Lecture Notes in Networks and Systems 770

Paulo Novais · Vicente Julián Inglada · Miguel J. Hornos · Ichiro Satoh · Davide Carneiro · João Carneiro · Ricardo S. Alonso *Editors*

Ambient Intelligence – Software and Applications – 14th International Symposium on Ambient Intelligence

Springer

Lecture Notes in Networks and Systems 770

Series Editor

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

Advisory Editors

Fernando Gomide, *Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil*

Okyay Kaynak, *Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Türkiye*

Derong Liu, *Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA*

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, *Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada*

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

Marios M. Polycarpou, *Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus*

Imre J. Rudas, *Óbuda University, Budapest, Hungary*

Jun Wang, *Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong*

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

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

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Paulo Novais · Vicente Julián Inglada ·
Miguel J. Hornos · Ichiro Satoh ·
Davide Carneiro · João Carneiro ·
Ricardo S. Alonso
Editors

Ambient Intelligence – Software and Applications – 14th International Symposium on Ambient Intelligence

Springer

Editors
Paulo Novais
ALGORITMI Centre/LASI
University of Minho
Braga, Portugal

Miguel J. Hornos
University of Granada
Granada, Spain

Davide Carneiro
CIICESI, ESTG
Politécnico do Porto
Felgueiras, Portugal

Ricardo S. Alonso
Deep tech lab
AIR Institute
Valladolid, Spain

Vicente Julián Inglada
Universitat Politècnica de València
Valencia, Spain

Ichiro Satoh
National Institute of Informatics
Tokyo, Japan

João Carneiro
ISEP/GECAD
Porto, Portugal

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-43460-0 ISBN 978-3-031-43461-7 (eBook)
https://doi.org/10.1007/978-3-031-43461-7

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2023

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

Paper in this product is recyclable.

Preface

Ambient intelligence (AmI) represents a cutting-edge paradigm in computing and artificial intelligence that aims to create an interconnected, adaptive and unobtrusive environment that seamlessly integrates technology into our daily lives. The core idea of AmI is to build intelligent environments capable of sensing, understanding and responding to human needs, preferences and behaviors without the need for explicit user commands. This technology envisions a world in which diverse devices, sensors and systems work together intelligently, creating a harmonious, context-aware ecosystem that enhances our overall experience and well-being.

The main application areas of ambient intelligence are diverse and far-reaching, spanning numerous sectors. In health care, ambient intelligence can revolutionize patient care by monitoring vital signs, detecting abnormalities and providing personalized medical care. In smart homes, AmI technologies can self-manage household tasks, adjust lighting and temperature according to user preferences and enhance security through smart surveillance systems. In addition, AmI has applications in retail, entertainment, education and many other sectors, all aimed at creating seamless, intuitive and efficient human–technology interaction.

In this regard, ISAmI is the *International Symposium on Ambient Intelligence*, which seeks to bring together scientists working in all areas related to AmI, as can be seen in the articles included in the conference. This year, we are proud to tell that ISAmI is in its 14th edition, including several papers focused on the different key aspects of the AmI. Among others, interested researchers can find works on the recognition of emotions and life conditions; smart applications for improving people's quality of life and safety; AmI for smart cities and territories; research in distributed architectures in AmI scenarios; as well as advances in data semantics, data organization and data privacy.

In total, 23 articles for the main track and 4 for the doctoral consortium have been selected from a total of 51 submissions (44 for the main track and 7 for the doctoral consortium) from many different countries (Japan, Italy, India, Nigeria, Lithuania, Ecuador, UAE, Palestine, Burkina Faso, Senegal, Cameroon, Spain and Portugal).

As in previous editions, there are special issues in JCR-ranked journals such as *Sensors*; *Electronics*; *Systems*; as well as *Advances in Distributed Computing and Artificial Intelligence Journal*.

The symposium is organized by the Universidade do Minho (Portugal), the Universitat Politècnica de València (Spain), the University of Granada (Spain), the National Institute of Informatics (Japan), the Instituto Superior de Engenharia do Porto (Portugal) and the Universidad de Salamanca (Spain). The present edition was hosted by the LASI and Centro Algoritmi of the University of Minho in Guimarães, Portugal, from July 12 to 14, 2023.

We would like to thank all colleagues who submitted their research, all the members of the Program Committee for their work, and the sponsors (AIR institute and Câmara

Municipal de Guimarães) for their support. Without their efforts, it would have been impossible to put together such an interesting and actual program.

Finally, we appreciate the efforts of local organizers, for hosting and organizing these events, without which ISAmI could not have taken place.

Paulo Novais
Vicente Julián
Miguel J. Hornos
Ichiro Satoh
Davide Carneiro
João Carneiro
Ricardo S. Alonso

Organization

General Chair

Paulo Novais	Universidade do Minho, Portugal

Program Committee Chairs

Vicente Julián	Technical University of Valencia, Spain
Miguel J. Hornos	University of Granada, Spain
Ichiro Satoh	National Institute of Informatics, Japan

Organizing Committee Chairs

Ricardo S. Alonso	AIR Institute, Spain
Davide Carneiro	ISEP/GECAD, Portugal
João Carneiro	ISEP/GECAD, Portugal

Advisory Board

Gianni Vercelli	Università di Genova, Italy

Program Committee

Ana Almeida	ISEP, Portugal
Cesar Analide	University of Minho, Portugal
Luis Blázquez	University of Salamanca, Spain
Jose Luis Calvo Rolle	University of A Coruña, Spain
Valérie Camps	University of Toulouse IIIIRIT, France
Javier Carbo	University Carlos III of Madrid, Spain
Davide Carneiro	ESTG—Escola Superior de Tecnologia e Gestao, Portugal
Fabio Cassano	Università degli Studi di Bari Aldo Moro, Italy
Alvaro Castro Gonzalez	University Carlos III of Madrid, Spain
Silvio Cesar Cazella	UFCSPA, Brazil

Pablo Chamoso	University of Salamanca, Spain
Stefano Chessa	University of Pisa, Italy
Stéphanie Combettes	University of Toulouse IIIIRIT, France
Angelo Costa	University of Alicante, Spain
Luis de la Fuente Valentín	Universidad Internacional de La Rioja, Spain
Patricio Domingues	ESTG-Leiria, Portugal
John Dowell	University College London, the UK
Dalila Alves Durães	Universidade do Minho, Portugal
Luiz Faria	Polytechnic of Porto, Portugal
Florentino Fernández Riverola	University of Vigo, Spain
Antonio Fernández Caballero	Universidad de Castilla-La Mancha, Spain
João Ferreira	ISCTE, Portugal
Lino Figueiredo	ISEP, Portugal
Daniela Fogli	Università di Brescia, Italy
Oscar Garcia	Universidad Internacional de La Rioja, Spain
Jose García Alonso	University of Extremadura, Spain
Marco Gomes	University of Minho, Portugal
Celestino Goncalves	Instituto Politecnico da Guarda, Portugal
Sérgio Gonçalves	University of Minho, Portugal
Alfonso González Briones	University of Salamanca, Spain
Junzhong Gu	East China Normal University, China
Esteban Guerrero	University of Vaasa, Finland
Guillermo Hernández	Universidad de Salamanca, Spain
Esteban Jove	UDC, Spain
Vicente Julian	Universitat Politècnica de València, Spain
Jason Jung	Chung-Ang University, South Korea
Alexander Kocian	University of Pisa, Italy
Tiancheng Li	Northwestern Polytechnical University, China
Afonso Lima	ABLN, Brazil
Raúl López Blanco	Universidad de Salamanca, Spain
José Machado	University of Minho, Portugal
João Paulo Magalhaes	ESTG—Escola Superior de Tecnologia e Gestao, Portugal
Goreti Marreiros	ISEP/IPP-GECAD, Portugal
Constantino Martins	Polytechnic of Porto, Portugal
Rene Meier	Lucerne University of Applied Sciences, Switzerland
Pedro Melo Pinto	CITABUTAD, Portugal
José Pascual Molina Massó	Universidad de Castilla-La Mancha, Spain
Imam Mukhlash	Institut Teknologi Sepuluh Nopember, Indonesia
Tatsuo Nakajima	Waseda University, Japan
Elena Navarro	University of Castilla-La Mancha, Spain

Andrei Olaru	University Politehnica of Bucharest, Romania
Jaderick Pabico	University of the Philippines Los Banos, Philippines
Hugo Peixoto	University of Minho, Portugal
Antonio Pereira	Escola Superior de Tecnologia e Gestão do IPLeiria, Portugal
Antonio Piccinno	University of Bari, Italy
António Pinto	ESTG, P.Porto, Portugal
Pedro Pinto	ESTG IPVC, Portugal
Tiago Pinto	Universidade de Trás-os-Montes e Alto Douro, Portugal
Francisco Pinto Santos	BISITE Research Group, Spain
Isabel Praça	GECAD/ISEP, Portugal
Abraham Prieto García	University of A Coruña, Spain
Héctor Quintián	University of A Coruña, Spain
Joao Ramos	Polytechnic Institute of Porto, Portugal
Manuel Rodrigues	Polytechnic Institute of Porto, Portugal
Sara Rodríguez	University of Salamanca, Spain
Ivan Rodriguez Conde	University of Arkansas at Little Rock, the USA
Albert Ali Salah	Utrecht University, the Netherlands
Altino Sampaio	Escola Superior de Tecnologia e Gestão, Portugal
Ricardo Santos	ESTG/IPP, Portugal
Ichiro Satoh	National Institute of Informatics, Japan
Fernando Silva	Polytechnic of Leiria, Portugal
Fábio Silva	University of Minho, Portugal
Radu-Daniel Vatavu	University Stefan cel Mare of Suceava, Romania
Fatos Xhafa	Universitat Politècnica de Catalunya, Spain
Marta Plaza-Hernández	Universidad de Salamanca, Spain
Ricardo S. Alonso	AIR Institute, Spain

Local Organizing Committee

Paulo Novais (Chair)	University of Minho, Portugal
José Manuel Machado (Co-chair)	University of Minho, Portugal
Hugo Peixoto	University of Minho, Portugal
Regina Sousa	University of Minho, Portugal
Pedro José Oliveira	University of Minho, Portugal
Francisco Marcondes	University of Minho, Portugal
Manuel Rodrigues	University of Minho, Portugal
Filipe Gonçalves	University of Minho, Portugal
Dalila Durães	University of Minho, Portugal

Sérgio Gonçalves	University of Minho, Portugal

Organizing Committee

Juan M. Corchado Rodríguez	University of Salamanca and AIR Institute, Spain
Fernando de la Prieta	University of Salamanca, Spain
Sara Rodríguez González	University of Salamanca, Spain
Javier Prieto Tejedor	University of Salamanca and AIR Institute, Spain
Ricardo S. Alonso Rincón	AIR Institute, Spain
Alfonso González Briones	University of Salamanca, Spain
Pablo Chamoso Santos	University of Salamanca, Spain
Javier Parra	University of Salamanca, Spain
Liliana Durón	University of Salamanca, Spain
Marta Plaza Hernández	University of Salamanca, Spain
Belén Pérez Lancho	University of Salamanca, Spain
Ana Belén Gil González	University of Salamanca, Spain
Ana de Luis Reboredo	University of Salamanca, Spain
Angélica González Arrieta	University of Salamanca, Spain
Angel Luis Sánchez Lázaro	University of Salamanca, Spain
Emilio S. Corchado Rodríguez	University of Salamanca, Spain
Raúl López Blanco	University of Salamanca, Spain
Beatriz Bellido	University of Salamanca, Spain
María Alonso	University of Salamanca, Spain
Yeray Mezquita Martín	AIR Institute, Spain
Sergio Márquez	AIR Institute, Spain
Andrea Gil	University of Salamanca, Spain
Albano Carrera González	AIR Institute, Spain

Contents

ISAMI'23 - MAIN TRACK

Sharing Human Seeing and Hearing and Their Annotated Portfolio 3
Risa Kimura and Tatsuo Nakajima

IoT Security and Privacy Challenges from the Developer Perspective 13
Yaqin Shaheen, Miguel J. Hornos, and Carlos Rodríguez-Domínguez

Flow Disaggregation: Underdetermined Non-negative Linear Systems 22
Javier Curto, Guillermo Hernández, María Alonso-García, Alicia Serrano-Ortega, Alberto Toledo-Garrote, and Pablo Chamoso

Ensuring Data Privacy in Open Data Platforms for Smart Cities 33
Juan Antonio González-Ramos, Francisco Pinto-Santos, Alesandro Gómez-Villar, Javier Curto, and Javier Parra-Domínguez

Intelligent Heart Rate Monitoring from a Social Robot 43
Arturo Martinez-Rodrigo, José Carlos Castillo, Sara Carrasco-Martínez, Daniele Padovano, and Angelo Costa

Review of Physical Aggression Detection Techniques in Video Using Explainable Artificial Intelligence 53
Pablo Negre, Ricardo S. Alonso, Javier Prieto, Angélica González Arrieta, and Juan M. Corchado

An Evaluation of the Performance of Convolution Neural Network and Transfer Learning on Face Gender Recognition 63
Ronke Seyi Babatunde, Akinbowale Nathaniel Babatunde, Roseline Oluwaseun Ogundokun, AbdulRahman Tosho Abdulahi, and Alfonso González-Briones

A Proposal for a Taxonomy of Augmented Humanity Based on Current Needs, Topics and Challenges 74
Graciela Guerrero, Fernando José Mateus da Silva, António Pereira, and Antonio Fernández-Caballero

A New Software Platform to Provide AI-Based Services to Smart Cities 84
Francisco Pinto-Santos, Juan Antonio González-Ramos, Raúl López-Blanco, and Javier Curto

Chatto: An Emotionally Intelligent Avatar for Elderly Care in Ambient Assisted Living 93
Carla Mendes, Rafael Pereira, José Ribeiro, Nuno Rodrigues, and António Pereira

Optimization of Traffic Light Controllers Using Genetic Algorithms: A Case Study in the City of Cádiz 103
Sara Balderas-Díaz, Andrés Muñoz, and Gabriel Guerrero-Contreras

WoA: An Infrastructural, Web-Based Approach to Digital Agriculture 113
Stefano Chessa, Giovanna Maria Dimitri, Marco Gori, and Alexander Kocian

Human-in-the-loop AAL Approach to Emotion Capture and Classification 123
Rafael Pereira, Carla Mendes, Roberto Ribeiro, José Ribeiro, and António Pereira

A Portuguese Mobile App for Diabetes Type 2 Self-Management: A Picture of One-Month Usage 133
Glória Conceição, Andreia Pinto, João Viana, Gonçalo Miranda, Paulo Santos, Cristina Santos, and Alberto Freitas

CERDL: Contextual Emotion Recognition Analysis Using Deep Learning 143
Aayushi Chaudhari, Chintan Bhatt, Achyut Krishna, and Juan M. Corchado

The Impact of Data Selection Strategies on Distributed Model Performance 157
Miguel Guimarães, Filipe Oliveira, Davide Carneiro, and Paulo Novais

AI-Powered Smart Book: Enhancing Arabic Education in Palestine with Augmented Reality 167
Murad Al Rajab, Suhail Odeh, Sherin Hazboun, and Emran Alheeh

How to Handle Fuzzy Aspects in African Sociocultural Ontology 179
Abdoul Azize Kindo, Sadouanouan Malo, Gaoussou Camara, Guidedi Kaladzavi, Théodore Marie Yves Tapsoba, and Kolyang

VRPrOE Toolbox for Virtual Pre-occupancy Evaluation: Proof of Concept on a BIM Model of a Conservatory Classroom 189
José L. Gómez-Sirvent, Desirée Fernández-Sotos, Francisco López de la Rosa, Alicia Fernández-Sotos, and Antonio Fernández-Caballero

Consumption Performance, of a Multi Agent System Used to Achieve Comfort Preferences 199
Pedro Filipe Oliveira, Paulo Novais, and Paulo Matos

Pleasure and Displeasure Identification from fNIRS Signals 209
Daniel Sánchez-Reolid, Roberto Sánchez-Reolid, Antonio Fernández-Caballero, and Alejandro L. Borja

Environmental Sound Recognition in Social Robotics 220
Sara Marques-Villarroya, Aythami Sosa-Aleman, Jose Carlos Castillo, Marcos Maroto-Gómez, and Miguel Angel Salichs

Intelligent Systems in Healthcare: An Architecture Proposal 230
António Chaves, Larissa Montenegro, Hugo Peixoto, António Abelha, Luís Gomes, and José Machado

ISAMI'23 - DOCTORAL CONSORTIUM

Teaching Innovation in Higher Education Through Virtual Reality 241
Fernando Lara-Lara, Juan-Carlos de la Cruz-Campos, Magdalena Ramos-Navas-Parejo, and Blanca Berral-Ortiz

Machine Learning Based Ambient Analysis of Railway Steel Bridges for Damage Detection 250
Muhammad Asad

Human Action Recognition in Uncontrolled Environments: Application from Artificial Intelligence to Contactless Interfaces 256
Vanessa Alvear, César Domínguez, and Gadea Mata

Analysis of Techniques for the Augmentation and Correction of Unbalanced Data in Binary Cybersecurity Datasets for Machine Learning . . . 262
Alberto Sánchez del Monte

Author Index 269

ISAMI’23 - MAIN TRACK



Sharing Human Seeing and Hearing and Their Annotated Portfolio

Risa Kimura(✉) and Tatsuo Nakajima

Department of Computer Science and Engineering, Waseda University, Tokyo, Japan
{risa.kimura,tatsuo}@dcl.cs.waseda.ac.jp

Abstract. The paper describes the use of an annotated portfolio for extracting their possibilities and challenges from "CollectiveEyes and CollectiveEars" to document abstract features of these digital platforms that are meant to engage human visual and auditory perception. This proposed annotated portfolio offers two different perspectives: the sociomateriality perspective and the pleasurability perspective. After introducing the background of our approach and providing an overview of CollectiveEyes and CollectiveEars, we present their annotations and the process to document them. Additionally, we share some early experiences with these annotations, which were used to analyze the platforms. These experiences demonstrate that this approach effectively highlights the abstract qualitative properties of the platforms, thereby facilitating the identification of potential opportunities for their future refinements.

Keywords: Material Aspects · Sharing Human Seeing and Hearing · Annotated Portfolio · Sociolmateriality · Pleasures

1 Introduction

In recent decades, research on digital technology has mainly focused on utilitarian advancements that prioritize functionality and practicality. However, there is growing concern about the limitations of this heavily utilitarian approach. Some researchers in the design-oriented human-computer interaction community argue that an excessive focus on utilitarian aspects may hinder efforts to understand the complex relationships between humans and technology, as well as how human experiences are shaped by technology. This overemphasis on practicality has raised questions about considering broader issues in the relationship between research and design. To address these concerns, innovative digital technologies have given rise to designerly approaches, encouraging human-computer interaction researchers to actively engage in design practices, exploring digital artifacts within real-world contexts [15]. Through this practical process, they gain new insights into domain use cases and design perspectives that go beyond mere functionality and utility. To bridge the gap and disseminate the knowledge acquired through the design of such artifacts, several proposals have been introduced to document the findings [6, 13]. One promising approach is the use of annotated portfolios [1, 4], which effectively captures the knowledge gained from the design of artifacts.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 3–12, 2023.
https://doi.org/10.1007/978-3-031-43461-7_1

This paper introduces annotated portfolios designed for digital platforms that cater to human senses - vision and hearing. These portfolios incorporate insights from two distinct disciplines to develop a structured annotation method. One discipline is the "sociomateriality" approach, which examines the interactions between humans and non-human on an equal level [11, 12]. Another discipline is the "four pleasures" approach, which goes beyond a rationalist perspective to design the non-functional pleasurable aspects of digital artifacts [7]. After providing the background of our approach, we present the annotations and explain how these two disciplines contribute to their specification. Furthermore, we share early experiences with these annotations in analyzing the platforms. These experiences demonstrate how the approach effectively highlights the platforms' essential properties, allowing for the identification and exploration of potential opportunities for their future refinements.

As modern digital platforms for smart space services become increasingly complex, it is crucial to understand their basic features and to refine and incorporate new requirements for the future. The abstract descriptive power of the annotations represented by the annotated portfolio method has proven promising for analyzing digital platforms and investigating a variety of complex issues among stakeholders with diverse expertise. These annotations are a powerful means to shed light on the complex interrelationships and requirements of digital platforms. As platforms evolve, new features and requirements are likely to be added. The annotations allow for a clear understanding of these requirements and the ability to effectively manage the evolution of the platform. The annotations can also be a useful tool when stakeholders with different expertise collaboratively explore a platform. The power of their abstract description enables the integration of the different perspectives of experts and users to gain insights into the operation and improvement of the platform. Furthermore, the annotations are important not only for analyzing the platform, but also for refining and incorporating new requirements for the future. As new technologies and usage scenarios may emerge, a flexible and sophisticated approach to the fundamental characteristics of the platform is required. The annotations allow new requirements to be understood quickly and appropriate improvements to be implemented. The abstract descriptive power of the annotations can be a very useful tool in understanding and improving digital platforms. Proactive use of the annotations in investigating complex issues and incorporating new requirements will contribute to the development of future smart space services. By leveraging the potential of the annotated portfolio method, we can ensure that digital platforms continue to evolve and meet the changing needs of users and stakeholders in the smart space environment.

2 Annotated Portfolio, Sociomateriality and Pleasures

Gaver and Bowers developed the annotated portfolio as a method to express novel knowledge acquired from research-oriented design practices [4]. This approach serves as a means to uncover the similarities and distinctions within a collection of artifacts [1]. Originally, the concept of this method involved selecting a set of artifacts, identifying suitable representations, and combining these representations with concise textual and pictorial annotations. Ultimately, the annotated portfolio serves as a methodology for publicly communicating extracted insights and knowledge.

The annotated portfolio method has been the subject of several previous studies [2, 5]. In [2], the authors extended the original concept of the annotated portfolio method to encompass designs for new domains. Another study demonstrated this extension by creating an annotated portfolio of Research through Design artifacts inquiries, specifically from the postphenomenological perspective [5], where Research through Design is a discipline that expands design practices and processes to generate new knowledge [15]. Design Exposès [3] employs an annotated portfolio method to extract knowledge. In the approach, a deliberate decision was made to omit the specified "value" layer in the reflection part. The primary emphasis of Design Exposès is on the "utility" aspect of the object, rather than assessing the specific value of the object.

Sociomateriality explores the interconnectedness of technology and society, seeking to comprehend how the social and material aspects are inextricably intertwined in everyday life. This perspective often employs the term "material" instead of "nonhuman" to refer to the tangible elements of this entanglement. It has emerged from various perspectives and discussions about technology, technology-in-use, and sociomateriality, all of which have contributed to its foundation [12, 14]. Sociomateriality is a perspective that emphasizes concrete practices integrating materiality with institutions, norms, discourse, culture, and other social phenomena. This approach is invaluable in understanding how nonhuman elements influence human attitudes and behaviors by focusing on the relationship between humans and technology. By exploring the interaction of human and nonhuman elements, sociomateriality allows for a deeper understanding of the complex interactions between humans and technology. This perspective is particularly relevant in today's technologically advanced world, where the influence of technology on human behavior and society is ever-increasing. Through sociomateriality, we gain insights into the intricate dynamics that shape human-technology relationships and how these interactions impact various aspects of our lives.

The pleasure that users derive from using goods and services is often based on the perceived benefits they provide. However, there is a need for a new approach to designing products and services from a deeper perspective than mere functional efficiency and ease of use. To provide such a perspective, Jordan introduces a framework that encompasses four types of pleasures that should be considered in the design of goods and services [7]. These pleasure categories are physio-pleasures, socio-pleasures, psych-pleasures, and ideo-pleasures. Physio-pleasure is derived from bodily sensations, including tactile and olfactory experiences related to goods and services. As socio-pleasure, enjoyment arises from social interactions and the status or identity that goods and services confer in social situations. Psycho-pleasure is of a mental nature, encompassing cognitive and emotional responses to goods and services. Finally, in ideo-pleasure, involves pleasure rooted in collective values, where goods and services reflect personal values, such as environmental consciousness or sustainable living. Each of these pleasure categories represents a unique and conceptual aspect of the user experience, providing a multifaceted pleasure that is not merely functional. Incorporating these pleasure elements in the design of products and services has the potential to create a more satisfying user experience.

3 Using Annotated Portfolio for Analyzing Digital Platforms for Sharing Human Seeing and Hearing

3.1 CollectiveEyes and CollectiveEars

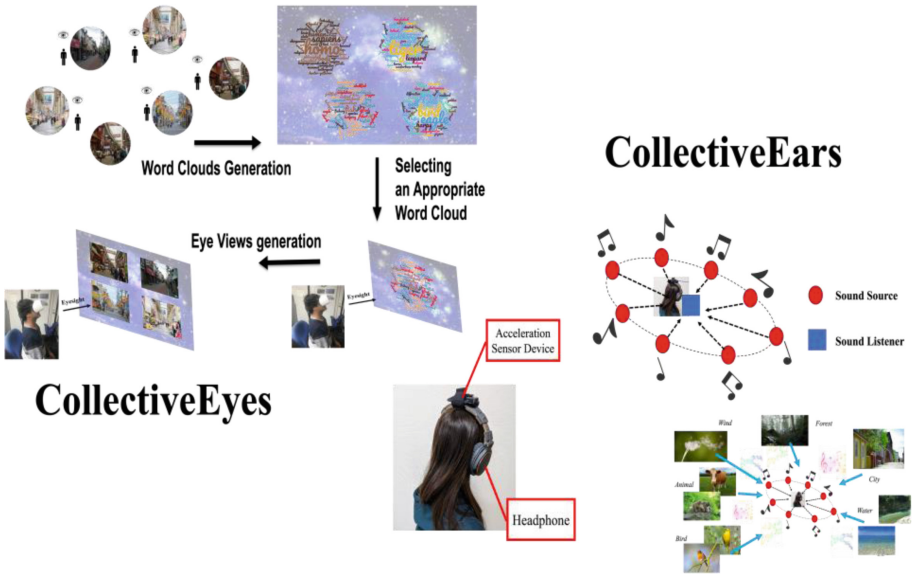


Fig. 1. CollectiveEyes and CollectiveEars

The digital platforms, CollectiveEyes and CollectiveEars, are centered around the collaborative sharing of specific human physical body parts, namely, eyes and ears, as depicted in Fig. 1.

CollectiveEyes [8] is a platform designed to capture and share collective human eyesight. The platform assumes that a user is equipped with a wearable display device, typically wearable glasses which equip cameras and microphones, allowing them to capture what they see and hear at any time and location. It enabling users to offer the ability to share some aspects of their visual perception with others. As wearable technology advances, the concept of virtually renting out body parts may become a reality. The CollectiveEyes platform uses a head-mounted display with a built-in camera and microphone positioned in front of it. Users can switch their own eyesight to view the current eyesight of others through their head-mounted displays. Additionally, they can view multiple users' eyesights simultaneously in a 3D virtual space, projecting them onto their HMDs. The platform facilitates searching for specific eyesights based on keywords and locations.

CollectiveEars is a digital platform designed to facilitate the sharing of a diverse collection of audio experiences, allowing users to listen to what others are currently hearing, thereby enriching their imagination of the world around them [9]. In the fast-paced and

stressful modern urban life, we often overlook various essential aspects that contribute to our well-being, such as connecting with nature or envisioning future possibilities. By capturing and sharing various sounds from around the world, CollectiveEars opens up new opportunities for us to appreciate the diversity of our world and expand our chances to experience it more fully. CollectiveEars gathers sounds heard by people worldwide, which users can then listen to. These sounds are positioned in a 3D sound space around the user and presented simultaneously. To interact with the platform, users can employ a novel method called "theme channels". This interaction technique allows users to select multiple sounds from the database, much like changing TV channels, using head gestures. Additionally, CollectiveEars implements a sound focusing function, enabling users to hear sounds more clearly by tilting their heads in the direction of interest.

3.2 Annotation Process

Developing digital platforms like CollectiveEyes and CollectiveEars comes with the challenge that their specific requirements and potential opportunities might not be clear until they are designed. As a result, an iterative approach involving designing, prototyping, and evaluating these platforms is crucial for refining or specializing them. However, traditional methods for developing and evaluating prototypes can be costly. To address these challenges, there is a need for a more efficient tool that can quickly reveal the possibilities and challenges of these platforms in a simple and cost-effective manner. As a solution, we proposed a novel annotation strategy to create annotation portfolios, which can effectively exploit and unveil the untapped opportunities of digital platforms. This annotation method offers visually structured annotations, facilitating the identification of diverse unexplored design aspects. The annotated portfolio serves to broaden the participants' perspectives in a participatory design process. To explore the abstract qualitative features of digital platforms, our annotation strategy for creating the annotated portfolio adopts two perspectives, one is representing the human aspect and the other is the nonhuman or material aspect of these platforms. By employing these two aspects also from the utilitarian perspective, we aim to gain deeper insights into the possibilities and challenges posed by these innovative digital platforms.

The first step in the process involves sociomateriality [12, 14], which prompts us to consider both human and nonhuman aspects. In this approach, we classify annotations into two categories: domain annotations and interaction annotations. The domain annotations, labeled as "Domain", encompass a set of fundamental items aimed at identifying domain use cases and defining the target goals of the platforms. On the other hand, the interaction annotations, labeled as "Interaction", consist of a set of basic items that describe the interaction methods the platforms offer to engage with users. An essential contribution of our strategy is explicitly introducing both perspectives, human and non-human, in our structural annotation approach. By doing so, we can effectively capture and highlight the interconnectedness between these perspectives, leading to a comprehensive understanding of the digital platforms' potential and capabilities.

In our annotation method, in the second step of the process, we also adopt the four pleasures to exploit the non-utilitarian perspective as described in Sect. 2. For our digital platforms, it is crucial to describe abstract values from both human and nonhuman visual and auditory perspectives. By investigating the potential values offered by digital

platforms based on the domain and interaction annotations, we can categorize them into the four pleasures and present them as annotations, referred to as the pleasure annotation.

The pleasure annotation is divided into two categories. One is the material side and the other is the human side. In Fig. 2 and Fig. 3, the left parts in the annotations represent the pleasure annotations on the material side, located below the interaction annotation. Conversely, the right parts of the annotations display the pleasure annotations on the human side, positioned under the domain annotation. To enhance visual clarity, material side annotations are emphasized in the red color, and human side annotations are denoted in the green color. Additionally, specific pleasure annotations are assigned distinct colors' texts, facilitating quick and easy identification.

3.3 Using Annotations for Demonstrating the Analysis of Digital Platforms for Sharing Human Seeing and Hearing

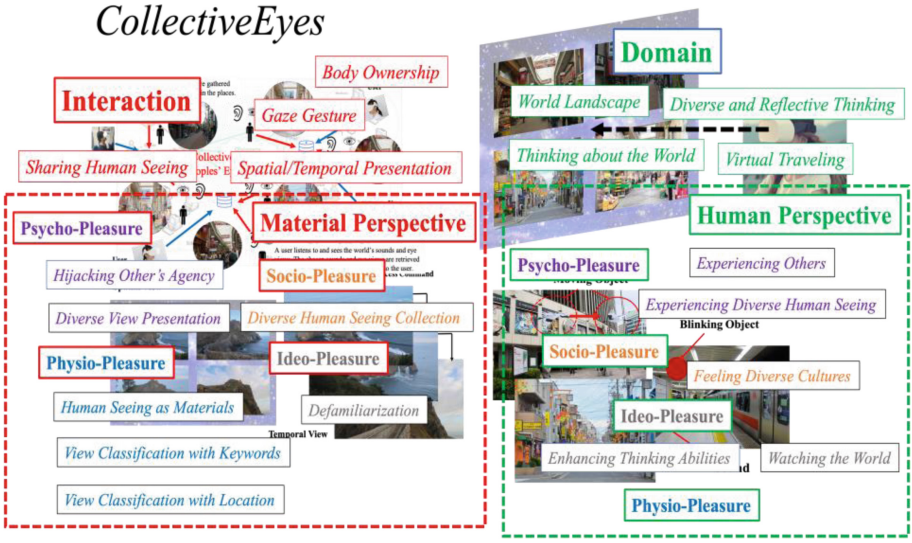


Fig. 2. Annotations for CollectiveEyes

Figure 2 illustrates the annotations for CollectiveEyes, while Fig. 3 displays the annotations for CollectiveEars. The pleasure annotation on the material side reveals the essential features offered by each platform. In the material perspective, psycho-pleasure represents the platform's features that cater to individual users. For instance, in CollectiveEyes, the feature "Diverse View Presentation" enables the presentation of multiple eye sights in a virtual space. On the other hand, socio-pleasure in the material perspective represents features related to multiple users. For example, in CollectiveEyes, the feature "Diverse Human Seeing Collection" allows the collection of people's current eye sights. Ideo-pleasure in the material perspective encompasses features that resonate with the collective users of the platforms. For example, in CollectiveEyes, the property

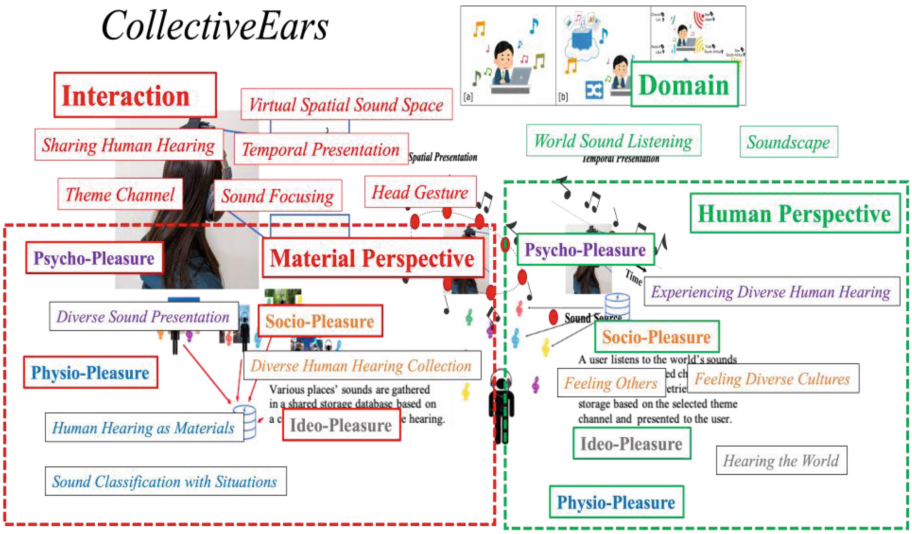


Fig. 3. Annotations for CollectiveEars

of "Defamiliarization" empowers all users [8]. Lastly, physio-pleasure in the material perspective represents the core abstract objects offered by the platforms. For instance, in CollectiveEyes, the feature "View Classification with Location" enables users to select eye sights presented in a virtual space simultaneously.

The pleasure annotation on the human side reflects the user experience provided by the platforms. Psycho-pleasure in the human perspective represents the individual user's experience. For example, in CollectiveEars, "Experiencing Diverse Human Hearing" signifies that the platform offers a wide range of human auditory experiences captured from various locations worldwide. Socio-pleasure in the human perspective represents the social user experience. For instance, in both platforms, "Feeling Diverse Culture" indicates that the platforms enable users to immerse themselves in diverse cultural atmospheres worldwide through human visual and auditory perspectives. Ideo-pleasure in the human perspective pertains to the collective user experience within the platforms [8]. For example, in CollectiveEyes, the annotation "Enhancing Thinking Abilities" signifies the platform's ability to encourage users to think critically by comparing multiple human visual perspectives. Physio-pleasure in the human perspective represents a user's physical experience. Currently, there are no items in this category in the annotations. However, the absence of items in this category presents new opportunities to explore potential physical experiences for users within the platforms.

3.4 Four Takeaways

Based on the annotations presented in Fig. 2 and Fig. 3, there following four key points that deserve attention to showcase the merits we discovered during the creation of these annotations. This highlights the effectiveness of our annotation strategy in revealing the potential opportunities and challenges of CollectiveEyes and CollectiveEars. The

current investigation of the annotations is still ad-hoc, and we need to explore new ways to extract more opportunities in a more systematic manner.

1) The four pleasures framework provides an opportunity to explore concealed aspects in our design process. For instance, during the first development of the platforms, we may not have placed strong emphasis on socio-pleasures. Our study's discovery prompted us to recognize the importance of exploiting socio-pleasures. We realized that incorporating collective visual and auditory perspectives could create new opportunities. When users feel that others also share the same visual or auditory experiences, a sense of closeness and connection is fostered. This aspect highlights the potential for enhancing the social experience within the platforms.
2) It is instructive to compare annotations across platforms. For example, CollectiveEyes prioritized the "Experiencing Others" annotation, which aims to provide users with a sense of becoming other through design. Conversely, CollectiveEars did not firstly focus on this particular annotation. While considering the annotation, we thought that it might be difficult to achieve the "Experiencing Others" annotation in CollectiveEars. As a result, however, we discovered the possibility of "Feeling Others" annotation, a newly discovered opportunity in the CollectiveEars design process that opens up new possibilities for improving the user experience.
3) The limited number of domain annotations for CollectiveEars suggests that during the platform's design phase, several domain use cases were initially considered. However, after the creation of CollectiveEyes, the designers became intrigued by the concept of collectively sharing human hearing experiences with a group of people. As a result, CollectiveEars was developed by extracting interaction methods and creating scenarios in various situations [9]. The fact that only a few items are included in the domain annotation highlights the opportunity to explore and find more fun in order to discover new domain use cases for CollectiveEars. Therefore, our focus was specifically on video pleasure to uncover new possibilities. The insights obtained from this study were integrated into the refined design of Artful CollectiveEars, where its aims ate to provide an artistic sound presentation [10].
4) Currently, the annotations do not include physio-pleasure from a human perspective. This is because both CollectiveEyes and CollectiveEars do not manipulate human's physical eyes and ears, and eliminate the need for physical movement to access others' current visual and auditory experiences. However, integrating physio-pleasure presents a promising opportunity to enhance human well-being. People's physical movement is an essential aspect to achieve a happy society, and providing location information of others' current visual and auditory perspectives might encourage users to explore what others are experiencing, especially when the locations are relatively close. These findings indicate that annotations can play a vital role in recognizing new opportunities to foster richer user experiences.

4 Conclusions

Our approach emphasizes the importance of adopting a hedonic perspective to effectively explore potential opportunities for digital platforms. While traditional approaches focus on the human experience through human-centered design, which is valuable for

developing improved digital platforms, they may not fully capture the impact of each aspect within the platform on the overall user experience. In contrast, our platforms, CollectiveEyes and CollectiveEars, provide a diverse range of user experiences, which in turn, enhances human thinking skills and opens up new opportunities. Throughout the process of creating annotations for both CollectiveEyes and CollectiveEars, we recognized the significance of the hedonic perspective that these digital platforms offer, considering both human and non-human perspectives. This understanding has unveiled novel opportunities and possibilities within these digital platforms. By embracing this hedonic perspective, we can better leverage the material aspect of human visual and auditory experiences to extract new opportunities and enhance the overall potential of these platforms.

