in this case, fear of heights [7]. People with Acrophobia tend to avoid situations that cause them suffering and anxiety, such as riding an elevator, climbing stairs, flying, walking on bridges, or even approaching windows [8].

For the treatment of Acrophobia, several therapeutic approaches are applied, the most commonly used being desensitization therapy, in vivo exposure therapy, virtual reality exposure therapy, and drug therapy in combination with one of the previous interventions [9].

Several studies report that the most used therapeutic method and which has the most positive results in this type of pathology is exposure therapy [10–15], framed in the principles of cognitive-behavioral therapy [11, 14]. Regarding in vivo exposure therapy, although very promising results regarding its use are described in the literature [12, 13, 16], the impossibility of rigorously guaranteeing adequate environmental control is mentioned as a critical element regarding the mobilization of this type of intervention [10, 17]. If the environment is not being therapeutically controlled, this generates some unpredictability and may trigger adverse and negative reactions to the treatment process [10, 16, 18]. Virtual Reality uses computer applications to create an immersive virtual experience in a therapeutic environment, making it possible to recreate real day-to-day situations that cause anxiety in individuals undergoing treatment, in a controlled and safe environment [13, 16, 19].

Duff, Miller & Bruce’s [19] studied the effects of using Virtual Reality as a therapeutic method compared to in vivo exposure therapy methods and concluded that these approaches have more positive results, because through Virtual Reality it is possible to obtain a more assertive control of the therapeutic environment, facilitating adaptability and problem solving and offering greater possibilities for customization, flexibility, and control of the therapeutic process [16, 17, 20, 21].

In addition to environmental control, Virtual Reality also allows for a very high degree of confidentiality, as exposure is made within a room and there is no risk of potential negative reactions from either the individual or possible observers [10]. This type of exposure seems to be very effective in the controlled and personalized systematic desensitization of the individual regarding the phobic stimulus and related impulsive responses, enhancing emotional self-regulation and associated cognitive restructuring [10, 11, 20–23].

The purpose of this study is to assess the impact of virtual reality in the treatment of acrophobia.

2 Methods

The present study is quasi-experimental because there is the interference of the variables under study and contains two sample groups, it is longitudinal because it evaluates the sample at two different times, that is, it followed a quantitative pre-and post-test methodology [25, 26], through the administration of instruments to assess anxiety levels and fear of heights and operationalization of psychophysiological procedures, centered on the principles of biofeedback. Questionnaires were also administered to characterize participants in terms of socio-demographic data and sensations caused by the use of Virtual Reality, namely the sense of presence in the virtual environment and cybersickness.

2.1 Participants

Through the non-probabilistic convenience technique, as individuals voluntarily agreed to participate in the study, they were selected according to our inclusion criteria, and to the more accessible contact [24, 25], with fourteen women and five men, 21 ± 1.8 years old, being recruited to participate in this study. Eleven women and 2 men, aged 22 ± 1.9 years with Acrophobia, were included in the experimental group, and the remainder, with no history of fear of heights, were included in the control group.

Inclusion criteria to participate in the study were fear of heights (at least 2 on the Likert scale when asked about the fear of being on a second-floor balcony, leaning against a 1-m high parapet), motivation and willingness to participate in the study and to go twice a week to LabRP/ESS.PPORTO facilities. Exclusion criteria were individuals with health problems that prevented exposure to Virtual Reality, namely labyrinthitis.

Regarding the control group, individuals were selected according to the availability and motivation to participate in the study. They were only considered as exclusion criteria, have some fear of heights, and feel discomfort during exposure to Virtual Reality, were met.

The participation of individuals in the study was formalized by completing the informed consent form [26], to ensure their rights and access to all information relevant to the decision to participate in the study. The privacy and confidentiality of the collected data were also attested. The study was approved by the Ethics Committee of the School of Health, Polytechnic of Porto.

2.2 Instruments

Clinical Interview: Built to collect sociodemographic information and survey brief clinical history related to fear of heights and other conditions contraindicated exposure to Virtual Reality. The semi-structured interview guide consisted of five questions related to sociodemographic data and seven open-ended questions.

Acrophobia Questionnaire: Split into two parts with various situations that can trigger anxiety due to fear of heights. The first part concerns the levels of fear that the participant feels in each situation and, therefore, must answer from 0 to 6 regarding the fear that would feel in those same situations. The second part is related to avoidance in the same situations, with the participant answering how much he avoided each of the situations, on a Likert scale from 0 to 2.

Behavioral Avoidance Test (BAT): Divided into two parts, one applied before in vivo exposure and one applied after it; in both moments the participants had to report their level of fear, anxiety, and danger felt for each situation (a second-floor balcony - exposure to 5 m or 10 m), on a Likert scale from 0 to 10.

Simulator Sickness Questionnaire (SSQ): Applied to assess the presence of cyber-sickness. Also, according to a Likert scale with 1 corresponding to “totally disagree” and 5 “totally agree”.

Intervention Procedure Evaluation Questionnaire: Applied at the end of the intervention, to understand participants’ level of satisfaction.

Psychophysiological Procedures: To collect physiological parameters that would allow the use of the biofeedback principles among the participants, the Biopac MP100 device connected to a Bionomadix 2-channel wireless system was used to obtain heart rate from the ECG signal (lead I), as well as the respiratory rate. These data were recorded during exposure in all the therapeutic sessions and were used to demonstrate to individuals their evolution over therapeutic intervention.

2.3 Procedures

The Virtual Reality Exposure Therapy program consisted of 8 sessions: an initial evaluation session, 6 biweekly Virtual Reality exposure therapy sessions, and the last reevaluation session corresponding to a total duration of the program of 4 weeks.

Each session, lasted approximately 50 min, following a formal structure consisting of the phases: Participant preparation and warm-up; Personalized and progressive exposure to virtual environments of systematic desensitization; Relaxation and feedback. This treatment strategy thus consisted of a combination of therapeutic ingredients, including exposure to fear-triggering stimuli, therapeutic instructions, monitoring of individual progress, performance feedback, and contingent performance enhancement.

The intervention focused on shaping appropriate approach behaviors through a process of successive approximations. Treatment was achieved by reducing the escape from the dreaded situation, assuming that the absence of consequences results in the extinction of fear, a common ingredient in exposure therapies [16, 17]. Approach behaviors were facilitated by reinforcing approaches and removing negative reinforcement from avoidance. Exposure grading was based on a length of time (between 15–30 min) or several practical exposures of varying complexity and intensity. Flowchart 1 summarizes the entire process of selection and follow-up of the participants.

This study was conducted at the Virtual Lab of the Psychosocial Rehabilitation Laboratory of the School of Health of the Polytechnic Institute of Oporto and the Faculty of Psychology and Educational Sciences of the University of Oporto. The immersion and presence in the virtual environment were achieved through the use of VirtualHTC Reality Live glasses. For greater involvement in the environment, a purpose-built simulator was also used, which reproduced the physical conditions of the virtual environment (board suspended between two fixed points and reproduction of the platform of an elevator and board suspended in a skyscraper). The software had various height-related environments that allowed each participant to be gradually and individually exposed, adjusted to their level of fear.

Collected data were analyzed using the IBM SPSS Statistics 25 software [27]. In terms of sociodemographic characterization of the participants, descriptive statistics were used, and taking into account the variables used, the mean and mode were calculated as a measure of central tendency, the standard deviation as a measure of dispersion, and the absolute frequencies and frequency of each characteristic under analysis. Regarding the verification of significant changes before and after the implementation of the intervention protocol, inferential statistical analysis procedures were used, assuming for all statistical tests a significance value (α) of 0.05 [28–30].

In this framework, the normality of all variables was tested in general and for each control and experimental group, and the variables were analyzed through the Shapiro-Wilk test, and it was concluded that all variables followed normality. After verifying normality, we tested data homogeneity through the covariance test (Levene test), and it was possible to assume data homogeneity. Regarding sphericity, this was assumed since only two evaluation moments were foreseen [28–30].

Once the assumptions were verified, ANOVA was applied to the repeated measures of the analyzed variables to verify if there were significant differences between the two groups and between the two evaluation moments. When statistical differences were obtained, t-tests for paired samples were applied to analyze the evaluation moments within groups, and t-tests for independent samples were used to analyze differences between groups [28–30].

3 Results

We will start the presentation of results, mobilizing descriptive statistics [28, 30] to characterize the sample, and then we will present the results for the various used instruments, allowing us to verify the impact of the exposure therapy program using Virtual Reality to reduce fear to heights.

Nineteen subjects participated in the study, divided by the control group ($n = 8$) and the experimental group ($n = 11$). Most of our sample consisted of female students (73.7%), single (100%), and who never had any kind of psychological intervention for the treatment of their fear of heights.

3.1 Acrophobia Questionnaire

Regarding the Acrophobia Questionnaire, there are significant differences between the experimental and control group values ($p = 0,001$). In this sense, the values of the experimental group are higher concerning the anxiety felt when exposed to heights, and the same happens about the adoption of avoidance behaviors.

When analyzing the pre-and post-test data, we found that in the control group there are no significant changes. On the contrary, in the experimental group, there is a significant statistical change caused by the exposure therapy, with anxiety ($p = 0,001$) and avoidance ($p = 0,001$) values being higher in the pre-test than in the post-test.

3.2 Psychophysiological Assessment Measures

Regarding respiratory rates, a statistically significant difference between the control group and the experimental group was observed, with values in the experimental group being lower than those of the control group ($p = 0,001$).

Regarding the post-test in the experimental group, the values of the respiratory rate are higher than in the pre-test ($p = 0,03$).

Only for the maximal respiratory rates, there are no statistically significant differences.

Regarding heart rates, there a statistically significant difference between the control group and the experimental group was observed, with values in the experimental group being higher than in the control group ($p = 0,001$). Regarding the experimental group, there are differences between the pre-and post-tests, with heart rate values in the post-test being lower ($p = 0,02$).

3.3 Behavioral Avoidance Test (BAT)

Regarding BAT, there is a statistically significant difference between the control and the experimental groups ($p = 0,001$). As for the experimental group, there are differences between pre-and post-test ($p = 0,001$).

3.4 Simulator Sickness Questionnaire (SSQ)

Regarding the SSQ, no significant differences were found in the interaction between the control and experimental groups at the initial and final moments, as participants only reported higher values for vertigo after exposure to the virtual environment, and even at the level of vertigo ($p = 0,62$).

4 Discussion

The present study aimed to verify whether exposure therapy using Virtual Reality has positive effects in the treatment of Acrophobia.

This study becomes relevant and innovative because in Portugal, to the best of our knowledge, no other research project with the same purpose has been carried out and published.

In the present study, as occurred in other research studies [10–13], Virtual Reality Exposure Therapy had a positive impact on the treatment of individuals with Acrophobia who were part of the experimental group, as evidenced by the results obtained by the administration before and after the intervention of the Acrophobia Questionnaire, the Behavioral Avoidance Test, and the psychophysiological measures, centered on the heart and respiratory rate.

The results obtained with the Acrophobia Questionnaire suggest that this Virtual Reality Exposure Therapy protocol had a positive effect, both in decreasing the anxiety level of the participants in the experimental group regarding situations of exposure to heights and with the possible need to adopt avoidance behaviors in the same situations.

Although the final score of this questionnaire does not directly reproduce the subject's Acrophobia level, its variation between pre-and post-test can be used to analyze the subject's perception of fear of heights given that high score values correspond to higher levels of fear and anxiety triggered by the situation, as well as its avoidance [7, 31]. Considering this assumption, the obtained data points to a decreased level of Acrophobia following the application of the experimental program and, therefore, supports the positive impact of this Virtual Reality Exposure Therapy protocol for the reduction of Acrophobia levels.

The Behavioral Avoidance Test (BAT) results also seem to show a positive evolution regarding the fear of heights reported by the participants of the experimental group. Although in vivo assessments were not sensitive enough to accurately characterize fear of subjects' heights, given that both pre-test and post-test subjects reported acceptable levels of self-control at both 5-m and 10-m exposure, the exposure to first-tier virtual environments before and at the end of program implementation allowed for a more sensitive and reliable assessment, illustrating a positive evolution of subjects in managing their fear of heights. These results seem to suggest and corroborate what is found in other studies, i. e., that virtual exposure under controlled environmental conditions may be more efficient than in vivo exposure for the assessment and treatment of acrophobia [10, 16].

About the psychophysiological measures used to assess the impact of the treatment protocol on respiratory and heart rate, the literature reports that varying respiratory and heart rate values can be used to characterize anxiety levels [5, 6]. Thus, when analyzing the psychophysiological measures, it is relevant to emphasize that in the exposure to the virtual environment the fear assessment at heights in the pre-test, generated lower values in the experimental group than in the control group, and higher as regards the second referenced measure. This finding seems to immediately show that the participants of the experimental group were effectively exposed to an anxiety-enhancing situation, reinforcing the diagnosis of Acrophobia [5, 6, 32]. The low respiratory rate values observed in this study appear to be related to respiratory apnea, as reported in several studies that concluded that individuals confronted with fearful situations suffer respiratory arrests [33], while its increased values in the experimental group at the post-test is strong evidence regarding the positive impact of the exposure therapy in virtual reality environment to treat acrophobia.