References

1. John Bowers, J.: The logic of annotated portfolios: communicating the value of "research through design". In: Proceedings of the Designing Interactive Systems Conference (DIS 2012), pp. 68–77 (2012). https://doi.org/10.1145/2317956.2317968
2. Culén, A.L., Børsting, J., Gaver, W.: Strategies for annotating portfolios: mapping designs for new domains. In: Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS 2020), pp. 1633–1645. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3357236.3395490
3. Frauenberger, C., Makhaeva, J., Spiel, K.: Designing smart objects with autistic children: four design exposès. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI 2016), pp. 130–139. Association for Computing Machinery, New York (2016). https://doi.org/10.1145/2858036.2858050
4. Gaver, W., Bowers, J.: Annotated portfolios. Interactions **19**(4), 40–49 (2012). https://doi.org/10.1145/2212877.2212889
5. Hauser, S., Oogjes, D., Wakkary, R., Verbeek, P.-P.: An annotated portfolio on doing postphenomenology through research products. In: Proceedings of the 2018 Designing Interactive Systems Conference (DIS 2018), pp. 459–471. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3196709.3196745
6. Höök, K., et al.: Knowledge production in interaction design. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2015), pp. 2429–2432. Association for Computing Machinery, New York (2015). https://doi.org/10.1145/2702613.2702653
7. Jordan, P.W.: Designing Pleasurable Products: An Introduction to the New Human Factors. CRC Press, Boca Raton (2000). https://doi.org/10.4324/9780203305683
8. Kimura, R., Nakajima, T.: Collectively sharing people's visual and auditory capabilities: exploring opportunities and pitfalls. SN Comput. Sci. **1**(5), 298:1–298:24 (2020)
9. Kimura, R., Nakajima, T.: A digital platform for sharing collective human hearing. J. Data Intell. **3**(2), 232–251 (2022)
10. Kimura, R., Nakajima, T.: Collectively sharing human hearing in artful CollectiveEars. In: Streitz, N., Konomi, S. (eds.) HCII 2021. LNCS, vol. 12782, pp. 224–234. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77015-0_16
11. Latour, B.: Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford University Press, Oxford, New York (2005)
12. Leonardi, P.M.: When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies. MIS Q. **35**(1), 147–167 (2011)

13. Löwgren, J.: Annotated portfolios and other forms of intermediate-level knowledge. Interactions **20**(1), 30–34 (2013). https://doi.org/10.1145/2405716.2405725
14. Orlikowski, W.J., Yates, J., Okamura, K., Fujimoto, M.: Shaping electronic communication: the metastructuring of technology in the context of use. Organ. Sci. **6**(4), 423–444 (1995)
15. Zimmerman, J., Forlizzi, J.: Research through design in HCI. In: Olson, J.S., Kellogg, W.A. (eds.) Ways of Knowing in HCI, pp. 167–189. Springer, New York (2014). https://doi.org/10.1007/978-1-4939-0378-8_8

IoT Security and Privacy Challenges from the Developer Perspective

Yaqin Shaheen[1,2(✉)], Miguel J. Hornos[2], and Carlos Rodríguez-Domínguez[2]

[1] Palestine Technical University-Kadoorie (PTUK), Tulkarm, Palestine
yaqin.shaheen@ptuk.edu.ps
[2] Software Engineering Department, Higher Technical School of Computer and Telecommunications Engineering, Aynadamar Campus, University of Granada, 18071 Granada, Spain
{mhornos,carlosrodriguez}@ugr.es

Abstract. Internet of things (IoT) is a highly-demand technological sector that can be applied to most of our life aspects, such as home, work, transportation, health, education and other important areas. At the same time, its heterogeneous nature, dynamism, intelligence, and mobility makes it vulnerable and unsafe in terms of security and privacy. Finding complete answers to the existing security and privacy concerns is a very challenging task for security experts due to the variety of platforms where the IoT is available. This paper demonstrates the major and critical challenges that facing developers through their journey of building and developing solutions for IoT-based systems (IoTSs). Developers are among the key players who directly engage with the IoTSs, along with manufacturers and users. Consequently, the goal of our study is to shed light on the main difficulties that developers have while attempting to manage the aspects related to security and privacy when they create and develop an IoTS.

Keywords: Internet of Things (IoT) · Security · Privacy · Developers

1 Introduction

The Internet of Things (IoT) is vying to become one of the information and communications technology (ICT) landscape's most prominent paradigms. The concept of integrating computing and communication capabilities into everyday objects or *things* encourages the creation of a wide range of solutions that might be applied in different scenarios like healthcare, smart homes, transportation systems, agriculture, industry, and tourism, to name only a few of them [1, 2]. Those great examples of IoT benefited from the latest advancements in technology, which offered a smart and intelligent sensor network to be used for collecting and transmitting data constantly over the internet for further analysis [3, 4].

Furthermore, IoT has profoundly revolutionized how information technology and communication environment operate [1]. However, it is still a developing and increasing paradigm. A recent research in privacy and security reveals there is limited unification

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 13–21, 2023.
https://doi.org/10.1007/978-3-031-43461-7_2

and integration of security and privacy into IoT, which may hinder user acceptance of the technology, due to concerns about the exposure of personal data [1].

Moreover, according to other study [5], security is among the most important problems and concerns in the IoT ecosystem that must be addressed and should be given top attention for many present and future IoT applications. Furthermore, enforcing security regulations and creating proper security standards for IoT have become not only a crucial issue but also a responsibility, given that main elements of the IoT network are essentially resource-constrained devices it terms of computation, power, memory, and bandwidth. Unfortunately, there is not an exhaustive list of specifications for an IoT security architecture that fully meets every demand of an IoT system [6, 7], making it extremely difficult to implement IoT applications and services in a safe, reliable, and scalable manner [8].

Additionally, other important factors including high performance, low power consumption, resistance to assaults, tampering with the data, and end-to-end security are used and considered to design and create security and privacy strategies for IoTSs [3].

Studying and investigating security and privacy challenges, difficulties and pain points of developing a safe IoTS can be done from different angles or viewpoints corresponding to different types of actors: (i) IoT device manufacturer; (ii) IoTS users; and (iii) IoTS developers. This study is mainly focusing on the developers' perspective. As a result, we will summarize the main challenges related to IoT privacy and security facing by IoTS developers, highlighting some of the top skills the developers should have, and the countermeasures for the attacks.

The rest of this paper is structured as follows: Sect. 2 briefly sketches out the background of the topic addressed. Section 3 talks about IoTS developers and the required skills for them. Section 4 presents our proposal and recommendations extracted from the analysis carried out. Finally, Sect. 5 outlines the conclusions and future work.

2 Background

Despite the spread of IoTSs across a range of industries, from embedded systems to smart homes, there has not been a thorough research to comprehend IoTS development issues from the point of view of developers [9, 10]. For that reason, many scholars started to study how to tackle with security and privacy concerns during the development of IoTSs. For instance, Macedo et al. [11] emphasized how crucial it is to follow a security by design approach to protect an IoTS. They also focused on having a holistic security approach in order to safeguard IoT building blocks. Moreover, they found that there is not a clear architecture that considers security and privacy while developing IoTS security solutions. Another important work has been done by Javed et al. [5], who examined applications of IoTSs, their potential in the future, and their biggest obstacles, to provide designers and developers with design considerations that will result in security by default. That study concludes that IoTS embedded security can increase the effectiveness of the security mechanisms and make control transparent to the end user. Moreover, embedded security offers increased performance, dependability, and cost.

Other scholars stressed on following the security by design approach for the development of IoTSs, such as Corno et al. [12], who firmly think that it is the greatest method

to increase the reliability of every software system. Moreover, they stated that fundamental security concepts must be in the thoughts of programmers from the start of the development cycle. Consequently, the demand for more skilled and knowledgeable IoTS developer will be increased.

Sequeiros et al. [13] talked about the absence of succinct techniques that developers may use to address key security vulnerabilities. Besides, there is no integration between different solutions and approaches that used to solve some security vulnerabilities.

For those reasons, our study intends to discuss security and privacy challenges and difficulties to be addressed in IoTSs, as well as skills required for developers building quality IoTSs.

3 Existing Techniques to Achieve Security and Privacy in IoTSs

Nowadays, an increasing number of IoTSs are being created and put into operation on open networks, without implementing proper security. A tiny user research indicated that inexperienced IoT programmers in particular had a tendency to ignore security risks [14]. For example, Arduino, due to its open-source design and user-friendly interface, it is now widely used by hobbyists and inexperienced programmers, particularly to develop IoTSs.

As many developers begin working on IoT platforms without having a thorough understanding of the most important security concepts in ICTs, there is a huge propagation of insecure IoT devices. As a result, IoT applications used at the open networks, which are frequently publicly accessible on GitHub (or other code-sharing sites), can be downloaded and used by novice users as well as other developers, spreading these vulnerabilities to other projects [15].

Programming Skills. Developer skills needed to design and implement IoTSs ranging from standard programming skills, like any other software developer, to the advanced and broad skills in different areas of IoTS development. Corneliu et al. [16] discovered that most application developers have had training in either web development or mobile development. Some of those developers tend to believe that their expertise would be immediately transferable to IoTS development [16]. This is not actually the case, since there are a number of very specific features of IoTSs that are usually not addressed when developing a mobile or web system. Along with being embedded, IoT devices and gateways also have highly distributed architectures, which have unique needs. Therefore, developers need training and expertise not only in web and mobile software development, but also in other important areas of software development, including embedded and cloud systems, in order to create safe and secure IoTSs. Moreover, distributed programming abilities are essential due to the ubiquitous nature of IoT devices [16].

Robust Cloud-Based Systems. Some researchers considered that appropriately using cloud-based IoT platforms, such as Amazon Web Services (AWS), should increase robustness and provide easier to implement, maintain and deploy back ends. But, even while such a platform could be secure on its own [17], inexperienced developers might have neglected to close some security holes that could be exploited by unauthorized users [12]. Therefore, in their opinion, the greatest technique to increase the dependability of

every system is security by design. To do this, programmers must start the development cycle with an understanding of the fundamental security concepts.

Security by Design. Macedo et al. [11] emphasized how crucial it is to take a holistic approach to security in order to safeguard IoT building blocks and provide security by design. They stressed there is a lack of a well-defined architecture that takes security into account and may be used as a model for creating IoT security solutions. Javed et al. [5] look at security by design as an architecture built on the foundation of safe hardware, secure software, and secure communications. Moreover, security by design must be achieved by designers and developers in order to provide protection, save costs, and increase performance and dependability.

Privacy Concerns. A key accusation against IoTSs might be privacy infringement, due to the risk that IoTSs cause to the privacy of personal data. When users learn about the information that has been gathered about them, a large portion of them change their minds about using a system or program. On this account, the development of IoTSs must also take privacy into consideration. One big obstacle is the lack of developer expertise about privacy [18]. It is commonly known that software developers focus the majority of their attention on system functionalities. However, privacy and security concerns, which are non-functional requirements, are frequently given secondary consideration [18].

Threat Modeling and Attack Modeling. Threat modeling, which is "an established way of enumerating and categorizing potential security implications for a given technology" [19], is another essential competency for developers. It is what enables developers to properly detect potential security threats during the design process, so that mitigation techniques can be applied in these early stages of development. Threat modeling should receive more attention than attack modeling, which is used in the detection of potential attacks and vulnerability exploitation [13]. Different thread modeling frameworks could be used, such as STRIDE, LINDDUN, and CORAS [19].

4 Proposal and Recommendations Extracted from the Analysis Carried Out

As aforementioned, along with the technological infrastructure, the three main pillars in IoT are manufactures, users, and developers. In addition, it is commonly known that developers play a key part in creating IoTS solutions that affects our daily life. Consequently, studying the major obstacles that developers face when they create and develop solutions is therefore crucial, since doing so will enable them to create more robust and secure IoTSs.

In this study, we analyzed a number of recent studies that mostly focused on the developers' perspective and were concerned with IoT security and privacy issues. Our analysis of those publications draws a comprehensive view that helps in formulating some recommendations, which is highly beneficial for IoTS developers and their businesses. Our recommendations are split into the following parts:

Part 1: General Recommendations. Based on our analysis, there are some immediate steps to take when developing an IoTS, which are the following ones:

- Provide instructional materials, such as example code, photos, and drawings, to developers to help them understand the fundamentals of security and privacy risks associated to IoTSs and how to prevent them.
- It is essential to shed the light on the core privacy and security issues related to IoTSs and help developers to understand them. Since most developers focus on functional requirements, they tend to neglect the non-functional ones, such as privacy and security.
- The heterogeneous nature of IoTSs makes it difficult to implement once solution for all cases. Developers must therefore thoroughly analyze the similarities and differences between each project before implementing their solutions.
- Novice developers should be closely work with senior ones, continuously discuss and share their ideas before being using any IoTS sample codes from open-source websites, as such code snippets could be vulnerable to attacks.
- Providing developers with regular and concise security recommendations and assistance both before starting and throughout the project.
- Keeping developers informed about new security flaws affecting the hardware they are using, as well as security fixes and upgrades.
- Developers should not be the only ones making the decision on which cloud platform to use for the development of a given IoTS. As a number of these platforms, like Google, Windows Azure, and Amazon, may be used currently, the attributes of each platform should be carefully compared in a complete project study.
- Developers should be given a detailed explanation of the primary security features in the selected cloud platform (e.g., Google, Windows Azure, or Amazon), and the correct usage and implementation of those features should be monitored and evaluated after the project conclusion.

Part 2: Challenges and Practical Recommendations. An Increased security issues in IoTS will be exacerbated by the non-standardization of IoT technologies with increased vulnerabilities. The following highlight some critical threats and their practical countermeasures that developers must be aware of:

Legislation Challenges. Despite the fact that regulations cannot guarantee the secure use of IoT data, it can reimburse for the destruction caused by data exploitation. The General Data Protection Regulation (GDPR) [20] and HIPPA [21] are only two examples of the efforts done by many countries to protect users' data. Therefore, app developers must follow such regulations while providing features to IoTS such as blood pressure, and other health insights [22].

Protecting the Device from Unauthorized Access. The use of built-in credentials and default passwords by the manufacturers is one of the primary security vulnerabilities popular nowadays. Similarly, developers purposefully leave insecure APIs available for remote access using default passwords. Thus, employing secure connections and complex passwords is a smart practice for developers [22]. In addition, developers can also use single-sign-on mechanisms in certain IoT contexts, so users just need to login once to utilize different devices. Users no longer need to input numerous passwords in order to access all resources for which they have access authorization [1]. Moreover,

IoT API security techniques including tokens, encryption, signatures, gateway limits, and throttling should be employed to safeguard the integrity of data transit [23].

Weak Application Security. Weak authentication and authorization mechanisms can lead to, brute force attacks, data leaks, elevated privileges, and data manipulation. Furthermore, according to the OWASP application security risk score, IoTS that may accessed through websites are in danger. For example, SQL injections attacks could be made against IoTS and databases, such as Belkin smart home which exploited by security researchers. In addition, WeMo android app that injected by malicious code to get a full control (root account) [22]. It's recommended for developers to use the common authentication techniques such as Role-based access control (RBAC), attribute-based access control (ABAC), and Authentication and Authorization for Constrained Environments (ACE) [1]. Moreover, strong authentication mechanisms must be implemented such as digital certificates, two-factor authentication, and biometrics on a single IoT device [23].

Poor Programming Techniques and Ineffective Software Updates. The authors of [1] came to the conclusion that the two primary IoTS vulnerabilities are (i) unnecessarily exposed ports and (ii) poor programming techniques combined with ineffective software updates capabilities. Consequently, Strong programming practices and the addition of security components may increase the IoTS's resilience, taking into account that software releases do not contain any vulnerabilities like backdoors, root users as prime access points, and a lack of Secure Socket Layer (SSL) implementation.

Improper Encryption. While it is obvious that encryption can assist to mitigate some IoTS vulnerabilities listed in [24], complicated cryptographic operations, such those included in the Advanced Encryption Standard (AES), can cause significant overhead for resource-constrained IoT devices. As a result, low-power hardware-optimized, ultra-light weight encryption methods are becoming more popular [1]. IoT PKI security techniques including cryptographic keys, X.509 digital certificates, and life-cycle features like private/public key creation, revocation, distribution, and administration may greatly help in ensuring a secure connection between an IoT device and app [23].

Sensor Based Attacks. To protect the devices from threats and assaults based on sensors, and to reduce the misuse of sensor data in applications, developers can adhere to the recommendations made available by the vendors. Moreover, developers may adhere to best practices for creating apps, such as using encrypted sensor data, a trusted data flow channel, and only requesting necessary sensor permissions, etc. [25].

Part 3: Building a Repository Tool. Our proposal is to build a tool considered as a repository of succinct information, direct actions, and procedures on building secure IoTSs. This tool would connect developers and manufactures, and allow them to share their experiences, which would certainly improve the level of security for the intended IoTS in each case.

The suggested tool is based on the security by design approach that is considered as a solution for most of IoTS security and privacy issues. This approach is recommend by most of the studies analyzed [5, 11, 12].

Our suggested tool, considered as a knowledge base that encompasses a huge data about IoTS devices, their security issues, and countermeasures that should be followed

by developers. Many studies emphasis on raising developers' security awareness and educating them more about security [1, 22]. For instance, developers working on smart agriculture app. Will be able to find all sensors information, their vulnerabilities and a practical security details for implementation.

5 Conclusions and Future Work

In this paper, we studied the major difficulties and challenges correlated to security and privacy that facing developers when addressing the development of IoTSs. In fact, we have first presented a brief background about how to achieve security and privacy in IoTSs. Then, we categorized and discussed the key security and privacy aspects from the perspective of IoTS developers.

In the central part of our analysis, we concluded that security by design should be pursued during the development of any IoTS to avoid critical security and privacy issues that may happened in the future, as well as to minimize the additional time and costs needed to fix security bugs later on. Additionally, a variety of abilities, including mastering programming, cloud-based systems, threat modeling and attack modeling, are fundamental for every IoTS developer. Finally, after this analysis, we were able to provide developers and businesses with a set of recommendations to have more secure IoTSs.

In our future work, we will go deeper into the primary security and privacy needs and risks of IoTSs, as well as how IoTS developers should address and approach them. In addition, we will implement our proposed repository tool to check and validate its ability to help developers in building more secure IoTSs.

Acknowledgments. The authors would like to thank Palestine Technical University - Kadoorie for their financial support to conduct this research. This research was also funded by the Spanish Ministry of Science and Innovation (State Research Agency), grant number PID2019-109644RB-I00, and Junta de Andalucía (Andalusian Regional Government), grant number B-TIC-320-UGR20.

References

1. Ogonji, M.M., Okeyo, G., Wafula, J.M.: A survey on privacy and security of internet of things. Comput. Sci. Rev. **38**, 100–312 (2020). https://doi.org/10.1016/j.cosrev.2020.100312
2. Corno, F., De Russis, L., Saenz, J.P.: Pain points for novice programmers of ambient intelligence systems: an exploratory study. In: Proceedings - International Computer Software and Applications Conference, vol. 1, pp. 250–255 (2017). https://doi.org/10.1109/COMPSAC.2017.186
3. Chanal, P.M., Kakkasageri, M.S.: Security and privacy in IoT: a survey. Wirel. Pers. Commun. **115**, 1667–1693 (2020). https://doi.org/10.1007/s11277-020-07649-9
4. Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of things: vision, applications and research challenges. Ad Hoc Netw. **10**, 1497–1516 (2012). https://doi.org/10.1016/j.adhoc.2012.02.016
5. Javed, B., Iqbal, M.W., Abbas, H.: Internet of things (IoT) design considerations for developers and manufacturers. In: 2017 IEEE International Conference on Communications Workshop, ICC Workshop 2017, pp. 834–839 (2017). https://doi.org/10.1109/ICCW.2017.7962762

6. Li, S., Tryfonas, T., Li, H.: The internet of things: a security point of view. Internet Res. **26**, 337–359 (2016). https://doi.org/10.1108/IntR-07-2014-0173
7. Sicari, S., Rizzardi, A., Grieco, L.A., Coen-Porisini, A.: Security, privacy and trust in internet of things: the road ahead. Comput. Netw. **76**, 146–164 (2015). https://doi.org/10.1016/j.comnet.2014.11.008
8. Pal, S., Hitchens, M., Rabehaja, T., Mukhopadhyay, S.: Security requirements for the internet of things: a systematic approach. Sens. (Switz.) **20**, 1–34 (2020). https://doi.org/10.3390/s20205897
9. Makhshari, A., Mesbah, A.: IoT development in the wild: bug taxonomy and developer challenges. In: Proceedings of the International Conference on Software Engineering, pp. 225–226 (2021). https://doi.org/10.1109/ICSE-Companion52605.2021.00103
10. Kühtreiber, P., Pak, V., Reinhardt, D.: A survey on solutions to support developers in privacy-preserving IoT development. Pervasive Mob. Comput. **85**, 1–31 (2022). https://doi.org/10.1016/j.pmcj.2022.101656
11. Macedo, E.L.C., et al.: On the security aspects of internet of things: a systematic literature review. J. Commun. Netw. **21**, 444–457 (2019). https://doi.org/10.1109/JCN.2019.000048
12. Corno, F., de Russis, L., Mannella, L.: Perception of security issues in the development of Cloud-IoT systems by a novice programmer. In: Intelligent Environments 2021: Workshop Proceedings of the 17th International Conference on Intelligent Environments, pp. 5–15 (2021). https://doi.org/10.3233/AISE210074
13. Sequeiros, J.B.F., Chimuco, F.T., Samaila, M.G., Freire, M.M., Inácio, P.R.M.: Attack and system modeling applied to IoT, cloud, and mobile ecosystems: embedding security by design. ACM Comput. Surv. **53**, 1–32 (2020). https://doi.org/10.1145/3376123
14. Corno, F., De Russis, L., Mannella, L.: Helping novice developers harness security issues in cloud-IoT systems. J. Reliab. Intell. Environ. **8**, 261–283 (2022). https://doi.org/10.1007/s40860-022-00175-4
15. Corno, F., Mannella, L.: Security evaluation of arduino projects developed by hobbyist IoT programmers. Sensors **23**, 27–40 (2023). https://doi.org/10.3390/s23052740
16. Corneliu, B., Oana, G., Marius, P.: Considerations on the development of IoT systems. In: 2019 International Conference on Sensing and Instrumentation in IoT Era, ISSI 2019, pp. 13–19 (2019). https://doi.org/10.1109/ISSI47111.2019.9043644
17. AWS: Introduction to AWS Security: AWS Whitepaper (2021)
18. Alkhariji, L., Alhirabi, N., Alraja, M.N., Barhamgi, M., Rana, O., Perera, C.: Synthesising privacy by design knowledge toward explainable internet of things application designing in healthcare. ACM Trans. Multimed. Comput. Commun. Appl. **17**, 1–29 (2021). https://doi.org/10.1145/3434186
19. Aufner, P.: The IoT security gap: a look down into the valley between threat models and their implementation. Int. J. Inf. Secur. **19**(1), 3–14 (2019). https://doi.org/10.1007/s10207-019-00445-y
20. Voigt, P., von dem Bussche, A.: The EU general data protection regulation (GDPR) (2017). http://dx.doi.org/10.1007/978-3-319-57959-7
21. Edemekong, P.F., Annamaraju, P., Haydel, M.J.: Health insurance portability and accountability act. Presented at the January (2023)
22. Iqbal, W., Abbas, H., Daneshmand, M., Rauf, B., Bangash, Y.A.: An in-depth analysis of IoT security requirements, challenges, and their countermeasures via software-defined security. IEEE Internet Things J. **7**, 10250–10276 (2020). https://doi.org/10.1109/JIOT.2020.2997651
23. Omolara, A.E., et al.: The internet of things security: A survey encompassing unexplored areas and new insights. Comput. Secur. **112**, 102494 (2022). https://doi.org/10.1016/j.cose.2021.102494

24. Wurm, J., Hoang, K., Arias, O., Sadeghi, A.-R., Jin, Y.: Security analysis on consumer and industrial IoT devices (2016). http://dx.doi.org/10.1109/aspdac.2016.7428064
25. Sikder, A.K., Petracca, G., Aksu, H., Jaeger, T., Uluagac, A.S.: A survey on sensor-based threats and attacks to smart devices and applications. IEEE Commun. Surv. Tutorials. **23**, 1125–1159 (2021). https://doi.org/10.1109/COMST.2021.3064507



Flow Disaggregation: Underdetermined Non-negative Linear Systems

Javier Curto[1(✉)], Guillermo Hernández[1,2], María Alonso-García[1], Alicia Serrano-Ortega[3], Alberto Toledo-Garrote[3], and Pablo Chamoso[1,2]

[1] Air Institute, IoT Digital Innovation Hub (Spain), 37188 Carbajosa de la Sagrada, Salamanca, Spain
{jcurto,guillehg,marialonsogar,chamoso}@air-institute.com
[2] BISITE Research Group, Dpto. Informática y Automática, Facultad de Ciencias, Universidad de Salamanca, Salamanca, Spain
[3] Airline Sciences, Airbus Operations SL, 28906 Getafe, Spain
{alicia.serrano-ortega,alberto.toledo-garrote}@airbus.com

Abstract. This article proposes and tests an iterative approximation algorithm to solve underdetermined non-negative linear systems of the form $A\mathbf{x} = \mathbf{b}$, where A is a binary matrix and all the elements of $\mathbf{x}$ and $\mathbf{b}$ are positive, while the system admits an infinite number of solutions. The algorithm converges (experimentally) to a solution based on the initial problem configuration, and is characterized by non-negativity and disaggregation invariance properties. The solution arises from the necessity to address the air transport problem of generating itinerary flows based on leg flows, and its effectiveness is demonstrated in this article. The algorithm exhibits positive results in this particular use case and is also applicable to other similar problems related to flows in areas such as transportation, migration, or economy.

Keywords: underdetermined non-negative linear systems · flow disaggregation · air transport

1 Introduction

Air traffic is an essential component of modern transportation and plays a crucial role in the global economic and social landscape. The creation of efficient itineraries and the management of air traffic flows are vital for ensuring the optimal operation of the aviation industry. However, the planning and management

The authors would like to express their sincere appreciation to the entire team who contributed to this project. Their invaluable support and dedication were instrumental in the successful completion of this research. Special thanks are due to Juan Villaseca, Rodrigo Álvarez, Albert Ruiz, Clara Argerich, Jesús Sanz (Airbus team), Iñaki A. Apaolaza, David Berrocal, Raúl Juan, Samuel Alonso, Jairo Calderón (AIR Institute team) whose contributions were particularly noteworthy.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 22–32, 2023.
https://doi.org/10.1007/978-3-031-43461-7_3

of these flows is a complex task due to the indeterminate nature of the problems that arise in this science.

Data disaggregation is a key aspect of air traffic planning and management. By disaggregating air traffic flows into different levels, more detailed information and valuable insights on traffic patterns can be obtained, allowing for more informed decision-making and enhancing overall system performance. A majority of air transport-related research typically conducts demand calculations on the basis of capacity, i.e., the number of seats available on each leg, and the unconstrained demand for Origin-Destination (OD) pairs [7], alongside other factors pertinent to air transport challenges like spill and recapture of passengers [2]. Nonetheless, this study do not take into account specific air transport concepts and approaches the problem in a more comprehensive manner, assuming that only information about real leg passengers is accessible, thereby allowing the algorithm to be adapted to domains beyond air transportation. This article approaches the particular problem of computing itinerary flows based on leg flows (explained in Sect. 2) which derives in underdetermined non-negative linear systems.

The article is structured as follows. Section 2 outlines the problem formulation, which serves as the basis for developing the proposed solution. In Sect. 3, the solution is presented, including the methodology, algorithms, and implementation details. Section 4 presents a real-world case study to demonstrate the effectiveness and applicability of the system. In Sect. 5, the main findings of the research are summarized, and the conclusions are drawn. Additionally, potential lines for future research are highlighted.

2 Problem Formulation

In this section, the problem of computing itinerary flows based on leg flow data is presented. This disaggregation issue results in linear equation systems possessing infinite solutions. Initially, the essential concepts required for the problem formulation will be elucidated, followed by a mathematical exposition of the problem as a system of equations and an analysis of its properties.

2.1 Flow Graph Definition

To comprehend the problem, it is crucial to define the term flow and the concepts related to it. As mentioned earlier, such problems predominantly arise in air transport science; therefore, a simple example based on air transport terminology will be employed to devise the solution algorithm. A flow can be defined as a collection of individuals traveling from one point to another. Consequently, if multiple interconnected points exhibit flows between them, a flow graph is obtained.

In air transport concepts, each vertex of the graph is considered an airport. A graph consisting of four airports is depicted in Fig. 1. Each directed edge in the graph signifies the passengers traveling on flights between the corresponding airports. This implies that only four segments are possible for flying, namely b_{AB}, b_{AD}, b_{BC}, and b_{BD}. The term "leg" is associated with each of these segments, and consequently, the flow graph in Fig. 1 is referred to as a "leg graph".

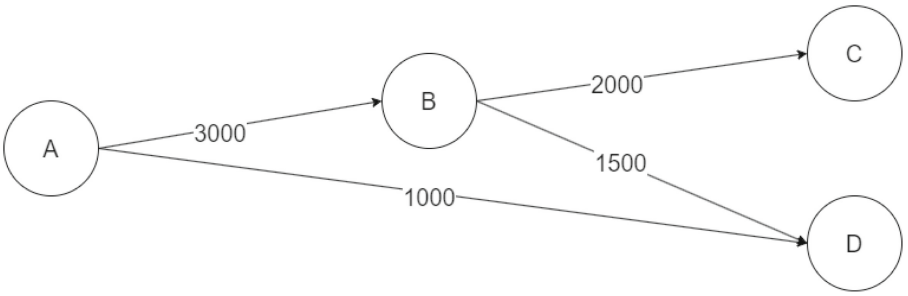


Fig. 1. Leg graph of 4 interconnected airports, and the 4 legs associated with the passengers of each leg

Nonetheless, some individuals may desire to travel from point A to C, necessitating traversing two segments of the graph in such instances. The complete journey of an individual is termed an "itinerary". Consequently, it is plausible that the same leg comprises individuals traveling on distinct itineraries (x_{AB}, x_{ABC}, x_{ABD}, x_{BC}, x_{BD}, x_{AD}). When a leg encompasses an itinerary, it will be denoted as $x \subset b$. Figure 2 serves as an illustration, wherein the legs are constructed from the itineraries as follows:

- b_{AB} is built with all passengers from x_{AB}, x_{ABC}, x_{ABD}
- b_{BC} is built with all passengers from x_{BC}, x_{ABC}
- b_{BD} is built with all passengers from x_{BD}, x_{ABD}
- b_{AB} is built with all passengers from x_{AD}

Finally, the final significant term in air transport demand problems is the OD, referring to individuals who commence and conclude their itineraries at identical points. In this case, there are 5 ODs constructed from the itineraries:

- OD_{AB} is built with all passengers from x_{AB}
- OD_{AC} is built with all passengers from x_{ABC}
- OD_{AD} is built with all passengers from x_{AD}, x_{ABD}
- OD_{BC} is built with all passengers from x_{BC}
- OD_{BD} is built with all passengers from x_{BD}

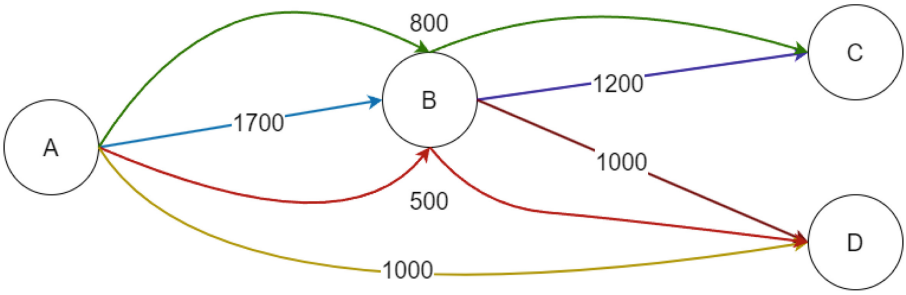


Fig. 2. Itinerary graph of 4 interconnected airports, and the 6 possible itineraries associated, with feasible passengers values for each itinerary (i.e. when aggregating the values they match with the leg values)

2.2 Mathematical Description

The aim of the algorithm is to calculate the itinerary values shown in Fig. 2 from the values of the leg graph in Fig. 1. Itinerary information is more granular than leg one, which means that by knowing the itinerary graph it is possible to know the exact leg graph by aggregation, however, for the inverse process it is not the same.

From this point, the terms "leg" or "itinerary" make reference to "passengers associated with that leg or itinerary". Let's start by defining the vector of known legs $\mathbf{b} = (b_1, b_2, ..., b_n)$ and the unknown itineraries $\mathbf{x} = (x_1, x_2, ..., x_m)$. A leg is composed of different itineraries, and this composition can be defined in terms of all the itineraries as:

$$b_i = a_{i1}x_1 + a_{i2}x_2 + ... + a_{im}x_m \tag{1}$$

$$a_{ij} = \begin{cases} 1 \text{ if } x_j \subset b_i \\ 0 \text{ if } x_j \not\subset b_i \end{cases} \tag{2}$$

extended to all the legs the problem is simplified to solving a linear system of equations:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{3}$$

An important point here is that for a leg between 2 points A and B (b_{AB}), there is always a possible itinerary x_{AB}. This leads to two important conclusions. Firstly, if at least one itinerary with more than one leg exists it is fulfilled that $m > n$. Then, the simplest solution to the problem can be found by giving 0

passengers to all the itineraries with more than one leg so that for the one-leg itineraries we have: $x = b$. These two conclusions imply that the linear system of equations is underdetermined and has infinite solutions.

Finally, it should be taken into account that the system is non-negative because all coefficients and variables meet the following condition $\{x_j, b_i, a_{ij}\} \geq 0\ \forall i, j$ and that all coefficients a_{ij} are constant and defined from the leg graph to solve the problem.

3 Solution with Weighted Numerical Method

As stated in the previous section the system has infinite solutions. There are many methods or algorithms designed to find specific solutions to underdetermined linear systems of equations, but most of them are parametrization methods [6] or optimization methods of the form $\min f(\mathbf{x}) | A\mathbf{x} = \mathbf{b}$ where $f(\cdot)$ measures the quality of the candidate solutions [3], for example, the minimum L_∞ solution [1,4] or the minimum norm one [6]. There are also specific algorithms implemented for specific problems like sparse linear systems [3,4,10,12] or non-negative linear systems [3,5,9,10]

In this case, the system to solve is defined as an underdetermined non-negative linear system where A is a binary matrix (i.e. its elements are either 0 or 1). The objective is not to look for an exact solution to the algorithm, instead, the aim is to find a solution that fits better an initial configuration of the problem; consequently, an iterative algorithm is proposed [4,8,11] to obtain successive solutions of the form:

$$A\mathbf{x^t} = \mathbf{y^t} \mid d(\mathbf{y^{t+1}}, \mathbf{b}) < d(\mathbf{y^t}, \mathbf{b}) \tag{4}$$

3.1 Initialisation

Continuing with the idea of the previous paragraph, the critical point of this method is the initialization of the problem. On each iteration t a weight matrix is going to be used

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1m} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nm} \end{bmatrix} \tag{5}$$

Each weight ω_{ij} represents the relative importance of the itinerary x_j in the leg y_i (i.e. the passengers ratio):

$$\omega_{ij} = \frac{x_j}{y_i} \tag{6}$$

For example, in Fig. 2, $\omega_{ABC-AB} = x_{ABC}/y_{AB} = 800/3000 = 0.267$. $\mathbf{x}^t$ is calculated using $\mathbf{y^{t-1}}$ and Ω^{t-1}. In that way, for $\mathbf{x}^1$ there is no previous

iteration so it is necessary to initialize the problem giving values to $\mathbf{x^0}$, where $x_j^0 > 0 \; \forall j$ so that all the values of A, $\mathbf{x^0}$, $\mathbf{y^0}$, Ω^0 are known. With this in mind, Ω^0 must be initialized as close to reality as possible, and the algorithm will search for a solution with a similar configuration. The definition of $\mathbf{x^0}$ depends on every particular problem and is out of the scope of this article, however, Sect. 4 shows an example where $\mathbf{x^0}$ is a set of itineraries with wrong data that must be corrected with the algorithm.

3.2 Algorithm

Once the initial configuration is defined, the algorithm can be developed. Let Fig. 1 serve as an example to illustrate the process, where it is known:

$$\mathbf{b} = (b_{AB}, b_{AD}, b_{BC}, b_{BD}) = (3000, 1000, 2000, 1500) \tag{7}$$

Now let's assume that on iteration t we have the $\mathbf{x^t}$ shown in Fig. 3:

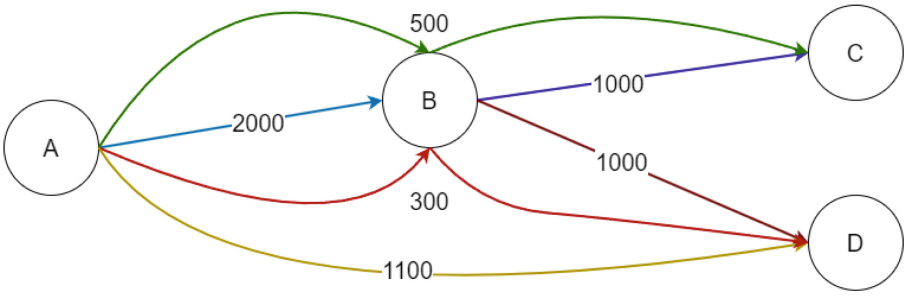


Fig. 3. Example of a possible $\mathbf{x^t}$ of the problem in Fig. 1

$$\begin{aligned} \mathbf{x^t} &= (x_{AB}^t, x_{ABC}^t, x_{ABD}^t, x_{AD}^t, x_{BC}^t, x_{BD}^t) \\ &= (2000, 500, 300, 1100, 1000, 1000) \end{aligned} \tag{8}$$

And so:

$$\mathbf{y^t} = (y_{AB}^t, y_{AD}^t, y_{BC}^t, y_{BD}^t) = (2800, 1100, 1500, 1300) \tag{9}$$

$$\mathbf{b} - \mathbf{y^t} = (200, -100, 500, 200) \tag{10}$$

The logic of the algorithm is as follows: $x_{ABC}^t = 500$, which represents the 17.8% of y_{AB}^t and 33.3% of y_{BC}^t. Looking at Eq. 10, b_{AB}^t should be incremented by 200 and b_{BC}^t should be incremented by 500. Proportionally to y_{AB}^t, x_{ABC}^t should be incremented by $0.178 * 200 = 35.6$ but proportionally to y_{BC}^t, x_{ABC}^t should be incremented by $0.333 * 500 = 166.5$. Therefore, x_{ABC}^t will increase

the mean value: $x_{ABC}^{t+1} = x_{ABC}^{t} + \Delta x_{ABC}^{t} = 500 + (166.5 + 35.6)/2 = 601.1$. Generalizing the logic for all the itineraries we have:

$$x_j^{t+1} = x_j^t + \Delta x_j^t \tag{11}$$

$$\Delta x_j^t = \frac{(\mathbf{a_j} \odot \mathbf{w_j^t}) \cdot (\mathbf{b} - \mathbf{y}^t)}{\mathbf{a_j} \cdot \mathbf{a_j}} \tag{12}$$

where $\odot$ represents the Hadamard product. The expression Eq. 12 is the core of the algorithm and is developed in order to use the importance of each itinerary to adjust the next value. As can be observed, this expression is applied on every itinerary and depends on Ω^t and $\mathbf{y^t}$, so that for each iteration it is necessary to calculate these parameters. By summarizing all calculations and eliminating vector notation from Eq. 12, the complete algorithm can be expressed as three successive calculations, each applied to all itineraries, during every iteration:

$$x_j^{t+1} = x_j^t + \frac{\sum_{i=1}^{n} a_{ij}\omega_{ij}^t (b_i - y_i^t)}{\sum_{i=1}^{n} a_{ij}} \tag{13}$$

$$y_i^{t+1} = \sum_{j=1}^{m} a_{ij} x_j^{t+1} \tag{14}$$

$$\omega_{ij}^{t+1} = \frac{x_j^{t+1}}{y_i^{t+1}} \tag{15}$$

A single expression using only A, $\mathbf{b}$, and $\mathbf{x^t}$ can also be employed.

$$x_j^{t+1} = x_j^t + \frac{\sum_{i=1}^{n} (a_{ij} x_j^t (b_i - \sum_{j=1}^{m} a_{ij} x_j^t) / \sum_{j=1}^{m} a_{ij} x_j^t)}{\sum_{i=1}^{n} a_{ij}} \tag{16}$$

3.3 Properties

Convergence. The convergence is not mathematically demonstrated, approaching this property is planned for future work. Nevertheless, it has been tested experimentally with positive results, i.e. in the experiments it is always fulfilled that $d(\mathbf{y^t}, \mathbf{b}) > d(\mathbf{y^{t+1}}, \mathbf{b})$. However, for the moment it is necessary to implement two stopping conditions for the algorithm, the first one regarding the minimum error required $\xi > 0$ (user input) and the other one regarding the not-convergent possibility; so the algorithm will stop when:

$$d(\mathbf{y^{t+1}}, \mathbf{b}) < \xi \tag{17}$$

or

$$d(\mathbf{y^t}, \mathbf{b}) < d(\mathbf{y^{t+1}}, \mathbf{b}) \tag{18}$$

Non-negative Solution. An additional noteworthy attribute of the algorithm is that it obviates the potential negative solutions, as represented by the condition $\exists; \mathbf{x}, j; |; A\mathbf{x} = \mathbf{b} \wedge x_j < 0$. Specifically, if the solution vector $\mathbf{x^t}$ is non-negative ($x_j^t > 0; \forall j$), then the subsequent vector $\mathbf{x^{t+1}}$ will always exhibit the same non-negative property, which is particularly advantageous for the flow problem. This property can be easily demonstrated by proving that $\forall i, j$ (taking into account that both y_i^t and x_j^t are positive):

$$\omega_{ij}^t(b_i - y_i^t) = \frac{x_j^t(b_i - y_i^t)}{y_i^t} > -x_j^t \tag{19}$$

This is trivial since solving the inequality yields $b_i > 0$ which is a requirement of the problem. Then we have:

$$\Delta x_j^t = \frac{\sum_{i=1}^n a_{ij}\omega_{ij}^t(b_i - y_i^t)}{\sum_{i=1}^n a_{ij}} > -x_j^t \tag{20}$$

so from Eq. 11:

$$\Delta x_j^t = x_j^{t+1} - x_j^t > -x_j^t \tag{21}$$

$$x_j^{t+1} > 0 \tag{22}$$

Itinerary Disaggregation Invariance. The last property of this algorithm is the itinerary disaggregation invariance. The process of itinerary disaggregation involves dividing a given physical itinerary (represented just by the airports involved) into multiple itineraries that follow the same route (same airports). Although this approach may appear counterintuitive, its rationale can be elucidated through the context of air transportation. As evidenced by Fig. 2, a hypothetical scenario may entail the presence of 800 passengers on a specific itinerary, denoted as x_{ABC}. However, in practical situations, such passengers may be allocated to different airlines while following the same route, thereby necessitating the creation of separate itineraries for each airline involved, denoted as $x_{ABC1}, x_{ABC2}, ..., x_{ABCk}$. This disaggregation process can be repeated for all itineraries, leading to an expansion in the number of possible itineraries and a consequent increase in the size of the equation system, represented as $A'\mathbf{x}' = \mathbf{b}$:

$$\begin{bmatrix} a'_{11} & a'_{12} & \cdots & a'_{1q} \\ a'_{21} & a'_{22} & \cdots & a'_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a'_{n1} & a'_{n2} & \cdots & a'_{nq} \end{bmatrix} \cdot \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_q \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{23}$$

with $q > m$. Itinerary disaggregation invariance means that if $A'\mathbf{x}' = \mathbf{b}$ is solved and $\mathbf{x'_t}$ is obtained, then aggregating the solution by physical itineraries yields the same result $\mathbf{x_t}$ as solving $A\mathbf{x} = \mathbf{b}$. Specifically, for a given t and $j = 1$ (indices to be removed for calculation), the itinerary can be disaggregated into k itineraries with the same path.

$$x_1 = x'_1 + x'_2 + ... + x'_k \tag{24}$$

and the weights

$$\begin{aligned} \omega'_{i1} &= x'_1/y_i \\ \omega'_{i2} &= x'_2/y_i \\ &\vdots \\ \omega'_{ik} &= x'_k/y_i \end{aligned} \tag{25}$$

so

$$\omega_{i1} = \sum_{p=1}^{k} \omega'_{ip} \tag{26}$$

As they follow the same paths (i.e. are involved in the same legs) is also true that:

$$a_{i1} = a'_{ip} \; \forall i, p \mid p = 1, ..., k \tag{27}$$

That can be developed:

$$\begin{aligned} \Delta x'_1 + ... + \Delta x'_k &= \frac{\sum_{i=1}^{n} a'_{i1}\omega'_{i1}(b_i - y_i)}{\sum_{i=1}^{n} a'_{i1}} + ... + \frac{\sum_{i=1}^{n} a'_{ik}\omega'_{ik}(b_i - y_i)}{\sum_{i=1}^{n} a'_{ik}} \\ &= \frac{\sum_{i=1}^{n} a_{i1}\omega'_{i1}(b_i - y_i)}{\sum_{i=1}^{n} a_{i1}} + ... + \frac{\sum_{i=1}^{n} a_{i1}\omega'_{ik}(b_i - y_i)}{\sum_{i=1}^{n} a_{i1}} \\ &= \frac{\sum_{p=1}^{k}\sum_{i=1}^{n} a_{i1}\omega'_{ip}(b_i - y_i)}{\sum_{i=1}^{n} a_{i1}} \\ &= \frac{\sum_{i=1}^{n} a_{i1}(b_i - y_i)\sum_{p=1}^{k}\omega'_{ip}}{\sum_{i=1}^{n} a_{i1}} \\ &= \frac{\sum_{i=1}^{n} a_{i1}\omega_{i1}(b_i - y_i)}{\sum_{i=1}^{n} a_{i1}} \\ &= \Delta x_1 \end{aligned} \tag{28}$$

Finally combining Eq. 24 and Eq. 28 the following statement is demonstrated:

$$x_1 + \Delta x_1 = x'_1 + ... + x'_k + \Delta x'_1 + ... + \Delta x'_k \tag{29}$$

4 Application to a Real Case: Data Correction

The present algorithm has been developed and evaluated within the context of a data correction problem where initially two data sources have been used, namely SABRE (https://www.sabre.com/), consisting of both itinerary ($\mathbf{x^0}$) and

leg data ($\mathbf{y^0}$), and OAG (https://www.oag.com/es/), which solely offers leg data. The leg data is expected to be the same in SABRE and OAG, however, this is not fulfilled as both sources present errors in the data. Consequently, the correction problem consists in the study and modification of these three datasets in order to obtain new itinerary and leg data with the best possible quality. The first part of the correction process is the comparison of SABRE leg data with that of OAG (this process is out of the scope of this article), yielding a new leg dataset ($\mathbf{b}$) but for sure this dataset does not match with the itinerary dataset from SABRE, so the point here is to adjust the itinerary dataset ($\mathbf{x^0}$) to the new leg dataset ($\mathbf{b}$) using the flow disaggregation algorithm.

The algorithm was applied to all routes between 2010 and 2020, encompassing $\approx 10^7$ legs and $\approx 4 \cdot 10^8$ itineraries in a pyspark environment (driverCores: 1, driverMemory: 6 GB, driverCPU: Intel(R) Xeon(R) Platinum 8375 CPU @ 2.90 GHz, executorCores: 4, executorMemory: 8 GB, executorCPU: Intel(R) Xeon(R) CPU E5-2686 v4 @ 2.30 GHz). With these characteristics, the execution time for each iteration is around 2 h. The convergence of the algorithm was assessed using the MAE.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |b_i - y_i^t| \tag{30}$$

Table 1. MAE of each iteration with real data

Iteration (t)	0	1	2	3	4	5	6	7	8
MAE	152.1	32.2	14.9	8.8	5.8	4.1	3.1	2.4	2.0

The results are shown in Table 1. Before applying the algorithm the average difference between the new leg dataset ($\mathbf{b}$) and the SABRE leg dataset ($\mathbf{y^0}$) is 152.1 passengers per leg. On every iteration new itinerary ($\mathbf{x^t}$) and leg ($\mathbf{y^t}$) datasets are calculated. It is possible to see that on every iteration the solution is closer to ($\mathbf{b}$), evidencing the experimental convergence of the algorithm.

Furthermore, the proposed algorithm has been tested in systems with no solutions, as evidenced by the removal of specific itineraries from the problem by giving them an initial value of 0 in $\mathbf{x^0}$) with positive results ($\mathbf{y_0}$), being this possible as the algorithm is applied independently to every itinerary. It is worth noting, however, that in such cases, the above mentioned properties have not been demonstrated, and the convergence of $\mathbf{y^t}$ toward $\mathbf{b}$ is unattainable (although it is not the convergence of $\mathbf{y^t}$ into $\mathbf{b'} \neq \mathbf{b}$), so that results are not exposed in this article.

5 Conclusion and Future Work

In this study, we present a novel approximation algorithm that is capable of effectively solving underdetermined non-negative linear systems of equations possess-

ing particular properties. The algorithm has demonstrated satisfactory performance in terms of experimental convergence, itinerary disaggregation invariance, and the production of non-negative solutions. Specifically, the proposed algorithm has been applied in the context of air transportation, yielding favorable outcomes. Despite these encouraging results, further research is warranted to identify opportunities for refinement and explore new features of the algorithm. In particular, a theoretical proof of convergence, study of the algorithm in systems with no solutions and the application of the algorithm in other contexts beyond air transportation, would be highly desirable lines for investigation.

Acknowledgements. Airbus Operations SL participates in the ERA - Environmentally Responsible Aviation project. This project has been financially supported by RED.ES (Ministry of Economic Affairs and Digital Transformation) with ID 2021/c005/00149494 through the call for grants 2021 for research and development projects in artificial intelligence and other digital technologies and their integration into value chains.

References

1. Abdelmalek, N.N.: Minimum L∞ solution of underdetermined systems of linear equations. J. Approx. Theory **20**(1), 57–69 (1977)
2. Barnhart, C., Kniker, T.S., Lohatepanont, M.: Itinerary-based airline fleet assignment. Transp. Sci. **36**(2), 199–217 (2002)
3. Bruckstein, A.M., Elad, M., Zibulevsky, M.: A non-negative and sparse enough solution of an underdetermined linear system of equations is unique. IEEE Trans. Inf. Theory **54**(11), 4813–4820 (2008)
4. Lai, M.J., Wang, Y.: Sparse Solutions of Underdetermined Linear Systems and Their Applications. SIAM (2021)
5. Landmann, S., Engel, A.: On non-negative solutions to large systems of random linear equations. Phys. A **552**, 122544 (2020)
6. Madych, W.: Solutions of underdetermined systems of linear equations. Lect. Notes-Monograph Ser. 227–238 (1991)
7. Phillips, R.L., Boyd, D.W., Grossman, T.A., Jr.: An algorithm for calculating consistent itinerary flows. Transp. Sci. **25**(3), 225–239 (1991)
8. Saad, Y., Van Der Vorst, H.A.: Iterative solution of linear systems in the 20th century. J. Comput. Appl. Math. **123**(1–2), 1–33 (2000)
9. Smith, C.B., Hernandez, E.M.: Non-negative constrained inverse eigenvalue problems–application to damage identification. Mech. Syst. Signal Process. **129**, 629–644 (2019)
10. Smith, C.B., Hernandez, E.M.: Non-negative and sparsity constrained inverse problems in damage identification-application to a full-scale 3D truss. Mech. Syst. Signal Process. **140**, 106648 (2020)
11. Stiros, S.C., Saltogianni, V.: Solution of underdetermined systems of equations with gridded a priori constraints. Springerplus **3**(1), 1–15 (2014). https://doi.org/10.1186/2193-1801-3-145
12. Tsiligianni, E., Kondi, L.P., Katsaggelos, A.K.: Preconditioning for underdetermined linear systems with sparse solutions. IEEE Signal Process. Lett. **22**(9), 1239–1243 (2015)

Ensuring Data Privacy in Open Data Platforms for Smart Cities

Juan Antonio González-Ramos[1], Francisco Pinto-Santos[2(✉)], Alesandro Gómez-Villar[2], Javier Curto[2], and Javier Parra-Domínguez[3]

[1] Servicios Informáticos, Universidad de Salamanca, Salamanca, Spain
juanan@usal.es
[2] Air Institute, IoT Digital Innovation Hub (Spain), 37188 Carbajosa de la Sagrada, Salamanca, Spain
{franpintosantos,agomez,jcurto}@air-institute.com
[3] BISITE Research Group, Universidad de Salamanca, Salamanca, Spain
javierparra@usal.es

Abstract. Smart cities generate vast amounts of data from various sources, such as IoT devices, sensors, and monitoring systems. Open data platforms provide an opportunity for governments, researchers, and private organizations to access and utilize this data for various purposes. However, ensuring data privacy is a critical concern in the context of smart cities and open data platforms. This article investigates the challenges and privacy issues related to open data platforms in Spain, reviews existing platforms that can be used for smart city data, and proposes an adaptation of one such platform to guarantee data security and privacy. The results demonstrate the feasibility of the proposed approach and its potential to enhance data privacy in open data platforms.

Keywords: Data Privacy · Open Data · Smart Cities

1 Introduction

Smart cities represent a vision for urban development that seeks to improve the quality of life for citizens through the efficient use of resources, data-driven decision-making, and technology integration. One of the critical aspects of smart cities is the vast amount of data generated from various sources, such as Internet of Things (IoT) devices [1], sensors, and monitoring systems. These data sources provide essential information that can be used to optimize urban services, enhance public safety, and foster sustainable development.

Open data platforms have emerged as a means for governments, researchers, and private organizations to access and utilize this data for various purposes. By making data publicly available, open data platforms aim to promote transparency, encourage innovation, and facilitate collaboration between different stakeholders. Examples of applications that can benefit from open data platforms include urban planning, traffic management, environmental monitoring, and public service delivery.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 33–42, 2023.
https://doi.org/10.1007/978-3-031-43461-7_4

However, ensuring data privacy is a critical concern in the context of smart cities and open data platforms [2]. With the increasing volume and diversity of data being collected, the risk of privacy breaches and unauthorized access to sensitive information becomes more significant. This concern is particularly relevant in countries with strict data protection regulations, such as Spain, which must comply with the European Union's General Data Protection Regulation (GDPR). The GDPR imposes strict requirements on organizations that process personal data, including the need to obtain consent from individuals, provide transparent information about data processing activities, and implement appropriate security measures to protect data from unauthorized access and breaches.

In this context, it is crucial to analyze the challenges and privacy issues related to open data platforms in Spain and propose strategies to ensure data security and privacy while maintaining the benefits of open data sharing. This article aims to address this challenge by reviewing existing open data platforms that can be used for smart city data, identifying the privacy issues and potential risks associated with these platforms, and proposing an adaptation of one such platform to guarantee data security and privacy.

The first part of the introduction provides an overview of the importance of data in smart cities and the role of open data platforms in facilitating data sharing and collaboration between different stakeholders. It highlights the potential benefits of open data platforms, such as promoting transparency, encouraging innovation, and enabling data-driven decision-making. The second part of the introduction focuses on the challenges and privacy issues related to open data platforms, with a particular emphasis on the legal and regulatory context in Spain. It discusses the GDPR requirements and their implications for organizations that process personal data in the context of smart cities and open data platforms.

The third part of the introduction presents the objectives of this article, which are to analyze the privacy issues related to open data platforms in Spain, review existing platforms that can be used for smart city data, and propose an adaptation of one such platform to guarantee data security and privacy. The article aims to contribute to the ongoing debate on data privacy in smart cities and provide practical guidance for policymakers, urban planners, and technology developers who seek to leverage open data platforms for sustainable urban development.

The remainder of this article is structured as follows: Sect. 2 provides background information on data privacy in smart cities and open data platforms, including a discussion of the challenges and privacy issues associated with these platforms in Spain. Section 3 presents the proposed approach for adapting an existing open data platform to ensure data security and privacy, including a review of available platforms and a description of the chosen platform and its adaptation. Section 4 discusses the results of the proposed approach, including an evaluation of the adapted platform's performance in terms of data privacy and a comparison with other open data platforms. Section 5 concludes the article

and suggests future research directions in the area of data privacy and open data platforms for smart cities.

2 Background

2.1 Data Privacy in Smart Cities

Smart cities leverage a wide array of technologies and data sources to improve urban services, enhance public safety, and foster sustainable development. IoT devices, sensors, and monitoring systems play a crucial role in generating valuable data that can be utilized to make informed decisions and optimize various aspects of city life [5]. However, the collection and processing of such data also raise concerns about data privacy and security, as it may include sensitive information about individuals, their activities, and their whereabouts [4].

Data privacy is particularly important in the context of smart cities because the data being collected often pertains to individuals' everyday lives, which can reveal personal habits, preferences, and behaviors [3,7,16]. In addition, the sheer volume and diversity of data being collected in smart cities increase the risk of privacy breaches, unauthorized access, and data misuse. As a result, ensuring data privacy and security is a critical concern for governments, city planners, and technology developers involved in smart city initiatives [6].

One approach to addressing these privacy concerns is to adopt privacy-by-design principles, which emphasize the need to incorporate privacy considerations throughout the entire data lifecycle, from collection to storage, processing, and sharing. Privacy-by-design principles advocate for data minimization, anonymization, and encryption as ways to mitigate the risks associated with data processing and sharing in smart cities [11].

2.2 Open Data Platforms

Open data platforms are repositories that enable the sharing and dissemination of data among various stakeholders, such as governments, researchers, private organizations, and the general public [15]. By making data publicly available, open data platforms aim to promote transparency, encourage innovation, and facilitate collaboration between different stakeholders. The use of open data platforms can lead to the development of new applications and services, enhance decision-making processes, and ultimately contribute to the improvement of urban life [8].

However, as mentioned earlier, the sharing of data through open data platforms also raises privacy concerns, especially when the data contains sensitive information about individuals [13,14]. Ensuring data privacy and security becomes a critical concern when managing open data platforms, as these platforms must strike a balance between promoting data sharing and protecting the privacy of the individuals whose data is being shared.

2.3 Privacy Issues and Challenges in Spain

Spain, as a member of the European Union, is subject to the GDPR, which imposes strict requirements on organizations that process personal data. The GDPR aims to harmonize data protection laws across the EU and provide individuals with greater control over their personal data. Key principles of the GDPR include obtaining consent from individuals before processing their data, providing transparent information about data processing activities, and implementing appropriate security measures to protect data from unauthorized access and breaches.

In the context of smart cities and open data platforms, complying with the GDPR poses several challenges. First, identifying personal data and obtaining consent from individuals can be difficult when dealing with large volumes of data from various sources. Second, ensuring transparency and providing clear information about data processing activities can be challenging, given the complexity and diversity of data processing activities in smart cities. Third, implementing appropriate security measures to protect data from unauthorized access and breaches requires continuous monitoring and updating of security protocols and technologies.

In addition to the GDPR, Spain has its own national data protection laws and regulations, such as the Organic Law 3/2018 on the Protection of Personal Data and the Guarantee of Digital Rights (LOPDGDD). The LOPDGDD complements the GDPR and provides additional guidance on specific aspects of data protection, such as the rights of individuals and the obligations of data controllers and processors. Ensuring compliance with both the GDPR and the LOPDGDD is critical for organizations involved in smart city initiatives and open data platforms in Spain.

Given the complex legal and regulatory landscape, it is essential to analyze the privacy issues and challenges associated with open data platforms in Spain and propose strategies to ensure data security and privacy while maintaining the benefits of open data sharing. This analysis requires a thorough understanding of the GDPR and LOPDGDD requirements, as well as a comprehensive examination of the privacy risks and potential threats related to open data platforms in the context of smart cities.

2.4 Review of Existing Open Data Platforms

There are several open data platforms available that can be used to share and disseminate smart city data. Some of these platforms are developed and maintained by governments or public organizations, while others are provided by private companies or research institutions. Examples of open data platforms include CKAN (Comprehensive Knowledge Archive Network), Socrata, and Data.World, among others.

CKAN is an open-source data management system that allows users to publish, share, find, and use datasets. It is widely used by governments and public organizations worldwide to create open data portals. CKAN provides a range of

features to ensure data privacy and security, such as data anonymization, access control, and encryption.

Socrata is a cloud-based open data platform designed for governments and public organizations. It offers a suite of tools and services to help users publish, manage, and analyze data, as well as create interactive visualizations and dashboards. Socrata incorporates privacy-by-design principles and provides features to ensure data privacy and security, such as role-based access control, data redaction, and secure data sharing.

Data.World is a collaborative data platform that allows users to share, discover, and work with data from various sources. It provides tools for data management, analysis, and visualization, as well as features to ensure data privacy and security, such as data access controls, data anonymization, and data encryption.

In addition to these platforms, there are other open data platforms that can be used for smart city data [9,12], each with its own set of features, capabilities, and limitations. The choice of an appropriate open data platform depends on factors such as the specific requirements of the smart city initiative, the types of data being collected and shared, and the legal and regulatory context in which the platform will be used [10].

2.5 Summary

The background section has provided an overview of the importance of data privacy in smart cities and the challenges associated with ensuring data privacy and security in the context of open data platforms. It has also discussed the legal and regulatory landscape in Spain, including the GDPR and LOPDGDD requirements, and their implications for organizations involved in smart city initiatives and open data platforms. Finally, the section has presented a review of existing open data platforms that can be used for smart city data, highlighting their features, capabilities, and limitations. The next section will propose an adaptation of one such platform to guarantee data security and privacy in the context of smart cities and open data platforms in Spain.

3 Proposal

3.1 Enhancing CKAN for Data Privacy and Security

To address the data privacy and security concerns in smart cities and open data platforms, we propose the development of a CKAN extension that can identify vulnerabilities and assess the risk of these vulnerabilities through a Key Performance Indicator (KPI). This extension aims to enhance the existing features of CKAN, ensuring that data privacy and security are maintained in accordance with the GDPR and LOPDGDD requirements.

3.2 CKAN Extension for Data Privacy and Security Assessment

The proposed CKAN extension will be designed to perform various checks on the data stored in the platform, as well as to identify potential vulnerabilities and risks associated with data privacy and security. The extension will incorporate the following features:

1. Data Anonymization: The extension will automatically detect and anonymize sensitive personal information in datasets, thereby reducing the risk of data breaches and ensuring compliance with data protection regulations.
2. Vulnerability Detection: The extension will analyze the datasets and metadata stored in the platform, as well as the CKAN instance configuration, to identify potential vulnerabilities and weaknesses that could compromise data privacy and security.
3. Risk Assessment: The extension will calculate a KPI for each detected vulnerability, which will indicate the severity of the vulnerability and the potential impact on data privacy and security. This KPI will help prioritize efforts to address the identified vulnerabilities and implement appropriate mitigation measures.
4. Reporting and Visualization: The extension will generate reports and visualizations of the identified vulnerabilities, their associated KPIs, and the recommended mitigation measures. These reports and visualizations will be easily accessible to platform administrators, enabling them to monitor the platform's data privacy and security status and take action as needed.

3.3 Developing a CKAN Extension

CKAN extensions are designed to be easily integrated into any CKAN instance, enabling the addition of new features, customizations, and improvements to the platform. Developing a CKAN extension involves the following technical steps:

1. Setting up the development environment: To create a CKAN extension, developers need to set up a local CKAN instance and install the necessary development tools, such as Python, pip, and virtualenv.
2. Creating the extension structure: Developers can use the CKAN Extension Toolkit (ckanext) to generate the basic structure of the extension, which includes the configuration files, templates, and static files required for the extension.
3. Implementing the extension functionality: Developers can write custom Python code to implement the desired features and functionality of the extension, such as data anonymization, vulnerability detection, and risk assessment.
4. Registering the extension: Once the extension has been developed and tested, developers need to register the extension with CKAN by updating the CKAN configuration file (ckan.ini) to include the extension's entry points.
5. Deploying the extension: The extension can be installed and activated on a CKAN instance by using the appropriate package manager (e.g., pip) and updating the CKAN configuration file to enable the extension.

3.4 Ease of Integration with CKAN Open Data Portals

The proposed CKAN extension for data privacy and security assessment is designed to be easily integrated into any CKAN-based open data portal. By leveraging the extensible architecture of CKAN, the extension can be seamlessly added to existing CKAN instances without requiring significant modifications to the platform's core functionality. This ease of integration ensures that the extension can be adopted by a wide range of open data portals, helping to enhance data privacy and security across various smart city initiatives and open data platforms in Spain and beyond.

4 Results

4.1 Successful Integration of the CKAN Extension

The CKAN extension for data privacy and security assessment has been successfully integrated into a CKAN-based open data portal. The integration process was straightforward and did not require any significant changes to the core functionality of the platform. The extension was installed and activated using the appropriate package manager and by updating the CKAN configuration file to enable the extension. The ease of integration demonstrates the extensibility and flexibility of the CKAN platform in accommodating new features and customizations.

4.2 Detection of Vulnerabilities and Risks

Upon integration, the extension performed various checks on the datasets stored in the platform and analyzed the CKAN instance configuration. The extension successfully detected a range of potential vulnerabilities and weaknesses that could compromise data privacy and security. Some of the detected vulnerabilities include:

- Inadequate data anonymization: The extension identified datasets containing sensitive personal information that had not been sufficiently anonymized, which posed a risk of data breaches and non-compliance with data protection regulations.
- Insecure data sharing: The extension detected instances of data sharing with external parties that lacked proper access controls and encryption, potentially exposing sensitive information to unauthorized users.
- Weak password policies: The extension discovered that some user accounts had weak password policies, increasing the likelihood of unauthorized access and data breaches.
- Outdated software components: The extension identified outdated software components and libraries used by the CKAN instance, which could contain known security vulnerabilities and pose a risk to the platform's overall security.

4.3 Risk Assessment and KPI Calculation

For each detected vulnerability, the extension calculated a KPI indicating the severity of the vulnerability and the potential impact on data privacy and security. The KPIs helped prioritize efforts to address the identified vulnerabilities and implement appropriate mitigation measures. Some examples of KPIs calculated by the extension include:

- High KPI for inadequate data anonymization: This KPI indicated a significant risk of data breaches and non-compliance with data protection regulations, necessitating immediate action to properly anonymize the affected datasets.
- Medium KPI for insecure data sharing: This KPI suggested a moderate risk of unauthorized data access and potential data breaches, warranting the implementation of enhanced access controls and encryption measures.
- Low KPI for weak password policies: This KPI pointed to a relatively low risk of unauthorized access, but still highlighted the need to enforce stronger password policies and user authentication mechanisms.
- Medium KPI for outdated software components: This KPI indicated a moderate risk of security vulnerabilities and potential data breaches, prompting the need to update the affected software components and libraries.

4.4 Reporting and Visualization

The extension generated comprehensive reports and visualizations of the identified vulnerabilities, their associated KPIs, and the recommended mitigation measures. Platform administrators were able to access these reports and visualizations through an intuitive dashboard, enabling them to monitor the platform's data privacy and security status and take action as needed. This streamlined reporting and visualization functionality facilitated effective decision-making and the efficient allocation of resources to address the identified vulnerabilities and enhance the overall data privacy and security of the CKAN-based open data portal.

5 Conclusion and Future Work

In this article, we have presented a comprehensive analysis of data privacy and security challenges in smart cities and open data platforms, with a focus on Spain. We have reviewed existing open data platforms and highlighted the importance of ensuring data privacy and security in accordance with the GDPR and LOPDGDD requirements. To address these challenges, we proposed the development of a CKAN extension that can identify vulnerabilities and assess the risk of these vulnerabilities through a Key Performance Indicator (KPI). Our results demonstrate the successful integration of the extension into a CKAN-based open data portal and its effectiveness in detecting various vulnerabilities and risks associated with data privacy and security.

The proposed CKAN extension offers a valuable tool for enhancing data privacy and security in smart city initiatives and open data platforms. By enabling the seamless integration of advanced data privacy and security assessment features, the extension can help ensure compliance with data protection regulations and reduce the risk of data breaches. Moreover, the extension's reporting and visualization capabilities provide platform administrators with valuable insights and actionable information to address identified vulnerabilities and enhance the overall data privacy and security of their platforms.

5.1 Future Work

While our proposed CKAN extension has demonstrated promising results, there are several avenues for future work and improvement. Some potential directions include:

1. Expanding the scope of vulnerability detection: Our current implementation focuses on a limited set of vulnerabilities and risks. Future work could expand the range of vulnerabilities detected by the extension to cover a broader spectrum of data privacy and security threats.
2. Incorporating machine learning and AI techniques: Advanced machine learning and AI algorithms could be integrated into the extension to improve its accuracy and effectiveness in detecting vulnerabilities and assessing risks.
3. Enhancing the KPI calculation: The current KPI calculation methodology could be refined and extended to better represent the severity and impact of identified vulnerabilities, as well as to incorporate additional contextual information and factors.
4. Customization and adaptability: Future work could focus on making the extension more customizable and adaptable to the specific needs and requirements of different smart city initiatives and open data platforms, allowing users to fine-tune the extension's features and functionality according to their unique contexts.
5. Integration with other data protection and security tools: The extension could be integrated with other data protection and security solutions, such as intrusion detection systems, data loss prevention tools, and encryption technologies, to provide a more comprehensive and holistic approach to data privacy and security in smart cities and open data platforms.

By exploring these potential directions, we believe that the proposed CKAN extension can be further enhanced and refined, ultimately contributing to a more secure and privacy-conscious smart city ecosystem and open data landscape.

Acknowledgments. This work has been partially supported by the Institute for Business Competitiveness of Castilla y León, and the European Regional Development Fund under grant CCTT3/20/SA/0002 (AIR-SCity project).

References

1. Bellini, P., Nesi, P., Pantaleo, G.: IoT-enabled smart cities: a review of concepts, frameworks and key technologies. Appl. Sci. **12**(3), 1607 (2022)
2. Bokhari, S.A.A., Myeong, S.: Use of artificial intelligence in smart cities for smart decision-making: a social innovation perspective. Sustainability **14**(2), 620 (2022)
3. Canizes, B., Pinto, T., Soares, J., Vale, Z., Chamoso, P., Santos, D.: Smart city: a GECAD-BISITE energy management case study. In: De la Prieta, F., et al. (eds.) PAAMS 2017. AISC, vol. 619, pp. 92–100. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61578-3_9
4. Casado-Vara, R., Martin-del Rey, A., Affes, S., Prieto, J., Corchado, J.M.: IoT network slicing on virtual layers of homogeneous data for improved algorithm operation in smart buildings. Futur. Gener. Comput. Syst. **102**, 965–977 (2020)
5. Casado-Vara, R., Prieto-Castrillo, F., Corchado, J.M.: A game theory approach for cooperative control to improve data quality and false data detection in WSN. Int. J. Robust Nonlinear Control **28**(16), 5087–5102 (2018)
6. Chamoso, P., González-Briones, A., Rodríguez, S., Corchado, J.M.: Tendencies of technologies and platforms in smart cities: a state-of-the-art review. Wirel. Commun. Mob. Comput. **2018** (2018)
7. Chamoso, P., González-Briones, A., Rivas, A., De La Prieta, F., Corchado, J.M.: Social computing in currency exchange. Knowl. Inf. Syst. **61**(2), 733–753 (2019). https://doi.org/10.1007/s10115-018-1289-4
8. Corchado, J.M., et al.: Deepint. net: a rapid deployment platform for smart territories. Sensors **21**(1), 236 (2021)
9. Corchado, J.M., Pinto-Santos, F., Aghmou, O., Trabelsi, S.: Intelligent development of smart cities: deepint.net case studies. In: Corchado, J.M., Trabelsi, S. (eds.) SSCTIC 2021. LNNS, vol. 253, pp. 211–225. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-78901-5_19
10. Dash, B., Sharma, P.: Role of artificial intelligence in smart cities for information gathering and dissemination (a review). Acad. J. Res. Sci. Publ. **4**(39) (2022)
11. García-García, L., Jiménez, J.M., Abdullah, M.T.A., Lloret, J.: Wireless technologies for IoT in smart cities. Netw. Protocols Algorithms **10**(1), 23–64 (2018)
12. Garcia-Retuerta, D., Chamoso, P., Hernández, G., Guzmán, A.S.R., Yigitcanlar, T., Corchado, J.M.: An efficient management platform for developing smart cities: solution for real-time and future crowd detection. Electronics **10**(7), 765 (2021)
13. González-Briones, A., Chamoso, P., de Alba, F.L., Corchado, J.M.: Smart cities energy trading platform based on a multi-agent approach. In: Ong, H.L., Doong, Ra., Naguib, R., Lim, C.P., Nagar, A.K. (eds.) Artificial Intelligence and Environmental Sustainability. Algorithms for Intelligent Systems, pp. 131–146. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-1434-8_7
14. Martí, P., Jordán, J., Chamoso, P., Julian, V.: Taxi services and the carsharing alternative: a case study of Valencia City (2022)
15. Rivas, A., Fraile, J.M., Chamoso, P., González-Briones, A., Sittón, I., Corchado, J.M.: A predictive maintenance model using recurrent neural networks. In: Martínez Álvarez, F., Troncoso Lora, A., Sáez Muñoz, J.A., Quintián, H., Corchado, E. (eds.) SOCO 2019. AISC, vol. 950, pp. 261–270. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-20055-8_25
16. Shoeibi, N., Shoeibi, N., Julian, V., Ossowski, S., Arrieta, A.G., Chamoso, P.: Smart cyber victimization discovery on Twitter. In: Corchado, J.M., Trabelsi, S. (eds.) SSCTIC 2021. LNNS, vol. 253, pp. 289–299. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-78901-5_25

Intelligent Heart Rate Monitoring from a Social Robot

Arturo Martinez-Rodrigo[1(✉)], José Carlos Castillo[2], Sara Carrasco-Martínez[2], Daniele Padovano[1], and Angelo Costa[3]

[1] University of Castilla-La Mancha, Cuenca, Spain
arturo.martinez@uclm.es
[2] University Carlos III, Madrid, Spain
jocastil@ing.uc3m.es, sacarras@pa.uc3m.es
[3] University of Alicante, San Vicente del Raspeig, Spain
angelogoncalo.costa@gcloud.ua.es

Abstract. The ageing of the population has increased the need for remote monitoring and support for older adults, especially those who live alone and may experience stress, fear, depression, and isolation. In this work, the Mini social robot implements an architecture designed to help the elderly live independently for longer. To achieve this, the robot uses information from a wearable device to monitor the heart rate of older adults in real-time using photoplethysmography signals. These signals are decoded, parsed, processed, and classified to identify potential abnormalities and respond accordingly. In the case of abnormally high or low pulse values, the robot interacts with the user to check for symptoms by asking a series of questions. If the user responds positively to any of the questions or does not respond at all, the robot alerts the contact persons via Telegram, a widely used and user-friendly messaging application.

Keywords: Social robotics · photoplethysmography · long-term monitoring

1 Introduction

The phenomenon of population ageing in developed countries is well-known within the scientific community. By 2030, one in six people worldwide is estimated to be 60 years of age or older [1]. Increased life expectancy presents new opportunities to contribute to society, but the extent of these contributions heavily relies on one aspect: health. Furthermore, people generally prefer to continue living in their homes as long as possible, which offers several benefits for them, such as proximity to family and friends, increased safety, and maintaining independence. However, prolonged home living also raises the issue of older adults living alone around the world. In European countries, 23–35% of people aged 60 or older live alone [2]. This situation can be considered as a social vulnerability, with studies revealing multiple factors associated with adverse health outcomes, such as isolation, loneliness, distress and depression [3].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 43–52, 2023.
https://doi.org/10.1007/978-3-031-43461-7_5

In these scenarios, social robots can play an important role by being equipped with monitoring, entertainment, and companionship features, allowing elderly people to live in their environments for extended periods. This has a dual aspect: on one hand, seniors can feel safer and better cared for, and on the other hand, family members can benefit from the monitoring and telepresence capabilities offered by a social robot. Moreover, continuous and non-invasive monitoring of physiological signals of older adults living alone at home can improve their quality of life, promote independent living, and contribute to better health outcomes [4]. More specifically, periodic monitoring of vital signs, such as heart rate (HR), blood pressure, or respiration, can help identify potential states of distress, fear, or health problems at an early stage [4].

Among these, HR is one of the most prominent parameters for observing cardiac activity, which is directly linked to the sympathetic nervous system. This system is primarily responsible for increasing HR in response to distress, perceived threats, and other factors. The relationship between HR and stress is well documented in the scientific literature. For example, distress has been reported to lead to a decrease in vagal tone, resulting in increased heart rate and a higher risk of cardiovascular disease [5]. Similarly, fear and distress have been confirmed to lead to increased HR due to the activation of the sympathetic nervous system [6]. Consequently, this work proposes a step toward achieving real-time pulse monitoring capabilities using a social robot acquiring information from a wearable device. The monitoring component running on the robot detects and processes photoplethysmography (PPG) signals and classifies them according to the resulting HR to identify potential abnormalities and respond accordingly. This may involve interacting with the user to verify the alarm or alerting caregivers if the user is unwell.

This paper is structured as follows. Section 2 describes how our user monitoring approach uses data from the wearable device to acquire and classify the user's pulse signals. Section 3 shows the integration within the social robot software architecture and the skill implemented for interacting with the users and warning the carers. Finally, Sect. 4 draws the main conclusion from our contribution.

2 Heart Rate Observer

The main module in charge of acquiring, preprocessing and analysing the HR signals is the *Heart Rate Observer* (HRO). This module acquires information from an external device, Polar Verity Sense (PVS), a wearable heart rate monitoring device designed to track training, control heart rate, and measure caloric expenditure [7]. The device is compact, comfortable, and non-invasive and can be placed on the arm or wrist using an adjustable elastic band (see Fig. 1(a)).

2.1 Measuring the Photoplethysmography Signals

In recent years, PPG has become increasingly popular, particularly in wearable systems, due to its convenience and affordability. It consists of a cost-

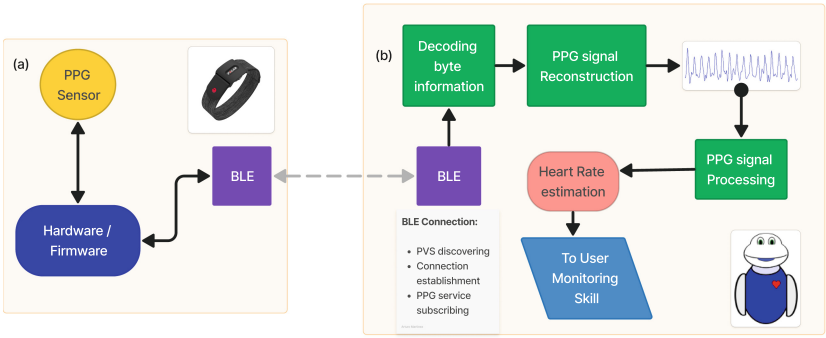


Fig. 1. Block diagram for capturing, parsing and processing PPG information.

effective, non-invasive opto-electronic method that detects blood volume changes in microvascular tissues using an infrared light sensor in contact with the skin. PPG signals, with a pulsating AC component, centred around $1Hz$, are highly correlated with cardiac activity, making them a valuable source of information about cardiovascular functioning [9].

PPG can be measured using the PVS device. It is compatible with third-party applications using the Bluetooth Low Energy (BLE) communication protocol. However, instead of utilising generic, standardised attribute profiles, PVS defines its own services and characteristics. By subscribing to these services, raw PPG signals and other variables can be obtained from PVS [7]. In this context, a third-party component was developed in this work to handle the connection and service subscription of BLE information. This information was later decoded and reconstructed to obtain the raw PPG signal, as shown in Fig. 1.

2.2 Photoplethysmography Signal Processing

Once the PPG signal is acquired and reconstructed, it must be processed to extract the HR at each time interval. PPG signals obtained from wearables can be affected by various noise issues, making it difficult to accurately interpret the data, especially in real-world situations outside of laboratory experiments. Some of these challenges include user motion, which can cause artifacts in the PPG signal, interference from electrical noise produced by other electrical fields, and ambient light that prevents sensors from accurately detecting changes in light absorption by blood vessels, among others.