Concerning the values of heart rate, it is noteworthy to mention that the observed in the experimental group are likely to be too high, even considering the exposure to an anxiety-enhancing situation, something possibly enhanced by the possibility of caffeine intake by participants near the time of exposure. This variable was not controlled in the investigation and may have influenced the results since college students have a history of high caffeine consumption [34, 35].

Overall, heart rate and respiratory rate values varied significantly from pre-test to post-test in the participants of the experimental group, both pointing to a positive program effect for the reduction of anxiety and fear at heights.

The application of the Simulator Sickness Questionnaire throughout the program did not reveal any noticeable difficulties or feelings of discomfort on the part of participants interacting with Virtual Reality, although nausea is commonly reported due to differences

in what is being experienced/visualized and what is happening to our body [18, 36, 37]. Only some sensations of vertigo, a sensation explained and associated more with the nature of the environment and the fear of heights, have been reported, rather than as a consequence of exposure to the Virtual Reality environment.

5 Conclusion

The results of this study seem to contribute to the evidence that exposure therapy using Virtual Reality has positive results and is a promising therapy in the treatment of individuals with Acrophobia.

However, further research in this area is suggested, to build more robust evidence in this area, either by conducting more experimental studies, with more significant samples, and more control of the variables under study.

This study has as limitations the possible lack of blindness of the investigator [38], given that the researcher who did the evaluations and reassessments was the same one who conducted the sessions, which may generate some bias to the investigation, as well as the fact that the questionnaires used in the evaluations are not validated for the Portuguese population.

References

- Carvalho, Á.: Depressão e outras Perturbações Mentais Comuns: enquadramento global e nacional e referência de recurso em casos emergentes, pp. 1–104 (2017). https://www.google.pt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiMxoX_64TUAhUDExoKHeycD0IQFggmMAA&url=https%3A%2F%2Fwww.dgs.pt%2Fficheiros-de-upload-2013%2Fdms2017-depressao-e-outras-perturbacoes-mentais-comuns-pdf.aspx&usg=AFQjCNGi1bcNm5X90lubf
- SNS. 2018 1. 2018. 1–88 p
- DGS: Saúde mental em números. Direção Geral da Saúde (2015). <https://www.dgs.pt/estatisticas-de-saude/estatisticas-de-saude/publicacoes/portugal-saude-mental-em-numeros-2015-pdf.aspx>
- Andreatta, M., Neueder, D., Glotzbach-Schoon, E., Mühlberger, A., Pauli, P.: Effects of context preexposure and delay until anxiety retrieval on generalization of contextual anxiety. *Learn Mem.* **24**(1), 43–54 (2017)
- Goessl, V.C., Curtiss, J.E., Hofmann, S.G.: The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. (2017), 2578–2586 (2018)
- Carnevali, L., Sgoifo, A., Trombini, M., Landgraf, R., Neumann, I.D., Nalivaiko, E.: Different patterns of respiration in rat lines selectively bred for high or low anxiety. *8*(5) (2013)
- DSM V: Manual Diagnóstico E Estatístico De Transtornos Mentais (2013). <https://blogs.sapo.pt/cloud/file/b37dfc58aad8cd477904b9bb2ba8a75b/obaudoeducador/2015/DSMV.pdf>
- Coelho, C.M., Waters, A.M., Hine, T.J., Wallis, G.: The use of virtual reality in acrophobia research and treatment. *23*, 563–574 (2009)
- Arroll, B., Henwood, S.M., Sundram, F.I., Kingsford, D.W., Mount, V., Humm, S.P., et al.: A brief treatment for fear of heights: a randomized controlled trial of a novel imaginal intervention (2017)
- Botella-Arbona, C., Osma, J., Garcia-Palacios, A., Quero, S., Banōs, R.M.: Treatment of flying phobia using virtual reality: data from a 1-year follow-up using a multiple baseline design. *Clin. Psychol. Psychother.* **11**(5), 311–323 (2004)

11. Gromer, D., Madeira, O., Gast, P., Nehfischer, M., Jost, M., Müller, M., et al.: Height simulation in a virtual reality CAVE system: validity of fear responses and effects of an immersion manipulation. **12**(Sept), 1–10 (2018)
12. Choy, Y., Fyer, A.J., Lipsitz, J.D.: Treatment of specific phobia in adults. **27**, 266–286 (2007)
13. Krijn, M., Hulbosch, A.M., De Vries, S., Schuemie, M.J.: Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. **40**, 509–516 (2002)
14. Morina, N., Ijntema, H., Meyerbr, K., Emmelkamp, P.M.G.: Behaviour research and therapy can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. **74**, 18–24 (2015)
15. Suso-ribera, C., Ferna, J., Ban, R.M.: A preliminary comparison of treatment efficacy in small animal phobia. 1–8 (2018)
16. Verkuyl, M., Romaniuk, D., Mastrilli, P.: Virtual gaming simulation of a mental health assessment: a usability study. Nurse Educ Pract. **31**, 83–87 (2018). <https://doi.org/10.1016/j.nepr.2018.05.007>
17. Landowska, A., Roberts, D., Eachus, P., Barrett, A., Pauli, P.: Within- and between-session prefrontal cortex response to virtual reality exposure therapy for acrophobia. **12**(November), 1–16 (2018)
18. Aldaba, C.N., White, P.J., Byagowi, A., Moussavi, Z., et al.: Virtual reality body motion induced navigational controllers and their effects on simulator sickness and pathfinding. **2**, 4175–4178 (2017)
19. Duff, E., Miller, L., Bruce, J.: Online virtual simulation and diagnostic reasoning: a scoping review. Clin. Simul. Nurs. **12**(9), 377–384 (2016). <https://doi.org/10.1016/j.ecns.2016.04.001>
20. Baus, O., Bouchard, S.: Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. Front. Human Neurosci. **8**(March), 1–15 (2014)
21. Hong, Y.-J., Kim, H.E., Jung, Y.H., Kyeong, S., Kim, J.-J.: Usefulness of the mobile virtual reality self-training for overcoming a fear of heights. Cyberpsychol. Behav. Soc. Netw. **20**(12) (2017). <https://online.liebertpub.com/doi/10.1089/cyber.2017.0085>
22. Schweizer, T., Schmitz, J., Plempe, L., Sun, D., Becker-Asano, C., Leonhart, R., et al.: The impact of pre-existing anxiety on affective and cognitive processing of a virtual reality analogue trauma. PLoS One. **12**(12), 1–19 (2017)
23. Botella, C., Fernández-álvarez, J., Guillén, V., García-palacios, A., Baños, R.: Recent progress in virtual reality exposure therapy for phobias: a systematic review (2017)
24. Andrews, J., Likis, F.E.: Study design algorithm **19**(4), 364–368 (2015)
25. de Oliveira, A.G.: Bioestatística, Epidemiologia e Investigação, 249 p. (2009)
26. Braga, R.: Ética na publicação de trabalhos científicos. Rev Port Med Geral e Fam (2013)
27. Released IC. IBM SPSS Statistics for Windows, Version 25.0., Armonk, NY (2017)
28. Marôco, J.: Análise estatística com o SPSS Statistics. Análise e Gestão da Informação (2014)
29. Pestana, M.H., Gageiro, J.N.: ANÁLISE DE DADOS PARA CIÊNCIAS SOCIAIS A Complementaridade do SPSS 6^a EDIÇÃO Revista, Atualizada e Aumentada MARIA HELENA PESTANA JOÃO NUNES GAGEIRO, pp. 1–2 (2014). https://www.researchgate.net/publication/272817141_ANALISE_DE_DADOS_PARA_CIENCIAS_SOCIAIS_A_Complementaridade_do_SPSS_6_EDICAO_Revista_Atualizada_e_Aumentada_MARIA_HELENA_PESTANA_JOAO_NUNES_GAGEIRO
30. Pereira, A.: SPSS - Guia prático de utilização, 243 p. Edições Sílabo (2008)
31. Dunsmoor JE, Paz R. Fear generalization and anxiety: behavioral and neural mechanisms. Biol. Psychiatry. (2015). <https://doi.org/10.1016/j.biopsych.2015.04.010>
32. Tovote, P., Fadok, J.P., Lüthi, A.: Neuronal circuits for fear and anxiety. Nat. Publ. Gr. **16**(6), 317–331 (2015). <https://doi.org/10.1038/nrn3945>
33. Yohannes, A.M., Junkes-cunha, M., Smith, J.: Management of dyspnea and anxiety in chronic obstructive pulmonary disease : A critical review. J. Am. Med. Dir. Assoc. (2017) <https://doi.org/10.1016/j.jamda.2017.09.007>

34. Pané-farré, C.A., Alius, M.G., Modeß, C., Methling, K., Blumenthal, T., Hamm, A.O.: Anxiety sensitivity and expectation of arousal differentially affect the respiratory response to caffeine 2014 (2009)
35. Coast, G., Coast, G., Sciences, M.: Caffeine Use and Alexithymia in, 37–41 (2014)
36. Kim, H.K., Park, J., Choi, Y., Choe, M.: Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* **69**(March 2017), 66–73 (2018). <https://doi.org/10.1016/j.apergo.2017.12.016>
37. Edward, M., Russell, B., Hoffman, B., Stromberg, S., Carlson, C.R.: Use of controlled diaphragmatic breathing for the management of motion sickness in a virtual reality environment (2014)
38. Waddington, H., et al.: Quasi-experimental study designs series d paper 6: Risk of bias assessment. *J. Clin. Epidemiol.* (2017). <https://doi.org/10.1016/j.jclinepi.2017.02.015>



FOMO Among Polish Adolescents. Fear Of Missing Out as a Diagnostic and Educational Challenge

Łukasz Tomczyk^(✉)

Pedagogical University of Cracow, Ingardena 4 Street, Cracow, Poland
tomczyk_lukasz@prokontro.pl

Abstract. Fear Of Missing Out (FOMO) continues to be an under-researched phenomenon and is a form of problematic use of the Internet (commonly mistaken for Internet addiction). FOMO is a behaviour connected with the avalanche of information and the development of new digital technologies. Young people in particular are exposed to FOMO. Age is a predictive factor for different types of problematic Internet use. Therefore, coverage by specific diagnostic and preventive activities is a special task for experts in media pedagogy. The aim of the research was to show the scale of the FOMO phenomenon among Polish adolescents. The research covered 979 adolescents aged 13–18 years. The research was conducted with the use of the questionnaire “Social Media Use and the Fear of Missing Out (FoMO)” throughout Poland in the school year 2018–2019. On the basis of the collected data it was noted that: 1) the most frequent symptom of FOMO is the use of social networking sites just before bedtime (about 70% of respondents) and just after waking up (about 50% of respondents); 2) About 30% of respondents are classified as possessing strong FOMO characteristics, about 43.5% have an average FOMO level, 26.5% have a problem-free use of SNS; 3) All FOMO factors coexist at medium or high levels; 4) Girls have a slightly higher FOMO level than boys; 5) Variables related to place of residence, type of school, grade average, and subjective sense of wealth are related to FOMO intensity. FOMO is an important challenge for education, the prevention of risky behaviour in cyberspace, and the development of the information society.

Keywords: FOMO · Internet addiction · Problematic internet use · Youths · Adolescents · Scale · Mechanism · Poland · Information society

1 Introduction and Related Works

Fear of Missing Out (FOMO) is the fear of disconnecting (or lacking the ability to access) from access to new information in cyberspace. It is a relatively new phenomenon, which arouses a lot of diagnostic controversy without being unambiguously and indisputably defined. FOMO is associated with the intense growth of information in the network media space [1]. In order to understand the phenomenon of the functioning of modern people in cyberspace, a few facts should be quoted. It is estimated that every minute, 500 h of videos

are uploaded to YouTube, nearly 300 trillion emails are sent, more than 4 million searches are carried out using Google, more than 300 million posts are sent via Twitter, more than 78 million posts are published using the CMS Wordpress, and 54 million messages are sent using Whatsapp [2]. In modern times, we face an avalanche of information. Over the past two decades, we have seen the development of several key processes that determine people's psychosocial functioning in information spaces. The development of new technologies has significantly accelerated the production and distribution of information. All information services, by which we can include social media, add new photos, posts, copies of media from other services, and entertainment materials on an ongoing basis [3, 4]. Moreover, users of new media are encouraged and supported to actively participate in creating new content for the consumption of other users. Being both the creator and recipient of information has never been so easy. In addition, the decreasing digital exclusion rate, access to unlimited data transfer, the universality of wifi networks, and the increasingly advanced capabilities of mobile devices make the technical layer of social networking and information services almost invisible. These are only selected mechanisms describing the permeation of the technical and human conditions accompanying FOMO [5].

For several years there has been an intense discussion on the issue of recognising a new disease, namely Internet addiction. The loss of self-control over the use of digital media leads to many negative consequences in the psychological, biological, and social sphere of individuals who are permanently online. At present, the term *Internet addiction* is one that experts seek to avoid, due to the lack of clear diagnostic criteria, as well as the lack of inclusion of addiction on the official list of diseases (ICD - International Statistical Classification of Diseases and Related Health Problems, DSM - Diagnostic and statistical manual of mental disorders), as is the case, for example, with addiction to games and gambling [6, 7]. FOMO for researchers is a type of problematic internet use, a phenomenon which is noticeable, but does not have the official status of a disease mediated by digital media, as well as clear and unquestioned diagnostic criteria. From the perspective of the features common to the various definitions of problematic internet use and the emerging difficulties and contradictions in the definitions, FOMO should be considered under the umbrella term of Problematic Internet Use [8, 9]. Among the key factors in the development of FOMO at present are: the intense development of social networking sites (SNS), the ubiquity of mobile devices with Internet access (especially smartphones), the hidden influence of SNS developers, and individual characteristics such as a lack of self-control. Among the many risk factors, there are also issues concerning the co-occurrence of FOMO with various problems occurring in bio-psycho-social functioning in the offline world. The online and offline spaces are interconnected in many ways, but most relevantly in terms of problematic situations [10]. FOMO is a challenge for civilisation, education, health, and prevention. FOMO has also become one of the typical cultural phenomena of the information society. The high intensity of FOMO may raise concerns in several situations: 1) it concerns people who are at the stage of developing skills related to self-control and digital competences, i.e. children and young people; 2) it causes negative consequences for social relationships; 3) it disturbs long-term mood and leads to the deterioration of the general state of health; 4) it generates conflicts related to the performance of typical family and professional

activities. Taking into account the peculiarities of FOMO, it should be emphasized that it is a state which manifests itself in continuously being online, permanently monitoring the content of various websites, especially social networking sites. FOMO also manifests itself in taking action related to the handling of new media inappropriate to the situation (e.g. during an offline meeting) and the time of day (e.g. at night - the time of sleep). FOMO is a phenomenon that should also be considered in the category of digital competence, in particular the skills related to accessing and processing information.

FOMO is not a variable that is often analysed in the literature, unlike other risky behaviours such as cyberbullying, media manipulation, or sexting. In Poland, research on FOMO among high-risk groups (e.g. young people) is rare [11–12]. Results obtained from representative samples are not often found in the literature. Therefore, it seems reasonable to mention several unique research results before considering further analyses. The research conducted by the Polish team *EU KIDS Online* on a sample of almost a thousand respondents shows that the problematic use of the Internet, of which FOMO is a component, concerns various areas (depending on the criterion adopted, up to several percent). All the diagnostic criteria in the above-mentioned studies meet only 0.1% of the total. However, using *EU KIDS* data, it can be noted that 13% log on to the Internet at least once a week without a specific purpose. Approximately 1.4% of respondents have daily conflicts related to the time spent using digital media (e.g. arguments between adolescents and parents). About 6% of the students surveyed claim that the time they spend on the Internet can cause problems [13]. However, studies of problematic Internet use (PIU) do not contain sufficient criteria to show the subtype of PIU related to FOMO. It seems justified to recall the research carried out by the team gathered around the University of Warsaw [14, 15]. The results of the research, which were collected on a representative sample of Poles using the tool *FOMO Scale* (author Andrew K. Przybylski) [16], offer a new perspective on the phenomenon. First of all, the researchers distinguished 3 levels of FOMO: a) low level of FOMO, to which they included 19% of the respondents, b) average level of FOMO - which matched with 67% of the respondents, and c) 14% were characterized by a high level of FOMO. Secondly, the occurrence of FOMO is not differentiated by gender. Internet users from larger cities are much more susceptible to FOMO than people from small towns and villages. The low and high FOMO levels are very much linked to age. Adolescents belong to the FOMO risk group. Those in the age range of 15 to 23 have a high level of FOMO [15]. Based on other data collected by NASK, it was noted that 60% of young people believe they should use their smartphones less. The data presented in the *Teenage 3.0* report offer surprising results: 31% of young people think that their lives would be empty without smartphones, 26% have the urge to use their smartphones immediately after putting them away, 27% cannot function without a smartphone, and 26% feel tired and neglect their duties due to being online all the time [17]. Of course it seems crucial to ask about the frequency of such feelings. One-off or rare or very rare declarations of this type do not indicate a massive problem. Much greater concern may be aroused by declarations that occur very frequently or exist in a permanent state.

The research carried out by P. Modzelewski on a sample of five hundred Internet users in the 9–71 age bracket shows that about 17.6% of the respondents have a high FOMO level. The variable age is a clear descriptor of predisposition to the problem of

FOMO [18]. A study conducted by the author of this study in Bosnia and Herzegovina in 2017 noted that among teenagers (average age of respondents 13 years old, sample of 717 adolescents), half of the respondents do not show an alarming intensity of FOMO symptoms, about 20% show several symptoms at full intensity, while 30% are at risk [19]. The studies carried out so far clearly show the dual nature of new technologies. ICT was created with the aim of improving the quality of life. However, many risks, such as FOMO, have been created as by-products of the development of the information society, and it is young people in particular who are vulnerable. In view of the research offered above, an analysis of the scale of FOMO from a pedagogical perspective and a discussion of the diagnostic criteria are certainly germane topics.

2 Methodology

2.1 Research Objectives

The aim of the research was to show the scale of FOMO phenomenon among Polish adolescents. Specifically, the research sought to present the diversity of FOMO depending on sociodemographic variables, i.e. age, gender, type of school, school grades, behavioural grades, place of residence, and subjective feeling of family prosperity. The research is diagnostic in nature and can also be used for the prevention in education of risky behaviours related to cyberspace.

2.2 Sample and Test Procedure

The research was conducted in the school year 2018–2019 in Poland. The sample selection was random. Using the IT National Education System (*System Informacji Oświatowej*), schools were invited to take part in the survey. The schools were selected at random. The diagnostic survey was conducted by qualified pedagogical staff with knowledge in the field of educational research. The research covered adolescents aged 13–18 years. Finally, after the rejection of incorrectly completed questionnaires (e.g. incomplete questionnaires), 979 questionnaires were obtained and subsequently analysed. The respondents were 56.58% girls and 43.42% boys, and lived in the following areas: rural areas (53.32%), a city with up to 50,000 inhabitants (24.20%), a city with between 50 and 100,000 (9.09%), a city with between 100 and 200,000 (8.06%), and a city with over 200,000 (5.31%).

2.3 Research Tool

The study used the tool “Social Media Use and the Fear of Missing Out (FoMO)” [20] aimed at measuring FOMO syndrome during a variety of typical adolescent situations. The tool consisted of five questions related to the use of social media in the following contexts: just after waking up, at breakfast, at lunch, at dinner, and before falling asleep. The answers were given on a 5-point Likert scale. The tool has satisfactory internal consistency parameters. The internal consistency of the tool are presented below in Table 1. The psychometric properties of the tool were as follows: Chi-Square = 221.278; df = 5; p.value = 0.000; RMSEA = 0.210; Lower CI RMSEA = 0.187; Upper CI RMSEA = 0.234; SRMR = 0.063.

Table 1. Internal consistency properties of the tool

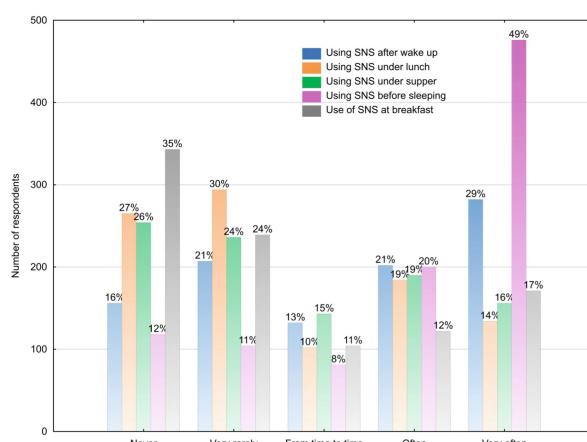
Estimate	McDonald's	Cronbach's	Guttman's 2	Greatest lower bound
Point estimate	0.880	0.882	0.884	0.917
95% CI lower bound	0.866	0.869	0.871	0.906
95% CI upper bound	0.892	0.893	0.896	0.929

2.4 Research Ethics

The research was carried out in accordance with the principles of the ethics of conducting diagnostic surveys. In order to carry out the research, permissions were obtained from: the management of the institution, teachers conducting the classes, and the young people who took part in the research. Participation in the research was voluntary and the tools ensured complete anonymity. The participants were able to withdraw their participation in the research at any time.

3 Results

Based on the data collected, it is noted that the most common symptoms of FOMO that occur *very often* or *frequently* are the use of SNS just before bedtime. About 70% of the respondents declare that such an action occurs *frequently* or *very often*. The second most common FOMO factor is the use of SNS just after waking up (50% of respondents). In itself, the use of SNS just before bedtime is not an unambiguous factor indicating FOMO. The data should be compared with the time spent using SNS at inappropriate times and the impact of such an activity on, for example, the quality of sleep. The percentage distribution of the other indications is presented in Fig. 1.

**Fig. 1.** Distribution of FOMO-related responses among the adolescents.

Using the k-means cluster analysis technique, three groups of SNS users were distinguished. Group number 1 is characterized by a high level of FOMO, where all five factors occur *frequently* or *very often*. In group number 1 there are 299 people (30.54%). In group number 2, those with low occurrence of selected factors, there are 258 cases (26.35%), while in group number 3 with an average level of FOMO symptoms 422 people (43.11%) were classified. The data collected in the Polish sample are consistent with earlier results for adolescents from other European countries, such as Bosnia and Herzegovina. It should be noted, however, that only one quarter of the respondents do not show any FOMO factors, while about one third, depending on the perspective adopted, have inconsistent FOMO levels. The results of cluster analysis are presented in Fig. 2.

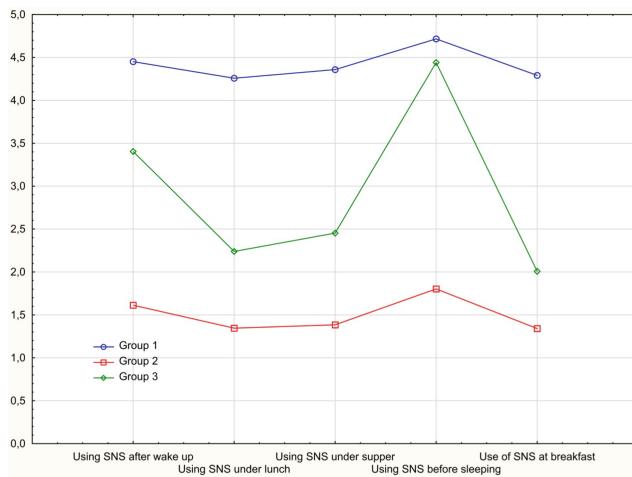


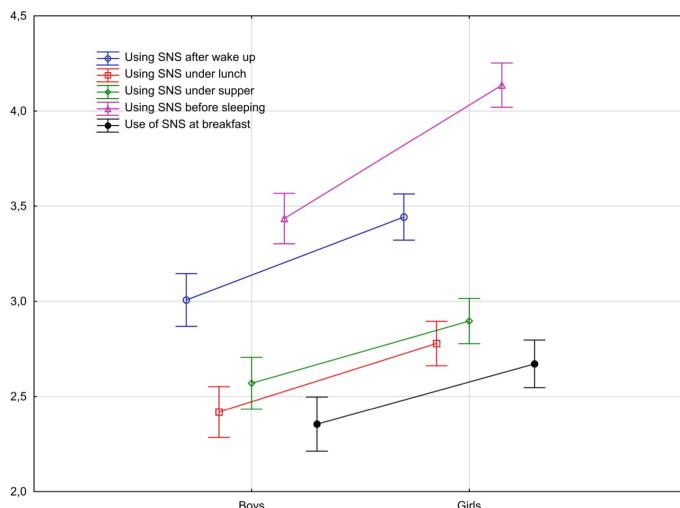
Fig. 2. FOMO cluster analysis by k-means for three groups of adolescents

The existence of each FOMO indicator is linked. The co-occurrence of indicators is always at medium or high levels. The most closely related are the FOMO symptoms occurring during meals. Adolescents who consume lunch and look at SNS content during this time also *very often* do so during other meals, such as dinner. Age is associated with all FOMO factors to a low but statistically significant degree. The relationship between FOMO factors does not assume a strong relationship for all factors. The results of the linear correlation coefficient are presented in Table 2.

Table 2. Coexistence of FOMO factors.

Variable	1	2	3	4	5
1. Using SNS after wake up	–				
2. Using SNS under lunch	0.573 ***	–			
3. Using SNS under supper	0.579 ***	0.804 ***	–		
4. Using SNS before sleepinf	0.640 ***	0.499 ***	0.533 ***	–	
5. Using SNS at breakfast	0.567 ***	0.666 ***	0.652 ***	0.460 ***	–
6. Age	0.246 ***	0.268 ***	0.268 ***	0.265 ***	0.199 ***

Gender is a statistically significant factor for two FOMO-related factors. First, it is the use of SNS before going to sleep ($F = 21,554$, $\eta^2 = 0,022$, $p < 0,001$) and just after waking up ($F = 60,989$, $\eta^2 = 0,059$, $p < 0,001$). In both cases it is girls who have a much higher level than boys. For the other indicators, girls are also slightly more likely to be affected. The differences between the FOMO indicators resulting are presented in Fig. 3.