To address these inconveniences, PPG data require intensive signal analysis, typically involving adaptive filtering, peak enhancement, and error detection. In this architecture, the raw PPG data obtained from the PVS (see Fig. 2(a)) was first refined with a Butterworth adaptive filter to remove the baseline, as shown in Fig. 2(b). Additionally, a band-pass filter ranging from 0.5 Hz to 5 Hz was applied to remove high-frequency noise and artefacts [8]. The PPG information was then segmented into rectangular 10-s windows and normalised to its maximum to enhance peak size. For each window, a moving average strategy with a

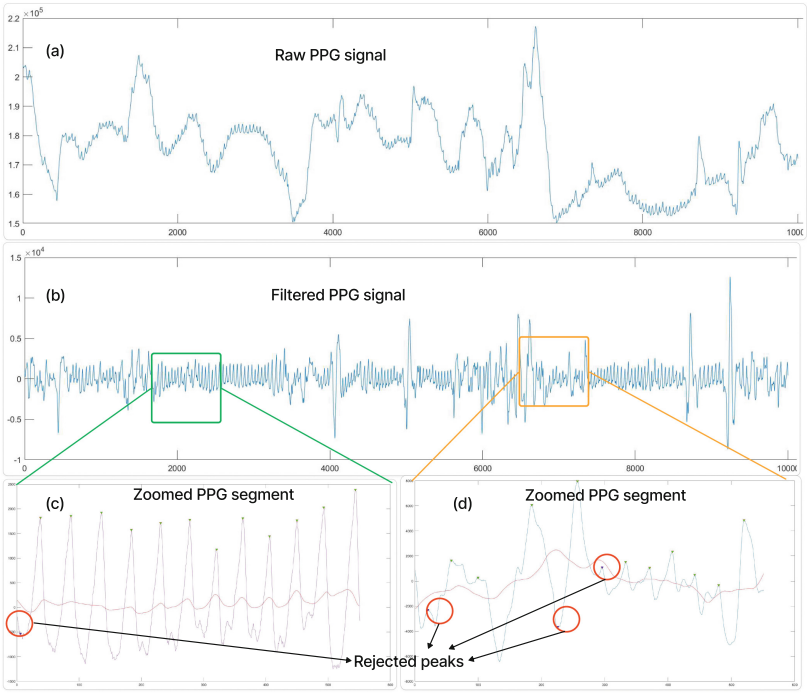


Fig. 2. Signal processor function blocks. (a) Raw PPG signal. (b) Filtered signal with baseline removed. (c) PPG segment correctly delineated. (d) PPG segment discarded.

100-sample size was computed, and peaks above that average were provisionally annotated. To reduce false positives, certain criteria were considered, such as a minimum distance between the peaks equal to half the sampling frequency and a minimum prominence of the evaluated peaks. In case those conditions are not satisfied, the proposed peaks are discarded. Figure 2(c) displays the peaks found in a random signal segment, where a correctly rejected peak can be observed. Finally, the beats per minute (BPM) are calculated as six times the sum of the non-rejected peaks for that window.

Despite filtering, it is common to encounter motion artefacts that cannot be removed. These artefacts often cause aberrant waves that result in poor detection of PPG peaks. In the proposed scheme, this results in an increase in the rejected peaks due to the high variability of the moving average. Therefore, if 30% of the peaks are rejected in an analysis window, the HR measurement for that window is discarded, as shown in Fig. 2(d).

2.3 Heart Rate Estimation

According to the American Heart Association, a typical resting heart rate range for adults is between 60 and 100 beats per minute [10]. Although this range is

just an approximation and may vary between individuals, it is usually considered a potential bradycardia situation when the heart beats less than 50 times per minute if the user is not at rest. Bradycardia can be a serious problem if the heart rate is too slow and the heart cannot pump enough oxygen-rich blood into your body. If this happens, symptoms may include dizziness, fatigue, weakness, and shortness of breath[1]. On the other hand, tachycardia or tachyarrhythmia usually corresponds to heart rates of more than 100 beats per minute without physical exercise. The general recommendation is to seek medical help if experiencing shortness of breath, weakness, dizziness, lightheadedness, fainting, and chest pain or discomfort[2].

In this initial approach, five categories are established based on heart rate to determine the input provided to the robot. When the PPG signal processing supplies valid BPM data, the heart rate estimation translates this information into a category for the robot. A heart rate below 50 BPM is deemed too low, while a range of 50 to 59 BPM is considered low. A patient with a heart rate between 60 and 100 BPM falls within the normal range. Conversely, a heart rate between 101 and 110 BPM is classified as high. Lastly, a resting heart rate exceeding 110 BPM is regarded as too high.

The HRO sends two kinds of messages, one with the current BPM (to allow continuous HR monitoring) and another one with the category. It is important to note that the HRO only informs of category changes, ensuring that if a person's state remains constant, the robot will not receive any new notifications and will assume the previous state persists. Furthermore, to prevent potential false alarms when the pulse is either too high or too low, the HRO requires three consecutive similar BPM readings from the processor before definitively relaying the message to the rest of the robot architecture.

3 Integration in a Social Robot

The proposal for long-time monitoring is integrated into the social robot Mini, developed by the RoboticsLab at Universidad Carlos III de Madrid.

3.1 Social Robot Mini

The Mini robot was designed to aid and entertain older people with moderate cognitive impairment [11]. Mini is a desktop robot that can convey its emotional state through its expressive eyes, LED heartbeat, cheeks, and moving arms, head, and base. More specifically, in terms of physical components, Mini features two OLED displays acting as its eyes, controlled with a Raspberry Pi, a VU-meter for a mouth, and LEDs in its heart and cheeks. Mini features a neck with two degrees of freedom that connects the head and body and allows the head to move

[1] https://www.mayoclinic.org/diseases-conditions/bradycardia/symptoms-causes/syc-20355474.

[2] https://www.mayoclinic.org/diseases-conditions/tachycardia/symptoms-causes/syc-20355127.

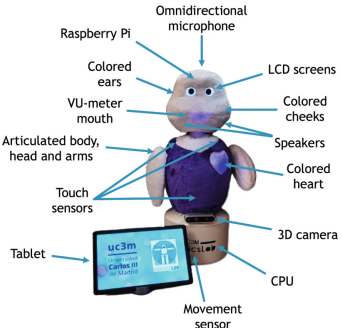


Fig. 3. Main elements of Mini social robot.

in both the vertical and horizontal axes. In addition, it includes a rotating base and two servomotor-driven arms with one degree of freedom. These components allow for natural vivacity. Moreover, Mini has capacitive sensors to recognize tactile stimuli and an external tablet to improve user engagement by showing multimedia content like video clips, photos, or custom-made game interfaces. Finally, the robot base is equipped with an RGB-D camera (see Fig. 3) and houses the main computer.

3.2 Robot's Software Architecture

The software architecture implemented for the robot Mini is composed of six main modules developed in ROS [12], each in charge of managing the principal operations of the robot (see Fig. 4). The robot's physical components are directly connected to the *Detectors* and *Expression Manager* blocks, which, respectively, collect data from the sensors and control the actuators. Among the detectors, the architecture integrates the HRO described in Sect. 2. The *Perception Manager* (PM) is in charge of gathering all the detections and merging them to provide a synchronised and unified response. As detectors implemented in the robot use different kinds of input information (e.g., touch, vision, sound, or interaction with tablet menus), their rates and type of responses tend to differ. For this reason, the work of the PM is important to ensure coherence and adequate rates for the *Human-Robot Interaction (HRI) Manager*.

The dialogues between users and robots are controlled by the HRI Manager. This component of the architecture examines perceptual data to determine whether it is pertinent to the interaction, such as when a skill requires input from the robot or when the robot is awaiting a response from the user. The robot can utilise voice communication or menus on the tablet to interact with the user; it also employs distinct motions and visual clues on the tablet to facilitate conversation. The *Decision-Making System (DMS)*, in the end, is in charge of deciding when to activate and when to deactivate a robot's functions. This part determines the robot's next move based on the surroundings

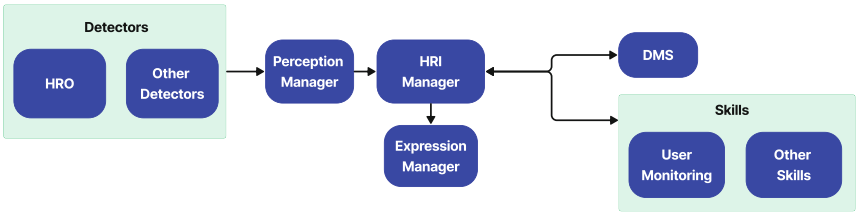


Fig. 4. Main modules of the Robot's software architecture with the user monitoring skill integrated.

Table 1. User Monitoring Skill actions depending on the pulse detected.

Pulse detected	Action	Response	Result
Normal	–	–	–
Low	Record pulse value	–	–
High	Record pulse value	–	–
Too low	Record pulse value Ask: - Are you dizzy? - Are you tired? - Do you feel short of breath? - Do you feel pain in your chest?	- Positive answer to any question - No response (possible fainting)	Warn relative-caregiver
		Negative answer to all questions	–
Too high	Record pulse value Ask: - Are you dizzy? - Do you feel short of breath? - Do you feel pain in your chest?	- Positive answer to any question - No response (possible fainting)	Warn relative-caregiver
		Negative answer to all questions	–

and internal information, such as the motivation and user requests for certain skills.

Finally, the *Skills* of the robot include all of the functionalities of the robot. These skills are grouped into categories according to their purpose (games, multimedia, information, cognitive stimulation). For example, the robot Mini integrates different games, including bingo, Akinator [13], tangram [14], and quiz games. According to the recommendations of psychologists and physicians, these games have been chosen to excite seniors. Multimedia skills use the robot's tablet to display content such as images, video clips, music, audiobooks or jokes. Through category information, the robot can tell news or weather reports [15]. In addition to these skills, the robot can perform cognitive stimulation exercises dedicated to memory, attention, planning, or comprehension, among others.

User Monitoring Skill: On the robot side, the User Monitoring Skill uses the data coming from the HRO and interprets the information to decide whether it has to interact with the user because an abnormal situation is detected or even alert the relatives or caregivers in case the situation is identified as potentially dangerous for the user.

Together with the current pulse value, the HRO also provides a classification depending on whether the pulse is considered normal, high, too high, low, or too low. To this respect, a very low detected pulse could be due to the patient suffering from bradycardia, while a very high or abnormally fast pulse could be because of some type of tachycardia or tachyarrhythmia, often related to states of restlessness or distress. With this information, the robot continuously evaluates the user pulse values, taking actions according to the classification result shown in Table 1. While the user's pulse rate remains within normal values, the robot will continue with its normal activities. If a rise or fall in the user's pulse rate is detected that is outside the normal range, though not alarming yet, the robot will store these values for later analysis. In contrast, if abnormally high or low pulse values are detected, the robot will interrupt its activity and, apart from continuing to record the abnormal pulse data, will start to interact with the user to check whether he/she is suffering from any of the usual symptoms. To do so, it will ask a series of questions, which may vary if the pulse is too high or too low. If the user responds to the robot and indicates that he/she is not suffering from any of the expected symptoms, the robot will not take any action (the pulse could be abnormally high if the user has performed physical effort or abnormally low if he/she is sleeping, for example). Alternatively, if the user responds positively to any of the questions or does not respond to the questions at all (this could be an indication of fainting), the robot will proceed to alert contact persons by Telegram as shown in Fig. 4. We selected Telegram because it allows the development of third-party applications that run within the messaging application (bot)[3]. The application is widespread and easy to use, which will reduce the need for training for carers in future tests.

Depending on how the user answers the robot's questions after detecting an abnormal pulse value, the carer receives a different type of message. Together with the warning message, two buttons appear on the screen so the carer can interact with the robot and tell it how to proceed. Pushing the "*Try to contact the user again*" button, the robot repeats the questions to the user to update the information. This option can be useful, for example, if the user did not respond previously. With the second button, "*Dismiss*", the robot will simply keep recording pulse values (Fig. 5).

[3] Telegram Bots: https://core.telegram.org/bots.



Fig. 5. Telegram screen examples of warnings to the contact person.

4 Conclusions

The ageing population has created a pressing need for innovative solutions to support and monitor older adults living alone. The Mini social robot, equipped with an integrated wearable device for real-time heart rate monitoring using PPG signals, offers a promising approach to address this challenge. The system effectively processes and classifies PPG signals to identify potential abnormalities, and the robot interacts with the users to assess their well-being based on their heart rate. By alerting caregivers or family members through a user-friendly messaging application, the robot can provide timely assistance and reassurance for older adults, ultimately improving their quality of life and enabling them to maintain their independence for longer. In future work, monitoring capabilities will be expanded, integrating additional wearable sensors for monitoring blood pressure, respiration rate, and body temperature. Furthermore, adaptive machine learning algorithms will be implemented to personalize the robot's responses and recommendations based on individual user preferences, needs, and health history. Although only the architecture is explained in this work, experiments will be performed with real older adults and caregivers in the near future.

Acknowledgements. The research leading to these results has received funding from the projects: PID2021-123941OA-I00, PID2021-128525OB-I00 and TED2021-130935B-I00 funded by Agencia Estatal de Investigación (AEI), Spanish Ministerio de Ciencia e Innovación. Also, this work was supported by Generalitat Valenciana CIGE/ 2021/136, Regional Development and "NextGenerationEU"/PRTR, SBPLY/21/ 180501/000186 from Junta de Comunidades de Castilla-La Mancha. Daniele Padovano holds a predoctoral scholarship 2022-PRED-20642.

References

1. World Health Organization. Ageing and health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed 20 Mar 2023
2. Kojima, G., Taniguchi, Y., Kitamura, A., Fujiwara, Y.: Is living alone a risk factor of frailty? A systematic review and meta-analysis. Ageing Res. Rev. **59**, 101048 (2020)
3. Xiu-Ying, H., Qian, C., Xiao-Dong, P., Xue-Mei, Z., Chang-Quan, H.: Living arrangements and risk for late life depression: a meta-analysis of published literature. Int. J. Psychiatry Med. **43**(1), 19–34 (2012)
4. Rashidi, P., Mihailidis, A.: A survey on ambient-assisted living tools for older adults. IEEE J. Biomed. Health Inform. **17**(3), 579–590 (2013). https://doi.org/10.1109/JBHI.2012.2234129
5. Thayer, J.F., Lane, R.D.: The role of vagal function in the risk for cardiovascular disease and mortality. Biol. Psychol. **74**(2), 224–242 (2007). https://doi.org/10.1016/j.biopsycho.2005.11.013
6. Shiota, M.N., Neufeld, S.L., Yeung, W.H., Moser, S.E., Perea, E.F.: Feeling good: autonomic nervous system responding in five positive emotions. Emotion **11**(6), 1368–1378 (2011). https://doi.org/10.1037/a0024278
7. Polar Electro Oy. Polar Verity Sense (n.d.). https://www.polar.com/en/products/accessories/verity_sense
8. Nilsson, L., Goscinski, T., Kalman, S., Lindberg, L.G., Johansson, A.: Filtering of non-invasive finger arterial photoplethysmogram signals. Med. Biol. Eng. Comput. **41**(2), 142–147 (2003)
9. Allen, J.: Photoplethysmography and its application in clinical physiological measurement. Physiol. Meas. **28**(3), R1–R39 (2007). https://doi.org/10.1088/0967-3334/28/3/R01
10. American Heart Association. All about heart rate (pulse) (2018). https://www.heart.org/en/health-topics/high-blood-pressure/the-facts-about-high-blood-pressure/all-about-heart-rate-pulse
11. Salichs, M.A., et al.: Mini: a new social robot for the elderly. Int. J. Soc. Robot. **12**, 1–19 (2020)
12. Quigley, M., et al.: ROS: an open-source Robot Operating System. In: ICRA Workshop on Open Source Software, vol. 3, no. 3.2, p. 5 (2009)
13. Velazquez Navarro, E., et al.: Social Robot Mini as a platform for developing multimodal interaction games. In: Medina, F.T., Garcia, O.R. (eds.) Proceedings of the Actas de las Jornadas Nacionales de Robótica. Spanish National Robotics Conference, pp. 214–220 (2019)
14. Menendez, C., Marques-Villarroya, S., Castillo, J.C., Gamboa-Montero, J.J., Salichs, M.A.: A computer vision-based system for a tangram game in a social robot. In: Novais, P., Vercelli, G., Larriba-Pey, J.L., Herrera, F., Chamoso, P. (eds.) ISAmI 2020. AISC, vol. 1239, pp. 61–71. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58356-9_7
15. González-Díaz, S., et al.: Social robot mini as information and entertainment platform. In: Medina, F.T., Garcia, O.R. (eds.) Proceedings of the Actas de las Jornadas Nacionales de Robótica. Spanish National Robotics Conference, pp. 92–97 (2019)

Review of Physical Aggression Detection Techniques in Video Using Explainable Artificial Intelligence

Pablo Negre[1(✉)], Ricardo S. Alonso[2,3], Javier Prieto[1], Angélica González Arrieta[1], and Juan M. Corchado[1]

[1] Grupo de Investigación BISITE, Universidad de Salamanca, 37006 Salamanca, Spain
{pablo.negre,javierp,angelica,corchado}@usal.es, corchado@air-institute.com

[2] AIR Insitute, Paseo de Belen 9A, 47011 Valladolid, Spain

[3] UNIR (International University of La Rioja), Av. de la Paz, 137, Logrono, Spain
ricardoserafin.alonso@unir.net

Abstract. Physical aggressions are a problem that affects our society throughout the world. This study is a review of the work done so far on the physical detection of aggressions in video using XAI techniques (XAI). In the Related Work section, other reviews have been presented that compile physical aggression detection, XAI techniques, and the combination of both: physical aggression detection using XAI techniques. The state of the art section, compiles papers on video assault detection, an explanation of XAI and the different ways of categorising and grouping existing techniques, and a review of papers dealing with video assault detection using XAI. The search methodology used consisted of using Google Scholar, reviewing articles with a publication date of no more than two years, with high quality and with a theme as close as possible to the subject matter of the article. As far as it has been possible to find, no works have been found on the detection physical aggressions or actions using XAI techniques; however, some works have been found in the context of object detection using XAI techniques. This opens the possibility of creating a research line on the detection of physical aggressions in video using XAI.

Keywords: Physical aggression detection · Explainable artificial intelligence · Computer vision

1 Introduction

Physical assaults are a serious and widespread problem in society, affecting individuals all over the world. Real-time detection of these situations can be a key factor in protecting and saving the lives of vulnerable people, thus ensuring the physical safety of the population. Supporting this assertion, according to a World

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 53–62, 2023.
https://doi.org/10.1007/978-3-031-43461-7_6

Health Organisation report, it is estimated that one in three women worldwide have been subjected to either physical and/or sexual intimate partner violence or non-partner sexual violence in their lifetime [23].

The use of security cameras has largely been used to provide evidence of a crime that can be used later to detect the culprits or serve as evidence for those affected in the eyes of insurance companies, police or judges. While it is a matter of debate, because of the balance between citizen security and individual rights, China is a country that has opted for drastic surveillance through the massive installation of security cameras; which aim, among other things, to detect and prevent crime [12]. The development of technologies in sectors such as electronics, big data, artificial intelligence (AI), IoT [19], etc., has made it possible to obtain and process a huge amount of data from the environment around us. This rise of big data and AI technologies [9] allows for real-time video analysis. This has opened up the possibility of developing security camera analysis for real-time crime detection through the use of computer vision algorithms. The development of optimal detection algorithms would allow for automatic real-time detection, which would make it possible to take immediate action such as calling the authorities [24]. All this added to the growing need for the generated algorithms to provide explainable results [8], so that the prediction is reliable, makes this a research topic of great interest.
The working methodology consisted of using Google Scholar, a well-known meta-search engine for articles. Articles with a publication date of no more than two years and whose subject matter was as similar as possible to the detection of video aggression using explainable artificial intelligence (XAI) were selected.

This work is a review of the state of art in the detection of physical aggressions in video using AI algorithms and from the approach of XAI. Section 2 and 3 present, respectively, related work and state of the art on the detection of video aggression using XAI techniques. Finally, Sect. 4 presents the conclusions as well as future lines of work.

2 Existing Reviews on the Detection of Video Aggression Using XAI Techniques

Section 2 provides an overview of related work on the detection of aggressions using computer vision techniques. It includes a presentation of previous research on aggression detection and an overview of reviews focusing on XAI. However, there is a lack of reviews discussing the detection of video aggression using XAI techniques.

There are multiple articles on the review of aggression detection techniques. A large part of them divide the detection process into three main parts: the detection method, the feature extraction and object detection, and the properties of the dataset and video used for model training [25]. Within the detection methods, methods based on surface features have to be designed manually and have higher demands on the video quality. On the other hand, deep learning based methods obtain from training with the set of videos to obtain the best

detection method [30]. Regarding the dataset used as training videos, it has been studied how the same detection method behaves with different databases of aggression videos. This is important since the methods used may vary in their detection quality depending on the type of video in question [18]. Other reviews studies have also compiled studies on aggression detection in crowded areas, where ensuring security becomes more complex [28].

There are numerous reviews that discuss explainable and trustworth AI. They are mostly based on the great development of AI in today's society and how its use has made it a concern to be able to understand its results (explainable) and to be able to trust them (trustworth) [10]. The review article [20], exposes the challenges that still exist in the field of XAI. Among these challenges are the lack of universal definitions, standards and measures for the explainability of AI systems, the lack of balance between explainability and performance, and the difficulty of making deep learning models explainable.

After a thorough search of the literature, no reviews have been found that jointly address video-based physical assault detection together with XAI techniques. Therefore, there is an interesting research opportunity to explore this area further, as both video-based assault detection and XAI are currently highly relevant topics.

3 State of the Art on the Detection of Video Aggression Using XAI Techniques

Section 3 is a review of the state of the art in aggression detection using XAI techniques. Section 3.1 collects the latest studies related to the detection of aggressions in video by using computer vision techniques. Section 3.2 outlines the need for XAI, different approaches and classifications of XAI methods. Section 3.3 consists of a compilation of recent papers focusing on object detection using XAI techniques, due to the lack of work on both aggression and action detection in general.

The methodology used consisted of the selection of quality articles, with a publication date of less than two years and with the closest proximity to the subject of this work, the detection of video aggression using XAI techniques. Google Scholar was used to select the articles.

3.1 Detection of Aggressions Using Computer Vision Techniques

The detection of aggression has been a growing research topic given its broad applications in public and industrial environments [27]. However, manual detection of violent behaviors is very labor-intensive and has a high personnel and time cost, which can result in loss of lives [16]. On the other hand, aggression detection is complex due to its ambiguity, as it is difficult to differentiate aggression from other types of actions [7], as well as factors such as changing lighting, complex background, and low resolution, among others [27]. All of this makes it difficult to describe aggression through simple rules or classifiers using trivial algorithms [7].

In [2], a traditional method has been used for the detection of aggression, which are primarily based on the analysis of visual and motion features present in the video motion features present in the video. Feature extraction is performed using the STIP and MoSIFT descriptors. A visual word vocabulary is created using the k-means algorithm, and visual word histograms are created using the feature descriptors and the created vocabulary. Finally, using an SVM classifier, events are classified as aggressive or non-aggressive. Both the use of STIP and MoSIFT had an accuracy close to 90%. The dataset used contained videos of hockey games, whose content is usually more homogeneous.

In addition to the use of traditional methods, a large number of works perform video aggression detection with methods based on deep neural networks (DNN). The multimodal function has been used for aggression detection with a focus on four multimodal fusion methods using audio, video, and text modalities, obtaining accuracies of 86%. The proposed architecture is the use of DNN (deep neural networks), which are effective not only in prediction tasks but also in feature extraction and dimensionality reduction [7]. A Multi-Scale Spatio-Temporal Network (MSTN) has also been used [31]. This consists of a spatiotemporal module that extracts features from the original video. Then, double pooling layers are used, which apply filters to the extracted features to create different scales of temporal information. Then, two modules STB (focused on object location and shape) and LTB (focused on object velocity and motion) are used to extract violence-specific features. Finally, the Trans module merges the results obtained from STS and LTB, obtaining a violence/non-violence result. Videos of actual assaults have been used, as well as videos of field hockey games and action scenes from movies. Results with an accuracy of 90.25% are obtained using the RWF-2000 dataset [3] (containing real assaults), obtaining a higher accuracy than with the use of algorithms used in other works. In [27], a convolutional neural network (CNN) model is used for pre-processing, searching for important objects such as humans and vehicles. Then it proceeds to the feature extraction phase to gather information and characteristics of the sequence. After that, features are extracted through another convolutional neural network (CNN) that is concatenated with the high-level features of the Darknet-19 (CNN) model, forming the final feature map. In the third step, an LSTM network is used that obtains sequence features and learns them for violence detection. Results with high accuracy are obtained for the three datasets used: 98.21%, 98%, and 74% on the Violent Flow, Hockey Fight, and Surveillance Fight datasets, respectively. Another study that uses convolutional neural networks (CNN) is [6]. CNNs have been compared with the models: MobileNet model (96.6%), AlexNet (88.99%), VGG (96.49%), GoogleNet (94.99%). The proposed MobileNet model has shown outstanding performance in the perspective of accuracy, loss, and computation time on the hockey fight dataset. In article [16] CNNs have also been used. The VGG-16 model (of CNN type) is used for feature extraction, the results of which feed a ConvLSTM for final classification. This architecture has obtained high accuracies in six different dataframes with values above 98% in five out of six of them.

All in all, there are several positive points in the works mentioned above. The use of feature extraction in the detection process allows relevant information to be captured. The creation of public databases containing examples of physical assaults has facilitated the training and evaluation of models. Also, the improvement in accuracy with the use of CNN has led to a clear improvement in accuracy compared to other types of algorithms. On the negative side, the wide variety of scenarios, lighting, recording quality, as well as the use of videos of non-real assaults (films, hockey games...) complicates the detection of real assaults.

3.2 Explainable Artificial Intelligence Techniques

XAI is growing rapidly due to the increasing need for connection between machines and human behaviour [21], developing interpretable algorithms that provide understandable explanations to humans of decisions made by the model [8]. Model explainability is not only targeted at AI experts but also at model end users and students [26]. In [1] and arguing against XAI, it is cited how some healthcare personnel (as end users of predictive model output) preferred factual information from trustworthy sources, rather than a complete understanding of how the information was generated as well as greater difficulty in detecting erroneous predictions. However, in [14] it is exposed how trustworthy AI requires XAI explainability, which is but a requirement to reach a high state of trustworthiness. It was found in [26] that while different explanatory models can give interpretable motivations, among them there are key differences in the results obtained by each. It was also noted how the chosen method of explainability influences the significance of the features more than the underlying model and the data. XAI methods can be classified according to several [5] factors: intrinsic or post hoc (depending on whether the model is interpretable for simplicity or not), model-specific or model-agnostic (depending on whether the XAI model is independent of the model used or not) and local or global (whether one seeks to understand the general behavior of the model or of a particular prediction). On the other hand, there are three approaches to obtain an explanatory AI using a deep neural network. The first is by visualizing or analyzing the internal behavior of DNNs (without modifying the existing structure). The second is to add an explanatory module to the DNN to analyze the input and generate an explanation of why a certain prediction has been produced. The third consists of creating a new interpretable model (from scratch) to make decisions through the explainable [22] representations.

In [17] an explainable-by-design model is proposed for object detection based on a combination of semantic and visual properties, so it would fall under the classification of intrinsic explainable model, since its results are understandable per se. In this case the XAI approach cannot be selected since the underlying model is not a DNN. Saliency maps are algorithms used to provide visual explanations for neural networks based on computer vision. These algorithms show the regions relevant to the decision that the AI model made (so they correspond to the first of the explanatory AI approaches). In [8] the Grad-CAM method and

its variants are used to produce high quality saliency maps (obtaining the best results Grad-CAM and XGrad-CAM).

In conclusion, the growing need for algorithms whose results are explainable has made XAI a growing field of study. The use of visual tools such as saliency maps to highlight important areas of an image makes it easier to understand the results. A better understanding of the outcome of the algorithms without the need for advanced technical knowledge is also an advance. On the other hand, the performance and computational complexity of XAI processes can still be an issue. While the clustering of XAI algorithms into different groups is an advance in their variety and development, there is a lack of metrics to assess the effectiveness of the applied method.

3.3 Explainable Artificial Intelligence in Aggresion Detection

Section 3.3 exposes existing articles that deal with explainable and trustworthy AI in the detection of aggressions. As will be shown below, as far as we know, there are no papers dealing with this topic, so we will present articles on explainable and trustworthy AI in the detection of actions (as a field of study closer to the detection of aggressions).

To the best of our knowledge, there is no paper addressing the detection of physical aggression by AI techniques using video cameras, from the point of view of XAI. Since physical aggression detection using video cameras is a growing research topic, as discussed throughout this paper, a whole avenue of research remains open to address XAI techniques that ensure that the detection of aggression by the model is understandable. This would allow trustworthy detection that avoids both confusion with other contact actions (e.g., hugging) and detection by people with different physical and clothing characteristics. Since no papers were found on XAI in aggression detection, we searched for papers on XAI in action detection, which is a broader set in which aggression detection is found. However, no papers could be found on XAI applied to action detection either. All in all, it was decided to look for papers on XAI in object detection, as this is a much more studied field within computer vision. In [15] the detection of autonomous vehicles has been studied using the "KITTI road dataset", using as detection algorithms: ResNet-18, ResNet-50 and SegNet; and as XAI methods Grad-CAM, Saliency maps and Intermediate DL model layer analysis. While the use of various XAI techniques provides greater clarity on the 'black box' operation of the model, the explainability of the model is not fully resolved; it is proposed to integrate NLP into the XAI system to provide explanations through the use of text. In [13], 10 different object classes of the autonomous driving environment are detected using the BDD100K dataset, the YOLOv4 model was used as the detection algorithm and RISE (which generates saliency maps) was used as the XAI tool; it is argued that the distance between vehicles and other objects should be taken into account in future work for correct detection. In [29], the MSCOCO and Flickr30K datasets (containing images from multiple different categories) are used and a CNN is used for object detection together with an RNN that associates the detected object with a word describing it,

making the model intrinsically explainable. In [11] multiple datasets have been used: COCO, KITTI, BDD, and OpenImages; YOLOv3 has been used as object detection algorithm and Grad-CAM has been integrated into the model. In [4] L-CRP, an XAI algorithm based on heat maps for the most used state-of-the-art object detection algorithms (UNet, DeepLab & Yolo), has been created using multiple public datasets such as: CityScapes, Pascal VOC 2012 and MS COCO 2017; demonstrating the fidelity of L-CRP.

In summary, as far as is known, there is no work on the detection of aggressions or actions using XAI algorithms. As far as object detection is concerned, the works carried out share the datasets used and widely used XAI algorithms, obtaining good results and facilitating comparison between them (Table 1).

Table 1. Summary table of the state of the art of object detection using XAI techniques.

Cite	Detection category	Dataset	Detection algorithms	XAI algorithms
[15]	Vehicle detection	KITTI road	ResNet-18, ResNet-50 and SegNet	Grad-CAM, Saliency maps and DL model middle layer analysis
[13]	Vehicle detection	BDD100K	YOLOv4	RISE (saliency maps)
[29]	Multiple categories	MSCOCO and Flickr30K	CNN + RNN	Intrinsically explainable
[11]	Multiple categories	COCO, KITTI, BDD and OpenImages	YOLOv3	Grad-CAM integrado en el modelo
[4]	Multiple categories	CityScapes, Pascal VOC 2012 and MS COCO 2017	UNet, DeepLab and Yolo	L-CRP

4 Conclusions and Future Work

Physical assaults are a serious and widespread reality in our society around the world. Real-time detection can be a key factor in ensuring the physical safety of the population in dangerous situations.

With the development of technology, it is possible to process information on a large scale, and for this to be done in real time. This opens up the possibility that physical aggressions can be detected in real time [24] with an algorithm that can send an alarm to the competent authorities. In addition to the great potential this offers, there is also a growing need for AI algorithms to be explainable for developers and users [8], making the combination of real-time physical assault detection and the use of AI a research topic of great interest.

This work is a review of the state of art in the detection of physical aggressions in video using AI algorithms and from the approach of XAI. Section 2 presents

related work on video assault detection using XAI, highlighting the importance of the chosen dataset on the accuracy of the algorithm, the complexity of detection in crowded areas and the need for greater understanding of the algorithm results.

Section 3 first discusses on Sect. 3.1 the state of the art on the detection of physical aggression in video using AI. Some positive points mentioned about the studied works are: the better results with the use of feature extraction, the creation of public databases tagged with aggression/non-aggression videos and the improvement in accuracy that the use of CNN has brought about. As negative points, the variety in video features and the use of non-real assault videos (movies, hockey games...), which make it difficult to detect real assaults. Section 3.2 presents work on XAI techniques, which is a growing field of study. Positive points to highlight in the works studied are the use of visual tools for the interpretation of the result, as well as the non-necessity of being a technical person to understand the result. On the downside, performance and computational complexity remain issues. In Sect. 3.3, to the best of our knowledge there are no works that bring together these fields of research, where the physical detection of aggressions through video by means of XAI is dealt with. While there are a number of papers that address XAI in object detection [11,15], where positive results have been presented with widely used XAI algorithms.

As future lines of research and as a continuation of this first review of the state of the art on the detection of physical aggressions in video by using XAI techniques, a systematic study will be carried out that will be relevant to structure the wide variety of works in which different algorithms and databases have been used, both in the detection of physical aggressions by using video and XAI techniques. Subsequently, it is proposed that, based on the information gathered in this article and this systematic study, algorithms for the detection of physical aggression using video XAI techniques, whose accuracy is at least as high as that of current algorithms, should be developed.

Acknowledgements. This work was supported by the project "XAI - Sistemas Inteligentes Auto Explicativos creados con Módulos de Mezcla de Expertos", ID SA082P20: financed by Junta Castilla y León, Consejería de Educación, and FEDER fund.

References

1. Amann, J., et al.: To explain or not to explain?-artificial intelligence explainability in clinical decision support systems. PLOS Digit. Health **1**(2), e0000016 (2022)
2. Bermejo Nievas, E., Deniz Suarez, O., Bueno García, G., Sukthankar, R.: Violence detection in video using computer vision techniques. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011. LNCS, vol. 6855, pp. 332–339. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23678-5_39
3. Cheng, M., Cai, K., Li, M.: RWF-2000: an open large scale video database for violence detection. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 4183–4190. IEEE (2021)

4. Dreyer, M., Achtibat, R., Wiegand, T., Samek, W., Lapuschkin, S.: Revealing hidden context bias in segmentation and object detection through concept-specific explanations. arXiv preprint arXiv:2211.11426 (2022)
5. Hall, S.W., Sakzad, A., Choo, K.K.R.: Explainable artificial intelligence for digital forensics. Wiley Interdisc. Rev. Forensic Sci. **4**(2), e1434 (2022)
6. Hussain, T., Iqbal, A., Yang, B., Hussain, A.: Real time violence detection in surveillance videos using convolutional neural networks. Multimedia Tools Appl. **81**(26), 38151–38173 (2022)
7. Jaafar, N., Lachiri, Z.: Multimodal fusion methods with deep neural networks and meta-information for aggression detection in surveillance. Expert Syst. Appl. **211**, 118523 (2023). https://doi.org/10.1016/j.eswa.2022.118523. https://www.sciencedirect.com/science/article/pii/S0957417422016013
8. Karim, M.M., Li, Y., Qin, R.: Toward explainable artificial intelligence for early anticipation of traffic accidents. Transp. Res. Rec. **2676**(6), 743–755 (2022)
9. Katta, P., Kandasamy, K., Raj, R.S.P., Subramanian, R., Perumal, C.: Regression based performance analysis and fault detection in induction motors by using deep learning technique. ADCAIJ: Adv. Distrib. Comput. Artif. Intell. J. **11**(3), 349–365 (2023). https://doi.org/10.14201/adcaij.28435. https://revistas.usal.es/cinco/index.php/2255-2863/article/view/28435
10. Kaur, D., Uslu, S., Rittichier, K.J., Durresi, A.: Trustworthy artificial intelligence: a review. ACM Comput. Surv. (CSUR) **55**(2), 1–38 (2022)
11. Kirchknopf, A., Slijepcevic, D., Wunderlich, I., Breiter, M., Traxler, J., Zeppelzauer, M.: Explaining yolo: leveraging grad-cam to explain object detections. arXiv preprint arXiv:2211.12108 (2022)
12. Kostka, G., Steinacker, L., Meckel, M.: Between security and convenience: facial recognition technology in the eyes of citizens in china, Germany, the United Kingdom, and the United States. Public Underst. Sci. **30**(6), 671–690 (2021)
13. Li, Y., et al.: A deep learning-based hybrid framework for object detection and recognition in autonomous driving. IEEE Access **8**, 194228–194239 (2020)
14. Ma, J., et al.: Towards trustworthy AI in dentistry. J. Dent. Res. **101**(11), 1263–1268 (2022)
15. Mankodiya, H., Jadav, D., Gupta, R., Tanwar, S., Hong, W.C., Sharma, R.: OD-XAI: explainable AI-based semantic object detection for autonomous vehicles. Appl. Sci. **12**(11), 5310 (2022)
16. Mugunga, I., Dong, J., Rigall, E., Guo, S., Madessa, A.H., Nawaz, H.S.: A frame-based feature model for violence detection from surveillance cameras using convlstm network. In: 2021 6th International Conference on Image, Vision and Computing (ICIVC), pp. 55–60 (2021). https://doi.org/10.1109/ICIVC52351.2021.9526948
17. Olszewska, J.I.: Snakes in trees: an explainable artificial intelligence approach for automatic object detection and recognition. In: ICAART (3), pp. 996–1002 (2022)
18. Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., Khassanova, M.: State-of-the-art violence detection techniques in video surveillance security systems: a systematic review. PeerJ Comput. Sci. **8**, e920 (2022)
19. Qader Kheder, M., Aree Ali, M.: Iot-based vision techniques in autonomous driving: a review. ADCAIJ: Adv. Distrib. Comput. Artif. Intell. J. **11**(3), 367–394 (2023). https://revistas.usal.es/cinco/index.php/2255-2863/article/view/28821
20. Rawal, A., McCoy, J., Rawat, D.B., Sadler, B.M., Amant, R.S.: Recent advances in trustworthy explainable artificial intelligence: status, challenges, and perspectives. IEEE Trans. Artif. Intell. **3**(6), 852–866 (2021)

21. Rodríguez Oconitrillo, L.R., Vargas, J.J., Camacho, A., Burgos, Á., Corchado, J.M.: RYEL: an experimental study in the behavioral response of judges using a novel technique for acquiring higher-order thinking based on explainable artificial intelligence and case-based reasoning. Electronics **10**(12), 1500 (2021)
22. Sakai, A., et al.: Medical professional enhancement using explainable artificial intelligence in fetal cardiac ultrasound screening. Biomedicines **10**(3), 551 (2022)
23. de la Salud, O.M.: Violence against women (2021). https://www.who.int/news-room/fact-sheets/detail/violence-against-women
24. Shah, N., Bhagat, N., Shah, M.: Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Vis. Comput. Ind. Biomed. Art **4**(1), 1–14 (2021). https://doi.org/10.1186/s42492-021-00075-z
25. Siddique, M., et al.: State-of-the-art violence detection techniques: a review. Asian J. Res. Comput. Sci. 29–42 (2022)
26. Swamy, V., Radmehr, B., Krco, N., Marras, M., Käser, T.: Evaluating the explainers: black-box explainable machine learning for student success prediction in MOOCs. arXiv preprint arXiv:2207.00551 (2022)
27. Ullah, F.U.M., et al.: An intelligent system for complex violence pattern analysis and detection. Int. J. Intell. Syst. **37**(12), 10400–10422 (2022). https://doi.org/10.1002/int.22537. https://onlinelibrary.wiley.com/doi/abs/10.1002/int.22537
28. Ullah, F.U.M., Obaidat, M.S., Ullah, A., Muhammad, K., Hijji, M., Baik, S.W.: A comprehensive review on vision-based violence detection in surveillance videos. ACM Comput. Surv. **55**(10), 1–44 (2023)
29. Wu, T., Song, X.: Towards interpretable object detection by unfolding latent structures. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6033–6043 (2019)
30. Yao, H., Hu, X.: A survey of video violence detection. Cyber-Phys. Syst. **9**(1), 1–24 (2023)
31. Zhou, W., Min, X., Zhao, Y., Pang, Y., Yi, J.: A multi-scale spatio-temporal network for violence behavior detection. IEEE Tran. Biometrics Behav. Identity Sci. **5**, 266–276 (2023)

An Evaluation of the Performance of Convolution Neural Network and Transfer Learning on Face Gender Recognition

Ronke Seyi Babatunde[1], Akinbowale Nathaniel Babatunde[1], Roseline Oluwaseun Ogundokun[2,3], AbdulRahman Tosho Abdulahi[4], and Alfonso González-Briones[5(✉)]

[1] Department of Computer Science, Kwara State University, Malete, Kwara State, Nigeria
{ronke.babatunde,akinbowale.babatunde}@kwasu.edu.ng
[2] Department of Multimedia Engineering, Kaunas University of Technology, Kaunas, Lithuania
rosogu@ktu.lt, ogundokun.roseline@lmu.edu.ng
[3] Department of Computer Science, Landmark University, Omu Aran, Nigeria
[4] Department of Computer Science, Institute of Information and Communication Technology, Kwara State Polytechnic, Ilorin, Nigeria
AbdulRahman.t@kwarastatepolytechnic.edu.ng
[5] Grupo de Investigación BISITE, Departamento de Informática y Automática, Facultad de Ciencias, University of Salamanca, Instituto de Investigación Biomédica de Salamanca, Calle Espejo 2, 24.2, 37007 Salamanca, Spain
alfonsogb@usal.es

Abstract. Gender detection has become an essential component of the biometric system and protection. It is used to recommend age-appropriate content and to broaden the reach of multilayer advertising and promotions, among others. Many challenges emanate from using contemporary machine learning algorithms for face gender recognition, ranging from low detection accuracy and overfitting. Most recognition tasks, including object identification, segmentation, and classification, are based on convolutional neural networks. The human Face is a nonlinear manifold in which slight input changes can result in the invariance of the output. This prompts the need for continuous research to achieve more accurate recognition, especially with experimenting on ground truth data. This research compares the performance of Convolution Neural Network (CNN) and Transfer Learning (TL) on faces acquired in real-time scenarios to detect gender. The empirical result based on the study demonstrates that TL outperforms CNN in terms of prediction accuracy, with TL having 95.83% and CNN having 83.11%, and also in terms of recognition time, with TL having 0.23 s and CNN having 0.98 s, amongst other evaluation metrics.

This research has been supported by the project "COordinated intelligent Services for Adaptive Smart areaS (COSASS), Reference: PID2021-123673OB-C33, financed by MCIN /AEI /10.13039/501100011033/FEDER, UE.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 63–73, 2023.
https://doi.org/10.1007/978-3-031-43461-7_7

The novelty of the research is in the recognition time, which is not a metric commonly reported in the literature. The work demonstrates the effectiveness of using pre-trained models based on the dataset used and evaluation metrics for the experiment.

Keywords: Gender detection · CNN · Transfer learning · Face gender recognition

1 Introduction

Gender recognition was first sighted as a challenge in psychological research. Several academics are now focusing on human detection, gender identification, age prediction, sentiment evaluation, and race identification. Although this gender identification is a binary classification issue, it might be considered the most direct face estimation challenge. In contemporary technology, determining a person's gender by face analysis has become very important. Face recognition assessment has become one of computer vision's greatest subfields, as it plays a crucial role in social contact. Automated gender categorization has significant applications in intelligent user interface design, biometric detection, interpersonal interactions, visual monitoring, collecting demographic information for advertisement, behavior recognition, and numerous other fields. These responsibilities include characterizing age, gender, facial characteristics, gestures, and personality. As a result, much research has gone into creating automatic systems that can classify genders. The primary goal of gender recognition is to distinguish between males and females founded on various human characteristics. The two eyes, nose, and mouth are essential facial traits used to determine between genders, as identified by González-Briones et al. (2018) and Lawal et al. (2019). Although this task has been extensively studied in the past, the reported performances are far from ideal, particularly in general settings. Furthermore, the application context and training procedures significantly impact how difficult this task is Sumi et al. (2021). In-dataset inference, also known as an inference from separate datasets, can be used to train and test gender classification models. Additionally, these datasets allow for capturing facial images in both controlled and uncontrolled/unconstrained contexts, which adds to the difficulty of the endeavour. Jia et al. (2016) achieved one of the most modern methods for gender classification by employing 4 million photos to train the approach and testing it on the unrestricted Labelled Face in the wide (LFW) dataset. A deep neural network, CNN, has recently emerged as the industry standard for object recognition. A CNN is a deep learning system that distinguishes one image from another while allowing an input image to take on several objects. One of the reasons why convolutional neural networks have emerged as one of the most popular algorithms with a strong reputation in picture feature extraction is that they require significantly less pre-processing than other classification techniques. Due to their outstanding performance in facial analysis, the design models of this deep-learning neural network have been used in many detection tasks. Nearly every aspect of computer vision has improved because of

it, including the ability to recognise facial expressions, handwritten numbers, and human actions. Zaman (2020). The development and maturation of facial recognition over time have been made possible by technological progress and the quick rise in the image and video quality. Finding missing children, combating human trafficking, improving healthcare, and other areas have benefited greatly. Healthcare facilities have employed facial recognition in many ways to safeguard the health and safety of patients and staff. These include protecting priceless lab equipment and enabling touchless access so that solitary skilled and approved staff can enter private spaces in a facility, lowering the danger of contamination. However, a user-friendly mechanism and timely and easy-to-interpret interface for gender recognition is still lagging. Hence, this research aims to compare the effectiveness of the face gender recognition scheme using CNN and Transfer learning.

2 Related Work

Babatunde et al. (2016) developed a real-time automatic system recognising face gender. The face images were captured in real-time via webcam, whereby an individual is made to position themselves in front of the system to acquire a frame. A green outline on the face region indicated a tracked face in the frame. Gender recognition takes effect in nanoseconds once the Face is detected. The work observed that the image quality degraded the system's resolution and varying illumination performance. Afifi and Abdelhamed (2019) conducted comparable cross-dataset tests on face images. According to the work, the face images were collected and assembled into individual datasets in various ways, with occlusions, changes in illumination, backdrops, etc. The paper recommended that a foggy face that combines discrete facial elements with a holistic attribute will be considered instead of treating the face image as a separate feature. This will lead to better classification performance. Hence, the work trained CNN on a cloudy face and combined it with AdaBoost to determine the gender category. For gender categorisation, a combination of local traits and contextual features (the clouded Face) were derived from the images. To extract the necessary attributes for the experiment, four CNNs were used, along with an Ada-boost-based classification. To build the system, the experiment employed five (5) popular datasets: Labelled face in the wide (LFW), Face Evaluation Recognition Test (FERET), Adience dataset, and image of groups dataset. The network was tested and assessed using Specs on Faces (SoF). Their approach produced good validation accuracy. Jewel et al. (2019) applied the CNN structured method and TL to recognise the ethnicity of Bengali public and did a systematic gender categorisation. To determine which transfer learning model is more practical for achieving the requisite accuracy, the work applied numerous models, including Visual Geometry Group 16 (VGG16), Mobile net, and Resnet50. Since there are many Indians that dress and appear like Bengalis, the investigation identified various issues. As a result, the model was created using the custom of their clothes and facial images. A complex model utilizing CNN and transfer learning was developed, showing that transfer learning led to better results. Moeini and Mozaffari (2017) suggested a tech-

nique for gender categorisation based on vocabulary acquisition. The three vocabularies they utilised were a male glossary, a female glossary, and a feature glossary. The three vocabularies were learned using two vocabulary learning methods: dictionary learning gender classification (DL-GC) and separate dictionary learning gender classification (SDL-GC). Furthermore, they used three (3) publicly available datasets: FERET, LFW, and Groups Dataset. Only two features were employed to extract the essential characteristics and were removed automatically. Sparse Representation Classification was then used to categorise the gender from the estimated values. Convincing results were obtained for gender classification when evaluated on the abovementioned datasets. Comparing other cutting-edge technologies revealed an improvement over prior gender classification accuracy. Sumi et al. (2021) developed a human gender recognition system from facial images utilising CNN. Two different datasets, each made of 100 males and females, were obtained from other sources for comparison. The first source contained up to 100 females and the same for males. The image size used was 96 $\times$ 96. There are 32 filters in the convolution layer, each having a 3 $\times$ 3 kernel and the pooling layer with a 3 $\times$ 3 kernel, thereby reducing the spatial dimension from 96 $\times$ 96 to 32 $\times$ 32. The convolution and Rectified Linear Unit (ReLU) were repeated to allow richer features. The work used Adam optimiser, and the result gave a high accuracy compared with similar systems. The work also discovered that a proper dataset and optimisers could improve recognition accuracy. Sharma et al. (2020) reviewed real-time sentiment, gender, and age recognition utilising CNN. The work used the Viola-Jones algorithm, CNN, and Histogram of Oriented Gradient (HOG). The pre-processing stage enhances the quality of the input image and generated data by removing noise and minimising redundancy. The result shows that the HOG+viola jones algorithm yields a good accuracy of 70% on FER13 datasets. At the same time, the CNN model achieved 97%. Greco et al. (2021) carried out Gender recognition in the wild using a robustness assessment over tainted imageries. Cropped images were used as the network's input, and a face detector based on Single Shot Detector (SSD) and ResNet-10 was used as the backbone technique. Then the images were rescaled to adapt to the native input size. A variety of data augmentation techniques were used to replicate changes that are likely to happen in real circumstances. The dataset consists of 9131 subjects and 3.31 million imageries with variations. Different convolutional neural networks were tested, and cross-dataset evaluation was performed. The Stochastic Gradient Descent (SGD) optimiser was used, and the result showed that VGG attained an accuracy of 96.6%, which is more advanced than the other types of CNN. The work discovered that a large dataset would increase processing time, which may necessitate using high-end systems for real-time implementation. YILDIZ et al. (2021) developed a CNN-founded gender recognition in unrestrained settings. 3.3 million images of more than 9000 unique individuals make up the VGGFace2 dataset, which uses 3.1 million images as train data and 200 000 images as test data. An average of 362 images represents each individual. The images in the dataset differ from each other. Bilinear interpolation was used to resize face images to $32 \times 32 \times 3$ before pre-processing. The Adience dataset was

used as a benchmark to evaluate the results. The classification was performed using the sigmoid classifier. The VGGNet architecture was scaled down and made more manageable. The suggested network has five convolution layers, three pooling levels, and one fully linked layer. The work achieved 96.29% test accuracy and 93.71% validation accuracy. However, comparing two different datasets may lead to bias in some instances. The deductions from the literature indicate that CNN is widely used in the recognition task. However, a face image's pose, illumination and expression challenges call for continual research to achieve an improved result.

3 Methodology

3.1 Dataset Collection and Pre-processing

The dataset used in this research consists of two hundred and forty (240) faces of forty (40) students taken from six distinct camera angles, poses and illumination to ensure adequate sample variations. Additionally, 4 of each face image were used to populate the training set to avoid overfitting, while the other two were used to aggregate the testing set. The photos were labelled, resized and converted to a greyscale to provide a suitable template for the next stage. The images were first scaled down to an arbitrary size of 32 × 32. Then the images were transformed to grayscale to reduce the channel of the idea from a 3-channel Red Green Blue/Yellow Cyan Magenta (RGB/YCM) image to a 1-channel image to prevent creating excessively large images. This process also removed noise and other undesired distortion. The normalised images were then passed on to the machine learning models. The CNN and TL approaches were trained on the human faces dataset. The images were split into a train set of 70% and a test set of 30%. Hence the training set consists of 168 pictures, and the testing set consists (of 72) images. The images were trained using a 15-piece batch size and 20-iteration training epochs. The choice of batch size was because a larger batch size reduces the network's generalization ability. Also, increasing the number of iterations beyond 20 in the experiment has little or no significance on the accuracy. Following model implementation, the accuracy and loss for each period were calculated.

3.2 Classification Using CNN

The resultant face images are in the form of matrices passed onto the CNN model as input into the input layer. Most of the framework's computational effort is carried out by the convolution layer, which is the essential element of CNN. The restricted province of the approachable field and the kernel, a collection of learnable parameters, merge two matrices in this layer to create a dot product. The kernel is smaller in space but more profound when using a face image. The output volume was computed utilising the formula below for an input of size W × W × D. D is the number of kernels with a spatial dimension of F, stride S, and amount of padding P Ogundokun et al. (2022).

$$W^{out} = \left(\frac{W - F + 2P}{S}\right) + 1 \tag{1}$$

The second layer, known as the pooling layer, substitutes a summary statistic produced from nearby outputs for the network's creation in certain areas. This helps reduce the spatial size of the representation, which minimises the amount of computation and weights required. Each slice of the model is subjected to the pooling operation separately. The output volume was determined using the formula below, with a pooling kernel of spatial size F, a pooling map of size W × W × D, and stride S.

$$W^{out} = \left(\frac{W - F}{S}\right) + 1 \tag{2}$$

Since convolution is a linear operation and Face images are nonlinear, the third layer, known as the non-linearity layer, is typically added straight after the convolutional layer to add non-linearity to the activation map. ReLU is more reliable than all the activation functions, as it accelerates convergence up to six times. As in a standard Fully Connected Neural Network (FCNN), all of the neurons in the fourth layer are fully connected to all the neurons in the layers above and below. This layer is known as the fully connected layer. Therefore, it can be determined by matrix multiplication, adding a bias effect. As seen in Fig. 1, the wholly connected layer makes it easier to map the representation between the input and output.

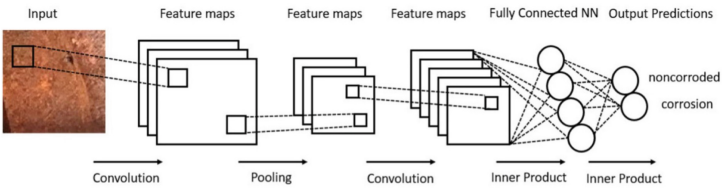


Fig. 1. The CNN architecture (Atha and Jahanshahi, 2018)

3.3 Classification Using VGG 16 TL

The facial images in this study were also exposed to the VGG16 algorithm, which uses a max pool layer of 2×2 filters with a stride of 2 and convolution layers of 3×3 filters with a stride of 1. Max pool and convolution layers are evenly spaced throughout the whole architecture. Two fully connected layers (FC) and a softmax for output are included as a conclusion. The 16 in VGG16 refers to the network's large size and the fact that it has 16 layers with weights. Figure 1 displays the architecture of the VGG16.

3.4 Testing and Validation

To measure the effectiveness of the models, after the training, the models were tested and validated by obtaining the accuracy of classification in terms of correctly and incorrectly classified images. Additional metrics of evaluation of the efficiency of the models are precision, recall, and F-measure values for CNN and TL.

3.5 Results and Discussion

The validation accuracy and loss of the CNN architecture implemented in this research work, after training the model with the acquired dataset, with an epoch of 20 and validation step of 20, based on the accuracy of the last epoch, is presented in Fig. 3.

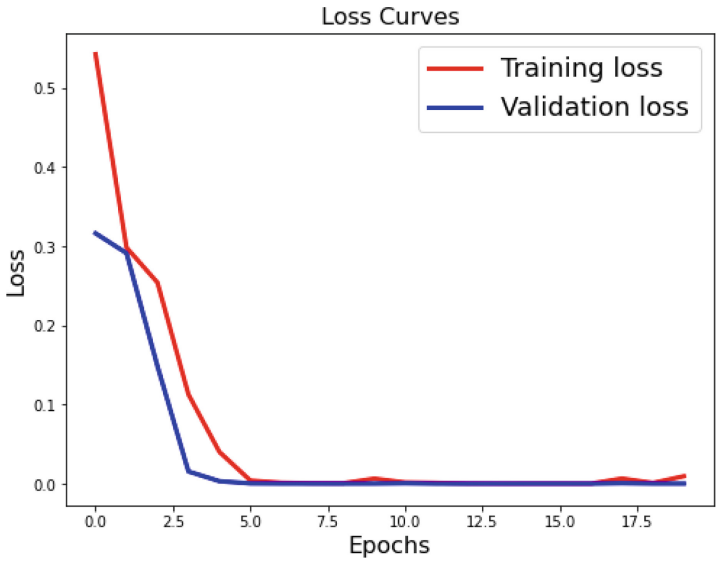


Fig. 2. The loss in training and validation

At the start of the training, there was a high training loss of over 0.5 and a validation loss of slightly over 0.3 was evident. As the epochs are further increased, the network can learn more, and the loss of the training and validation experiences declines as it gets closer to 0.

The accuracy rises to a good range near 0.75 at the beginning of training, and the validation accuracy rises to well above 0.80. As the number of epochs increases, the network experiences an increase and gets close to the range of 0.97 and above. Similarly, the TL VGG16 architecture shows the loss and accuracy curves in Fig. 2 and Fig. 3.

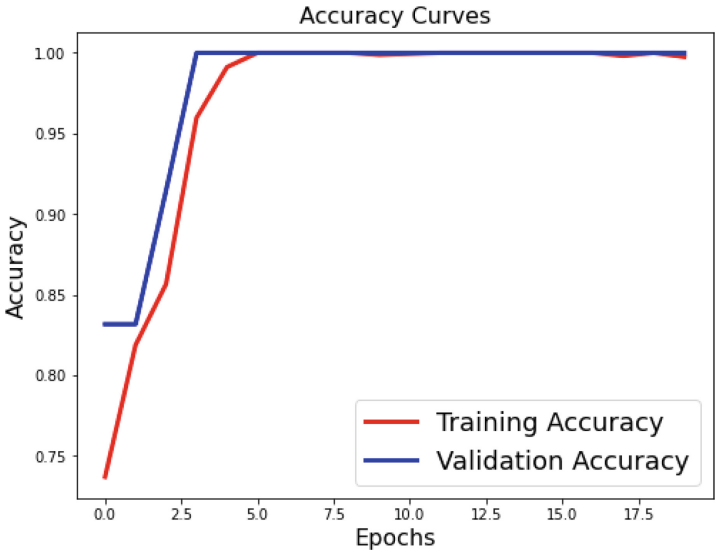


Fig. 3. The accuracy in training and validation

The training starts with a loss of up to 0.175, but there was no consistent decrease as the epoch increased. It ended up at a value slightly above 0.050. The validation loss started well and ended at the same weight as the training loss. This means that the TL model's performance will keep increasing with an increase in the number of epochs.

The validation accuracy was steady until the last epoch when the model's performance dropped as a result of the early stopping callback included to avoid overfitting. The result of implementing using the CNN, as well as TL models for the face gender recognition, is further presented in the below tables. Table 1 reveals the result of the time taken to recognise the gender of the queried faces and the recognition accuracy.

Table 1. Performance result of recognition.

Model	Recognition time (sec)	Nº. of tested samples	Correct Classification	Incorrect Classification	Accuracy (%)
CNN	0.98	72	62	10	86.11
TL	0.23	72	69	3	95.86

The result in Table 1 reveals that the TL model outperforms the CNN model regarding faster recognition time, higher accuracy and a limited number of incorrect classifications. This result could be because only the early and central layers are employed in transfer learning, while the latter layers are only retrained. The

overall effect is a shorter training time, and as can also be observed, TL takes 0.23 s to recognise the test image, unlike CNN, which takes 0.98 s. This indicates that the TL that involves pre-trained models to carry out the new task is faster than the existing CNN.

Furthermore, the TL has a higher accuracy in recognising the queried image. A possible reason is that CNN loses information in the pooling layers due to changes in pose and orientation, which makes it challenging to handle ambiguity in terms of a slight shift in face orientation.

The contingency table for the CNN and TL models are presented in Table 2 and Table 3, respectively. The letters TP, FP, FN, and TN denote the true positives, false positives, and false negatives, respectively.

Table 2. Contingency table for CNN.

Model	Accuracy (%)
TP (36)	FN (6)
FP(4)	TN (26)

Table 3. Contingency table for TL.

Model	Accuracy (%)
TP (48)	FN (1)
FP(2)	TN (21)

These values were used to obtain the precision, recall and f-measure based on the widely adopted relations indicated in Eqs. 3, 4 and 5 Ogundokun et al. (2022).

$$Precision = \left(\frac{TP}{TP + FP}\right) \tag{3}$$

$$Recall = \left(\frac{TP}{TP + FN}\right) \tag{4}$$

$$F - measure = 2 * \left(\frac{ecall * precision}{recall + precision}\right) \tag{5}$$

The measure of the efficiency of the models based on the experiment carried out in this research is presented in Table 4 in terms of recall, precision and f-measure, which are computed using the values from the contingency tables above for CNN and TL.

It can be observed from Table 4 that the TL outperforms the CNN in terms of the F-measure, which is a result of the very low false negative (FN) outcome. This implies that the model will not deny authentic users (as obtained from the

Table 4. Result of precision, recall and f-measure.

Classifier	Recall	precision	F-measure
CNN	0.86	0.90	0.88
TL	0.98	0.96	0.97

recall). The low false positive (FP) outcome indicates that the model will not wrongly accept unauthorized users.

It is worthy of note that much of the reviewed research related have not satisfactorily reported information on the evaluation metrics such as recall, precision, f-measure and recognition time, which are important determinants of the efficiency of a machine learning model. Therefore, this work recommends that recognition accuracy alone is insufficient to measure a model's performance.

4 Conclusion

In this study, a performance comparison between CNN and TL on classifying human face gender was carried out. The experimental results indicate that both algorithms performed optimally. However, TL outperforms CNN regarding recognition accuracy, recognition time, precision, recall and f-measure. The TL approach proves effective and efficient based on the performance metrics and dataset. Furthermore, the dataset used for training is relatively small. This work intends to extend the experiment to a larger and augmented dataset. This technique is essential to address the problem of distributed coordination and decision making in hostile environments with heterogeneous physical IoT (Internet of Things) devices operating with images in Edge-Computing, to achieve their goals more efficiently. This enables solutions to flexible, adaptive and seamless decision-making problems at the edge of the cloud, where complex predictions are processed in the cloud in an optimised way and operational decisions are made locally at the edge. in the cloud and operational decisions are made locally at the edge.

References

Afifi, M., Abdelhamed, A.: AFIF4: deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces. J. Vis. Commun. Image Represent. **62**, 77–86 (2019)

Atha, D.J., Jahanshahi, M.R.: Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection. Struct. Health Monit. **17**(5), 1110–1128 (2018)

Babatunde, R., Abdulsalam, S., Yusuff, S., Babatunde, A.: Gender recognition using local binary pattern and Naive Bayes classifier. J. Comput. Sci. Appl. **23**(1), 65–74 (2016)

González-Briones, A., Villarrubia, G., De Paz, J.F., Corchado, J.M.: A multi-agent system for the classification of gender and age from images. Comput. Vis. Image Underst. **172**, 98–106 (2018)

Greco, A., Saggese, A., Vento, M., Vigilante, V.: Gender recognition in the wild: a robustness evaluation over corrupted images. J. Ambient. Intell. Humaniz. Comput. **12**, 10461–10472 (2021)

Jewel, M., Hossain, M. I., Tonni, T.H.: Bengali ethnicity recognition and gender classification using CNN & transfer learning. In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 390–396. IEEE (2019)

Jia, S., Lansdall-Welfare, T., Cristianini, N.: Gender classification by deep learning on millions of weakly labelled images. In: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 462–467. IEEE (2016)

Lawal, C.O., Akinrinmade, A.A., Badejo, J.A.: Face-based gender recognition analysis for Nigerians using CNN. In: Journal of Physics: Conference Series, vol. 1378, p. 032014. IOP Publishing (20190

Moeini, H., Mozaffari, S.: Gender dictionary learning for gender classification. J. Vis. Commun. Image Represent. **42**, 1–13 (2017)

Ogundokun, R.O., Maskeliūnas, R., Damaševičius, R.: Human posture detection using image augmentation and hyperparameter-optimized transfer learning algorithms. Appl. Sci. **12**(19), 10156 (2022)

Sharma, D.J., Dutta, S., Bora, D.J.: REGA: real-time emotion, gender, age detection using CNN-a review. Ann. Comput. Sci. Inf. Syst. **24**, 115–118 (2020)

Sumi, T.A., Hossain, M.S., Islam, R.U., Andersson, K.: Human gender detection from facial images using convolution neural network. In: Mahmud, M., Kaiser, M.S., Kasabov, N., Iftekharuddin, K., Zhong, N. (eds.) AII 2021. CCIS, vol. 1435, pp. 188–203. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-82269-9_15

Yildiz, K., Güneş, E., Anil, B.: CNN-based gender prediction in uncontrolled environments. Düzce Üniversitesi Bilim ve Teknoloji Dergisi **9**(2), 890–898 (2021)

Zaman, F.: Gender classification using custom convolutional neural networks architecture. Int. J. Electr. Comput. Eng. (IJECE) **10**(6), 5758–5771 (2020)



A Proposal for a Taxonomy of Augmented Humanity Based on Current Needs, Topics and Challenges

Graciela Guerrero[1,2(✉)], Fernando José Mateus da Silva[3], António Pereira[3,4], and Antonio Fernández-Caballero[2,5,6]

[1] Departamento de Ciencias de la Computación, Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
rgguerrero@espe.edu.ec

[2] Neurocognition and Emotion Unit, Instituto de Investigación en Informática de Albacete, Albacete, Spain

[3] School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal

[4] INOV INESC INOVAÇÃO, Institute of New Technologies-Leiria Office, Leiria, Portugal

[5] Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Albacete, Spain

[6] Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain

Abstract. In recent years, the term Augmented Humanity (AH) has been increasingly used in the research community. AH is a term that refers to the use of advanced technologies to enhance human capabilities, whether physical, cognitive, or social. Technologies used to achieve AH range from wearable devices such as smart watches to cybernetic implants and advanced prostheses. The purpose of this article is to present a proposal for a taxonomy of AH. After describing the main research topics that have been worked on in this area, the reasons why such a taxonomy is needed are explained. Then, this article presents three main approaches to AH: the enhancement of physical, intellectual, and social capabilities. In addition, several challenges of providing a taxonomy of AH are introduced. Finally, it may be concluded that this proposed AH taxonomy will help to organize the field of research and establish a common knowledge base and terminology for future studies.

Keywords: Augmented Humanity · Taxonomy · Enhanced abilities

1 Introduction

A variety of terms related to human augmentation can be found in the literature. Human augmentation is a broad field that spans several disciplines, from electrical engineering or mechanics to genetics. Human augmentation [2] and the related terms Human Augmented [27] and Human 2.0 [10] refers to technologies that enhance human productivity or capabilities, or otherwise augment the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 74–83, 2023.
https://doi.org/10.1007/978-3-031-43461-7_8

human body or mind. Modern advances in science and technology have led to a wide variety of technologies that could be classified as human augmentation. Augmentation is the most commonly used term in the interdisciplinary research community that focuses on interactive digital extensions of human capabilities.

In our recent research [15], we conducted a systematic mapping to define the terminology of Augmented Humanity (AH), and arrived at the following: "*Augmented humanity is a human-computer integration technology that proposes to improve capacity and productivity by changing or increasing the normal ranges of human function, through the restoration or extension of human physical, intellectual and social capabilities.*"

The objective of this new research is to propose a taxonomy based on the main topics in which AH work has been done [15]: Augmented Reality, Computer Vision, Wearable Computing, Human-Robot Interaction, Human-Computer Interaction, Mixed Reality, Smart Devices, Virtual Reality, Cognitive Science, Human-Centered Design, Human Intention, Interactive System, and Mobile Computing. From the aforementioned topics, and considering the current challenges faced by AH, this paper proposes AH focus groups to achieve a classification for future research work.

2 The Need for a Taxonomy

The purpose of a taxonomy is to provide an orderly classification on which to focus theoretical debates, evaluate developments, conduct research, and compare data in a meaningful way. According to recent studies [8], human augmentation as a field is still so young that a definition derived from a compilation of a representative number of articles and books has barely been developed, and a taxonomy on which to structure the various proposals of researchers is still lacking, even though there has been increasing conversation about AH in the last few years.

Having already defined the concept of AH and some of its various forms, here are three pertinent points regarding the need for a taxonomy:

- Researchers may encounter some confusion regarding the terminology of Augmented Reality (AR) and Augmented Humanity. AH is a discipline related to Augmented Reality, but the difference between the two may not be clear [40], so it is necessary to establish a taxonomy of AH.
- The use of the term AH shows that it can be considered in several fields of knowledge, such as computer science, engineering, robotics, automation and control systems, telecommunications and others [15]. Therefore, the applications developed by the different fields of knowledge and strongly related to the definition of AH can be framed in the taxonomy to be proposed.
- As AH technologies become more advanced and accessible, there is a need for a taxonomy that provides a clear understanding of the different categories of technologies used to enhance human abilities.

3 Most Important Topics of AH

In the aforementioned systematic mapping review on AH, keywords from a total of 133 relevant articles dealing with AH terminology were clustered. By extracting the identifying words from each of the articles, the total number of keywords was 200. The keywords selected for clustering were those that were repeated more than twice. Among these keywords, those that mentioned the technologies used in the scientific articles studied were selected. From this grouping and selection of the main technologies, it is highlighted that Augmented Reality is the most popular, below is a description of each of the most relevant technologies that are directly related to AH:

- **Augmented Reality:** Augmented Reality (AR) has been defined by two approaches, one broad and one narrow [25]. In the broad sense, AR is defined as "augmenting natural feedback to the operator with simulated cues", while the narrow approach emphasizes and defines AR as "a form of virtual reality in which the participant's head-mounted display is transparent, allowing a clear view of the real world". A similar definition is in terms of its characteristics, where real and computer-generated information are combined in a physical world, interactively in real time, and the display of the virtual object is intrinsically aligned with the real world orientation.
- **Wearable Computing:** The concept of Wearable Computing (WC) emerged in the mid-1990s, when wearing a computer combined with a head-mounted display and control interface became a practical possibility [28]. DARPA (Defense Advanced Research Projects Agency) defined WC as data collection and dissemination devices that allow the user to operate more efficiently. WC is an effort to make computers part of our everyday lives through our clothing (e.g., in shoes) or by creating wearable artifacts that can be worn like clothing [31].
- **Human-Robot Interaction:** Human-Robot Interaction (HRI) is currently a very large and diverse design and research activity. HRI is a field of study dedicated to the understanding, design, and evaluation of robotic systems for use by or with humans [29]. This interaction, by definition, requires communication between robots and humans. The literature is growing rapidly, with hundreds of publications each year and the activity of many different professional societies and ad hoc meetings, primarily in the technical disciplines of mechanical and electrical engineering, computer science, control engineering, and artificial intelligence.
- **Computer Vision:** Computer Vision (CV) has a dual goal [18]. From a biological point of view, CV aims to build computational models of the human visual system. From an engineering point of view, CV aims to build autonomous systems that can perform some of the tasks that the human visual system can perform (and in many cases surpass). Many vision tasks are related to the extraction of three-dimensional and temporal information from time-varying two-dimensional data (e.g., from one or more cameras) and, in general, to the understanding of such dynamic scenes [11]. The applications

of CV are innumerable in the fields of automated manufacturing and robotics, where problems such as image motion resolution and three-dimensional analysis can be addressed.

- **Human-Computer Interaction:** Human-Computer Interaction (HCI) is a term that has been around since the early 1980s [9]. The concept of HCI or interface was automatically introduced with the advent of the computer or, more generally, the machine. The approach is based on the assumption that sophisticated machines are of no value if they cannot be used properly by humans. To this end, the design of HCI must strike an appropriate balance between functionality and usability [26]. The SIGCHI (Special Interest Group on Human-Computer Interaction) of the ACM (Association for Computing Machinery), one of the leading international HCI societies, has developed the following working definition: *Human-Computer Interaction is a discipline concerned with the design, evaluation, and implementation of interactive computer systems for human use, and with the study of the major phenomena surrounding them* [17].
- **Mixed Reality:** Mixed Reality (MR) was first defined as the fusion of the physical and virtual worlds [24]. In its definition, Augmented Reality (AR) and Augmented Virtuality (AV) are considered special cases of MR. MR refers to the process of merging computer-generated graphics and a real-world scene to create new environments and visualizations in which physical and digital objects coexist and interact in real time.
- **Virtual Reality:** Virtual Reality (VR) is defined as the computer-generated simulation of a three-dimensional image or environment with which a person can interact in a seemingly real or physical way, using special electronic equipment such as a helmet with a screen inside or gloves equipped with sensors [23].
- **User-Centered Design:** For more than two decades, User-Centered Design (UCD) has been the guiding philosophy and process in the field of design, both practically and pedagogically [12]. UCD has evolved across technical disciplines, drawing on human-computer interaction, artificial intelligence, human-machine systems, and social and cognitive sciences It is based on the same goal as systems engineering, but UCD brings concepts and terminologies that are a broader, trans-disciplinary language that supports a new, broader approach to systems design.
- **Smart Devices:** Smart Devices (SDs) help people be context-aware, run errands or perform tasks, and support people's daily activities [5]. SDs promise benefits including: (i) increased convenience and capability - people can control more of their world when smart objects help them; if designed correctly, we can control more smart devices in less time, leading to a radical change in productivity and (ii) remote control - devices can be controlled remotely. Sometimes this provides time savings and convenience, sometimes safety or security. The Internet of Things adds new interfaces to smart devices so that they can have memory, sensors, actuators, computing and communications.

- **Cognitive Computation:** Cognitive science is defined as the interdisciplinary study of mind and intelligence, encompassing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology [38]. Cognitive Computation and Artificial Intelligence share the view that information processing is the central activity of intelligent behavior [36]. In addition, networks of interconnected nodes have long played a key role in cognitive science, from artificial neural networks to the popularization of models of semantic memory activation [4].
- **Artificial Intelligence:** Artificial Intelligence (AI) has been used to focus on many different aspects, and there is no widely accepted definition [16]. One possible definition is: *Artificial intelligence is designed to run computer programs that can replicate some aspect of human behavior* [37]. A similar definition states that AI has the ultimate goal of understanding intelligence and building intelligent systems.

4 Proposed Taxonomy of Augmented Humanity

The main technologies described above have been included to establish the AH taxonomy. In Fig. 1, we have chosen to leave a section open for the reception of new technologies that can make use of AH. In addition, the human is shown in the center of the figure because it is the core around which this taxonomy must revolve. A relevant part of this taxonomy are the physical senses (taste, touch, hearing, smell and sight) and the human intellect has also been considered because it is the one that can analyze, understand, reason, communicate, collaborate, etc. on different scales. All these physical and intellectual aspects can be augmented with various AH applications. Finally, the figure shows that the AH taxonomy focuses on three broad areas of application or use. Specifically, Table 1 defines the increase in physical, intellectual, and social skills.

Table 1. Augmented abilities in Augmented Humanity

Focus on the human	Definition
Augmentation of physical abilities	Enhanced physical abilities are achieved through the interpretation of the enhanced senses and the actions they produce. Sight, taste, touch, smell, and hearing can be physically enhanced.
Augmentation of intellectual abilities	Intellectual abilities are achieved through the acquisition of knowledge, information processing, and enhanced reasoning. Numerical aptitude, verbal comprehension, perceptual speed, inductive reasoning, deductive reasoning, spatial visualization, memory, and willpower can be intellectually improved.
Augmentation of social abilities	Enhanced social skills are considered through the interpretation of basic and complex social skills. Empathy, emotional intelligence, assertiveness, listening skills, the ability to communicate emotions, the ability to define a problem and evaluate solutions, negotiation, can be socially enhanced

The above taxonomy provides a guide for classifying applications that focus their development on AH. This classification is also important in light of the

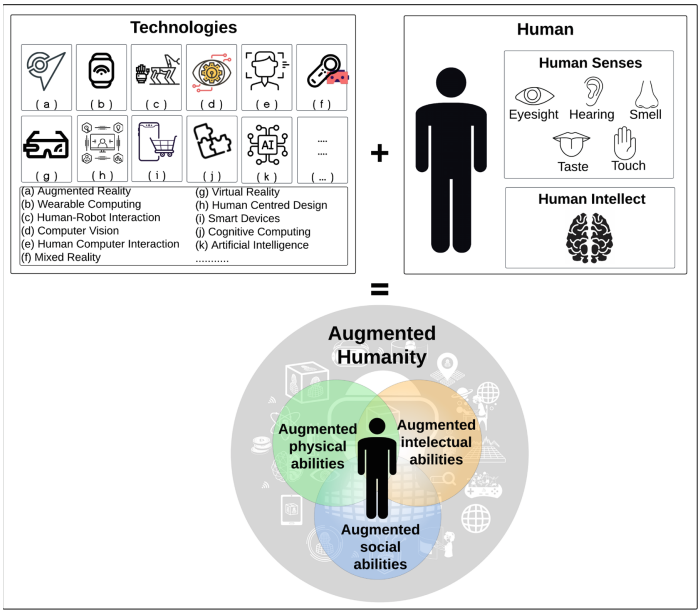


Fig. 1. Towards an open taxonomy of Augmented Humanity

rapid development of new applications. The classification of focus areas that has been presented clearly illustrates the diversity of applications related to developments in different AH topics. It is important to note that the topics covered may not be all of them, and there may be a need to add an additional topic. Having defined the proposed main categories of AH (see Table 1), we now provide a brief description of just a few of the subcategories, together with examples of their use. However, it is important to note that these subcategories are not exhaustive, as the field of technology is constantly evolving and new subcategories may emerge at any time. Therefore, the following examples serve to illustrate the diverse range of applications within Augmented Humanity, but do not encompass the full potential.

– **Augmentation of physical abilities**: Augmented vision: Enhancing visual perception, acuity, or field of view through devices like smart glasses or contact lenses [13]; Augmented hearing: Enhancing auditory perception, such as amplifying sounds or filtering background noise using hearing aids or earbuds with built-in audio processing capabilities [21]; Augmented touch: Enhancing tactile feedback or sensitivity through haptic feedback devices that provide vibrations or pressure sensations [33]; Augmented smell: Enhancing olfactory perception by simulating scents through electronic devices that release specific odors or use chemical sensors to stimulate the sense of smell [19]; Augmented taste: Enhancing gustatory perception by altering the perception of flavors through the use of electrical or thermal stimulation of the tongue [19].

- **Augmentation of intellectual abilities**: Augmented learning: Enhancing the process of acquiring knowledge and skills through personalized adaptive learning systems, virtual tutors, or brain-computer interfaces for direct knowledge transfer [20]; Augmented problem-solving: Enhancing problem-solving capabilities through AI-powered decision support systems, cognitive computing, or augmented reality-based problem-solving environments [32]; Augmented creativity: Enhancing creative thinking and ideation through AI-generated suggestions, creative collaboration platforms, or brain stimulation techniques like transcranial magnetic stimulation [41]; Augmented focus and attention: Enhancing concentration and attention span through neurofeedback training, brain stimulation, or wearable devices that monitor and provide real-time feedback on cognitive states [39]; Augmented memory: Enhancing memory encoding, retention, and recall through memory prosthetics, brain implants, or cognitive training programs tailored to memory enhancement [35].
- **Augmentation of social abilities**: Augmented communication: Enhancing communication skills through natural language processing, voice assistants, or augmented reality-based communication aids for individuals with speech or language impairments [3]; Augmented empathy: Enhancing empathic understanding and emotional resonance through virtual reality-based empathy training simulations, affective computing, or emotion recognition technologies [30]; Augmented collaboration: Enhancing collaborative work and teamwork through virtual collaboration platforms, remote presence technologies, or shared augmented reality environments [34]; Augmented social perception: Enhancing the ability to interpret social cues and non-verbal communication through augmented reality glasses with facial recognition or emotion detection capabilities [7]; Augmented negotiation: Enhancing negotiation skills through AI-powered negotiation assistants, virtual reality-based negotiation simulations, or real-time feedback systems during negotiation scenarios [1].