**Fig. 3.** Gender differences in FOMO.

It is also noticeable that for all categories of adolescents living in small and medium-sized towns, such young people have a higher FOMO level than young people living in either rural or very large towns. A slightly higher level of FOMO occurs among people declaring that their family lives affluently than among peers declaring subjectively low and average family income status. Using a one-factor analysis of variance, it was noted that people with lower behavioural ratings achieve slightly higher average FOMO

values. Also young people with lower grades (last semester's average grades) have higher FOMO values. The type of school further determines FOMO level. Students from high schools and technical schools (secondary schools) show a slightly higher level of FOMO symptoms. However, this pattern can be linked to previous data, where FOMO level increases with the metric age. The differences between the average FOMO value and independent (sociodemographic) variables are presented in Table 3.

Table 3. Coexistence of FOMO factors.

Variable	Mean	SD	N	
Place of residence				
City of 100–200 thousand inhabitants	3.408	1.160	79	Sum of Squares = 71.099; df = 4; Mean Square = 17.775; $F = 13.113$; $p = < .001$; $\eta^2 = 0.051$
City of 50–100 thousand inhabitants	3.124	1.285	89	
A city of up to 50 000 inhabitants	3.343	1.066	237	
A city of more than 200 thousand inhabitants	2.865	1.111	52	
Village	2.769	1.191	522	
Behaviour grade				
5	2.992	1.188	334	Sum of Squares = 5.703; df = 5; Mean Square = 1.141; $F = 0.801$; $p = < 0.549$; $\eta^2 = 0.004$
4	2.934	1.260	198	
1-the lower	2.778	1.412	9	
2	3.435	1.283	23	
3	3.014	1.261	87	
6 – the highest	3.010	1.125	328	
Average school grade from the previous semester				
Between 1 and 2 (lowest)	3.323	1.397	13	Sum of Squares = 17.085; df = 4; Mean Square = 4.271; $F = 3.027$; $p = < 0.017$ $\eta^2 = 0.012$
Between 2 and 3	3.339	1.257	111	
Between 3 and 4	2.983	1.211	299	
Between 4 and 5	2.921	1.168	369	
Between 5 and 6 (highest)	2.943	1.131	187	
Type of school				
Middle school	3.131	1.150	295	Sum of Squares = 88.467; df = 4; Mean Square = 22.117; $F = 16.577$; $p = < 0.001$; $\eta^2 = 0.064$
High school	3.416	0.906	139	
The elementary	2.717	1.208	475	
Technical high school	3.611	1.270	57	
Professional	3.022	1.151	9	

4 Discussion and Summary

FOMO, like other types of problematic internet use, can be diagnosed in different ways. The various approaches to FOMO have in common the characteristic of the inappropriate situations in which smart phones and other devices with internet access are used. FOMO is an increasingly common phenomenon due to the aforementioned changes in the information society. On the basis of the data collected, it can be seen that FOMO is the most common phenomenon among adolescents compared to other threats mediated by digital media, such as cyberbullying, sexting, and computer piracy. FOMO in its strongest version (the occurrence of all factors at a very high level) included in the *Social Media Use* and the *Fear of Missing Out* tool concerns about a third of adolescents questioned in Poland. This is a group of young people who require special preventive support related to the development of critical thinking, support of self-control, and developing hobbies and interests not related to cyberspace. FOMO is a form of problematic internet use, which means that people with underdeveloped defense mechanisms against various forms of addiction are particularly vulnerable to the bio-psycho-social effects of this phenomenon.

At present, no uniform definition has been developed for the diagnostic criteria of FOMO, Internet addiction, and PIU. It is known, however, that with the continued development of the information society, the time being spent online is growing ever longer. However, time by itself is not a criterion on which to build a conclusion. It is more important to consider how appropriate it is to be online at given times, whether being online matches any real need, and what the purpose is of being submerged to such an extent in social networking sites. Researchers analysing risky behaviour on the Internet also draw attention to the mechanisms associated with the use of SNS. The creators of this type of e-service use many technical and socio-technical measures to attract users' attention. Many smartphone apps 'push' notifications to the user's phone to remind them to use that application again, especially if the app has been programmed to notice the user's absence. Other activities, such as the profiling of the content displayed on SNS, the ease of content sharing, the wide availability of multimedia, the opportunity to 'rate' images by way of 'likes', and the ease with which video can be shared, all encourage users to browse constantly, and to be perpetually online.

Digital competence is currently not just about the ability to view and share digital content. It is one of the key skills, which also includes the ability to recognise and defend oneself against FOMO or other forms of Problematic Internet Use. Due to the fact that young people have a high level of FOMO and are at the same time the group that uses new media the most, there is a need to strengthen the development of digital competence among adolescents in "soft areas". Such digital competence areas are the ability to control the time spent in networked media spaces, the ability to recognize and classify hidden mechanisms that cause uncontrolled immersion in networked media, and the ability to select information. Given the dynamics of change and the transformation of the information society, FOMO is becoming an increasingly visible threat resulting from the spread of new media, and requires appropriate educational measures to be taken.

Acknowledgements. The text was produced with the support of the research network COST CA16207 - European Network for Problematic Usage of the Internet. The author received the ITC

Conference Grant (ECOST-CONFERENCE_GRANT-Request-CA16207-2265), which enables the presentation of the research results.

References

1. Tomczyk, Ł.: FOMO (fear of missing out) – wyzwanie diagnostyczne i edukacyjne. Lubelski Rocznik Pedagogiczny **37**(3), 139 (2019)
2. Chaffey, D.: What happens online in 60 seconds? SmartInsights Homepage (2020). <https://www.smartsights.com/internet-marketing-statistics/happens-online-60-seconds/>
3. Eger, L., Tomczyk, Ł., Klement, M., Pisoňová, M., Petrová, G.: How do first year university students use ict in their leisure time and for learning purposes? Int. J. Cogn. Res. Sci. Eng. Educ. (IJCREE) **8**(2), 35–52 (2020)
4. Eger, L., Klement, M., Tomczyk, Ł., Pisoňová, M., Petrová, G.: Different user groups of university students and their ict competence: evidence from three countries in central Europe. J. Baltic Sci. Educ. **17**(5), 851 (2018)
5. Riordan, B.C., Cody, L., Flett, J.A., Conner, T.S., Hunter, J., Scarf, D.: The development of a single item FoMO (fear of missing out) scale. Curr. Psychol. **39**(4), 1215–1220 (2018). <https://doi.org/10.1007/s12144-018-9824-8>
6. Dempsey, A.E., O'Brien, K.D., Tiamiyu, M.F., Elhai, J.D.: Fear of missing out (FoMO) and rumination mediate relations between social anxiety and problematic Facebook use. Addict. Behav. Rep. **9**, 100150 (2019)
7. Nasr, H.E., Rached, K.S.B.: The problematic use of smartphone and fomo as antecedents of Facebook addiction. Acad. Mark. Stud. J. **23**(4), 1–10 (2019)
8. Kuss, D.J., Lopez-Fernandez, O.: Internet addiction and problematic Internet use: a systematic review of clinical research. World J. Psychiatry **6**(1), 143 (2016)
9. Anderson, E.L., Steen, E., Stavropoulos, V.: Internet use and problematic internet use: a systematic review of longitudinal research trends in adolescence and emergent adulthood. Int. J. Adolesc. Youth **22**(4), 430–454 (2017)
10. Tomczyk, Ł., Wąsiński, A.: Risk behaviors among youths in a two-aspect approach: using psychoactive substances and problematic using of internet. J. Child Adolesc. Substance Abuse, **29**(1), 27–45 (2020). <https://doi.org/10.1080/1067828x.2020.1805839>
11. Smahel, D., MacHackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Olafsson, K., Hasebrink, U.: EU Kids Online 2020: Survey results from 19 countries. LSE (2020)
12. Tomczyk, Ł., Szyzka, M., Stošić, L.: Problematic internet use among youths. Educ. Sci. **10**(6), 161 (2020)
13. Pyżalski, J., Zdrodowska, A., Tomczyk, Ł., Abramczuk, K.: Polskie badanie EU Kids Online 2018. Najważniejsze wyniki i wnioski. UAM, Poznań (2019)
14. Jupowicz-Ginalska, A., Jasiewicz, J., Kisilowska, M., Baran, T., Wysocki, A.: Fear of missing out a korzystanie z urządzeń umożliwiających dostęp do mediów społecznościowych na podstawie badań polskich internautów. Forum Socjologiczne, vol. 9, pp. 219–247 (2019)
15. Jupowicz-Ginalska, A., Jasiewicz, J., Kisilowska, M., Baran, T., Wysocki, A.: FOMO. Polacy a lęk przed odłączeniem—raport z badań. Uniwersytet Warszawski, Warszawa (2018)
16. Przybylski, A.K., Murayama, K., DeHaan, C.R., Gladwell, V.: Motivational, emotional, and behavioral correlates of fear of missing out. Comput. Hum. Behav. **29**(4), 1841–1848 (2013)
17. Bochenek, M., Lange, R.: NASTOLATKI 3.0. Raport z ogólnopolskiego badania uczniów. NASK, Warszawa (2019)
18. Modzelewski, P.: FOMO (fear of missing out)—an educational and behavioral problem in times of new communication forms. Konteksty Pedagogiczne **1**(14), 215–232 (2020)
19. Tomczyk, Ł., Selmanagic-Lizde, E.: Fear of missing out (FOMO) among youth in Bosnia and Herzegovina—scale and selected mechanisms. Child Youth Serv. Rev. **88**, 541–549 (2018)
20. Hetz, P.R., Dawson, C.L., Cullen, T.A.: Social media use and the fear of missing out (FoMO) while studying abroad. J. Res. Technol. Educ. **47**(4), 259–272 (2015)



Elderly Monitoring – An EPS@ISEP 2020 Project

Julian Priebe¹, Klaudia Swiatek¹, Margarida Vidinha¹,
Maria-Roxana Vaduva¹, Mihkel Tiits¹, Tiberius-George Sorescu¹,
Benedita Malheiro^{1,2}, Cristina Ribeiro¹, Jorge Justo¹, Manuel F. Silva^{1,2(✉)},
Paulo Ferreira¹, and Pedro Guedes^{1,2}

¹ Instituto Superior de Engenharia do Porto,
Rua Dr. António Bernardino de Almeida, 431, 4249-015 Porto, Portugal
² INESC TEC, Campus da Faculdade de Engenharia da Universidade do Porto,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
<http://www.eps2020-wiki5.dee.isep.ipp.pt/doku.php>

Abstract. In the spring of 2020, six undergraduate students from diverse countries and engineering fields decided to design together a solution to monitor the elderly. This project was performed as part of the European Project Semester (EPS) programme at Instituto Superior de Engenharia do Porto (ISEP). The EM-BRACE solution encompasses two interconnected devices (a home station and a bracelet) and mobile/Web twin applications. The bracelet measures and transmits vital user data (pulse, temperature and impacts) to the home station, whereas the latter measures home environment parameters (temperature, humidity and pressure) and sends local and bracelet data to an Internet of Things (IoT) platform. This way, these data become accessible via the mobile/Web application. Thereby, EM-BRACE monitors the health and environment of the elderly and timely notifies caregivers about problems, contributing to the well-being of the elderly and their families.

Keywords: Elderly monitoring · European project semester · Health · Innovative solutions · Internet of Things · Sustainability

1 Introduction

The European Project Semester (EPS) is a one semester programme offering students from different countries the opportunity to study at another university to develop the bachelor capstone/internship project within a team. It adopts a project based learning framework, with a strong emphasis on multicultural and multidisciplinary teamwork, and on the development of hard and soft skills. The syllabus is composed of a central project module supported by a set of complementary seminars, all taught in English. The programme has, since its beginning in 1995, been implemented by 19 European universities, called EPS providers,

including the Instituto Superior de Engenharia do Porto (ISEP) of the Polytechnic of Porto. EPS@ISEP has, in addition to EPS concept features, a strong focus on ethically aligned and sustainability driven design and development [4].

In the spring of 2020, six students from Mechanical Engineering (Germany), Production Engineering and Management (Poland), Biomedical Engineering (Portugal), Industrial Design (Romania), Transportation and Logistics (Estonia), and Telecommunications Technologies and Systems (Romania), chose to address the problem of elderly monitoring. Each member contributed with his/her own and different skills, culture and vision to create an innovative solution.

Humankind is facing global population ageing, with many old people living alone, suffering from health problems, and in need of frequent assistance [21]. Several accidents are not reported timely because elderly are simply unable to ask for help, leading to delayed treatments or even fatalities. Through non-stop monitoring, problems and accidents can be quickly detected and the repercussions substantially diminished. This made the team consider the design of a privacy preserving product, focused on improving the well-being of the elderly user and on helping caregivers.

One of the biggest challenges arising from caregiving is being able to monitor patients with few staff. Due to technological evolution and constant cost reduction of electronic components, affordable remote monitoring has become a reality. This allowed the team to idealise a system capable of measuring vital signs and recognising patient falls. The final solution targets elders, living mostly indoors, with some (albeit reduced) degree of mobility. The health focus of the project introduced numerous product design constraints, such as respecting ethical requirements, sustainability, and environmental protection. The maximum budget of 100 € required the use/reuse of low-cost materials and components.

This paper is structured in six more sections, presenting the performed background studies, the EM-BRACE solution, namely its design, development and tests, discussing the results and, finally, summarising the outcomes of this EPS@ISEP project.

2 Background

The project started by researching related products, marketing, sustainability, ethics and deontology. The aim was to identify the mandatory values, principles, and criteria to be respected and, then, derive the requirements of the solution.

2.1 Related Solutions

Five different categories of monitoring devices were considered: (*i*) wearable elderly monitors; (*ii*) unwearable elderly monitors; (*iii*) domestic hand-held vital signs monitors; (*iv*) fitness trackers; and (*v*) smartwatches.

Each category has different specifications and characteristics. Wearable elderly monitors have integrated fall detectors and emergency buttons but they

do not monitor vital signs [10, 17]. Unwearable elderly monitors are based on artificial intelligence, record behaviour patterns, and detect irregularities. Being unwearable, they do not have the ability to monitor vital signs [2, 9, 23]. Domestic vital sign monitors, such as electrocardiograms, blood pressure or heart beat readers, target people in general and monitor only a small number of vital signs. They are particularly suitable for sports enthusiasts or for people with known diseases [3, 16, 18]. Fitness trackers are wearable devices that monitor daily physical activities and fitness-related metrics, such as steps, running distance, heart rate, sleep patterns, swimming laps or calories burned [7, 8, 12]. Smartwatches are wearable computers, integrating wristwatch, smartphone and fitness tracker functions. However, they suffer from reduced battery autonomy due to the many functions performed [1, 6, 20]. Smartwatches and fitness trackers were not designed having elderly in mind and, therefore, provide only partial vital monitoring and are difficult to use and understand. Products for the elderly must include, as a prerequisite, safety features, such as emergency buttons or drop sensors [19].

As a result, the team decided to design a solution integrating both vital sign monitoring and safety requirements, as well as caregiver automatic notification and access to indoor data. The innovation of this idea lies in the combination of individual features to fill a market gap. Thus, the team decided to measure and remotely monitor environmental and vital parameters as well as falls. Also, as the modularity of unwearable monitors opens a large market and wearable devices facilitate customisation, the adopted concept includes both types.

2.2 Ethics, Marketing and Sustainability

Concerning the ethical aspects, the goal of the study was to find an ethically aligned solution to monitor the elderly and improve the quality of human life. The project has a profound ethical goal: to contribute to the well-being of consumers by monitoring the health of aged users, while adopting safe components and processes, *i.e.*, which have been previously tested, validated, and certified. The gathered data must be encrypted and stored, ensuring restricted access and compliance with European Union (EU) General Data Protection Regulation. Ethically, the design and development of the device should follow the engineering profession ethical values [15] and comply with the applicable EU directives. This way, the health, security, safety and privacy requirements of the consumer are fulfilled, boosting elderly confidence and independence. Furthermore, and considering environmental ethics, the device should present an eco-friendly design, easy maintenance and, at the end of the product life cycle, generate minimal waste by reusing or recycling materials and components. In terms of marketing ethics, the promotion of the product should focus on the real benefits and its true purpose, namely, the well-being of the user. Finally, the product name EM-BRACE was checked against existing trademarks to avoid infringing intellectual property, and the adopted logo elements were not found in other brands.

Regarding marketing, the contemplated strategy was based on the steady worldwide growth of elderly living alone over the last years. In northern coun-

tries, over 30% of the elderly live alone [5], suggesting a major market potential for the product. Unaccompanied elderly are prone to accidents and the main objective of the product is to ensure the safety and independence of this segment of people. As a result, the target was set to “men and women over 60 years with an average income, individuals and families interested in a healthy lifestyle”. Furthermore, the analysis of the Political, Economic, Social, Technological, Legal, and Environmental factors highlighted as opportunities: (i) increasing life expectancy rates; (ii) increment of the global demand for wireless monitoring devices; (iii) willingness to share personal and health data; (iv) higher pension expenditure trend and pension values; (vi) expected expansion of the health monitoring market; and (vii) remote monitoring versus hospital care costs.

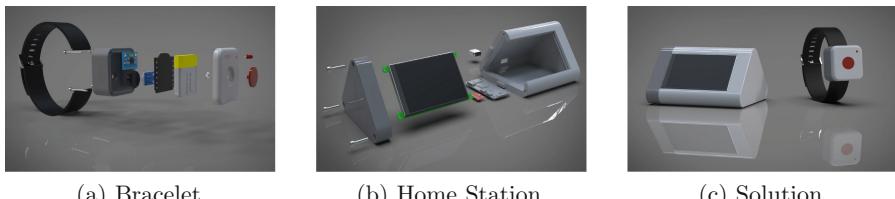
In terms of sustainability, the United Nations identified 17 Sustainable Development Goals (SDG) to ensure a better and more sustainable future for all [22]. This project contributes to the promotion of good health and well-being (SDG 3) and, to a lesser extent, to responsible consumption and production (SDG 12).

3 Design

The EM-BRACE concept is a solution for semi-autonomous elderly who live alone. It comprises two paired devices: (i) a wearable bracelet to measure vital signs, and (ii) a home station to measure room pressure, temperature, and humidity. The data collected by the bracelet will be transmitted via Bluetooth to the home station, where it will be stored on a memory card, and, then, encrypted and transmitted to an Internet of Things (IoT) platform. A dedicated mobile/Web application will allow authorised persons to monitor non-stop the status of the user and home environment. The packaging will be recyclable and fully reusable as a pill organiser and medication storage box.

The design was kept simple and easy to use. The bracelet has an emergency button, which is recessed in the lid to avoid false triggers caused by accidental activation, and a light emitting diode (LED) to indicate when to recharge. The upper part of the casing is light grey, and the lower one is dark grey, as can be seen in Fig. 1a. These neutral colours are suitable for any age and clothing, and direct the user attention to the button (nevertheless, the design of the bracelet and the home station contemplates different colour options, taking into account the preferences and needs of the user). The home station has a triangular profile where the display is mounted at an angle to facilitate data reading (Fig. 1b).

The designed monitoring system (Fig. 1c) comprises the bracelet and home station. The bracelet has a battery, an Arduino BLE with a Bluetooth communication interface, a LED to indicate when to recharge, an emergency button, a gyroscope to detect falls and vital data sensors (oxygen saturation, pulse and body temperature). The system reads and stores emergency and sensor data in the local flash memory and, once paired with the home station, shares all data. The home station includes a Raspberry Pi 3 single-board computer, a battery and a temperature, humidity and pressure sensor. The Raspberry collects, interprets and displays room and bracelet data, as well as encrypts and sends all



(a) Bracelet

(b) Home Station

(c) Solution

Fig. 1. EM-BRACE design: exploded axonometric views and 3D model

data to the IoT database, where they become available for the mobile and Web applications. Emergency events trigger the notification of the caregiver or family.

4 Development

The development encompassed three components: (*i*) Monitoring & Control System, (*ii*) Mobile App, and (*iii*) Web App. Due to the imposed COVID-19 constraints, the decision was to develop what was possible and simulate the rest.

Monitoring and Control System. The team assembled a reduced version of the designed monitoring & control system using a Raspberry Pi 3, a DHT22 humidity and temperature sensor, an Arduino Uno, two buttons, an HC-05 Bluetooth Low-Energy module and a LED.

The Raspberry Pi together with the DHT22 sensor are responsible for the home station control, while the Arduino Uno together with the HC-05 Bluetooth Module perform the wearable-side control. The “Simulation Button” triggers emergency events on the bracelet-side, indicating a potential fall. If this trigger is not reset by the user in 5 s, the emergency event is confirmed and sent to the database, where the notification and email push system are activated.

The Arduino produces pseudo-random values within the expected bracelet sensor ranges. Once the Arduino (bracelet) and the Raspberry (home station) are paired, all bracelet data are shared with the Raspberry (emergency state, pulse, peripheral oxygen saturation and body temperature). Next, the Raspberry processes bracelet and room data. To calculate the pulse rate, it discretises the signal and, then, applies a digital gradient peak detector to several heart beats. Finally, the heart rate, level of oxygenation, emergency and average temperatures (using a 10-sample filter) are uploaded to the Firebase IoT platform, where it becomes available to the mobile and Web applications. Firebase is used since it allows the simple provision, management, and automation of connected devices [14].

Mobile App. The mobile application has a registration function to maintain a database of users and their roles (patient, doctor, caregiver, family) (Fig. 2a), an authentication function for accessing the app (Fig. 2b) and four tabs: (*i*) Home to provide the welcome screen and user notifications; (*ii*) Monitor to display real-time measurements; (*iii*) Calendar to schedule and check appointments (Fig. 2c);

and (iv) Account to show/set account information (such as email and name), notification and calendar synchronisation permissions (Fig. 2d).

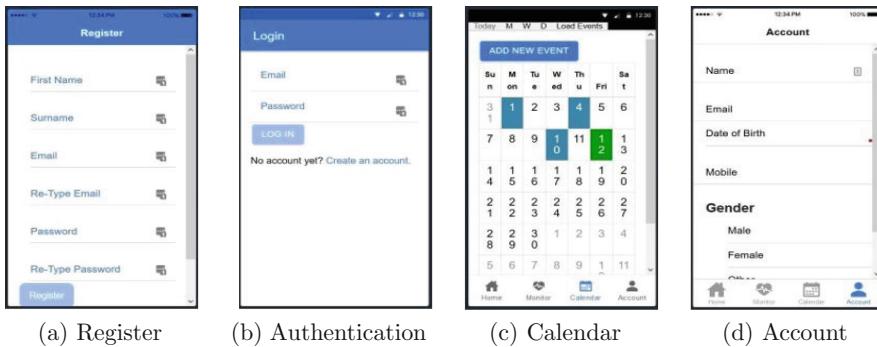


Fig. 2. Mobile application

Web App. The Web application provides cross-platform access from Android, iOS, Linux, MacOS or Windows. It encompasses the front-end, that defines the User Interface (UI), developed with the Ionic Framework and AngularJS, and the back-end, that offers an Application Programming Interface (API) to store and retrieve data, supported by Google’s Firebase. Firebase provides analytics, databases, messaging, and crash reporting, making it perfect for testing and simulation [11]. The code tree structure is similar to that of any AngularJS website, combining HyperText Markup Language (HTML), Cascading Style Sheets (CSS), and Angular components.

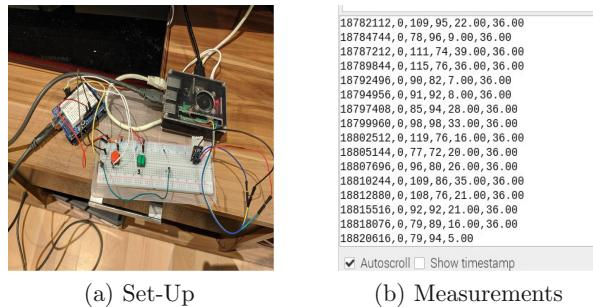
5 Tests

The tests performed validated code rather than the control system operation. The latter was not fully implemented due to the undergoing pandemic. Nevertheless, it was possible to verify the proper functioning of the Mobile app, Web App and parts of the control code, both on the bracelet and home station.

The mobile application functional tests included registration, authentication, calendar synchronisation and profile updating. All user data is stored in Firebase. The registration is a validation form which checks the email format, the password size (at least six characters) and a re-type field. After authentication, the home tab opens to greet the user and show upcoming appointments. The calendar tab provides an interface to view, add, or remove personal events. A random event generator was used to verify if the information was correctly displayed. The “Add new event” button adds an event for the next hour and synchronises with the database. The account tab shows and allows the updating of the user profile, *e.g.*, age, weight or gender. The monitor tab provides, depending on the

user role, access to different real time sensor data. The tests with the Mobile App validate also the Web App since Ionic is a cross-platform framework.

The monitoring & control system tests were performed with the physical set-up presented in Fig. 3a. The bracelet is controlled by the Arduino and the home station by the Raspberry. This set-up allowed testing the: (i) creation and communication of environment and vital data to the database; (ii) processing of pulse data; and (iii) generation and storage of emergency events in the database.



(a) Set-Up

(b) Measurements

Fig. 3. Monitoring & control system tests

Although the Arduino and Raspberry Pi were paired, the Bluetooth connection systematically failed regardless of the connection method (Bluetooth serial communication and socket communication). Therefore, communication was established via USB. The Secure Shell (SSH) protocol was used to establish the communication between a master computer and the Raspberry Pi, allowing remote management of the system. This code, running on the master computer, was also able to successfully interact with the IoT platform.

The Arduino code was tested using the Terminal/Command Line Interface, as shown in Fig. 3b. It successfully generated and printed the random sensor values to the serial connection, including, from left to right, the timestamp (ms), emergency state (true or false), pulse rate (beat/min), peripheral oxygen saturation (SpO_2) and body and environment temperature ($^{\circ}\text{C}$).