5 Challenges of Taxonomizing Augmented Humanity

As discussed above, research in AH has increased significantly in recent years, leading us to develop an AH taxonomy to classify and organize the various approaches and applications of this technology. While an AH taxonomy facilitates understanding in the research and development of related technologies, it also raises important ethical and social challenges that need to be considered and addressed appropriately. One of the main ethical challenges of an AH taxonomy is the concern that the classification and organization of AH technologies may lead to discrimination and exclusion of certain groups of people [6]. For example, if AH technologies are classified and organized according to their ability to enhance the physical, intellectual, or social capabilities of individuals, there is a risk that those who cannot afford access to these technologies will be left behind and excluded.

In terms of social challenges, the implementation of an AH taxonomy could widen the social and economic gap between developed and developing countries,

as countries with greater resources would have an advantage in the development and use of AH technologies [14]. In addition, the implementation of AH technologies could also affect the relationship between individuals and society in general, as well as the very nature of what it means to be human.

In summary, the creation of a taxonomy of AH poses significant ethical and social challenges that must be approached with caution. It is necessary to ensure that this technology does not create divisions in society or discrimination, and that the privacy and rights of individuals are respected [22].

6 Conclusions

This article has demonstrated the need for a taxonomy to organize the various terms and technologies involved. It has explained that human enhancement involves several disciplines, from electrical or mechanical engineering to genetics, and has mentioned concepts such as Augmented Human and Human 2.0. It was also explained that there is some confusion between the terms Augmented Reality and Augmented Humanity, and the importance of a taxonomy to mark the differences between the two disciplines was described.

In summary, this research has presented a taxonomy of Augmented Humanity that provides a guide for the scientific community, consisting of the classification of applications that focus their development on human enhancement. It has listed the topics of interest that are closest to the AH terminology, leaving open the possibility of adding other terms of use as the contributions grow.

In the field of Augmented Humanity, there is a growing need to develop a taxonomy to organize the different technologies and approaches. While this could facilitate research and development of more advanced technologies, it also raises significant ethical and social challenges. A taxonomy could lead to discrimination and exclusion of certain groups of people and widen social and economic gaps. These issues need to be carefully addressed to ensure that technologies do not lead to social division or discrimination, and that privacy and individual rights are respected.

Acknowledgments. Grants PID2020-115220RB-C21 and EQC2019-006063-P funded by MCIN/AEI/10. 13039/501100011033 and by "ERDF A way to make Europe". Grant 2022-GRIN-34436 funded by Universidad de Castilla-La Mancha and by "ERDF A way of making Europe". This work has been partially supported by Portuguese Fundação para a Ciência e a Tecnologia - FCT, I.P. under the project UIDB/04524/ 2020 and by Portuguese National funds through FITEC - Programa Interface, with reference CIT "INOV - INESC Inovação - Financiamento Base". This research was also supported by CIBERSAM, Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación.

References

1. Adewole, S., et al.: Dialogue-based simulation for cultural awareness training. arXiv preprint arXiv:2002.00223 (2020)

2. Alicea, B.: An integrative introduction to human augmentation science. arXiv preprint arXiv:1804.10521 (2018)
3. Almurashi, H., Bouaziz, R., Alharthi, W., Al-Sarem, M., Hadwan, M., Kammoun, S.: Augmented reality, serious games and picture exchange communication system for people with ASD: systematic literature review and future directions. Sensors **22**(3), 1250 (2022)
4. Baronchelli, A., Ferrer-i Cancho, R., Pastor-Satorras, R., Chater, N., Christiansen, M.H.: Networks in cognitive science. Trends Cogn. Sci. **17**(7), 348–360 (2013)
5. Bastos, D., et al.: SmartWalk BAN: using body area networks to encourage older adults to perform physical activity. Electronics **10**(1), 56 (2021)
6. Bostrom, N., Roache, R.: Ethical issues in human enhancement. In: New Waves in Applied Ethics, pp. 120–152 (2008)
7. Chantaranimi, K., Sugunsil, P., Natwichai, J.: An approach to enhance academic ranking prediction with augmented social perception data. In: Barolli, L., Chen, H.-C., Miwa, H. (eds.) INCoS 2021. LNNS, vol. 312, pp. 84–95. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-84910-8_9
8. De Boeck, M., Vaes, K.: Structuring human augmentation within product design. Proc. Des. Soc. **1**, 2731–2740 (2021)
9. Ebert, A., Gershon, N.D., van der Veer, G.C.: Human-computer interaction. Künstliche Intelligenz **26**(2), 121–126 (2012)
10. Eriksen, A.: The human version 2.0: AI, humanoids, and immortality. Soc. Anal. **65**(1), 70–88 (2021)
11. Fernández, M.A., Fernández-Caballero, A., López, M.T., Mira, J.: Length-speed ratio (LSR) as a characteristic for moving elements real-time classification. Real-Time Imaging **9**(1), 49–59 (2003)
12. Friess, E.: The sword of data: does human-centered design fulfill its rhetorical responsibility? Des. Issues **26**(3), 40–50 (2010)
13. Fürstenau, N., Schmidt, M.: Remote tower experimental system with augmented vision videopanorama. In: Fürstenau, N. (eds.) Virtual and Remote Control Tower: Research, Research Topics in Aerospace, pp. 109–137. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-93650-1_6
14. Giubilini, A., Sanyal, S.: The ethics of human enhancement. Philos. Compass **10**(4), 233–243 (2015)
15. Guerrero, G., da Silva, F.J.M., Fernández-Caballero, A., Pereira, A.: Augmented humanity: a systematic mapping review. Sensors **22**(2), 514 (2022)
16. Górriz, J.M., Ramírez, J., et al.: Artificial intelligence within the interplay between natural and artificial computation: advances in data science, trends and applications. Neurocomputing **410**, 237–270 (2020)
17. Hewett, T.T., et al.: ACM SIGCHI Curricula for Human-Computer Interaction. ACM, New York (1992)
18. Huang, T.: Computer vision: evolution and promise. CERN (1996)
19. Kerruish, E.: Arranging sensations: smell and taste in augmented and virtual reality. Sens. Soc. **14**(1), 31–45 (2019)
20. Kidd, S.H., Crompton, H.: Augmented learning with augmented reality. In: Churchill, D., Lu, J., Chiu, T.K.F., Fox, B. (eds.) Mobile Learning Design. LNET, pp. 97–108. Springer, Singapore (2016). https://doi.org/10.1007/978-981-10-0027-0_6
21. Korzepa, M.J., Johansen, B., Petersen, M.K., Larsen, J., Larsen, J.E., Pontoppidan, N.H.: Learning preferences and soundscapes for augmented hearing. In: IUI Workshops (2018)

22. Lima, V., Belk, R.: Human enhancement technologies and the future of consumer well-being. J. Serv. Mark. **36**(7), 885–894 (2022)
23. McMillan, K., Flood, K., Glaeser, R.: Virtual reality, augmented reality, mixed reality, and the marine conservation movement. Aquat. Conserv. Mar. Freshw. Ecosyst. **27**, 162–168 (2017)
24. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **77**(12), 1321–1329 (1994)
25. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: a class of displays on the reality-virtuality continuum. Telemanipulator Telepresence Technol. **2351**, 282–292 (1995)
26. Nielsen, J.: Usability Engineering. Morgan Kaufmann, Burlington (1994)
27. Papagiannis, H.: Augmented Human: How Technology is Shaping the New Reality. O'Reilly, Sebastopol (2017)
28. Randell, C.: Wearable computing: a review. Research Gate Department of Computer Science, University of Bristol, Bristol, UK (2005)
29. Roda-Sanchez, L., Garrido-Hidalgo, C., García, A.S., Olivares, T., Fernández-Caballero, A.: Comparison of RGB-D and IMU-based gesture recognition for human-robot interaction in remanufacturing. Int. J. Adv. Manuf. Technol. **124**, 3099–3111 (2023)
30. Rueda, J., Lara, F.: Virtual reality and empathy enhancement: ethical aspects. Front. Robot. AI **7**, 160 (2020)
31. Sánchez-Reolid, R., Martínez-Rodrigo, A., López, M.T., Fernández-Caballero, A.: Deep support vector machines for the identification of stress condition from electrodermal activity. Int. J. Neural Syst. **30**(7), 2050031 (2020)
32. Sarkar, P., Kadam, K., Pillai, J.S.: Learners' approaches, motivation and patterns of problem-solving on lines and angles in geometry using augmented reality. Smart Learn. Environ. **7**(1), 1–23 (2020)
33. Scherfgen, D., Schild, J.: Estimating the pose of a medical manikin for haptic augmentation of a virtual patient in mixed reality training. In: Symposium on Virtual and Augmented Reality, pp. 33–41 (2021)
34. Sereno, M., Wang, X., Besançon, L., Mcguffin, M.J., Isenberg, T.: Collaborative work in augmented reality: a survey. IEEE Trans. Vis. Comput. Graph. **28**(6), 2530–2549 (2020)
35. Shang, J., Xiao, C., Ma, T., Li, H., Sun, J.: GAMENet: graph augmented memory networks for recommending medication combination. In: The AAAI Conference on Artificial Intelligence, vol. 33, pp. 1126–1133 (2019)
36. Skinner, B.F.: Cognitive science and behaviourism. Br. J. Psychol. **76**(3), 291–301 (1985)
37. Song, D., et al.: Predicting human intention in visual observations of hand/object interactions. In: 2013 IEEE International Conference on Robotics and Automation, pp. 1608–1615. IEEE (2013)
38. Thagard, P.: Introduction to Cognitive Science. MIT Press, Cambridge (1996)
39. Vortmann, L.M., Putze, F.: Attention-aware brain computer interface to avoid distractions in augmented reality. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–8 (2020)
40. Yasumura, M.: Retrospect of augmented reality. Augmented Hum. **2012**(21), 1–5 (2012)
41. Zünd, F., et al.: Augmented creativity: bridging the real and virtual worlds to enhance creative play. In: SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications, pp. 1–7 (2015)



A New Software Platform to Provide AI-Based Services to Smart Cities

Francisco Pinto-Santos[1(✉)], Juan Antonio González-Ramos[2], Raúl López-Blanco[1], and Javier Curto[1]

[1] Air Institute, IoT Digital Innovation Hub (Spain), Carbajosa de la Sagrada, 37188 Salamanca, Spain
{franpintosantos,rlopez,jcurto}@air-institute.com
[2] Servicios Informáticos, Universidad de Salamanca, Salamanca, Spain
juanan@usal.es

Abstract. Smart cities require the integration of various data sources and the application of artificial intelligence (AI) to enhance decision-making, operations, and planning. However, the complexity of AI solutions often poses a barrier for non-experts. This paper proposes a user-friendly software architecture and AI-based solution that allows end-users to harness the power of AI without delving into its theoretical aspects. The solution features an AI-driven assistant, which helps users create interactive dashboards and visualizations by importing data from multiple heterogeneous sources, with a focus on the Internet of Things (IoT). The proposal is evaluated within the context of smart cities, and the results show a promising direction for future development.

Keywords: Smart cities · Internet of Things · Software platform

1 Introduction

Smart cities represent a vision of urban development that leverages advanced technologies, such as artificial intelligence (AI) and the Internet of Things (IoT), to improve the quality of life for residents, optimize urban operations, and enhance overall decision-making. With the rapid growth of urbanization and the increasing demands placed on city infrastructures, the development and implementation of smart cities have become essential to ensure sustainable urban living. As a result, numerous efforts are being made by governments, municipalities, and private organizations to integrate innovative technologies into the urban landscape, ultimately transforming the way cities are managed and experienced.

The integration of AI and IoT in smart cities allows for the collection and analysis of large volumes of data from various sources [8,9], including sensors [3], devices [1], and social media platforms [16]. This wealth of information can be utilized to optimize resource allocation [11], transportation [14], energy consumption [13], public safety [12], and numerous other aspects of urban living [10]. Furthermore, AI can be used to predict and model complex urban phenomena,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 84–92, 2023.
https://doi.org/10.1007/978-3-031-43461-7_9

enabling decision-makers to identify potential issues and implement proactive solutions.

However, despite the significant potential of AI to transform smart cities [6], the widespread adoption of AI solutions is often hindered by their inherent complexity. Many AI tools and techniques require a deep understanding of the underlying algorithms and a high level of technical expertise to be effectively utilized [4,5]. This presents a significant barrier for non-experts, such as urban planners, city officials, and other stakeholders, who may not possess the necessary technical knowledge to harness the full potential of AI in their decision-making processes.

In response to this challenge, this paper proposes a user-friendly software architecture and AI-based solution that aims to simplify the process of applying AI in smart cities. The proposed solution features an AI-driven assistant that guides end-users through the steps required to create interactive dashboards and visualizations, enabling them to harness the power of AI without delving into its theoretical aspects.

The primary goal of this research is to democratize the use of AI in smart cities, making it accessible and easy to use for a wide range of stakeholders [2], regardless of their technical expertise. By providing an intuitive, user-friendly interface, the proposed solution seeks to bridge the gap between complex AI algorithms and their practical applications in urban planning and management.

To achieve this goal, the software architecture integrates several components, including a data importer for ingesting data from multiple heterogeneous sources, an AI-driven assistant that uses natural language processing (NLP) and machine learning algorithms to guide users through the process of creating visualizations and dashboards, a visualization engine for rendering interactive and dynamic visualizations based on input data and user preferences, and a dashboard designer that enables users to create customizable dashboards using drag-and-drop techniques.

The proposed solution is evaluated within the context of smart cities, focusing on the integration of AI-driven insights into urban planning and resource allocation. The evaluation aims to assess the effectiveness of the AI-driven assistant in guiding users through the process of importing data from various sources, including IoT devices, and creating visualizations that facilitate decision-making. Moreover, the evaluation examines the user-friendliness of the dashboard designer in allowing users to create customizable dashboards without programming.

In addition to making AI more accessible to non-experts, the proposed solution offers several benefits in the context of smart cities [7,15]. First, by simplifying the process of creating visualizations and dashboards, the solution enables users to quickly and efficiently analyze large volumes of data, leading to more informed and effective decision-making. Second, the AI-driven assistant can help users identify patterns, trends, and correlations in the data that may not be readily apparent, further enhancing their understanding of the urban environment. Finally, the intuitive interface provided by the dashboard designer facilitates collaboration and communication among stakeholders, as it allows users to easily share and discuss their findings.

In summary, this paper presents a user-friendly software architecture and AI-based solution that aims to democratize the use of AI in smart cities, making it accessible to a broad range of stakeholders, regardless of their technical expertise. By providing an intuitive, user-friendly interface, the proposed solution seeks to bridge the gap between complex AI algorithms and their practical applications in urban planning and management.

The remainder of this paper is organized as follows: Sect. 2 outlines the proposed software architecture, detailing its various components and their roles in simplifying the application of AI in smart cities. Section 3 presents the results of the evaluation of the proposed solution within the context of smart cities, focusing on its effectiveness in guiding users through the process of importing data, creating visualizations, and designing dashboards. Finally, Sect. 4 concludes the paper and outlines potential directions for future work in the development of user-friendly AI-based solutions for smart cities.

Through the development and evaluation of the proposed solution, this research contributes to the ongoing efforts to make AI more accessible and easy to use in the context of smart cities. By lowering the barrier to entry for non-experts, the proposed solution has the potential to significantly enhance the adoption of AI in urban planning and management, ultimately leading to more sustainable, efficient, and livable cities. Moreover, the insights gained from this research may be applicable to other domains that face similar challenges in harnessing the power of AI, paving the way for further innovations in the democratization of AI.

2 Proposal

The proposed solution consists of a software architecture that integrates an AI-driven assistant to guide users through the process of creating interactive dashboards and visualizations using drag-and-drop techniques. The AI assistant helps users import data from multiple heterogeneous sources, including IoT devices, and create dynamic visualizations without programming. The assistant supports the application of various classical machine learning algorithms, allowing users to derive insights from their data without the need for extensive technical knowledge.

The software architecture consists of the following components:

- **Data Importer:** Enables users to import data from various sources, including IoT devices, databases, and files, by providing an intuitive interface. The Data Importer supports multiple data formats and handles data preprocessing, ensuring that the data is ready for analysis.
- **AI-driven Assistant:** Guides users through the process of creating visualizations and dashboards using natural language processing (NLP) and machine learning algorithms. The assistant allows users to apply a wide range of classical machine learning algorithms, such as regression, classification, clustering, and dimensionality reduction techniques, to their data. Additionally, the AI-driven Assistant provides explanations and recommendations on the most

suitable algorithms based on the characteristics of the data and the user's goals.
- **Algorithm Library:** A comprehensive library of classical machine learning algorithms, such as linear regression, logistic regression, support vector machines, decision trees, random forests, k-means clustering, and principal component analysis, among others. The library provides a robust foundation for the AI-driven Assistant to offer users various options for analyzing their data and generating insights.
- **Visualization Engine:** Renders interactive and dynamic visualizations based on the input data and user preferences. The engine supports a wide range of visualization types, including bar charts, line charts, scatter plots, heatmaps, and geographic maps, allowing users to effectively communicate their findings and explore their data.
- **Dashboard Designer:** Allows users to create customizable dashboards using drag-and-drop techniques, incorporating the visualizations created with the AI-driven assistant. The designer provides a variety of layout options and widgets, enabling users to create visually appealing and informative dashboards that facilitate data-driven decision-making.

The proposed solution has been designed with usability and flexibility in mind. By providing a wide range of classical machine learning algorithms and an intuitive interface, the AI-driven assistant empowers users to derive valuable insights from their data without needing to understand the intricacies of each algorithm. Moreover, the assistant offers guidance on the most appropriate algorithms to use based on the data and the user's objectives, ensuring that the generated insights are both relevant and reliable.

The inclusion of an extensive library of classical machine learning algorithms in the software architecture allows users to explore their data from various perspectives and uncover hidden patterns, trends, and relationships. This flexibility is particularly important in the context of smart cities, where the data is often diverse and complex, requiring the application of multiple techniques to fully understand the urban environment and make informed decisions.

By seamlessly integrating the AI-driven assistant with the visualization engine and dashboard designer, the proposed solution enables users to effectively communicate their findings and explore their data in an interactive and engaging manner. This, in turn, facilitates collaboration and discussion among stakeholders, promoting the widespread adoption of data-driven decision-making in smart cities.

Overall, the proposed software architecture represents a significant step towards democratizing the use of AI in smart cities, making it accessible to a broad range of stakeholders and facilitating the development of more sustainable, efficient, and livable urban environments.

3 Results

The proposed software architecture and AI-based solution were implemented and evaluated within the context of smart cities, focusing on the ease of use for end-

users and the effectiveness of the AI-driven assistant in guiding users through the process of importing data, creating visualizations, and designing dashboards. The AI-driven assistant and the algorithm library were implemented in Python, a popular programming language for machine learning applications, while the user interface was developed using TypeScript, Vue.js, and Node.js. The solution was deployed on a cloud-based architecture, ensuring scalability and accommodating the large volumes of data typically generated by smart cities.

This section presents the results of the evaluation, highlighting the effectiveness of the proposed solution in achieving its objectives of democratizing the use of AI in smart cities and making it accessible to a wide range of stakeholders, regardless of their technical expertise. The results are organized into the following subsections:

3.1 AI-Driven Assistant and Algorithm Library

The AI-driven assistant, implemented in Python, was found to be effective in guiding users through the process of applying various classical machine learning algorithms to their data. The assistant's natural language processing capabilities enabled users to interact with the system using conversational language, making the process more intuitive and user-friendly. Additionally, the assistant provided explanations and recommendations on the most suitable algorithms based on the characteristics of the data and the user's goals, further enhancing the user experience.

The algorithm library, which included a comprehensive collection of classical machine learning algorithms, proved to be a valuable resource for users, enabling them to explore their data from various perspectives and uncover hidden patterns, trends, and relationships. The flexibility offered by the library allowed users to tailor their analyses to their specific needs and objectives, ensuring that the insights generated were both relevant and reliable.

3.2 Visualization Engine and Dashboard Designer

The visualization engine, implemented using TypeScript and Vue.js, was effective in rendering interactive and dynamic visualizations based on the input data and user preferences. The engine supported a wide range of visualization types, allowing users to effectively communicate their findings and explore their data. The visualizations were found to be easy to customize and manipulate, enabling users to interact with their data in a more engaging and immersive manner.

The dashboard designer, also implemented using TypeScript and Vue.js, allowed users to create visually appealing and informative dashboards using drag-and-drop techniques. The designer provided a variety of layout options and widgets, ensuring that users could create dashboards tailored to their specific needs and preferences. The intuitive interface of the designer facilitated the process of designing and sharing dashboards, promoting collaboration and discussion among stakeholders.

3.3 Cloud-Based Deployment and Scalability

The proposed solution was deployed on a cloud-based architecture, ensuring scalability and accommodating the large volumes of data typically generated by smart cities. The cloud deployment utilized a load balancer to distribute incoming requests across multiple server instances, guaranteeing efficient performance and resource utilization. This approach allowed the system to handle a large number of concurrent users and process vast amounts of data without experiencing performance degradation or downtime.

To further enhance the performance and scalability of the system, a caching mechanism was implemented to store the results of frequently executed queries and computations. This reduced the overall response time of the system and minimized the computational overhead associated with complex analyses. Additionally, the cloud-based architecture enabled the system to automatically scale the number of server instances based on the current workload, ensuring that resources were efficiently allocated and that the performance of the system was maintained even during periods of high demand.

3.4 Performance Evaluation

The performance of the proposed solution was evaluated using a combination of quantitative and qualitative methods. The quantitative evaluation focused on measuring the system's response time, throughput, and resource utilization, while the qualitative evaluation involved gathering feedback from users regarding their experience with the system.

The quantitative evaluation results showed that the system was able to effectively handle large volumes of data and maintain efficient performance, even during periods of high demand. The load balancer and caching mechanisms were found to be particularly effective in ensuring resource utilization and response time remained at optimal levels. This demonstrated the system's capability to accommodate the vast amounts of data generated by smart cities and support a large number of concurrent users.

The qualitative evaluation involved collecting feedback from users, who included urban planners, city officials, and other stakeholders involved in the management of smart cities. Overall, the users reported a positive experience with the system, praising its intuitive interface, the AI-driven assistant, and the flexibility offered by the algorithm library. Users found the process of importing data, creating visualizations, and designing dashboards to be straightforward and user-friendly, with the AI-driven assistant providing valuable guidance and recommendations throughout the process.

Many users also reported that the system facilitated collaboration and communication among stakeholders by making it easy to share and discuss their findings. The interactive visualizations and customizable dashboards were particularly well-received, as they allowed users to explore their data in an engaging and immersive manner. This, in turn, contributed to more informed and effective decision-making in the context of smart cities.

3.5 Summary of Results

The results of the evaluation demonstrate the effectiveness of the proposed software architecture and AI-based solution in achieving its objectives of democratizing the use of AI in smart cities and making it accessible to a wide range of stakeholders, regardless of their technical expertise. The AI-driven assistant, algorithm library, visualization engine, and dashboard designer were found to be effective in guiding users through the process of importing data, applying machine learning algorithms, creating visualizations, and designing dashboards. Moreover, the cloud-based deployment and scalability mechanisms ensured the system could efficiently handle large volumes of data and support a large number of concurrent users.

4 Conclusion and Future Work

This paper presents a user-friendly software architecture and AI-based solution that simplifies the application of AI in smart cities by providing an AI-driven assistant. The assistant guides users through the process of creating interactive dashboards and visualizations by importing data from multiple heterogeneous sources, with a focus on IoT devices. The results show that the proposed solution is effective in helping users create customizable dashboards without programming, providing an accessible interface for managing and analyzing data within the context of smart cities.

Future work will focus on the following aspects:

- **Improved AI-driven Assistant:** Enhancing the AI-driven assistant by incorporating more advanced NLP techniques and machine learning algorithms to better understand user needs and provide more accurate guidance.
- **Expanded Data Source Integration:** Developing methods to seamlessly integrate additional data sources, such as social media and streaming data, to further enrich the available information for decision-making.
- **Advanced Visualization Techniques:** Investigating and incorporating more advanced visualization techniques to provide users with a broader range of options to visualize and analyze data.
- **Scalability:** Addressing the scalability of the software architecture to accommodate the growing volume of data generated by smart cities and the increasing complexity of AI algorithms.
- **User Experience:** Continuously refining the user interface and overall user experience to ensure that the software architecture remains intuitive and easy to use for end-users.

By addressing these aspects, the proposed solution can continue to evolve and better serve the needs of smart cities, empowering end-users to harness the power of AI without the need for extensive technical knowledge. Ultimately, this will contribute to the advancement of smart cities and the improvement of urban life for their inhabitants.

Acknowledgments. This work has been partially supported by the Institute for Business Competitiveness of Castilla y León, and the European Regional Development Fund under grant CCTT3/20/SA/0002 (AIR-SCity project).

References

1. Bellini, P., Nesi, P., Pantaleo, G.: IoT-enabled smart cities: a review of concepts, frameworks and key technologies. Appl. Sci. **12**(3), 1607 (2022)
2. Bokhari, S.A.A., Myeong, S.: Use of artificial intelligence in smart cities for smart decision-making: a social innovation perspective. Sustainability **14**(2), 620 (2022)
3. Canizes, B., Pinto, T., Soares, J., Vale, Z., Chamoso, P., Santos, D.: Smart city: a GECAD-BISITE energy management case study. In: De la Prieta, F., et al. (eds.) PAAMS 2017. AISC, vol. 619, pp. 92–100. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61578-3_9
4. Casado-Vara, R., Martin-del Rey, A., Affes, S., Prieto, J., Corchado, J.M.: IoT network slicing on virtual layers of homogeneous data for improved algorithm operation in smart buildings. Futur. Gener. Comput. Syst. **102**, 965–977 (2020)
5. Casado-Vara, R., Prieto-Castrillo, F., Corchado, J.M.: A game theory approach for cooperative control to improve data quality and false data detection in WSN. Int. J. Robust Nonlinear Control **28**(16), 5087–5102 (2018)
6. Chamoso, P., González-Briones, A., Rodríguez, S., Corchado, J.M.: Tendencies of technologies and platforms in smart cities: a state-of-the-art review. Wirel. Commun. Mob. Comput. 2018 (2018)
7. Chamoso, P., González-Briones, A., Rivas, A., De La Prieta, F., Corchado, J.M.: Social computing in currency exchange. Knowl. Inf. Syst. **61**, 733–753 (2019)
8. Corchado, J.M., et al.: Deepintnet: a rapid deployment platform for smart territories. Sensors **21**(1), 236 (2021)
9. Corchado, J.M., Pinto-Santos, F., Aghmou, O., Trabelsi, S.: Intelligent development of smart cities: Deepint.net case studies. In: Corchado, J.M., Trabelsi, S. (eds.) SSCTIC 2021. LNNS, vol. 253, pp. 211–225. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-78901-5_19
10. Dash, B., Sharma, P.: Role of artificial intelligence in smart cities for information gathering and dissemination (a review). Acad. J. Res. Sci. Publishing **4**(39) (2022)
11. García-García, L., Jiménez, J.M., Abdullah, M.T.A., Lloret, J.: Wireless technologies for IoT in smart cities. Netw. Protoc. Algorithms **10**(1), 23–64 (2018)
12. Garcia-Retuerta, D., Chamoso, P., Hernández, G., Guzmán, A.S.R., Yigitcanlar, T., Corchado, J.M.: An efficient management platform for developing smart cities: solution for real-time and future crowd detection. Electronics **10**(7), 765 (2021)
13. González-Briones, A., Chamoso, P., de Alba, F.L., Corchado, J.M.: Smart Cities Energy Trading Platform Based on a Multi-agent Approach. Artificial Intelligence and Environmental Sustainability: Challenges and Solutions in the Era of Industry 4.0, pp. 131–146 (2022)
14. Martí, P., Jordán, J., Chamoso, P., Julian, V.: Taxi services and the carsharing alternative: a case study of Valencia city (2022)
15. Rivas, A., Fraile, J.M., Chamoso, P., González-Briones, A., Sittón, I., Corchado, J.M.: A predictive maintenance model using recurrent neural networks. In: Martínez Álvarez, F., Troncoso Lora, A., Sáez Muñoz, J.A., Quintián, H., Corchado, E. (eds.) SOCO 2019. AISC, vol. 950, pp. 261–270. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-20055-8_25

16. Shoeibi, N., Shoeibi, N., Julian, V., Ossowski, S., Arrieta, A.G., Chamoso, P.: Smart cyber victimization discovery on twitter. In: Corchado, J.M., Trabelsi, S. (eds.) SSCTIC 2021. LNNS, vol. 253, pp. 289–299. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-78901-5_25

Chatto: An Emotionally Intelligent Avatar for Elderly Care in Ambient Assisted Living

Carla Mendes[1], Rafael Pereira[1], José Ribeiro[1], Nuno Rodrigues[1], and António Pereira[1,2(✉)]

[1] Computer Science and Communications Research Centre, School of Technology and Management, Polytechnic of Leiria, 2411-901 Leiria, Portugal
{carla.c.mendes,rafael.m.pereira,jose.ribeiro,nunorod,apereira}@ipleiria.pt

[2] INOV INESC Inovação, Institute of New Technologies, Leiria Office, 2411-901 Leiria, Portugal

Abstract. In 2020, the number of people aged 60 years and older outnumbered children younger than 5 years and this trend is growing at a rapid pace. Therefore, elders are at higher risk of social isolation and loneliness due to the busy lifestyle of families in the current society. According to recent studies, conversational agents have a significant impact on the emotional and physical well-being and independence of elders in Ambient Assisted Living (AAL) environments. Chatto, a conversational agent with an animated avatar, aims to enhance the emotional well-being of elders by establishing an empathetic connection through interactive conversations.

This paper focuses on the implementation and benefits of Chatto for the elder population. The proposed solution offers numerous benefits for elders, including increased social interaction, emotional support, and real-time monitoring of their well-being. Utilizing advanced techniques to recognize facial expressions and analyze voice sentiment, Chatto offers a dynamic and interactive avatar. A key innovation of the proposed solution is the ability for caregivers to personalize the user experience by selecting which of Chatto's animations are displayed and when resulting in a highly personalized and engaging experience. Usability tests with diverse elder users demonstrate that Chatto is user-friendly, easy to navigate, and offers an enjoyable experience. Future improvements, informed by user feedback, will further enhance the avatar's intuitiveness and impact on the emotional well-being of elders.

Keywords: Ambient Assisted Living (AAL) · Animated avatar · Elderly well-being · Positive emotion reinforcement · Facial expression recognition · Voice sentiment analysis

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 93–102, 2023.
https://doi.org/10.1007/978-3-031-43461-7_10

1 Introduction

According to recent studies from the World Health Organization [1], by 2050, the proportion of the world's population over 60 years is expected to nearly double from 12% in 2015 to 22% in 2050, whereas the number of people aged 60 years and older already outnumbered children younger than 5 years in 2020.

Furthermore, elders who are ill or dependent and living at home often experience social isolation, as their physical and mental abilities decline and they become hesitant to venture outside. Even those residing in sheltered accommodations or hospitals may go through periods of loneliness when they are left unattended or when caregivers are unable to provide round-the-clock care. As the population ages and the cost of hiring full-time caregivers for every dependent individual at home becomes a financial burden for families, the increasing availability of advanced assistant conversational systems appears to offer a promising solution for the future [8].

Ambient Assisted Living (AAL) has the potential to enhance the quality of life of elders by offering personalized care and assistance while using information and communication technologies combined with Artificial Intelligence (AI). In recent years, the use of avatars has emerged as a promising approach to improving the psychological well-being of individuals in AAL environments. By establishing an empathetic connection with the user, avatars can facilitate the expression of emotions and promote positive emotional reinforcement [5,7].

This paper presents an innovative solution that combines the power of AAL systems with the benefits of avatars for emotional support. The proposed solution, Chatto, is an interactive avatar that is capable of detecting emotions and responding accordingly. By providing a friendly and engaging interaction, Chatto aims to improve the emotional state of elders and reduce feelings of loneliness and monotony.

The main contributions of this research are as follows:

- Easy interaction: Chatto, an avatar that offers a user-friendly and intuitive interaction for elders. It promotes engagement, reinforces positive emotions, and alleviates loneliness and monotony.
- Multisource message interpretation: Chatto interprets and processes messages from multiple sources like chatbot conversations, RGB cameras, and microphone audio, therefore harnessing a diverse range of inputs to enhance the understanding of the user's emotions.
- Personalized animations: Development of customized animations for Chatto, an innovative feature, where caregivers can customize which animations Chatto can perform. Therefore, providing personalized features to the user accordingly to their preferences.

Understanding the potential of avatars in the AAL context [10], the focus consists of the interaction between elders and Chatto. Therefore, this paper aims to investigate the impact of this interaction on the emotional state of elderly individuals and explore the benefits of using avatars to support their emotional well-being.

The remainder of this paper is organized as follows: Sect. 2 provides information regarding related work in the field. Section 3 describes the proposed solution in detail. Section 4 discusses the implementation and testing of the solution. Section 5 presents the analysis and discussion of the results. Finally, Sect. 6 concludes the paper and offers directions for future work.

2 Related Work

This section presents a summary of previous research on conversational agents (CA) designed for elderly users within the context of AAL or healthcare emphasising CA with a visual interface of an avatar.

The authors in [10] discuss the potential of using avatars and virtual agents as interfaces for the elder. The authors explain how these technologies can improve communication and interaction between the elders and their caregivers and help alleviate loneliness and social isolation. They also highlight some challenges, such as personalized interfaces and maintaining privacy and security. Overall, the article suggests that avatars and virtual agents have great potential as relationship interfaces for elder people.

Wan et al. [11] describes the development of a chatbot named Midoriko, which utilizes a three-dimensional (3D) avatar to convey emotional states. The chatbot is powered by a long short-term memory (LSTM) network, which helps it to respond to user inputs in a more natural and human-like way. The 3D avatar of Midoriko has several advantages. Firstly, it provides a more engaging and personalized experience for users. The avatar is designed to be visually appealing and can be customized to suit individual preferences. Secondly, the avatar can convey emotions more effectively than a traditional chatbot interface, as it can display facial expressions, body language, and other nonverbal cues, which can help to build a stronger emotional connection between the user and the chatbot.

Noh et al. [3] presents a new approach for designing mobile user interfaces that are more accessible and user-friendly for elderly users. The authors propose a virtual companion that can assist elderly users in navigating and using mobile devices. The virtual companion uses Natural Language Processing (NLP) and speech recognition to allow users to interact using voice commands. The authors also conducted a usability study to evaluate the effectiveness of the virtual companion interface, which was well-received by elderly users, who found it more intuitive and easier to use than traditional mobile interfaces. The paper concludes that the virtual companion-based mobile user interface has the potential to improve the accessibility and usability of mobile devices for elderly users, thus enhancing their quality of life.

The authors in [2] describes the development of the NESTORE e-Coach, a tangible and embodied conversational agent designed to support healthy ageing and independent living in elders. The e-Coach is based on the principles of user-centred design, incorporating feedback from elders and healthcare professionals. The article outlines the technical details of the e-Coach's development, including

its hardware components and software architecture. The e-Coach's functionalities, such as reminders, suggestions, and motivational messages are described. A pilot study with elders was conducted evaluate the conversational agent and the study showed promising results regarding user engagement and perceived usefulness.

Compared to existing research, Chatto stands out with its personalization feature, allowing caregivers to customize the avatar's expressions and animations, providing a tailored, emotionally supportive experience for elderly users by addressing their unique needs and preferences.

3 The Proposed System

This section presents the proposed system architecture, Chatto, an empathetic conversational agent designed to enhance the emotional well-being of elders in AAL environments. The system is structured to seamlessly capture the auditory and visual context of a conversation between the elder user and Chatto, process the data for emotion recognition, and generate appropriate expressions and responses for the avatar. Figure 1 illustrates the flow of information and communication between the various modules and components within the system.

The process commences with data acquisition from the conversation between the elder user and Chatto. Utilizing sensors embedded in the mobile device, the

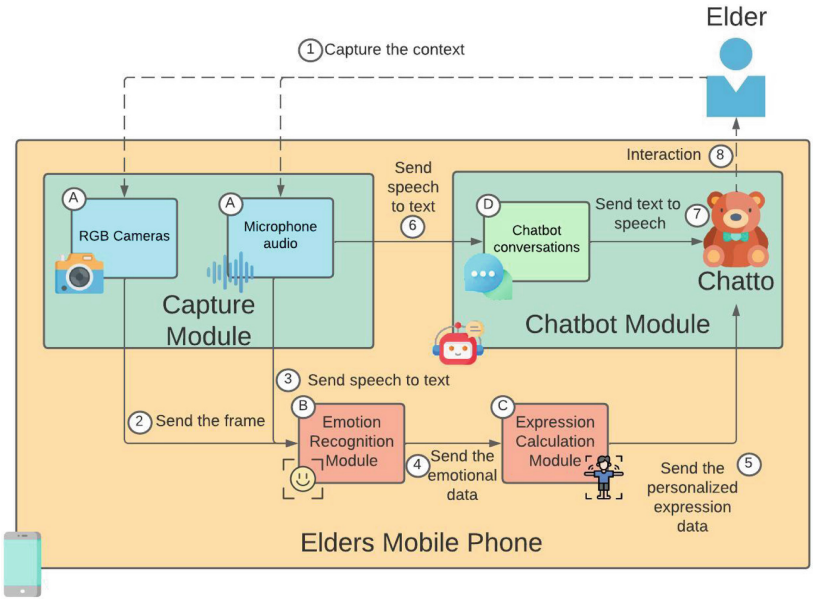


Fig. 1. Overview of the Chatto system, illustrating the flow of information and communication between data acquisition, emotion recognition, expression calculation, and conversational interaction modules.

system (1) captures both visual and auditory context through (A) RGB cameras and the microphone in the capture module. In the current implementation, only the dialogue and the elder's facial expressions are used; however, this system could be scaled to incorporate additional information, such as the vocal spectrogram of the dialogue or the elder's body poses.

Once the visual and auditory data are captured, the system processes the information and (2) sends the frame and the (3) speech-to-text data to the emotion recognition module (B). It analyzes these inputs to identify the elder user's emotional state during the interaction.

Subsequently, the emotional data are sent to the (4) expression calculation module (C), which determines the appropriate expressions for Chatto based on the detected emotions and the elder's personalized expressions. The calculated expression is then (5) sent to Chatto, enabling the avatar to accurately represent the elder's emotional state.

In addition to emotion detection and response generation, Chatto offers a unique feature that allows caregivers to customize the avatar's expressions to better cater to the individual needs of each elderly user. This personalization process involves the caregivers selecting and adjusting specific animations and gestures for Chatto to perform, based on their knowledge of the elder's preferences and emotional triggers. By enabling caregivers to directly influence the avatar's behavior, Chatto ensures a more engaging and supportive experience tailored to the elder user's emotional requirements, ultimately contributing to their emotional well-being.

The calculation of the expressions is necessary due to the presence of multiple data sources, and a weighted sum formula is used to combine the information from different modalities. The formula is defined as:

$$y = (Ws * AccS * WEmoS) + (Wf * AccF * WEmoF) \tag{1}$$

The variables in Eq. 1 are defined as follows:

- **y** - A decimal value ranging between [-100;100] that will be converted into an expression;
- **Ws** - The weight assigned to the sentiment analysis module based on voice;
- **AccS** - The confidence value (in percentage) of the emotion detected from the voice data;
- **WEmoS** - The weight assigned to the emotion detected from the voice, ranging from -1 to 1. Values close to -1 correspond to negative emotions, values close to 1 signify positive emotions, and values near 0 represent neutrality;
- **Wf** - The weight assigned to the sentiment analysis module based on facial expressions;
- **AccF** - The confidence value (in percentage) of the emotion detected from facial expressions;
- **WEmoF** - The weight assigned to the emotion detected from facial expressions, ranging from -1 to 1. Values close to -1 correspond to negative emotions, values close to 1 signify positive emotions, and values near 0 represent neutrality.