Figure 4a displays real pulse readings – infrared (IR) and red light measurements acquired by a pulse sensor and obtained from [13] – and Fig. 4b the random values generated by the Arduino and processed by the Raspberry Pi. Concerning Fig. 4a, the first plot represents the values read by the internal IR receiver and the second shows the red light emitted. While the red/infrared modulation ratio indicates the SpO_2 level, the processing of the IR signal provides the pulse rate (beat/min).

To obtain a meaningful pulse rate, several signal processing algorithms were explored on the Raspberry side. The first Python sketch reads and saves the serial data, removes noise, saves the filtered data to a Comma Separated Value file, as well as plots and stores the resulting graph as a Portable Network Graphics (PNG) file. The second sketch performs a Fast Fourier Transform (FFT) analysis

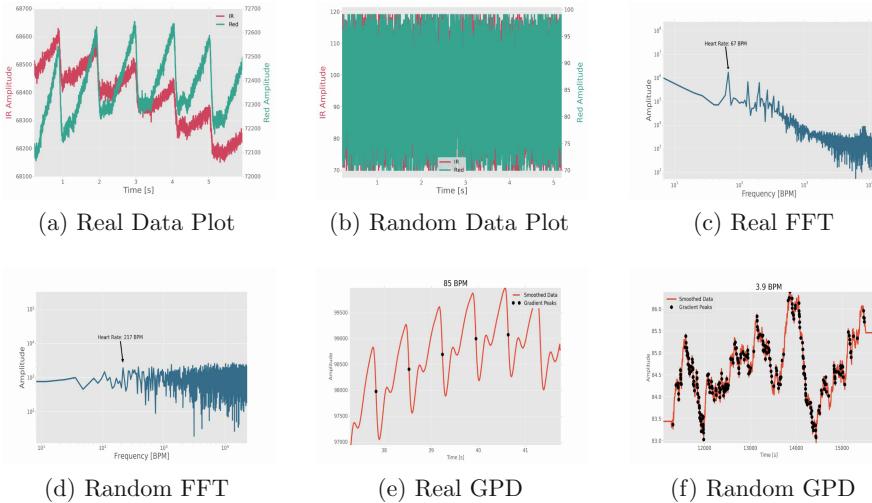


Fig. 4. Signal processing experiments

on the saved samples to determine the fundamental frequency (Fig. 4c), which corresponds to the pulse rate (beat/min). However, the FFT requires several heart cycles to determine the correct rate. Finally, a second-order gradient function was used to approximate the pulse rate. Since the steepest point in the circulatory cycle is the systolic point (heart contraction), the method applies a gradient peak detection (GPD) algorithm to identify the systolic gradient peak. Figure 4e displays the promising results obtained.

6 Discussion

EM-BRACE allows the remote monitoring of the health and environment of the elderly, and is in line with recent remote health monitoring trends. The performed tests allowed to perceive the potential of this solution, leading to a successful future implementation.

The development of a full proof-of-concept prototype would allow real testing, debugging and refinement. Also, a larger budget would permit choosing higher quality components, *e.g.*, providing longer bracelet battery autonomy, better accuracy and more frequent readings or reducing dimensions. Smaller and lighter components would improve the product design and daily comfort. Moreover, the team would need to continue to pursue high-performance solutions for the privacy and encryption of data.

Despite the difficulties encountered due to the COVID-19 outbreak, the team managed to fulfil the initial objectives of the project. The members adapted to the new requirements, but, unfortunately, it was impossible to make the prototype. Nevertheless, the control system was simulated, the mobile application was

developed and tested, and mock data was stored in the IoT platform. To illustrate the actual dimensions and appearance, the cases of the two devices were 3D-printed (Fig. 5). Without a prototype, the functional tests and simulations were carried out exclusively online, diminishing the relevance of the results.

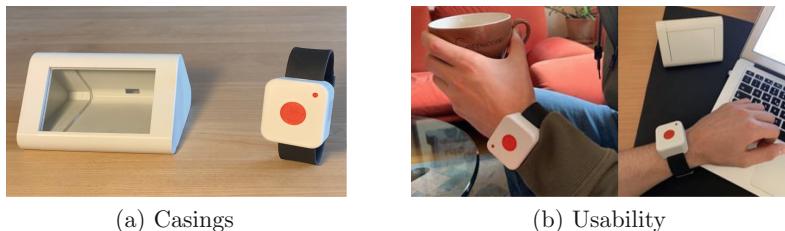


Fig. 5. Printed components

Future steps will focus on the assembly and test of the EM-BRACE prototype, leading to the refinement and, possibly, to the addition of new features.

7 Conclusion

This paper presented the development by an international team of EPS@ISEP students of a solution for the remote monitoring of the health and environment of the elderly. The initial objectives were accomplished, except for the full construction and physical tests of the prototype, which were not performed due to the suspension of face-to-face activities imposed by the COVID-19 pandemic constraints. The team successfully designed the EM-BRACE solution (device, control system, mobile/Web applications and packaging) due to its inner organisation and the support of the teaching staff. The design meets all requirements initially identified.

Acknowledgement. This work was partially financed by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project UIDB/50014/2020.

References

1. Apple: Buy Apple Watch Series 5 (2020). <https://www.apple.com/shop/buy-watch/apple-watch>. Accessed September 2020
2. Casenio, A.G.: Casenio für Betreutes Wohnen (2020). <https://www.casenio.eu/geschaefsfelder/betreutes-wohnen/>. Accessed September 2020
3. Chronolife: Home (2020). <https://www.chronolife.net/>. Accessed September 2020
4. Duarte, A.J., Malheiro, B., Arnó, E., Perat, I., Silva, M.F., Fuentes-Durá, P., Guedes, P., Ferreira, P.: Engineering education for sustainable development: the European project semester approach. IEEE Trans. Educ. **63**(2), 108–117 (2020)

5. Eurostat: People in the EU - statistics on household and family structures (2020). https://ec.europa.eu/eurostat/statistics-explained/index.php/People_in_the_EU_-_statistics_on_household_and_family_structures. Accessed September 2020
6. Fitbit: Fitbit Versa 2 Watch (2020). https://www.fitbit.com/us/products/smart_watches/versa. Accessed September 2020
7. Fitbit: Our most advanced tracker ever (2020). <https://www.fitbit.com/be/charge3>. Accessed September 2020
8. Garmin Ltd.: Garmin vivosmart 4: Fitness Activity Tracker (2020). <https://buy.garmin.com/en-US/US/p/605739>. Accessed September 2020
9. Gigaset: Gigaset smart care (2020). https://www.gigaset.com/de_de/gigaset-smart-care/. Accessed September 2020
10. Society: GoLiveClip (2020). <https://www.goliveclip.eu/solutions/goliveclip/>. Accessed September 2020
11. Google: Firebase helps mobile and web app teams succeed (2020). <https://firebase.google.com/>. Accessed September 2020
12. Huawei: HUAWEI Band 3 Pro (2020). <https://consumer.huawei.com/en/wearables/band3-pro/>. Accessed September 2020
13. Hrisko, J.: Arduino heart rate monitor using MAX30102 and pulse oximetry (2020). <https://makersportal.com/blog/2019/6/24/arduino-heart-rate-monitor-using-max30102-and-pulse-oximetry>. Accessed September 2020
14. KaaIoT Technologies: What is the Internet of Things Platform (2020). <http://www.kaaproject.org/what-is-iot-platform>. Accessed September 2020
15. National Society of Professional Engineers: Code of Ethics (2019). <https://www.nspe.org/resources/ethics/code-ethics>. Accessed September 2020
16. Omron: HeartGuide Wearable (2020). <https://omronhealthcare.com/products/heartguide-wearable-blood-pressure-monitor-bp8000m/>. Accessed September 2020
17. Panion: M-GUARD (2020). <https://www.panion24.com/>. Accessed September 2020
18. Qardio: Smart Wearable ECG EKG Monitor - QardioCore (2020). <https://www.getqardio.com/qardiocore-wearable-ecg-ekg-monitor-iphone>. Accessed September 2020
19. O'Dea, S.: Smartphone users worldwide 2020 (2020). <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>. Accessed September 2020
20. Samsung: Samsung Galaxy Watch (2020). <https://www.samsung.com/global/galaxy/galaxy-watch/>. Accessed September 2020
21. United Nations: World Population Prospects - Population Division (2019). <https://population.un.org/wpp/Download/Standard/Population/>. Accessed September 2020
22. United Nations: About the Sustainable Development Goals – United Nations Sustainable Development (2020). <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>. Accessed September 2020
23. Vitracom: Safe@home (2020). <http://vitracom.de/en/assisted-living/safe-home.html>. Accessed September 2020



Emotion Recognition in Children with Autism Spectrum Disorder Using Convolutional Neural Networks

Rodolfo Pávez¹, Jaime Díaz¹ , Jeferson Arango-López² , Danay Ahumada³ , Carolina Méndez⁴, and Fernando Moreira⁵

¹ Depto. Cs. de la Computación e Informática, Universidad de la Frontera, Temuco, Chile
jaimeignacio.diaz@ufrontera.cl

² Depto. de Sistemas e Informática, Universidad de Caldas, Manizales, Colombia

³ Depto. Procesos Diagnósticos y Evaluación, Universidad Católica de Temuco, Temuco, Chile

⁴ Hospital Hernán Henríquez Aravena, Temuco, Chile

⁵ REMIT, IJP, Universidade Portucalense and IEETA, Universidade de Aveiro, Aveiro, Portugal

Abstract. Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder, and it is defined as the persistent difficulty in maturing the socialization process. By applying different therapies, health professionals have made advances that promise to improve patients' conditions. Taking advantage of technologies like Artificial Intelligence (AI) techniques, improvements in therapies have been obtained. This article proposes creating a smart mirror to recognize five basic emotions: Angry, Fear, Sad, Happy, and neutral. The above is based on Convolutional Neural Networks (CNN), which can support therapies performed by health professionals to children with ASD.

Keywords: Autism spectrum disorders · Emotion recognition · Human-computer interaction · Convolutional neural networks

1 Introduction

The continuous technological development has become fundamental support in different medical areas. One particular area is therapies for people with Autism Spectrum Disorder (ASD), where facial emotion recognition interventions using technology are particularly promising [1].

The recognition of facial emotions is a component of social cognition [2] and is essential for effective communication and social interaction [3, 4]. In ASD people, who have a neurodevelopmental disorder characterized by deficiencies in social communication and unusual restricted and repetitive behaviors [5], there is difficulty recognizing others' facial emotions, making social interaction difficult [6].

Helping their emotion recognition skills through intervention tools could significantly improve children's social interactions with ASD [7]. In this area, the recognition of emotions through facial expressions has had an upward trend. Emotions could be detected by inferring them from the different movements of muscles in the face. Research

is carried out using technology to intervene in children with autism spectrum disorder [8]. In this article, the development of an intelligent mirror for recognizing emotions, based on Convolutional Neural Networks (CNN), is proposed. It would continuously support the health professional, being a tool with various images that train emotions to patients with ASD. Besides, unlike other tools such as using cards [6], The Smart Mirror reduces the exposure to Covid-19 because it has less manipulation by the health professional and the ASD user.

2 Background

2.1 Recognition of Emotions as Therapy for Children with ASD

The recognition of emotions plays a relevant role in a social environment. Children with ASD have deficiencies in initiating and responding to social or emotional interaction [5]. Consequently, professionals who work with children with ASD can use different techniques to assist in the training for emotion recognition; Examples may include using Microsoft-Kinect, webcams to capture emotions in a controlled environment, games on mobile devices, expressive robots, among others.

2.2 Neural Networks

Deep Learning (DL) has emerged as a promising model for solving various problems, such as natural language processing, speech recognition, and visual recognition [9]. This is due to increased research on convolutional neural networks, achieving promising results in various tasks [9]. By definition, neural networks are a class of mathematical algorithms inspired by the brain's structure and natural functioning to classify information and make decisions. This approach works very well in the construction of prediction models using computer vision [10]. Some examples of its use can be autonomous cars, facial recognition, the classification of data in bioinformatics, or the recognition of facial emotions.

2.3 Smart Healthcare

The concept of Smart healthcare is directly associated with the use of technologies in the medical area, where different actors are involved. Generally, this concept is implemented to support the prevention, diagnosis, monitor diseases, and possible treatments [11]. Consequently, the smart healthcare concept uses technological advances to transform traditional methods into more efficient and personalized to improve the results obtained [11].

In the therapies carried out by health professionals to children with ASD, different activities are implemented so that children can imitate and recognize emotions. Professionals must show them the activity's meaning to better perform therapy [11, 12]. An alternative is the one we propose in this research, where a mirror is used to validate and show in their face the meaning of the emotion they are trying to imitate. Because a mirror is an essential element in people's daily routines, it is unlikely that children find it strange; therefore, they should feel comfortable developing activities.

For this reason, and under the previously exposed context, we propose to develop a smart mirror for the support of facial emotion recognition therapies for children with ASD. We implemented a smart mirror according to the specifications presented in Fig. 1: (a) a 3 mm bidirectional mirror, (b) a Raspberry Pi 4 Model B of 4G ©, a refurbished LCD screen, and a (d) Raspberry Pi Camera Module v2.

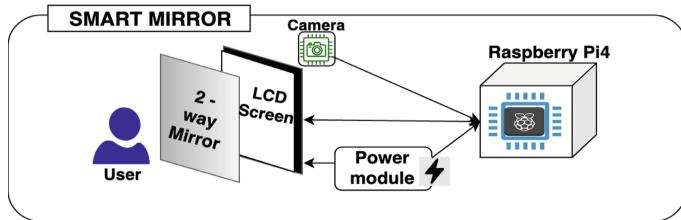


Fig. 1. Physical components of the smart mirror proposal

Therefore, by joining the parts and pieces described above and together with the software that we detail in this research, it was possible to obtain concrete results associated with recognizing emotions based on computer vision. This proposal serves as an automated input for a professional who works in therapies with children with ASD to obtain real-time relevant information.

3 Related Works

The use of technological tools in a therapeutic context with children with ASD generally turns out to be beneficial by increasing motivation, decreasing inappropriate behavior, and increasing their attention, which can be translated -on some occasions- into better learning compared to the traditional methods [13].

Therefore, when analyzing research related to recognizing facial emotions with the use of technologies in children with ASD, the results show an upward trend in this field research during the last ten years [8]. In turn, there is an increase in children's motivational levels in developing activities mediated by the use of technologies [14]. Similarly, according to a work that aimed to verify, analyze, and categorized research results related to the use of technologies to train the recognition of facial emotions in children with ASD, the most used technique was artificial intelligence, together with information systems as a type of contribution [8].

This research shows some initiatives that address emotion recognition in children with ASD using similar technologies (see Table 1). The first column of Table 1 gives (i) the emotion recognition technique applied. The second column shows the associated technologies used for recognition, and finally, we present the main findings of the selected articles. This last one summarizes if the contribution was a “game” or a traditional information system (IS).

It is important to note that none of the listed investigations have their source codes available to be analyzed and thus compare or try to improve the results obtained at the model level. After analyzing the recommendations and results from related works, it is

possible to generate a prototype based on the idea of a “*smart mirror*” focused on the recognition of facial emotions through a camera.

Table 1. Related research of emotion recognition in children with ASD using technology

Recognition technique	Associated technologies	Main findings
Artificial intelligence (convolutional neural networks)	Desktop computer	(Game) Fan et al. proposes that images used in the game should be in a familiar context, such as school, home, or park [15]
Artificial intelligence (personal algorithm)	Mobile devices	(Game) Harrold et al. mention that children show a high level of motivation while playing. They had problems capturing emotion under specific scenarios related to lighting [14]
Artificial intelligence (SVM and Logistic Regression Classifier)	Portable motion camera (Google Glass) & Mobile device	(IS) Voss et al., with his real-time facial recognition system, describes that children respond better to auditory and visual interactions [16]
Artificial intelligence (SVM)	Desktop computer	(IS) Chu et al. propose a form of SVM-based facial emotion recognition with transition detection using a webcam [17]
Artificial Intelligence (Machine Learning Algorithms: Bayes Network, Naïve Bayes, ANN, kNN, Random Forest, Decision Tree, SVM)	Neurofeedback (Emotiv EPOC neuroheadset)	(IS) Fan et al. uses electroencephalography (EGG) together with machine learning algorithms for the recognition of emotions in children. They conclude that using EGG it is possible to interpret the brain process associated with emotional expressions [18]
Artificial Intelligence (Machine Learning Algorithms: Decision Trees, Random Forest, SVM, K- NN and AdaBoost)		(IS) Uluyagmur-Ozturk et al. focus on classifying participants based on their performances during an emotion recognition experiment [19]
Artificial intelligence (SVM)	Multimedia (Videos, Images)	(IS) Tamil et al. mention some promising techniques for capturing facial emotion, improving accuracy to 87% after classification [20]

(continued)

Table 1. (continued)

Recognition technique	Associated technologies	Main findings
Artificial intelligence (SVM)	Camera RGB-D	(IS) Zhao et al. propose using an RGB-D camera to correct the head's angle or posture for improving the recognition of facial emotion. The general average of recognition was 91.6% [21]
Artificial Intelligence (Machine Learning Algorithms)	Portable motion camera (Google Glass) & mobile devices	(IS) Washington et al. mentioned that children tend to respond better to audio comments after her facial emotion recognition experiment than to visual feedback [22]
Artificial Intelligence (DTW Classification Algorithm)	Desktop computer	(IS) Adams et al. concluded that based on a video containing facial expressions, the child tried to imitate what he is visualizing while that is happening. With this activity, the algorithm analyzes and compares both tables, providing feedback on the result [23]

This approach meets the specification as intuitive, customizable, and offers visual stimuli when capturing a facial emotion. The recognition of emotion is achieved using convolutional neural networks (CNN) and an information system that is a crucial piece to generate an interaction between the child and the mirror.

4 Methods and Experiments

The next section describes the training dataset and then the proposed technology architecture.

4.1 Dataset and Features

In computer vision (CV), image databases are fundamental support to validate or innovate new techniques for the automatic detection of facial emotions. Along these lines, there are various associated image databases such as Extended Cohn-Kanade Dataset (CK +) [24], FER2013 [25], AffectNet, Japanese Female Facial Expressions (JAFFE), among others.

As a data set for the generation and subsequent comparison of the models, FER2013 [25] and a CK + [24] were selected. The previous, since both options are publicly

accessible, thus facilitating access to each of the resources. FER2013 contains more than 35,000 images distributed among seven categories associated with facial expressions, while the second includes a total of 593 images, distributed among eight categories. The categories are created based on the facial emotion shown in the image, being labeled in one of the following base emotions: happiness, neutrality, sadness, anger, surprise, disgust, fear. In the case of CK +, it contains an extra category associated with contempt.

In previous studies, children on the autism spectrum have been shown to more readily understand the link between the following four raw emotions: fear, anger, happiness, and sadness [26]. Therefore, based on this information, it was decided to construct a model that can respond to these four emotions. This includes neutrality to recognize transitions between emotions and to be used for cases where instructions could be received in therapy.

The data set was divided into three categories, as shown in Table 2, where 80% of the total images are left available for a training process, 10% for validation, and 10% for evaluation.

Table 2. Summary of images used.

	FER-2013	CK +
Training	25071	442
Validation	3.133	56
Evaluation	3.134	57
Total of images	31.338	555



Fig. 2. Example images of FER 2013 according to categories

These images were categorized into one of the five emotions described above (See Fig. 2). It is essential to highlight that images provided by FER2013 are found in gray scales and with a dimension of 48×48 pixels, while those of CK + are mostly found in gray scales and with a dimension of 640×480 pixels.

4.2 Architecture

It was decided to evaluate 2 of the classic convolutional neural network (CNN) architectures: VGG 16 and ResNet50. VGGNet was invented by the Visual Geometry Group of the University of Oxford, positioning itself as the first finalist of ImageNet's Large Scale Visual Recognition Challenge in 2014 on the classification area (ILSVRC-14)

[27]. On the other hand, Deep residual networks, better known as ResNet, are among the most recent architectures and one of the most used neural networks for image recognition. ResNet won the ImageNet Large Scale Visual Recognition Challenge in 2015 (ILSVRC-15) [28].

Regardless of the architectures or dataset used, the evaluation dataset was considered to inform the model's precision for each of the tests performed. Each model was trained from zero to 100 epochs on a Google Colab instance. It was necessary to use similar methods when training the models to compare the results.

It is worth mentioning that the FER2013 source is based on a CSV, so it is necessary to reconstruct them as an image with the default dimensions (48×48 px). The training proceeds when the images from both datasets are obtained digitally (FER2013 and CK +).

For both cases, when generating the model, a resize is defined in the images of 200×200 pixels. Stochastic Gradient Descent is an optimization algorithm that is usually used on large data sets. By definition, the examples to be considered are drawn randomly from the sample and processes during the experiment [29]. For this reason, it was decided to use Stochastic Gradient Descent (SGD) with momentum at a learning rate of 0.01. Different optimizers were tested, including Adam, where SGD seemed to perform slightly better.

Under the same scheme, two additional activation layers of the RELU type were added with a dropout value of 0.5. Besides, to train the model with a more significant number of images and that the result obtained could be prepared to evaluate small variations at the entrance, the models were trained by increasing the data, mainly based on flipping the images horizontally specific configuration parameters.

Due to the difference in the number of images, it takes approximately 2 to 4 h to train the model using the FER2013 dataset and about 10 min using the CK + dataset.

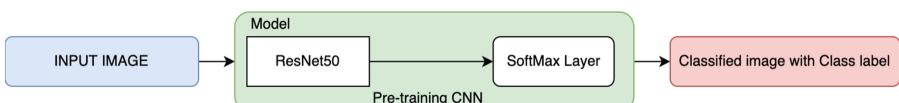


Fig. 3. Process for obtaining the associated emotion overview.

Some of the investigations included in the systematic mapping of the use of technologies in children with ASD [8] that used a camera to capture emotions mention that one of the problems detected to capture the child's emotion correctly was the position of the camera. In the case of the innovation project presented in this research, it is expected that the child is always positioned facing the mirror seeing its reflection. Consequently, the camera was placed on top of it to capture its image in real-time completely.

The proposed solution included a personalized welcome message for the child, reflected in the mirror, to obtain the child's attention. Subsequently, the health professional selected the emotion to work on in the session from the available options: Angry, Fear, Happy, Sad, and neutral. Once an emotion was selected, a set of associated images was displayed to recognize the emotion projected by the mirror and later

manage to imitate it to perfection. If this was not achieved, the mirror gave the possibility of moving to the next emotion or generating a pause (leaving this as a regular mirror), which gave the professional the opportunity of working with the face's image, reinforcing weak aspects. The smart mirror acts as a support in therapy, capturing and evaluating the child's emotion reflected in the mirror, generating emotional recognition and expression training. Then, once the child's image was captured, only her facial area was extracted using the object detection method proposed by Viola-Jones [30]. Additionally, once the child's facial image was obtained, it entered the classifier to decide what type of emotion matches. This process is displayed step by step in Fig. 3.

As it was a real-time capture activity, the time it may take for the device to classify an image must be considered. For the case of raspberry PI 4 used in this study, the average classification time was 0.05 s. Once the match between the selected emotion and the child's emotion was obtained, a hand-clapping appears, representing favorable results. Otherwise, a message is displayed, encouraging them to try again. It is important to note that this coincidence must remain for at least 2 s before showing the respective feedback.

5 Results

Finally, due to the results obtained from the trained model, considering the data corresponding to evaluation, it was decided to use the ResNet50 architecture as a CK + dataset for the generation of the model included in the Raspberry Pi 4.

Table 3. Final results using the evaluation data

Heading level	FER-2013	CK +
ResNet50	72.3%	89.9%
VGG16	70.9%	93.3%

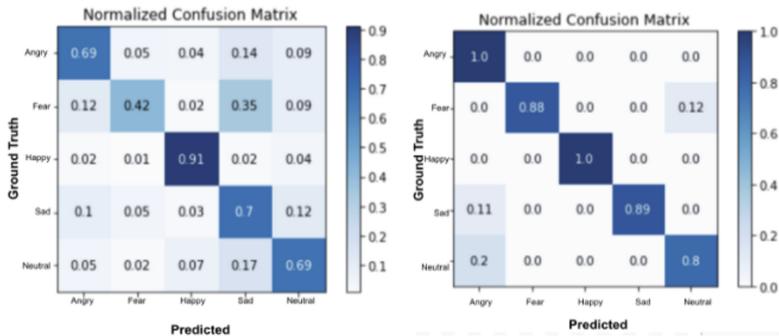


Fig. 4. Example of the confusion matrix using VGG16 with FER 2013 (left) and CK + (right).

In Table 3, we summarize the differences in the results. The differences appear not only due to the architecture but also due to the dataset used. In some cases, a difference

exciding 22% between different datasets but the same architecture. On the other hand, for both architectures, a maximum precision value was achieved when the training was close to 40 epochs. For example, the precision for the case of the VGG16 architecture and the FER2013 dataset tends to have very similar values between training and validation close to epoch number 40.

We used a confusion matrix to know the algorithm's performance in supervised learning (see Fig. 4). These results allowed us to compare the values obtained in each category using the evaluation data.

6 Conclusions

In this article, we presented a technological prototype of smart healthcare that uses artificial intelligence. This approach supports the therapies of children with ASD in emotion recognition.

We propose a smart mirror to achieve the necessary interaction between the child and the recognition software, together with an information system in charge of recording the results obtained in each of the therapies to measure children's progress in recognizing emotions over time.

The preceding is supported by the evolution and trend that the use of technologies in children with ASD has had in recent years, due to the growing knowledge in artificial intelligence and the development of new or improved electronic devices. Consequently, it is essential to highlight the upward trend in research associated with human-computer interaction and the conclusions obtained by different authors about recognizing emotions in children with ASD, which were applied in the proposed prototype. A limitation that other authors declared about light and shadow when using a camera was reflected in the tests carried out with the prototype exposed here. It is suggested that it be used in an environment with good light.

Regarding the results in the generation of the model that is in charge of classifying the emotion delivered by the image, the different tests carried out to achieve the highest level of accuracy among some of the most used architectures when working with images reached 93.3% using the CK + dataset and the VGG16 architecture.