In parallel with the emotion detection and expression calculation processes, the text derived from the audio captured by the microphone (A) is (6) sent to the chatbot conversation module (D). This module processes the text and generates a suitable response for Chatto, ensuring a meaningful and engaging conversation with the elder user. The generated response is then (7) sent to Chatto, which delivers the response audibly, maintaining an interactive dialogue with the elder user.

With the information about the appropriate expression and response, Chatto (8) interacts with the elder user, effectively conveying empathy and providing emotional support. This comprehensive process ensures that Chatto can adapt its behavior in real time, considering the elder user's emotions during the conversation.

4 Implementation and Testing

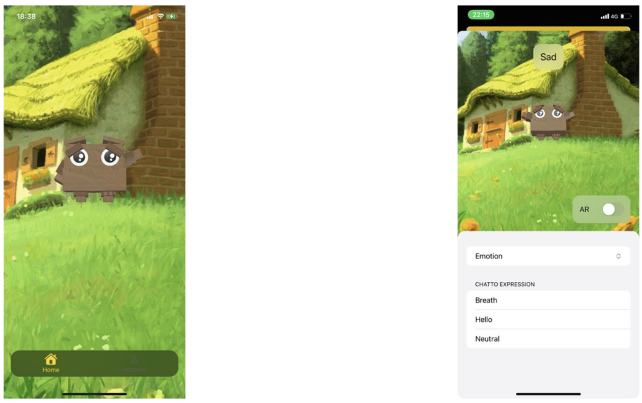
In this section, the implementation and testing of Chatto is presented, the animated avatar created to improve the emotional well-being of elders in AAL environments. The implementation is based on the system architecture and modules described in the previous section, which includes data acquisition (Module A), emotion recognition (Module B), expression calculation (Module C), and conversational interaction (Module D).

Initially, an 3D animated avatar was developed, as displayed in Fig. 2a, utilizing a combination of modeling and animation tools to ensure that Chatto's expressions were visually appealing and easily comprehensible. Afterwards, facial expression recognition (Module B) was integrated using voice and image data (Module A) to identify the emotions of elders and adapt the avatar's behaviour accordingly. This integration involved employing the DeepFace model [9] for facial expression analysis and the SentimentPolarity model [4] for voice sentiment analysis, both trained on extensive datasets to enhance their accuracy in emotion detection.

For the conversational interaction (Module D), DialogFlow [6] was used as the chatbot engine to enable seamless and natural conversations between Chatto and the elder users. DialogFlow facilitated the creation of robust, context-aware, and engaging conversations tailored to the needs of the elderly population in AAL environments. This implementation ensured that the avatar could provide not only emotional support but also a reliable and enjoyable conversational experience.

Next, a user interface that facilitated seamless interaction between elders and Chatto was designed. The interface was optimized for readability, ease of use, and accessibility, considering the specific needs and preferences of the target user group.

For the testing phase, comprehensive usability tests were conducted and represented in Table 1, with a diverse group of elder users in an AAL setting. Participants engaged in conversations with Chatto, during which their facial expressions and voices were analyzed to detect emotions. The detected emotions were sent



(a) Chatto avatar (b) Personalized animations

Fig. 2. Visual representation of the Chatto avatar and its personalized animations: (a) The Chatto avatar engaged in conversation with an elder user; (b) A range of Chatto's expressions and gestures tailored to the user's emotions

to the expression calculation module (Module C), which adjusted the avatar's expressions based on the weighted sum formula described in the previous section and according to the caregiver's preferences. This approach aimed to establish an empathetic connection and provide emotional support.

Additionally, Chatto's real-time monitoring capability was assessed, which allowed caregivers to be informed of the user's emotional state and intervene when necessary.

To further evaluate the effectiveness of Chatto, quantitative measures were incorporated in the testing process and pre-and post-test questionnaires were utilized to assess changes in the participants' emotional states and satisfaction levels, as well as semi-structured interviews to gather detailed feedback on their experiences. This approach enabled us to identify areas for improvement and optimize the avatar's performance based on user input.

The testing phase sought to evaluate the user-friendliness, intuitiveness, and overall effectiveness of the proposed solution. Participants provided feedback on their experiences interacting with Chatto, including any difficulties they faced and suggestions for improvement.

An important aspect of the testing phase was to assess the effectiveness of the personalization feature that allows caregivers to customize Chatto's expressions and animations. This customization aimed to enhance the emotional support provided to the elder users by tailoring the avatar's behavior to their specific needs and preferences. Feedback from caregivers was essential to understanding the impact of this feature and identifying potential improvements to better support the emotional well-being of elderly users.

Multiple tests and iterations were conducted to gather valuable insights and refine the avatar's performance. The results demonstrated that Chatto was well-

Table 1. Usability test results indicating the ease with which participants completed various tasks while interacting with Chatto.

Task	Failed to complete	Finished with many struggles	Finished with some struggles	Finished with some ease	Finished with much ease
Interact with Chatto	–	–	3	3	10
Observe last detected emotions	–	–	2	3	11
Observe last sent notifications	–	–	–	1	15
Observe last sent notifications	–	–	–	1	15
Observe emotion shown in the DT	–	–	–	4	12
Associate an emotion to a Chatto's animation	–	–	3	5	8
Observe Chatto's animation with AR	–	–	–	4	12

received by the majority of users, who found the avatar enjoyable, easy to navigate, and emotionally supportive.

Furthermore, caregivers appreciated the real-time monitoring feature, which facilitated more timely and responsive care. The combination of user feedback and quantitative assessments enabled us to establish a comprehensive understanding of Chatto's impact on the emotional well-being of elder users. This understanding will be instrumental in guiding future development and enhancements to the system, ultimately ensuring that Chatto continues to provide valuable emotional support to the elder population in AAL environments.

5 Analysis and Discussion

This section discusses and evaluates the Chatto system as an interactive and empathetic avatar for elderly users. Moreover, detailing challenges found during implementation.

Chatto is a promising system that aims to enhance the emotional well-being of elderly users where the integration of facial emotion recognition and voice sentiment analysis enables the system to detect emotions. Furthermore, by allowing the caregiver to adjust the avatar's animation behavior accordingly, the systems creates a more personalized and supportive experience when the animation is joined with the textual response obtained from DialogFlow.

Overall Chatto was well-accepted by the families involved in testing and overall elder people were pleased with the visual appearance and interaction modes of the system. Furthermore, the caregivers also provided positive opinions regarding the features related to the emotional accompanying.

Some struggles found during the implementation were choosing Chatto's visual style form e.g. animal, character, human, plant, among others and animations to best match elderly users' preferences [10], how to map the emotional results from both models, DeepFace and SentimentPolarity, and how to aggregate the results into a final expression.

Another of the challenges encountered during the implementation was the delay in obtaining a response from DialogFlow after the elderly user interacted with Chatto. If this delay exceeds a few seconds, it can result in an unpleasant user experience, as the user may feel that the conversation is not flowing naturally. This challenge highlights the importance of minimizing response times in chatbots to ensure that the user remains engaged and satisfied with the interaction.

6 Conclusions and Future Work

This paper presented Chatto, an interactive and empathetic avatar designed to enhance the emotional well-being of elderly users in AAL environments. Chatto offers easy interaction through a user-friendly and intuitive design, multisource message interpretation by processing messages from sources like chatbot conversations, RGB cameras, microphone audio, and personalized animations that allow caregivers to customize the avatar's behavior.

During the implementation phase, a 3D animated avatar was developed, integrated facial emotion recognition, and voice sentiment analysis, and designed a user interface optimized for readability, ease of use, and accessibility. Some challenges were encountered, such as selecting an appropriate visual style, mapping the emotional results from both the DeepFace and SentimentPolarity models, and aggregating the results into a final personalized expression. Additionally, response time delays from DialogFlow posed a challenge, emphasizing the need to minimize response times for a more natural conversation flow.

Following the implementation, comprehensive usability tests were conducted with a diverse group of elderly users in an AAL setting, in which quantitative measures were employed to evaluate the user-friendliness, intuitiveness, and overall effectiveness of Chatto. The participants provided valuable feedback on their experiences interacting with the avatar, including any difficulties they faced and suggestions for improvement. Multiple tests and iterations allowed us to refine the avatar's performance, leading to a majority of users finding Chatto enjoyable, easy to navigate, and emotionally supportive. Caregivers also appreciated the real-time monitoring feature, which facilitated more timely and responsive care.

Future research and development efforts will focus on addressing the challenges encountered during the implementation: decreasing the time delay between the user interaction and Chatto response from DialogFlow by training NLP models to be used on the device. Furthermore, to ensure a more informed choice of weights for the weighted sum formula, experts could be contacted to integrate knowledge from other areas, such as psychology and tests regarding the formula could be performed to assess its effectiveness.

Acknowledgements. This work was supported by national funds through the Portuguese Foundation for Science and Technology (FCT), I.P., under the project UIDB/04524/2020 and was partially supported by Portuguese National funds through FITEC-Programa Interface with reference CIT "INOV-INESC Inovação-Financiamento Base". The authors would also like to acknowledge FlatIcon for providing the icons used in the architecture diagram image.

References

1. Ageing and health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health
2. Towards the NESTORE e-coach: a tangible and embodied conversational agent for older adults. In: Proceedings of the 2018 ACM International Joint Conference. Singapore. https://doi.org/10.1145/3267305.3274188
3. Virtual Companion Based Mobile User Interface: An Intelligent and Simplified Mobile User Interface for the Elderly Users, shorttitle = Virtual Companion Based Mobile User Interface, doi = 10.1109/ISUVR.2017.10, author = Noh, Sangwoo and Han, Juyeon and Jo, Jinwoo and Choi, Ahyoung, month = jun, year = 2017, keywords = Animation, intelligent user interface, Meteorology, Mobile communication, Monitoring, Senior citizens, Speech recognition, User interfaces, Virtual companion, visualization in 3D, pages = 8-9
4. Callan, J.: Homework: Sentiment analysis (2016). http://boston.lti.cs.cmu.edu/classes/95-865-K/HW/HW3/
5. Chang, Y.C., Hsing, Y.C.: Emotion-infused deep neural network for emotionally resonant conversation. Appl. Soft Comput. **113**, 107861 (2021). https://doi.org/10.1016/j.asoc.2021.107861
6. Muhammad, A.F., Susanto, D., Alimudin, A., Adila, F., Assidiqi, M.H., Nabhan, S.: Developing English conversation chatbot using dialogflow. In: 2020 International Electronics Symposium (IES), pp. 468–475 (2020). https://doi.org/10.1109/IES50839.2020.9231659
7. Pinto, M.D., Greenblatt, A.M., Hickman, R.L., Rice, H.M., Thomas, T.L., Clochesy, J.M.: Assessing the critical parameters of eSMART-MH: a promising avatar-based digital therapeutic intervention to reduce depressive symptoms. Perspect. Psychiatr. Care **52**(3), 157–168 (2016). https://doi.org/10.1111/ppc.12112
8. Sansen, H., et al.: The Roberta IRONSIDE project a cognitive and physical robot coach for dependent persons. Modell. Meas. Control C **77**(2), 169–181 (2016)
9. Serengil, S.I., Ozpinar, A.: Lightface: a hybrid deep face recognition framework. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 23–27. IEEE (2020). https://doi.org/10.1109/ASYU50717.2020.9259802
10. Shaked, N.A.: Avatars and virtual agents - relationship interfaces for the elderly. Healthc. Technol. Lett. **4**(3), 83–87 (2017). https://doi.org/10.1049/htl.2017.0009. https://onlinelibrary.wiley.com/doi/abs/10.1049/htl.2017.0009
11. Wan, Y.T., Chiu, C.C., Liang, K.W., Chang, P.C.: Midoriko chatbot: LSTM-based emotional 3d avatar. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), pp. 937–940 (2019). https://doi.org/10.1109/GCCE46687.2019.9015303. iSSN: 2378-8143



Optimization of Traffic Light Controllers Using Genetic Algorithms: A Case Study in the City of Cádiz

Sara Balderas-Díaz(✉), Andrés Muñoz, and Gabriel Guerrero-Contreras

Department of Computer Science and Engineering, University of Cádiz, Avda. de la Universidad de Cádiz 10, Puerto Real, 11519 Cádiz, Spain
{sara.balderas,andres.munoz,gabriel.guerrero}@uca.es

Abstract. Traffic congestion is a significant problem in many cities, leading to delays, increased fuel consumption, and environmental pollution. In recent years, optimization algorithms have been used to improve traffic control systems and reduce congestion. This paper proposes an optimization algorithm based on genetic algorithms to optimize traffic signal controllers in Cádiz, Spain, using the Simulation of Urban MObility (SUMO) traffic simulator. The proposed optimization algorithm is a promising solution for the optimization of traffic control systems, which can lead to more efficient traffic flow, reduced congestion, and improved environmental conditions.

Keywords: Genetic Algorithms · Traffic Light Controllers · Optimization · SUMO

1 Introduction

Traffic congestion is a major problem in many cities around the world, causing delays, increased fuel consumption, and environmental pollution [8]. One alternative to mitigate traffic congestion is the proper use of traffic control systems, such as traffic lights and the adjustment of traffic light timing. However, traditional traffic control systems are often not optimized, leading to inefficient traffic flow and increased congestion.

In recent years, there has been growing interest in the use of optimization algorithms to improve traffic control systems [1]. These algorithms can be used to find optimal traffic signal timings, which can reduce congestion and improve traffic flow. In this context, the present paper proposes an algorithm based on genetic algorithms to optimize the traffic signal waiting times at traffic lights in the city of Cádiz, Spain, and to reduce CO_2 emissions from road traffic. To evaluate the proposed algorithm, the Simulation of Urban MObility (SUMO) traffic simulator [3] has been used. The road network of Cádiz has been designed based on OpenStreetMap data, and the traffic lights have been manually added using the SUMO netedit tool.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 103–112, 2023.
https://doi.org/10.1007/978-3-031-43461-7_11

This paper is structured as follows: Section 2 provides an overview of previous proposals that have tackled this problem. Section 3 explains the materials utilized in this work. Section 4 outlines the methodology applied in this study, including the proposed optimization algorithm's particulars. The experimental results obtained are presented in Sect. 5. Finally, Sect. 6 concludes the paper and suggests future research directions.

2 Related Work

The field of smart mobility has seen numerous advancements and innovations aimed at improving traffic flow and reducing congestion on roads. One such proposal is presented by [4], which introduces a mechanism that adapts itself for regulating traffic, utilizing cooperative agents. This mechanism focuses on reducing traffic congestion at intersections by using three types of agents: Viewagents, Trafficlightagents, and Intersectionagents. Viewagents are responsible for monitoring a specific view and detecting the number of cars in that area. Trafficlightagents control the traffic lights and adjust the duration of green and red lights accordingly. The adjustment of a particular traffic light's duration is managed by Intersectionagents through the use of Adaptive Value Tracking.

Another interesting proposal is presented by [12], where a system for preempting traffic lights is proposed specifically for emergency vehicles and public transportation. The main objective of this system is to minimize the average response time during emergencies and improve traffic flow at junctions. The authors evaluate the system using a simulator in real-life scenarios and explore how communication between multiple intersections can be used to minimize the time spent in traffic. Furthermore, [6] assesses the effectiveness of an emergency vehicle in a simulated traffic situation with traffic lights, using both V2X communications and the conventional siren. The research also proposes a collaborative algorithm that employs V2X communications for vehicles and traffic lights to act accordingly and minimize the delay in emergency response.

As for signal control methods, [7] introduces a signal control method for improving the control efficiency of signalized intersections based on deep reinforcement learning and value distribution. The approach uses a Distributional DQN algorithm to acquire knowledge of the estimated distribution of phase actions in the intersection setting. The authors of [15] introduce a technique for intelligent traffic signal control that combines multiagent reinforcement learning and Artificial Intelligence of Things through an information fusion approach. They treat each traffic intersection as an agent and utilize Q-learning to update the Q-factors of each agent, enabling the controllers to collaborate on signal control actions with adjacent intersections. Meanwhile, [2] presents a solution to the "green wave" problem, which involves synchronizing traffic signals on a route to enable vehicles to traverse multiple intersections without stopping. This solution reduces the green wave problem on busy routes to a Job Shop Scheduler problem, which can be solved using optimization techniques, with the objective of minimizing delays and enhancing traffic flow, especially during peak periods.

Moreover, [16] proposes a centralized route management solution for autonomous vehicles, which can effectively mitigate the effects of traffic congestion and improve traffic flow in cities. The authors conducted an experimental study in Valencia, Spain, which analyzed traffic congestion in an area with high traffic density. They found that their proposed solution resulted in significant benefits compared to the traffic flow on a typical day. Finally, [13] presents a methodological approach that uses microsimulation and refined Origin-Destination matrices to estimate greenhouse gas and pollutant emissions from traffic.

Overall, these studies contribute to the development of efficient traffic management systems that can improve traffic flow and reduce congestion, thereby improving the overall mobility of people and goods. Nevertheless, there is a lack of focus on optimizing traffic in cities where geographic conditions pose unique challenges. Cities located near marshes, isthmuses, or small peninsulas are especially affected, with limited space, winding and narrow roads, high population density, and other factors that lead to traffic congestion or prevent proper management.

3 Material

This section describes the traffic simulator tool and the traffic network employed for the experiments performed in this work.

3.1 Simulator

To perform the evaluation of the proposed system, the SUMO traffic simulator has been selected. It comprises a set of open-source tools for traffic simulation (vehicular, pedestrian, and public transport). It provides information related to travel times, traffic congestion, emissions, and road safety. In addition, it offers a set of tools for scenario creation, facilitating the simulation of realistic scenarios of existing cities. SUMO version 1.16.0 has been used in this work.

3.2 Traffic Network

Cádiz is a city in Spain, located in the autonomous community of Andalusia. It has 113,066 inhabitants in 2022, according to data provided by the Spanish National Institute of Statistics [10]. The city has a particular geographical location (Fig. 1). It is located on an island, linked by a tombolo to the island of León. It has an area of 12.3 km^2 and a population density of 9192 inhabitants/km^2. There are only three entrances to the city, two bridges that connect it with the municipality of Puerto Real and a road that connects it with the city of San Fernando.

Within the city, there are a total of 13,022 daily trips by private vehicles in 2022, according to the Spanish National Institute of Statistics [9]. The city has numerous junctions regulated by traffic lights. These are controlled by the

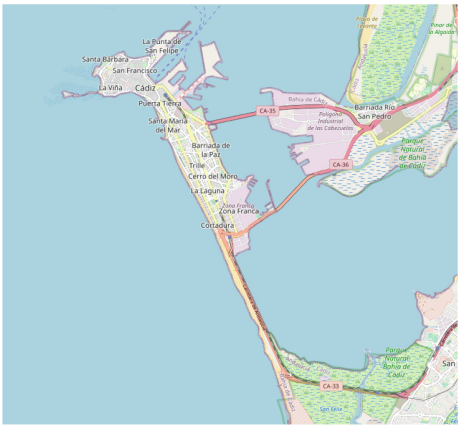


Fig. 1. Map of the city of Cádiz. Source: [11].

Traffic Control Centre of the Cádiz City Council [5]. This center has a network of the vehicle and pedestrian detectors that provide real-time information on the state of traffic. Furthermore, the network includes thirteen traffic cameras installed in the main avenues of the city and several vehicle counting points with electromagnetic detectors.

4 Methodology

This paper aims to evaluate the proposed system using the SUMO traffic simulator, and the city of Cádiz, Spain as a case study. In this section, the methodology of the study is presented, including the design of the road network and the optimization algorithm.

4.1 Design of the Road Network

The road network has been designed using the information provided by OpenStreetMaps. For this purpose, the `netconvert` tool has been used, which allows importing the information provided by OpenStreetMaps about a region and convert it to a SUMO road network. Although the OpenStreetMaps information is considerably complete, there are several errors that have been corrected manually through the `netedit` tool. Specifically, it has been necessary to include the city's traffic light network. A total of 223 traffic lights have been included, focusing specifically on the city's main avenues and streets. It should be noted that Cádiz counts with a historic center in which traffic is restricted.

4.2 Design of the Optimization Algorithm

The problem of optimizing traffic light signals is a complex and challenging task. It involves balancing multiple variables and constraints, such as traffic volume

and road network layout. One of the key challenges is coordination, as traffic signals must be timed and synchronized to ensure smooth and safe traffic flow along an entire route or network. To address the problem, an approach based on genetic algorithms has been chosen.

In SUMO, all traffic lights at the same junction are controlled by the same program (i.e., controller). This is because, for the safety of the users, the traffic lights must be synchronized with each other. In other words, traffic lights controlling conflicting lanes cannot be green at the same time.

The traffic light controller is defined in the SUMO network file, an XML configuration file. This file defines the road network for a given scenario. The network file contains information about the geometry of the road network, including the location, shape, and properties of individual road segments, intersections, and traffic signals. The logic of a specific traffic light is defined through a `tlLogic` element, which comprises a set of phases. Each of the phases, defined with the `phase` element, has a duration and a set of associated states for each of the traffic lights at the junction. `'G'` indicates that the traffic light will remain green, `'g'` that the vehicle can turn left, `'r'` that the traffic light will be red, and `'y'` that it will be yellow.

```
<tlLogic id="1341845187" type="static" programID="0" offset
    ="0">
    <phase duration="40" state="GGrr"/>
    <phase duration="5"  state="yyrr"/>
    <phase duration="40" state="rrGG"/>
    <phase duration="5"  state="rryy"/>
</tlLogic>
```

Listing 1.1. Definition of the traffic light controller of a junction in SUMO.

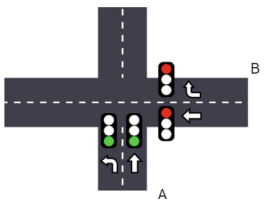


Fig. 2. Example of a traffic light controlled junction (Color figure online).

In Listing 1.1, the example of a traffic light controller for a four-lane one-way junction is given (see Fig. 2). The controller consists of four phases. In the first phase, the traffic lights of the two lanes of road A in Fig. 2 will be green, while the two lanes of road B will be red. This is indicated by the property `state="GGrr"` in the XML file. This phase will last for 40 s. Subsequently, for 5 s the two traffic lights on road A will turn yellow, warning drivers of a change. After this, in the third phase, it will be the lanes on road B that will be green for 40 s, followed by a 5-second yellow phase. This cycle will be repeated indefinitely.

The proposal adheres to the basic structure of genetic algorithms (Algorithm 1). It begins with the initialization of a population of potential solutions, which are randomly generated. Each of the traffic light controllers will respect the four phases of the controller, but the duration of each phase will be randomly assigned between 1 and 100 s. The status of the traffic lights will also be randomly generated within the possible combinations that do not cause conflicts at the junction. The initial population has been set at 100 candidate solutions.

SUMO is used for the evaluation of candidate solutions (Fig. 3). The traffic light configuration is written to the traffic network file and the simulation is launched. SUMO will provide information regarding the CO_2 emission of the vehicles in the simulation. The solution fitness will be determined by summing the emissions of all vehicles.

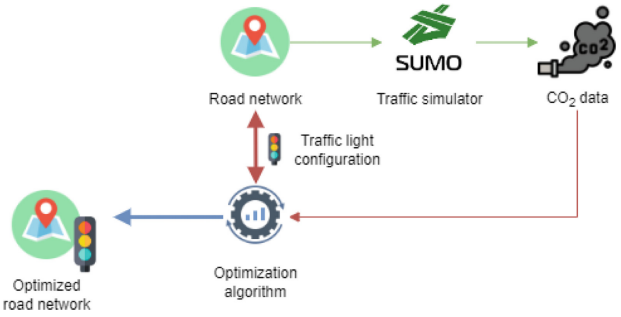


Fig. 3. Overview of the optimization process.

From the resulting population, the top 10% will be selected to generate new candidate solutions. To that end, two candidates will be randomly selected from them, and for each traffic light controller, the configuration of one of the parents will be selected to generate a child solution. This child solution will have a mutation probability of 30%. In this mutation, for each traffic light controller, the duration of each phase will be modified with a random value between -10 and 10. The state of one of the traffic lights will also be randomly permuted within the constraints of the junction.

4.3 Vehicle Fleet

The simulated vehicles are of the HBEFA3\PC_G_EU4 class. This class represents passenger cars with petrol engines and emissions corresponding to the Euro 4 level of the European standard [14]. According to this standard, these are vehicles manufactured from 2006 to 2009 with a CO_2 emission of 0.5 g/km. It should be noted that these values have been maintained for vehicles manufactured in subsequent years.

Algorithm 1. Pseudocode of the genetic algorithm.

```
function geneticAlgorithm(population)
    populationSize ← size(population)
    newPopulation ← []
    bestFitness ← getAverageFitness(population)
    stallCount ← 0
    maxStallCount ← 10
    while stallCount < maxStallCount do
        while size(newPopulation) < populationSize do
            eliteSize ← populationSize × 0.1
            elite ← selectElite(population, eliteSize)
            newPopulation ← newPopulation + elite
            parent1 ← randomSelection(population)
            parent2 ← randomSelection(population)
            child ← crossover(parent1, parent2)
            if random() < 0.3 then
                child ← mutate(child)
            end if
            childFitness ← evaluateFitness(child)
            newPopulation ← newPopulation + [(child, childFitness)]
        end while
        population ← newPopulation
        newPopulation ← []
        newFitness ← getAverageFitness(population)
        if newFitness > bestFitness then
            bestFitness ← newFitness
            stallCount ← 0
        else
            stallCount ← stallCount + 1
        end if
    end while
    return population
end function

function selectElite(population, eliteSize)
    sortedPopulation ← sort(population, descending order of fitness)
    return sortedPopulation[: eliteSize]
end function
```

In the simulation, 10.000 vehicles have been introduced, to approximate the daily traffic of the city according to the Spanish National Institute of Statistics. Vehicle routes have been randomly generated through the `randomTrips` utility.

5 Results

The present study reveals that the application of an optimization system results in a noteworthy decrease of 55.43% in CO_2 emissions compared to the default

solution generated in SUMO. The optimization system displays fast convergence within a mere 25 generations, as shown in Fig. 4.

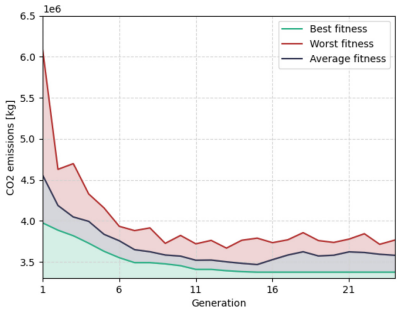


Fig. 4. Convergence of the population in the optimization system.

It is found that the total CO_2 emissions from vehicles in the optimized scenario with traffic signal controllers are notably lower, i.e., 3,374,990 kg, as opposed to 6,088,216 kg in the scenario without optimized controllers. This is attributed to the protracted waiting time of vehicles at traffic signals.

Further analysis of emissions by zones reveals that areas with a higher number of lanes, such as the avenues, and the large intersections where different avenues of the city converge are the most adversely impacted by emissions. However, the optimized scenario significantly curtails CO_2 emissions at these critical points, as demonstrated in Fig. 5.

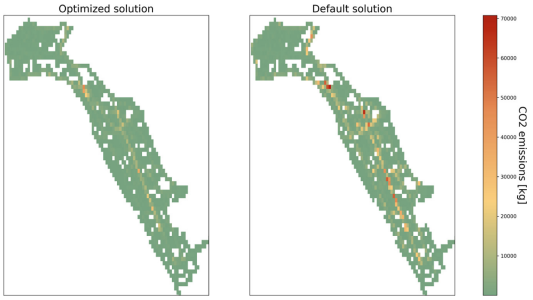


Fig. 5. Map of CO_2 emissions in the city of Cádiz. Comparison of the best configuration obtained versus the default configuration.

6 Conclusions

In this paper, an optimization algorithm based on genetic algorithms has been proposed to optimize the traffic signal controllers in the city of Cádiz, Spain. The

experimental results show that the proposed algorithm significantly reduces congestion and improves traffic flow. The proposed algorithm can adapt to changing traffic conditions and find optimal signal timings, thereby reducing delays, fuel consumption, and environmental pollution. This approach is applicable to any road network, regardless of the particular city under consideration. The only requirement for implementation is the creation of the road network within the simulation framework. The use of the SUMO traffic simulator has facilitated the design and evaluation of the proposed algorithm. The road network of Cádiz has been designed based on OpenStreetMap data, and the traffic lights have been manually added using the SUMO netedit tool.

As a future plan, we suggest applying the implemented method to real traffic data from the city of Cádiz, including information about public transport and pedestrian traffic. Additionally, the approach used may be expanded to optimize the current crossing control methods (traffic lights, roundabouts, stop or yield signs) and the direction of the streets. The use of crowdsensing applications will be also analyzed to detect traffic problems in real-time through online social networks.

Acknowledgments. Financial support for this research has been provided under grant PID2020-112827GB-I00 funded by MCIN/AEI/10.13039/501100011033. It is also partially granted by the "EMERGIA" programme, funded by the Junta de Andalucıa through the grant EMC21 004171.

Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.

References

1. Abbasi, M., Shahraki, A., Taherkordi, A.: Deep learning for network traffic monitoring and analysis (NTMA): a survey. Comput. Commun. **170**, 19–41 (2021)
2. Almeida, A.M., de Macêdo, J.A., Machado, J.C.: Optimization of urban semaphore times turning into jssp. In: BiDu-Posters@ VLDB (2018)
3. Behrisch, M., Bieker, L., Erdmann, J., Krajzewicz, D.: Sumo-simulation of urban mobility: an overview. In: Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation. ThinkMind (2011)
4. Belbachir, A., El Fallah-Seghrouchni, A., Casals, A., Pasin, M.: Smart mobility using multi-agent system. Procedia Comput. Sci. **151**, 447–454 (2019)
5. Cádiz City Council: Traffic Delegation (2022). https://institucional.cadiz.es/. Accessed 22 Mar 2023
6. Gama, O., et al.: Performance evaluation of an emergency vehicle in a simulated traffic scenario. In: 2018 16th International Conference on Intelligent Transportation Systems Telecommunications (ITST), pp. 1–7. IEEE (2018)
7. Li, J., Chen, T., Zhou, F., Lv, X., Peng, F.: Research on signal control method of deep reinforcement learning based on value distribution. In: Journal of Physics: Conference Series, vol. 2330, p. 012019. IOP Publishing (2022)
8. Lu, J., Li, B., Li, H., Al-Barakani, A.: Expansion of city scale, traffic modes, traffic congestion, and air pollution. Cities **108**, 102974 (2021)

9. National Statistics Institute. Spanish Statistical Office: Daily mobility: Persons aged 16 and over by main mode of transport used for commuting to work/study by sex and age group (2022). https://www.ine.es/dynt3/inebase/es/index.htm?padre=8983&capsel=8987. Accessed 22 Mar 2023
10. National Statistics Institute. Spanish Statistical Office: Official population figures of Spanish municipalities in application of the 'Ley de Bases del Régimen Local' (Art. 17) (2022). https://www.ine.es/jaxiT3/Datos.htm?t=2864. Accessed 22 Mar 2023
11. OpenStreetMap contributors: Planet dump (2023). https://planet.osm.org. https://www.openstreetmap.org
12. Petrică, B.G., Ciobanu, R.I., Dobre, C.: Automatic traffic light preemption for intelligent transportation systems. In: 2021 20th International Symposium on Parallel and Distributed Computing (ISPDC), pp. 1–8. IEEE (2021)
13. Quichimbo, J.E.L., et al.: Estimation of green house gas and contaminant emissions from traffic by microsimulation and refined origin-destination matrices: a methodological approach. In: SUMO Conference Proceedings, vol. 1, pp. 27–37 (2020)
14. Singh, S., Kulshrestha, M.J., Rani, N., Kumar, K., Sharma, C., Aswal, D.: An overview of vehicular emission standards. MAPAN **38**, 1–23 (2022)
15. Yang, X., Xu, Y., Kuang, L., Wang, Z., Gao, H., Wang, X.: An information fusion approach to intelligent traffic signal control using the joint methods of multiagent reinforcement learning and artificial intelligence of things. IEEE Trans. Intell. Transp. Syst. **23**(7), 9335–9345 (2021)
16. Zambrano-Martinez, J.L., Calafate, C.T., Soler, D., Cano, J.C.: Towards a centralized route management solution for autonomous vehicles. In: 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE (2019)

WoA: An Infrastructural, Web-Based Approach to Digital Agriculture

Stefano Chessa[1], Giovanna Maria Dimitri[2], Marco Gori[2], and Alexander Kocian[1(✉)]

[1] Department of Computer Science, University of Pisa, Pisa, Italy
{stefano.chessa,alexander.kocian}@unipi.it
[2] Department of Information Engineering and Mathematics (DIISM), University of Siena, Siena, Italy
{giovanna.dimitri,marco.gori}@unisi.it

Abstract. The Agritech project is a large Italian project funded with the National Recovery and Resilience Plan to contribute, among other objectives, to the digitization of Agriculture in Italy and to the implementation of an information technology platform and its web-based abstraction (called Web of Agriculture - WoA) that will support research, production and commercial activities in the agrifood sector in Italy, by leveraging on innovative artificial intelligence tools. We present here the design of the platform and of the WoA and its respective case studies.

Keywords: Digital Agriculture · Data space · Web of Agriculture

1 Introduction

The European Green Deal [2] is the response of the EU to the call for a sustainable development of the United Nations 2030 Agenda [11] and it sets, among others, specific goals for the agricultural sector to achieve climate neutrality by 2050. These goals address all sectors of agriculture and give a key role to the adoption of digital technologies. The objectives of the European Green Deal fall on the partner countries that are responsible for their implementation. In the case of Italy, this is included in the National Recovery and Resilience Plan (PNRR) and, in the specific case of the agrifood sector, implemented by the National Research Centre for Agricultural Technologies (AGRITECH) [6].

The AGRITECH center itself is structured into 9 Spokes, each covering a thematic area, and among these Spoke 9 addresses "new technologies and methodologies for traceability, quality, safety, measurements and certifications to enhance the value and protect the typical traits in agrifood chains". It is within this Spoke that rises the idea of building a technological platform, called METRIQA, to support the digitization of the agrifood sector in Italy.

The main purpose of METRIQA is tow-fold. On one hand it supports the entire chain of data and information production, storage and analysis typical of an intelligent environment. The data, mostly in form of experimental datasets

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 113–122, 2023.
https://doi.org/10.1007/978-3-031-43461-7_12

(together with associated reports) produced by means IoT sensors and created by the on-site experimental activities of the researchers in the Spoke, are stored and processed by using AI tools offered by the platform to support the creation of innovative decision support systems. On the other hand, METRIQA implements the abstraction of the"Web of Agrifood" (WoA) which provides a web-like access to data (digital resources) and services to its users. According to this abstraction, WoA glues data lakes of different subjects by connecting their resources despite of their internal technology and structure, so that different providers may make available any kind of unstructured resources, either relevant for research or production and concerning the agrifood sector in Italy, and METRIQA user can rely on AI-based retrieval services to obtain relevant information about this sector. We expect that METRIQA and its abstraction of the WoA will contribute both to research and productive stakeholders of the agrifood sector. In particular, it will support researchers by the implementation of a "widespread laboratory" offering them the opportunity for collecting and sharing digital resources and state-of-the-art Artificial Intelligence (AI) tools, and a mean to make their results available to non-researchers stakeholders by leveraging on a web-style search engine that support integration and searches over unstructured data. At the same time, METRIQA will support private companies in the sharing of their digital resources by the creation of a decentralized data space supporting services like those related to traceability and certification. The integration in the same platform of resources coming from both the research and the productive worlds will also pave the ways for innovative agrifood services, for example those concerning the verification of the authenticity the genuineness of the products, as it is discussed in a use case Sect. 4.3.

2 Background and State of the Art

In this section we will briefly give an overview of the main references for the state of the art for what concerns the components of the WoA.

Digital information drives business innovation and transforms our daily life into prosperity with an incredible speed. Data generated at exponential pace, is one driving force to achieve this goal [10].

Data platforms form a market place that connects data providers and data users. To depend less large, quasi monopolistic players, the EU strategy for data is on the way to adopt multiple data spaces until 2027 Data spaces are a distributed data integration concept which is taken up by consortia. The open, transparent and secure digital ecosystem GAIA-X and International Data Spaces (IDSs) specify reference architectures for distributed data infrastructures and data spaces, respectively. International Data Spaces (IDS) enable the sovereign and self-determined exchange of data via a standardized connection across company boundaries. They address the many challenges such as interoperability, transparency, trust and security of data. The IDS develops a standardized open-source reference architecture model [9]. Similarly, the association GAIA-X with more than 334 members across 20 countries aims to increase the adoption of

cloud services and accelerate data exchange by European businesses. To foster cybersecurity, the Information Security (NIS) directive and the European cybersecurity framework are the pillars on top of the requirements on technical and organizational measures related to personal data. Finally, we want to point out that the combined use of IDS technology and blockchain enables supply chain traceability.

Focusing on data space technologies [10], semantic World Wide Web Consortium (W3C) standards, such as the resource description framework (RDF) and Web Ontology Language (OWL) have been developed over two decades, to ensure a common understanding of the data to be shared in the data space. The IDS information model is based on a graph-based RDF ontology, to connect the conceptual functionalities, roles, and processes with the implementations in the Connector interfaces and endpoints. MY DATA Control Technologies can be sued to implement the ability of data sovereignty. To share data at the same time, there must be some sort of usage control enforcement. This concept builds on a certification process for components and environments based on a public key infrastructure. With data federation, integration takes place on the fly, Data is only accessed and integrated in the moment when the consumer ask for. The rise of trustworthy federated digital ecosystems will not only improve data availability but also handle the large volume of training data required to train machine learning and artificial intelligence, which are crucial for the competitiveness of entire business branches [10].

Over the past few years, several EU H2020 infrastructure projects aimed at digitize agriculture. Among them is the project "DESIRA" comprising an intervention planning platform along with other elements of the infrastructure (sensors, drones, geographic information system). The objectives are to i) reduce response times to citizens requesting interventions, ii) enhance the traceability and security aspects of the products, iii) foster communication, exchange of knowledge and joint activities, and iv) to provide training opportunities using gamification techniques [1]. "OPEN DEI", in contrast, proposes a 6C reference architecture framework for cross-domain digital transformation [4]. Finally, "i4Trust" lets share vineyard data with the wine value chain. Wine consumers, cellar personnel, viticulture advisors or, marketing companies can access relevant, secure, and voluntarily shared data from the vineyard [3]. To the best knowledge of the authors, this is the only solution on data spaces for agriculture at Technical Readiness Level (TRL) 9 at the time of writing.

3 The Agritech Project

Agritech, the National Center for the development of New Technologies in Agriculture, is one of the 5 National Centers dedicated to frontier research financed through the MUR by the PNRR Plan as part of Mission 4 *"Education and research" - Component 2 "From research to business"*. The project objective is to promote climate change adaptation, reduction of the environmental impact in agrifood, marginal areas development, safety and traceability of the supply

chains. Santa Chiara Lab (University of Siena) was selected by the Federico II University of Naples, National Center hubs leader, among the 9 Spokes, research nodes, to deliver the project objectives, jointly with the main agrifood companies sector in Italy.

Spoke 9 goal is to conduct research activities on *"Measurement, certification and enhancement of the quality, origin, typicality, safety and sustainability of products, processes, supply chains and agrifood companies"*, and was assigned an overall budget of around 40 million euro[1].