Advances in technology and new techniques are expected to help continue advancing research overtime on this topic. Future steps involve carrying out a validation of the proposed prototype by experts in autism spectrum disorder, evaluating the perceived usefulness, perceived ease of use, and intention to use it. Besides, we expect to test the smart mirror on children with ASD. Finally, we plan to integrate narratives in the mirror that allows the generation of spontaneous emotions in children, complementing images' imitations.

Funding. This work was supported by the FCT – Fundação para a Ciência e a Tecnologia, I.P. [Project UIDB/05105/2020].

References

1. Wainer, A.L., Ingersoll, B.R.: The use of innovative computer technology for teaching social communication to individuals with autism spectrum disorders. *Res. Autism Spectr. Disord.* **5**(1), 96–107 (2011)

2. Adolphs, R.: Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* **12**(2), 169–177 (2002)
3. Ekman, P.: An argument for basic emotions. *Cogn. Emot.* **6**(3), 169–200 (1992)
4. Wang, A.T., Dapretto, M., Hariri, A.R., Sigman, M., Bookheimer, S.Y.: Neural correlates of facial affect processing in children and adolescents with autism spectrum disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **43**(4), 481–490 (2004)
5. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub. (2013)
6. Wieckowski, A.T., Flynn, L.T., Richey, J.A., Gracanin, D., White, S.W.: Measuring change in facial emotion recognition in individuals with autism spectrum disorder: a systematic review. *Autism* **24**(7), 1607–1628 (2020)
7. Abirached, B., et al.: Improving communication skills of children with ASDs through interaction with virtual characters. In: 1st International Conference on Serious Games and Applications for Health (SeGAH) Proceedings, pp. 1–4 (2011)
8. Pavez, R., Diaz, J., Vega, D.: Emotion recognition in children with ASD using technologies: a systematic mapping study. In: 38th International Conference of the Chilean Computer Science Society (SCCC) Proceedings (2019). <https://doi.org/10.1109/sccc49216.2019.8966449>
9. Gu, J., et al.: Recent advances in convolutional neural networks. *Pattern Recognit.* **77**, 354–377 (2018)
10. Bisong, E.: What is deep learning? Building machine learning and deep learning models on Google cloud platform, pp. 327–329 (2019). https://doi.org/10.1007/978-1-4842-4470-8_27
11. Tian, S., Yang, W., Grange, J.M.L., Wang, P., Huang, W., Ye, Z.: Smart healthcare: making medical care more intelligent. *Glob. Health J.* **3**(3), 62–65 (2019)
12. Rogers, S.J., Bennetto, L., McEvoy, R., Pennington, B.F.: Imitation and pantomime in high-functioning adolescents with autism spectrum disorders. *Child Dev.* **67**(5), 2060–2073 (1996)
13. Goldsmith, T.R., LeBlanc, L.A.: Use of technology in interventions for children with autism. *J. Early Intensive Behav. Intervent.* **1**(2), 166–178 (2004)
14. Harrold, N., Tan, C.T., Rosser, D., Leong, T.W.: CopyMe: a portable real-time feedback expression recognition game for children. In: Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 1195–1200, April 2014. Accessed 09 July 2019
15. Fan, M., Fan, J., Jin, S., Antle, A.N., Pasquier, P.: EmoStory: a game-based system supporting children's emotional development. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, April 2018. Accessed 09 July 2019
16. Voss, C., et al.: Superpower glass: delivering unobtrusive real-time social cues in wearable systems. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (2016). Accessed 09 July 2019
17. Chu, H.-C., Tsai, W.W.-J.M., Liao, J., Chen, Y.-M.: Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. *Soft Comput.* **22**(9), 2973–2999 (2018)
18. Fan, J., Bekele, E., Warren, Z., Sarkar, N.: EEG analysis of facial affect recognition process of individuals with ASD performance prediction leveraging social context. In: Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) Proceedings, pp. 38–43 (2017)
19. Uluyagmur-Ozturk, M., et al.: ADHD and ASD classification based on emotion recognition data. In 15th IEEE International Conference on Machine Learning and Applications (ICMLA) Proceedings, pp. 810–813 (2016)
20. Tamil, P.S., Vyshnavi, P., Jagadish, R., Srikumar, S., Veni, S.: Emotion recognition from videos using facial expressions. In: Dash, S.S., Vijayakumar, K., Panigrahi, B.K., Das, S. (eds.) Artificial Intelligence and Evolutionary Computations in Engineering Systems, vol. 517, pp. 565–576. Springer, Singapore (2017)

21. Zhao, X., Zou, J., Li, H., Dellandrea, E., Kakadiaris, I.A., Chen, L.: Automatic 2.5-D facial landmarking and emotion annotation for social interaction assistance. *IEEE Trans. Cybern.* **46**(9), 2042–2055 (2016)
22. Washington, P., et al.: A wearable social interaction aid for children with autism. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2348–2354 (2016). Accessed 09 July 2019
23. Adams, A., Robinson, P.: Expression training for complex emotions using facial expressions and head movements. In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) Proceedings, pp. 784–786 (2015)
24. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended Cohn-Kanade dataset: a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops (2010). <https://doi.org/10.1109/cvprw.2010.5543262>
25. Goodfellow, I.J., et al.: Challenges in representation learning: a report on three machine learning contests. *Neural Netw.* **64**, 59–63 (2015)
26. Rieffe, C., Meerum, M.T., Kotronopoulou, K.: Awareness of single and multiple emotions in high-functioning children with autism. *J. Autism Dev. Disord.* **37**(3), 455–465 (2007)
27. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
29. Bottou, L.: Stochastic gradient descent tricks. In: Lecture Notes in Computer Science, pp. 421–436 (2012). https://doi.org/10.1007/978-3-642-35289-8_25
30. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR (2001). <https://doi.org/10.1109/cvpr.2001.990517>

Author Index

A

- Abelha, António, 511
Acosta, Laura, 435
A-Farsi, Manal H., 499
Ahumada, Danay, 585
Alvarado, Héctor Gómez, 14
Alves, P., 207
Alves, Victor, 186
Andaluz, Víctor H., 80
Andrade, Francisco, 425
Arango-López, Jeferson, 585
Archibald Taylor, George Washington, 293
Author, Astrid Jaime, 293
Ayuso, Mercedes, 123

B

- Bădică, Costin, 92
Baptista, Luís, 468
Baptista, Luís M. T., 166
Barba-Guaman, Luis, 176
Batista, Josias G., 113
Bernardes, Marciele, 425
Bezerra, Ismael S., 113
Błoszko, Yurii, 29
Bossard, Antoine, 415
Bouras, Christos, 344
Bravo, Jorge M., 123
Brusius, Carlos, 396
Burguillo, Juan C., 260

C

- Cajas, Viviana, 455
Canão, J., 207
Cano, Sandra, 435

Carneiro, Davide, 156

- Carvalho, Mariana, 156
Castillo-Salazar, David, 14
Chaim, Ricardo Mattos, 69
Coelho, Helder, 396
Córdova, Paulo Roberto, 396
Cuervo, Mauro Callejas, 375

D

- Daranda, Andrius, 49
de Almeida, Raquel Simões, 555
de la Higuera Amato, Cibelle Albuquerque, 532
de Sousa, Sara, 555
Dias, Rafael, 468
Díaz, Jaime, 585
Dimitrovs, Emils, 3
Djehiche, Raid, 357
Dovleac, Raluca, 522
Durães, Dalila, 104, 488
Dzemyda, Gintautas, 49

E

- Espinosa, C. Alexandra, 293
Estrada, Carlos Andrés, 375

F

- Farina, Pedro Giuliano, 532
Felgueiras, Sérgio, 217
Fernandes, Jorge, 425
Ferreira, Diana, 511
Ferreira, Paulo, 575
Fonseca, Joaquim, 104
Fortes, Inês, 313, 323

Franco, Tiago, 207
 Fuertes, Walter, 375

G

García Arango, David Alberto, 385
 García Pereáñez, José Antonio, 385
 Ginters, Egils, 3
 Gkamas, Apostolos, 344
 Gomes, Marco, 104
 Gonçalves, Filipe, 104
 Gonçalves, Rui, 238
 González García, Ignacio, 136
 Gonzalez, Claudio C., 405
 Gonzalez, Juan Guillermo Pinzon, 176
 Granollers, Antoni, 405
 Guedes, Pedro, 575
 Guevara, Cesar, 14
 Guimarães, Miguel, 156

H

Habib, Sami J., 333
 Hamukwaya, Shemunyenge T., 478
 Hamukwaya, Shindume L., 478
 Haoud, Mohamed, 357

I

Idri, Ali, 544
 Idrissi, Touria El, 544
 Ionica, Andreea, 302
 Ionica, Andreea Cristina, 522

J

Jorque, Byron S., 80
 José, Rui, 313, 323
 Júnior, Antônio B. S., 113
 Júnior, José N. N., 113
 Justo, Jorge, 575

K

Katsampiris Salgado, Spyridon Aniceto, 344
 Kimura, Herbert, 425
 Kokkonen, Tero, 197

L

Laato, Samuli, 365, 478
 Lanzarini, Laura, 14
 Leal, Fátima, 260
 Leba, Monica, 302, 522
 Li, Yuchen, 39
 Lima, Ana Carolina Oliveira, 445
 Lin, Hai, 270
 Ludeña-González, Patricia, 283

M

Machado, Gabriel F., 113
 Machado, Jorge, 166
 Machado, José, 104, 511
 Machado, Samuel, 249
 Major, Laszlo, 478
 Malheiro, Benedita, 260, 574
 Maravalhas, Vanessa, 555
 Marcondes, Francisco, 104
 Marimuthu, Paulvanna N., 333
 Marmelo, Daniel, 468
 Marques, António, 555
 Martins, Ana Isabel, 445
 Martins, Mónica V., 166
 Martins, Valéria Farinazzo, 532
 Mateos, Alfonso, 136
 Mendes-Moreira, João, 146
 Méndez, Carolina, 585
 Méndez, Gonzalo Gabriel, 455
 Méndez, Yenny A., 405
 Mijas-Abad, Thalia, 283
 Millán, Andrés F., 405
 Moghrabi, Issam A. R., 499
 Molloanca, Jéssica D., 80
 Monteiro, Pedro, 555
 Moreira, Fernando, 405, 435, 585
 Morgado, Sónia M. A., 217
 Moro, Sérgio, 227
 Mourato, João, 468

N

Naranjo, Jose Eugenio, 176
 Neto, Cristiana, 511
 Nogueira, Fabrício G., 113
 Novais, Paulo, 104, 156, 425, 488

O

Olar, Oksana, 29
 Oliveira e Sá, Jorge, 249
 Oliveira, Tiago Gil, 186
 Ortega, Oscar, 293
 Ortiz, Anthony, 176
 Ortiz, Jessica S., 80

P

Papachristos, Nikolaos, 344
 Páscoa, Paulo, 468
 Pávez, Rodolfo, 585
 Pedrosa, T., 207
 Peñenory, Victor, 435
 Pereira, Fernando Lobo, 238
 Pessoa, Ricardo, 313
 Pineda, Jhon, 293
 Pinto, Diana, 313, 323
 Priebe, Julian, 575

R

- Ramirez V., Gabriel M., 405
Rauti, Sampsaa, 365
Realinho, Valentim, 166, 468
Ribeiro, Cristina, 575
Ribeiro, Vitor Miguel, 238
Risteiu, Marius, 522
Riurean, Simona, 302
Rocha, Álvaro, 302
Rocha, Ana Paula, 238
Rocha, Nelson Pacheco, 445
Rodríguez, Andrés, 455
Rossi, Gustavo, 455
Royo, Álvaro Arribas, 375
Rubio, Manuel Sánchez, 375

S

- Sandoval, Francisco, 283
Santana, Ronny, 455
Santos, Flávio, 104
Saoudi, Lalia, 357
Sarmiento, Román, 293
Seca, Diogo, 146
Silva, Fátima Solange, 186
Silva, Manuel F., 575
Sipola, Tuomo, 197
Sorescu, Tiberius-George, 575
Soto, Jonathan, 435
Sousa, Emanuel, 323
Souza, Darielson A., 113
Su, Hailong, 59
Suraj, Zbigniew, 29
Swiatek, Klaudia, 575

T

- Tianyuan, Zhang, 227
Tiits, Mihkel, 575
Toala, Rámon, 488
Tolledo, Daniel, 166
Tomczyk, Łukasz, 565
Torres, Rommel, 283
Torrico, Bismark C., 113
Toulkeridis, Theofilos, 375

V

- Vaduva, Maria-Roxana, 575
Varanda Pereira, M. J., 207
Veloso, Bruno, 260
Venturini, Fabíola Cristina, 69
Vicari, Rosa Maria, 396
Vidinha, Margarida, 575
Vieira, Eduarda, 511
Vieira, Joana, 313
Vilas-Boas, Vera, 323
Vultureanu-Albiști, Alexandra, 92

W

- Wadghiri, Mohamed Zaim, 544
Wang, Patrick Shen-Pei, 39, 59, 270
Wingbermuehle, Jochen, 104

Y

- Yang, Lina, 39, 59, 270

Z

- Zhang, Fengqi, 59