We will now describe the Spoke 9 objectives for the WP5. When developing enabling Information and Communication Technologies (ICT) to be applied for the AGRITECH world, it is important to understand how networking policies will impact production processes, system durability and resilience. Each solution will have to deal with the operating conditions, which could be typically described as extreme. Objectives of the projects will be minimizing the number of devices in the field, obtaining the absence of inline calculation, absence of local orchestration nodes, minimization of "intelligent" components allocated directly to the sensor nodes, off-grid and low-power energy production, use of low-cost field components, sustainable life cycle assessment and certification of operation in extreme conditions. Two main application areas are present in this field of research. The first one concerns the cultivation systems in a protected environment (horticultural and nursery plants in greenhouses). The second one relates to the open field cropping systems (horticulture and cash crops). The process phases, in fact, controllable by Decision Support System (DSS)-type reactive agents are characterized by certain processes to be checked (for instance irrigation at greenhouse level, single master mix fertigation, fertigation with mother mixes and a high content of single macroelements, etc) [7]. Variability in terms of measurements and open field environment should induce some considerations on the networking model to be used. For instance, the need of sending measurement data from medium/low frequency sensors (f.i. every hour). Further tasks might entail sending ending of request for state change by high frequency actuator nodes (minutes), medium/high data packet size (kbytes), local decision making processes only in case of network absence and for limited times (hours), very low network coverage distances (metres) and low operating powers (12/24 V). Same considerations could be done for DSS-type reactive agents, in the open field. Here, processes such as precision irrigation, fertigation and monitoring might be needed. Moreover problems related to energy supply and signal coverage might become critical since the sensor nodes typically operate on an island basis and should not be connected to power energy networks.

4 The METRIQA Platform Design and Architecture

4.1 METRIQA and the Web of Agrifood Abstraction

The initial objective of METRIQA is to provide a mean to handle all the data of the AGRITECH center by easing the storage and sharing of non-structured, het-

[1] https://santachiaralab.unisi.it/agroalimentare/progetto/agritech-spoke-9.

erogeneous information (hereafter *d*igital resources) produced by all the project partners, but it is expected to grow, progressively enriched by other resources data and meta-data that can either created by experts or by the platform AI-engine, and by services built over those data and made available to all stakeholders (researchers, citizens, private companies, public bodies etc.), so to become a permanent asset in the Italian agrifood sector even after the end of the PNRR financial support. The key aspects of METRIQA are:

- It is neutral with respect to stakeholders: the platform is maintained by the center as a support to the stakeholders operating in the sector but no one of them controls it.
- It is technology neutral: operators in the Agrifood sector may connect their Information technology (IT) systems to the platform irrespective of their inner technologies.
- it provides a universal repository of unstructured, heterogeneous agrifood data (such as reports, documents or datasets), as those that are expected to be produced by the experimental activity of the center and by agrifood industries in Italy that enter into the platform.

To achieve all these objectives METRIQA adopts a model that closely reminds the spirit of the web: it is completely decentralised, and it makes use of most recent advances in AI to enrich with metadata the digital resources in the repository so to enable web-style search/anwser engines and the construction of application-specific AI-based decision support systems. In other terms, METRIQA provides the mean to implement the so-called "Web of Agrifood" (WoA) abstraction.

It should be observed that the need for handling unstructured digital resources comes from the need of supporting the activities of a large number of research laboratories and companies operating in the agrifood sector, each with its own objectives, tools, procedures and technologies, for which there is no standard and perhaps a common agreement on the structure of these resources and on protocols and procedures is not even possible, at least in the short/medium term. On the other hand, the idea of the project is to invest instead on artificial intelligence technologies to provide the needed structure and connection among these resources to overcome the problem by letting the machines to deal with that, as it currently happens on the web, especially considering the last developments in conversational agents such as Chat GPT [8] or Bart [5].

4.2 METRIQA Architecture

From an architectural point of view, the METRIQA platform comprises one *authority* node and several, decentralized nodes one of which is the *public* node developed and maintained by the Center itself). It also defines four roles for the participants, namely service (and data) providers, end users, WoA nodes managers, and a governance body (WoA authority). A high-level view of the platform is represented in Fig. 1. The nodes may be to servers (or clusters of

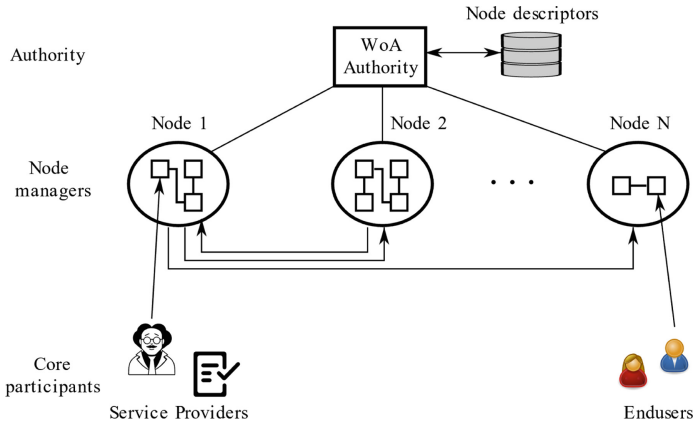


Fig. 1. A high-level view of the METRIQA platform.

servers) on the cloud or even at the edge of the network that are run and managed by any public or private entities. Each node provides hosting for data of the participants that belong to that node and possibly a set of services of the WoA. There is not a standard for the nodes, as they may host any arbitrary set of services and they may be implemented by using different technologies. In general, the services hosted by a node respond to the needs of its participants. In the case of the public node managed by the center, it offers hosting of the research data produced by the center, with services suitable to let the researchers to upload, share and publish their data, and private workspaces to implement AI-based analysis of the data. At the same time, the public node also hosts the search engine (which is publicly available) that operates over all public data of the entire platform. Private nodes may implement additional services specific for companies or production. For example, companies along a production and distribution chain may share their traceability data, even if each of them operates with its own systems and technologies, so to implement on top of this food traceability services, "from grower to consumer". The implementation of such services will require the integration of data from different sources and exported by means of different services (in Fig. 1 such services are represented by square boxes within the nodes).

Even if there are no strict requirements on the nodes, they should be registered at the centralized authority node that plays a role similar to the Domain Name Server on the Internet, but at a higher level. The authority node keeps a database of descriptors of the nodes in the platform that, when relevant, contains also a specification of the services that are present in the the nodes (e.g. a publishing services for their digital resources). The information about how to reach the resources published by each node, which is available in the authority node, is used by the search engine to retrieve the public resources of the platform and to respond to the relative queries. Concerning the description of the services of each node, we consider three cases. The first is that of publishing digital resources

service, which happens in a web-style way, and for which the descriptor is just the conventional one for the web (address of the web server and port, while the protocol is the usual HTTP). There may be however proprietary services for which the description is not available/public, and in this case the authority node just keeps an informal, textual description of those services, along with a reference to the address and port of the server. The third case is that of services that are meant as public, but that require specific ways of interaction among clients and servers. For these we adopt the same approach of the International Data Spaces initiative (IDS) [9], so that, in parallel to the web-style access to digital resources METRIQA also offers an interface compliant with that of IDS.

Concerning the roles, they reflect different classes of participants to the platform, where a participant may take more than one role at a time.

The service and resources providers are those that populate the platform with resources or services. For example, they can be researchers creating new knowledge in the agrifood sector that they publish in the WoA in form of reports and datasets, or private companies in the agrifood sector that offer public resources (e.g. food traceability data), or services to other companies, but they can also be independent professionals (e.g. agronomists) or even public bodies.

The end users are any potential stakeholder of the WoA. They may be just citizens, researchers, companies operating at any level in the agrifood sector, professionals, public bodies, etc. They are consumers of data, information and services offered by the WoA, and they interact with it either by using the search/answer engine of WoA, or by connecting directly to the interfaces of the services offered by the platform.

The node managers are entities running nodes. In the case of the public node the manager is the center itself, but managers of other nodes may be IT companies with an interest in the agrifood business that run their WoA node to host their own services or even resources or services provided by third parties (e.g. independent service providers). In some cases node managers may even be individual researchers hosting (possibly a limited number of) services and resources in their own servers.

Finally, the WoA authority is the Agritech center and runs the Authority node. It receives request of accreditation of new nodes of the platform and updates accordingly the database of node descriptors.

As discussed above, beyond their ability of publishing in web-style their resources there are no other technical requirements on the nodes. On the other hand, it is possible instead to provide a functional description of the public node managed by the Agritech center. Specifically, this node plays two roles: on one hand it implements the web-style search engine over the entire resources available in the platform. On the other hand, it provides services to research in terms of an individual workspace of researchers to host their digital resources, to share them with other researchers to create the so-called "widespread laboratory", to process them by using AI engines available in the platform and finally to publish in web style their resources so to make them publicly available in the WoA. A sketch of the functional components of the public node is shown in Fig. 2.

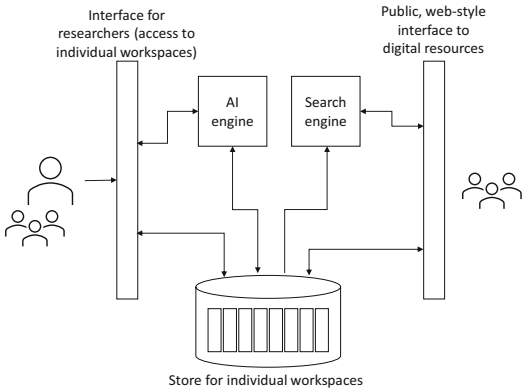


Fig. 2. A sketch of the functional components of the public node.

4.3 State of the Project and Use Cases

The first part of the project has been devoted to the definition of requirements and use cases related to both research and production, based on interviews with researchers and companies operating in the sector. So far, we had 31 meetings including 7 with individual researchers and 5 with companies, to filter out the requirements. This activity is still undergoing and the first tangible result of the project is expected in September 2023 with the delivery of the public node of the platform with an initial set of services. Most of the interviews with researchers aimed at finding research-related use cases put emphasis on the fact that researchers have, in general, very well established methodologies in terms of acquisition of data, in their analysis even based on complex machinery's as it is the case for the chemical and biological analysis of samples, and for the reporting and the publication of results. At the same time, being non-IT experts, they look with interest at the exploration of AI technologies provided this results in a complementary with respect to the research activity that they normally carry out. Beyond confirming the need for a platform that deals with unstructured and heterogeneous data, these interviews highlighted the need of services to upload data in spots or continuously, even in streaming from Internet of Things sensors, and cooperatively among researchers from different side. Interviews with companies instead highlighted a strong interest in food traceability services and for the certification of chains. Under this respect many companies expressed an interest in blockchain technologies.

To illustrate one of the expected behaviour of the platform we briefly report here a research-oriented use case. Some teams of researchers scattered in different locations in Italy (e.g. the Chianti area of Tuscany or area of Squinzano in Apulia) perform sampling of the soil and biological material (fruits, leaves) to perform chemical analyzes in the laboratory to be associated with the wine supply chains of those locations and identify specific characteristics of each grape production site. These characteristics can be used to certify production and/or

to its quality and for the traceability of the products. Each team performs several samples over a long sampling campaign. Each sample is associated to the location, altitude and other metadata, and it is analyzed by different machines (e.g. spectrometers) along a series of steps along which different types of intermediate data (e.g. Electron Paramagnetic Resonance spectroscopy, HyperSpectral Imaging - HSI) are produced. Each team then produces a final report at the end of the campaign that reports in natural language the details of the campaigns, the procedure used for the analysis, and the results in terms of tables, graphs etc. Although all these data are treated in a digital format (digital resources) they have many different formats, that may even differ between different teams, and some of these information are available as unstructured information in the reports. Returning to our Chianti example, the HSIs are plugged into an machine learning algorithm that assesses grape variety, water and nutritional statuses, disease-causing viruses, and the grape maturity. All these data, determining the figure-of-merit for the quality of this particular wine, is available in the accompanying report. The researchers use the platform to upload their material, each in his own workspace, and decide which resources are to be shared with the other teams, and which documents or reports to make available publicly on their website. The intermediate data, which have been processed by using established methodologies, are however available in the platform to experiment other methods of analysis by using the AI tools available in the platform, even cooperating in this with other IT teams. Some months later, a wine center in another country has to test the authenticity of a wine delivery, which is supposed to be produced with grapes of the Chianti area. They access to the WoA search engine to look for methodologies to check the authenticity of wines of that area, and they discover both the methodology (type of chemical tests) and the parameters that the tested wines should meet. When the wine center wishes a report on quality of the particular delivery, Generative Pre-Trained Transformer (GPT) technology forms a reply.

5 Conclusions

The Agritech project is in its early stage, and consequently also the activity of design of the METRIQA platform and of its WoA abstraction is still in a preliminary phase, in which we collected and analysed interviews with various research and production stakeholders. The platform itself is designed as a scalable data space that can grow over time with the contribution of independent resources and service providers, following a data space model similar to that of IDS. At the same time, however, the platform provides the WoA abstraction that frees from the need of rigid, predefined structures of the digital resources by leveraging on advanced AI-based metadata enrichment tools and search engines. Future work will be the delivery of a public call for the implementation of the first "public node" to be made available to the project community on the short term, so to focus the work of the rest of the activity on the population of the platform with digital resources, on the development digital services for the agrifood sector based on of metadata enrichment tools and decision support systems.

Acknowledgements. This work is funded within the project AGRITECH Spoke 9 - Codice progetto MUR: AGRITECH "National Research Centre for Agricultural Technologies" - CUP CN00000022, of the National Recovery and Resilience Plan (PNRR) financed by the European Union "Next Generation EU".

References

1. Bacco, F.M., Ferrari, A., Brunori, G.: Co-design of technological solutions for agriculture and rural areas: methodology and cases for responsible innovation. In: IEEE 8th World Forum on Internet of Things (2022)
2. European Commission and Directorate-General for Communication: European green deal : delivering on our targets. Publications Office of the European Union (2021). https://doi.org/10.2775/373022
3. Goncharuk, A.: Wine value chains: challenges and prospects. J. Appl. Manage. Investments **6**(1), 11–27 (2017)
4. Kung, A., Gusmeroli, S., Monteleone, G., Dognini, A., Nicolas, L., Polcaro, C.: Reference architectures and interoperability in digital platforms. techreport, H2020 Open Dei (2022)
5. Lewis, M., et al.: BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.acl-main.703, https://aclanthology.org/2020.acl-main.703
6. Ministero dell'Università e della Ricerca (MIUR): Agritech - national research centre for agricultural technolgies (2022). https://www.mur.gov.it/sites/default/files/2022-10/Schedadiprogetto-CN2.pdf. Accessed 16 Mar 2023
7. Naud, O., et al.: Agricultural Internet of Things and Decision Support for Precision Smart Farming, Chap. Support to Decision-making, pp. 183–224. Academic Press (2020). https://doi.org/10.1016/B978-0-12-818373-1.00004-4
8. OpenAI: GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. https://openai.com/product/gpt-4. Accessed 16 Mar 2023
9. Otto, B., Steinbuss, S., Teuscher, A., Lohmann, S.: IDS reference architecture model (version 3.0). International Data Spaces Association (2019). https://doi.org/10.5281/zenodo.5105529, https://internationaldataspaces.org/download/16630/
10. Otto, B., ten Hompel, M., Wrobel, S. (eds.): Designing Data Spaces. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-93975-5
11. United Nations: The 2030 Agenda and the Sustainable Development Goals: An Opportunity for Latin America and the Caribbean. Santiago (2018)

Human-in-the-loop AAL Approach to Emotion Capture and Classification

Rafael Pereira[1], Carla Mendes[1], Roberto Ribeiro[1], José Ribeiro[1], and António Pereira[1,2(✉)]

[1] Computer Science and Communications Research Centre, School of Technology and Management, Polytechnic of Leiria, 2411-901 Leiria, Portugal
{rafael.m.pereira,carla.c.mendes,roberto.ribeiro, jose.ribeiro,apereira}@ipleiria.pt
[2] INOV INESC Inovação, Institute of New Technologies, Leiria Office, 2411-901 Leiria, Portugal

Abstract. In Ambient Assisted Living (AAL), emotion detection plays a crucial role in offering personalized and effective support to users. However, the lack of labeled data for emotion detection poses a significant challenge for developing intelligent solutions. To tackle this issue, we propose a human-in-the-loop (HITL) approach that engages users in acquiring and classifying their facial images, which can be utilized to augment existing emotion detection datasets. Our system is composed of three primary phases: facial image acquisition, emotion recognition model image classification, and user-based classification. By involving users in the data annotation process, we ensure the quality of the labeled data and also facilitate the collection of diverse and context-specific emotional expressions. We conducted experiments in a semi-supervised environment, where users catalogued frames according to the corresponding emotions, aiming to improve the model's prediction accuracy for their emotions. The experiments supported the introduction of the user in the image classification process, promoting the creation of reliable labeled for later model training.

Keywords: Ambient Assisted Living (AAL) · Emotion detection · Human-in-the-loop (HITL) · Data acquisition · Data annotation · Intelligent solutions · Human Labeling

1 Introduction

Ambient Assisted Living (AAL) has gained significant attention in recent years due to its potential to improve the quality of life of individuals with decreased autonomy, such as elders, people with disabilities, and those with chronic illnesses. Through technologies supported by the Internet of Things, AAL systems can provide personalized care and assistance, fostering independence and well-being. However, accurately detecting and responding to users' emotional states remains a significant challenge for AAL systems.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 123–132, 2023.
https://doi.org/10.1007/978-3-031-43461-7_13

Emotion detection plays a critical role in achieving these goals, as understanding users' emotional states enables more personalized and effective care. By recognizing and responding to users' emotions, AAL systems can adapt their assistance and support, resulting in better user satisfaction and improved quality of life. The importance of context-specific emotion data becomes evident in this scenario, as personalized emotion detection ensures that the AAL system can accurately recognize and respond to users' emotional states in diverse situations [3,5].

Various approaches have been employed to detect emotions [9,11], including:

- Physiological signals: Emotions inferred from physiological data such as heart rate, skin conductance, and body temperature. However, these methods typically require specialized equipment and continuous monitoring, which may not be practical or comfortable for users.
- Audio analysis: Emotions detected from speech patterns and sentences, tone, and pitch. While this method has shown promise, it can be susceptible to background noise and can be ineffective for individuals with speech impairments.
- Computer vision (CV): Emotion recognition (ER) through facial expressions has gained considerable attention recently. By analyzing facial features and movements, CV algorithms can identify emotions with relatively high accuracy. Nevertheless, CV can struggle with differentiating overlapping objects, demands meticulous calibration and consistent illumination, and is susceptible to lighting variations over time [15].

Data acquisition implies dealing with sensitive data, such as facial images and emotional states. Therefore, this creates some challenges: obtaining high-quality and diverse samples and addressing ethical concerns such as informed consent, biased data, security and privacy in personal data [8]. The data annotation process, known as labeling or classification, involves assigning meaningful labels to the collected data, which can be time-consuming and labour-intensive [4,7].

Despite the advancements in ER techniques, the lack of sufficient labeled data remains a significant challenge, particularly for AAL systems. This data scarcity hinders the development of accurate and reliable models, as Deep Learning (DL) algorithms typically require large amounts of labeled data for training. Furthermore, collecting diverse and context-specific emotional expressions is essential for ensuring the generalizability and effectiveness of ER models in AAL scenarios [7]. Human-in-the-loop (HITL) [6] approaches have emerged as a potential solution to these data acquisition and annotation challenges since these approaches ensure the quality of the labeled data by involving humans during the annotation phase [16]. Involving users in the data annotation process offers several advantages, such as increased accuracy in classification and reduced reliance on external classification systems [13].

To address this challenge and considering the mentioned benefits, we propose an emotion detection and labeling system that follows a HITL approach that leverages user engagement and the quality of labeled data, in terms of the

reliability and consequent precision of the ER model. Additionally, this approach facilitates the collection of diverse and context-specific emotional expressions, as users can provide insights into their personal emotional experiences confirming and manually classifying expressions. By overcoming the challenge of data scarcity and incorporating users' unique emotional experiences, our approach has the potential to support the development of more accurate and personalized ER models for AAL systems.

In this paper, we focus on the older population, as they often need continuous emotional monitoring to ensure their well-being. However, a potential challenge arises when considering the practicality of having older individuals classifying their own emotions. In cases where elders may have difficulties performing this task, external agents, such as caregivers, can be involved in the classification process [2]. As an already identified disadvantage, this approach may also require more effort and time from the users or caregivers when compared to automated AI-based classification methods [16].

The main contributions of this paper are as follows:

- Proposal of a HITL system for acquiring and classifying facial images in the context of AAL, focusing on elders.
- Proposal of an architecture that combines facial image acquisition, preprocessing, and storage with user-based classification to enable effective and personalized emotion detection in various AAL contexts.
- Proof of concept, consisting of the description, implementation and testing of a low-cost prototype for facial ER, supported with a platform located in the cloud.

The remainder of this paper is organized as follows: Section 2 provides information regarding related work in the field. Section 3 describes the proposed solution in detail. Section 4 discusses the implementation and testing of the solution. Section 5 presents the analysis and discussion of the results. Finally, Section 6 concludes the paper and offers directions for future work.

2 Related Work

In this section, we provide an overview of the topic through the presentation of related work regarding data acquisition and labeling in AAL or healthcare contexts.

The authors in [18] propose a privacy-preserving scheme for healthcare that uses a HITL approach. They designed a three-stage process: data preprocessing, model training, and inference. In the data preprocessing stage, the patient data is anonymized and encrypted using homomorphic encryption. In the model training stage, a privacy-preserving DL model is trained on the encrypted data using federated learning. The HITL assistance belongs in the inference stage, where the encrypted patient data is sent to a trusted human auditor who verifies the accuracy of the model's output. The HITL assistance provides an extra layer of

accuracy verification and privacy protection, making the proposed scheme more trustworthy and reliable for healthcare applications.

A control system for a meal assistance orthosis device was proposed in [17] to assist elderly or disabled people who have lost the ability to move their upper limbs to take meals by themselves independently. Their system splits into two components, the neural network (NN) and HITL. The neural network is responsible for learning the mapping between the sensory input signals from the orthosis device and the control inputs. Meanwhile, the HITL component involves a human operator that uses a graphical interface to provide real-time feedback on the status of the orthosis device. Therefore, the HITL component can modify the control inputs generated by the neural network as necessary based on real-time feedback from the orthosis device.

The authors in [10] describe a fall detection alert system to help support the independent living of elders in their own homes. Their system uses thermal vision cameras and an edge device for image processing and classification to detect falls. The system sends alerts to a remote monitoring station where human operators review and confirm the fall event before taking action. This HITL approach improves the accuracy and reliability of the system while preserving the privacy of the individuals being monitored and reducing the potential for false positives or negatives, which can be critical in a fall detection alert system.

Vorobyov et al. [14] devised a HITL interface to enhance wheelchair control. This cutting-edge interface utilizes electroencephalography, myoelectric sensors, and a gyroscopic-accelerometric mix, contributing to an improved user feedback system in real-time. An embedded artificial neural network works in unison with these components to understand and adapt to the unique behaviors of each user, thereby fortifying the effectiveness of the HITL approach.

3 The Proposed System

To address the identified issues, we propose a flexible and adaptable HITL emotion classification system composed of multiple modules.

The architecture of this system is depicted in Fig. 1, which illustrates the following steps: (1) capturing the frame using the Camera Module (A), (2) sending the frame to the Facial Recognition module (B) inside the processing unit, (3) sending facial data to the ER module (C), (4) transmitting AI classified emotion and frame data into the centralized platform Human Labeling module (E), (5) Elder or Caregiver classify a frame with the respective emotion in the Human Labeling Module (E) inside a centralized cloud-hosted web platform, and (6) storing emotion and frame data from human-labeling into the processing unit Data Persistence module (D). In this architecture we consider the elder's home as the elder's own house or a rest home where they live. Also, although this proposal focuses on emotion detection, and this system can be adapted to accommodate other use cases with varying classification requirements or contexts. For example, it could be employed to identify individuals' faces for security purposes or to detect signs of fatigue. The modular design of the system enables customization for specific contexts and use cases.

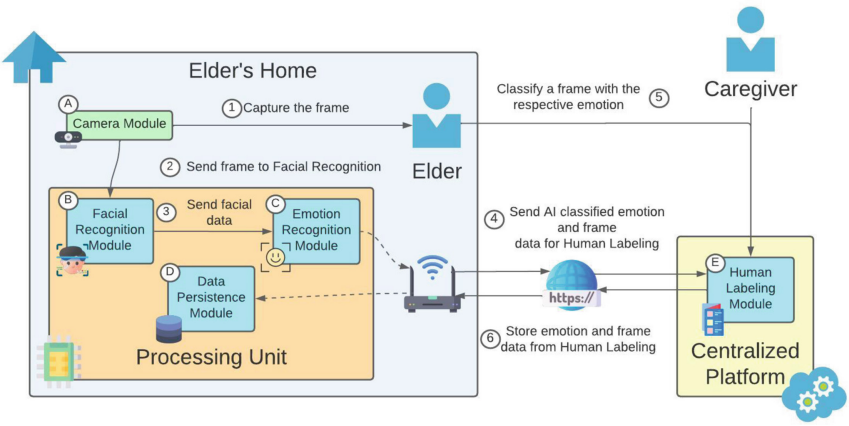


Fig. 1. Architecture of the proposed HITL emotion classification system.

The facial image acquisition process is initiated by the Camera Module (A), which captures images of the user's environment. These images are then sent to the Facial Recognition Module (B) inside the processing unit for preprocessing, which includes face detection, cropping, and filter application. The preprocessed images are securely stored in the processing unit Data Persistence Module (D), ensuring data preservation and easy access for further processing and analysis.

The ER Module (C) plays a critical role in selecting images to be classified by humans. This module employs an AI-based model to perform an initial classification of the captured facial images by their emotional content. By providing a preliminary classification, the module streamlines the human labeling process, directing users to focus on images where the AI model's confidence is lower than a predefined threshold or where further clarification is needed. As a result, the module accelerates the data labeling process and ensures that human expertise is effectively utilized to refine the model's understanding of complex emotional expressions.

A centralized web platform is provided for users to classify facial images. This platform enables users to classify their images with corresponding emotions using the Human Labeling Module (E). By involving users in the labeling process, the HITL approach guarantees the quality of the labeled data and encourages the collection of diverse and context-specific emotional expressions. As users continue to provide feedback and classify new emotions, the model may adapt and improve its results, ultimately enhancing its overall performance in detecting users' emotional states.

The HITL emotion classification system architecture aims to incorporate the concepts of human labeling and human-in-the-loop, fostering a collaborative and interactive environment for refining the ER model. By integrating the Human Labeling Module within the centralized web platform, the architecture enables users to provide feedback on AI-generated classifications, potentially enriching

the collection of diverse and context-specific emotional expressions. This iterative process forms a continuous feedback loop, which may drive model enhancement and contribute to the overall accuracy and context-specificity in emotion detection. As a result, the proposed system seeks to leverage the strengths of both AI and human expertise to achieve improved performance and adaptability, ultimately aiming to create a more personalized and reliable emotion detection system for a wide range of applications.

4 Implementation and Testing

In this section, we present the implementation of the HITL emotion classification system, its components, and the testing conducted in semi-controlled environments following the Declaration of Helsinki guidelines [1] for user testing.

The proposed system prototype, as shown in Fig. 2, was implemented using a Raspberry Pi Compute Module 4 as the processing unit, housed in a custom-designed enclosure.



Fig. 2. Photo of the HITL emotion classification system prototype in a real scenario.

Addressing ethical considerations and privacy preservation is crucial. Therefore, the system was designed with user privacy in mind, ensuring identifiable features in images are blurred and any data regarding Personal Identifiable Information (PII) or Personal Health Information (PHI) is stored securely.

The implementation of the system aligns with the modules proposed in Sect. 3, as follows:

- Capturing the frame: The Xiaomi Imilab Full HD camera functions as the Camera Module (A), responsible for capturing the facial expressions of the users.
- Sending the frame to the Facial Recognition module inside the Raspberry Pi: The Raspberry Pi computer module 4 receives the captured frames and processes them using the Facial Recognition Module (B).
- Sending facial data to the ER module: The identified facial data is then passed to the ER Module (C) for further analysis. The AI-based model employed in the ER Module (C) is an adaptation of the Deepface model [12].

– Transmitting AI-classified emotion and frame data to the centralized platform for human labeling: The system communicates with the centralized web platform, which houses the Human Labeling Module (E) and facilitates human labeling of the emotion data.

To conduct the tests of the implemented system, we obtained informed consent from the participants, provided study information, and protected participants' privacy and confidentiality.

The testing phase encompassed deploying the system in two residential settings, capturing frames over consecutive days. Family members offered invaluable feedback on the usability of the centralized web platform. This platform enables users to catalog images based on pre-detected emotions, as shown in Fig. 3. The emotion detection module classifies emotions into three primary categories: negative, positive, and neutral. In this work, negative emotions include anger, fear, sadness, and pain, while positive emotions encompass happiness, joy, and contentment. If the module identifies negativity, the available emotion options for classifying the image, depending on the use case, are drawn from the negative category. Human Labeling facilitated the cataloging of emotional states, generating a labeled emotions dataset for future research while adhering to ethical guidelines.

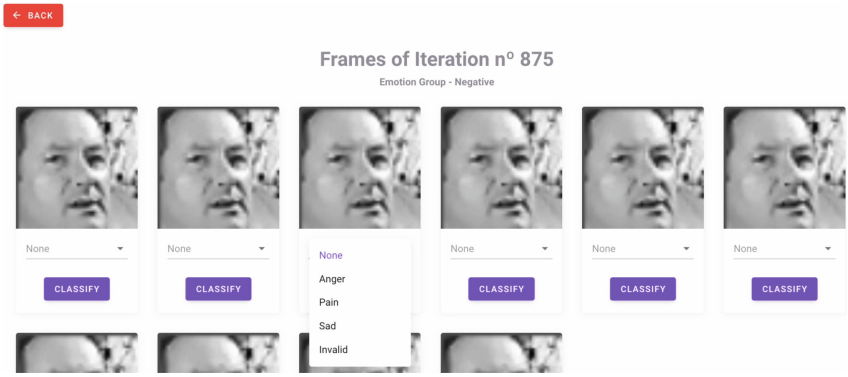


Fig. 3. Hitl emotion classification page in the centralized platform.

The HITL emotion classification system's implementation and testing demonstrated its potential in acquiring and classifying emotional data in ambient assisted living contexts. The semi-controlled tests and human labeling provided insights into the system's performance and usability, laying a foundation for future refinements.

Aligning the implementation with Sect. 3 demonstrated the system's capacity for acquiring and classifying emotional data. The modular design ensures flexibility, adaptability, and continuous improvements as more user data becomes available.

5 Analysis and Discussion

In this section, we delve into the analysis and discussion of the HITL emotion classification system based on the data obtained from the implementation and testing phase. We address potential ethical considerations and privacy preservation, exploring the limitations and challenges faced during the system's deployment in real-life scenarios, and suggest possible avenues for future improvements.

Given the limited dataset for training the ER model, a comprehensive assessment of the system's efficacy is challenging. However, the positive feedback from participants during semi-controlled tests indicates the system's potential. This feedback was obtained through a semistructured interview, performed after the tests, focusing on the usability of the HITL interface and functionalities. As the dataset grows and the model improves, the implemented system has potential to effectively assist elders and their caregivers in monitoring the elders' emotional states.

The overall user experience was positive, with family members feeling more confident and comfortable and elders finding the system non-intrusive and helpful. The emotion detection system provided valuable insights into users' emotional states, enabling caregivers to provide better support and care.

Apart from the ethical considerations mentioned during implementation, open communication with users and their families and ongoing monitoring of the system's impact can help maintain trust and ensure that ethical considerations are addressed appropriately.

Limitations and challenges during the implementation included the limited dataset for ER, challenges in capturing nitid facial expressions in varying light conditions and ensuring user privacy. To address these limitations, it is advisable to collect more labeled data, implement advanced image processing techniques to handle different lighting conditions, and use anonymization strategies to protect user privacy.

6 Conclusions and Future Work

In this paper, we presented the design, implementation, and testing of the HITL emotion classification system: an AAL system that aims to detect and classify the emotional states of elders using a HITL approach. The system consists of a Raspberry Pi computer module 4 and a Xiaomi Imilab Full HD camera module, that communicates with a centralized cloud platform that facilitates the human labeling of facial emotional data. We conducted testing in semi-controlled environments, following the Declaration of Helsinki guidelines, to ensure the ethical treatment of research participants.

The positive feedback from users during testing and the system's modular design indicate its potential to effectively support the development of AAL solutions to assist elders and their caregivers. Despite the challenges faced during implementation, such as the limited dataset for ER, difficulties in capturing

facial expressions in varying light conditions, and ensuring user privacy, the proposed system holds promise as a valuable tool for understanding and predicting emotional states in AAL contexts.

Future work includes selecting more labeled data to improve the ER model and implementing advanced image processing techniques to handle different lighting conditions. Moreover, future work also includes using anonymization strategies to protect user privacy. Further refinements to the system's design and functionality will be explored, such as incorporating personalized ER models based on specific user data and demographics. These refinements are expected to enhance the system's overall effectiveness and therefore ensure that it will better serve the needs of individual users and their caregivers. These improvements should be supported with further testing with elders and their caregivers. Finally, the labeled data should be used to train different emotion recognition models in order to analyze the benefits of personalized datasets.

Acknowledgements. This work was supported by national funds through the Portuguese Foundation for Science and Technology (FCT), I.P., under the project UIDB/04524/2020 and was partially supported by Portuguese National funds through FITEC-Programa Interface with reference CIT "INOV-INESC Inovação-Financiamento Base". The authors would also like to acknowledge FlatIcon for providing the icons used in the architecture diagram image.

References

1. World medical association declaration of helsinki: ethical principles for medical research involving human subjects. The Journal of the American College of Dentists **81**, 14–18 (2014). https://doi.org/10.1093/acprof:oso/9780199241323.003.0025
2. Blažič, B.J., Blažič, A.J.: Overcoming the digital divide with a modern approach to learning digital skills for the elderly adults. Educ. Inf. Technol. **25**(1), 259–279 (2019). https://doi.org/10.1007/s10639-019-09961-9
3. Dohr, A., Modre-Opsrian, R., Drobics, M., Hayn, D., Schreier, G.: The internet of things for ambient assisted living. In: 2010 Seventh International Conference on Information Technology: New Generations, pp. 804–809 (2010). https://doi.org/10.1109/ITNG.2010.104
4. Feng, Q., He, K., Wen, H., Keskin, C., Ye, Y.: Rethinking the data annotation process for multiview 3d pose estimation with active learning and self-training, pp. 5684–5693 (2023). https://doi.org/10.1109/WACV56688.2023.00565
5. Hervas, R., Baños, O., Villalonga, C.: Ambient intelligence for self-care and self-management. J. Ambient. Intell. Humaniz. Comput. **14**(1), 1–3 (2023). https://doi.org/10.1007/s12652-022-04474-9
6. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. **3**, 119–131 (2016). https://doi.org/10.1007/S40708-016-0042-6
7. Hung, S.K., Gan, J.Q.: Boosting facial emotion recognition by using GANs to augment small facial expression dataset. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2022). https://doi.org/10.1109/IJCNN55064.2022.9892096

8. Mainsant, M., Solinas, M., Reyboz, M., Godin, C., Mermillod, M.: Dream net: a privacy preserving continual leaming model for face emotion recognition (2021). https://doi.org/10.1109/ACIIW52867.2021.9666338
9. Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans. Affect. Comput. **10**(1), 18–31 (2019). https://doi.org/10.1109/taffc.2017.2740923, https://ieeexplore.ieee.org/document/8013713
10. Naser, A., Lotfi, A., Mwanje, M.D., Zhong, J.: Privacy-preserving, thermal vision with human in the loop fall detection alert system. IEEE Trans. Hum. Mach. Syst. **53**(1), 164–175 (2023). https://doi.org/10.1109/THMS.2022.3203021
11. Pantic, M., Rothkrantz, L.: Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1424–1445 (2000). https://doi.org/10.1109/34.895976
12. Serengil, S.I., Ozpinar, A.: Lightface: a hybrid deep face recognition framework. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 23–27. IEEE (2020). https://doi.org/10.1109/ASYU50717.2020.9259802
13. Teichman, A., Thrun, S.: Tracking-based semi-supervised learning. Int. J. Robot. Res. **31**, 804–818 (2012). https://doi.org/10.1177/0278364912442751, https://journals.sagepub.com/doi/10.1177/0278364912442751
14. Vorobyov, M., Galkin, I.: Concept of cost-effective power-assisted wheelchair: human-in-the-loop subsystem. In: Proceedings of the 5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering, AIEEE 2017 2018-January, 1–5 (2017). https://doi.org/10.1109/AIEEE.2017.8270559
15. Wu, D., Sun, D.W.: Colour measurements by computer vision for food quality control - a review. Trends in Food Sci. Technol. **29**, 5–20 (2013). https://doi.org/10.1016/J.TIFS.2012.08.004
16. Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., He, L.: A survey of human-in-the-loop for machine learning. Future Gener. Comput. Syst. **135**, 364–381 (2022). https://doi.org/10.1016/j.future.2022.05.014, https://arxiv.org/abs/2108.00941, arXiv:2108.00941 [cs]
17. Zhang, T., Nakamura, M.: Neural network-based hybrid human-in-the-loop control for meal assistance orthosis. IEEE Trans. Neural Syst. Rehabil. Eng. **14**(1), 64–75 (2006). https://doi.org/10.1109/TNSRE.2005.863840
18. Zhou, T., Shen, J., He, D., Vijayakumar, P., Kumar, N.: Human-in-the-loop-aided privacy-preserving scheme for smart healthcare. IEEE Trans. Emerg. Top. Comput. Intell. **6**(1), 6–15 (2022). https://doi.org/10.1109/TETCI.2020.2993841

A Portuguese Mobile App for Diabetes Type 2 Self-Management: A Picture of One-Month Usage

Glória Conceição[1(✉)], Andreia Pinto[1,2], João Viana[1,2], Gonçalo Miranda[3], Paulo Santos[1,2], Cristina Santos[1,2], and Alberto Freitas[1,2]

[1] Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine, University of Porto, Porto, Portugal
glorialmeida6100@gmail.com
[2] Center for Health Technology and Services Research (CINTESIS), Faculty of Medicine, University of Porto, Porto, Portugal
[3] Department of Digital Services, Biometrid (FYI), Porto, Portugal

Abstract. Digital health interventions are promising in improving self-management of different health conditions, such as type 2 diabetes mellitus (T2DM). Diabetes is one of the most prevalent worldwide chronic diseases, and self-management can reduce the risk of diabetes-related complications. The Food Friend app was developed to promote the self-management of T2DM through lifestyle modification. Diabetic patients were recruited to use the Food Friend app over three months to prove the impact of this app on diabetes management. This paper describes the user's interaction with the Food Friend app over the first month. User engagement is followed by UXCAM software. The usability of 49 participants (≈56 years old; 69% males) showed that the number of daily active users is lower than the number of users recruited for the study, and the app's use decreases on the weekend. Furthermore, meal registrations or updates were the most performed tasks. However, we observed that the meal summary or meals history are the screens less visualized, which shows that the meals feedback is not used. In summary, the description of one month of interaction with the Food Friend app demonstrates that it's hard to guarantee participants' commitment and ensure that Food Friend helps in diabetes self-management.

Keywords: Type 2 Diabetes Mellitus · mHealth · Usability · Digital health interventions

1 Introduction

Diabetes is a chronic disease characterized by high blood glucose levels due to insulin resistance and/or beta-cell dysfunction [1]. Type 2 Diabetes Mellitus (T2DM) is the most prevalent form of diabetes associated with ageing populations, economic development, overweight, unhealthy diet, and reduced physical activity [2, 3].

T2DM has a significant economic impact due to its association with several complications [4, 5]. Diabetes management needs strict glycemic control, which requires

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 133–142, 2023.
https://doi.org/10.1007/978-3-031-43461-7_14

lifestyle adjustments, glucose monitoring, and/or pharmacotherapy [6]. Diabetes self-management and education are crucial to reducing the risk of diabetes-related complications and maintaining acceptable health standards [6]. Clinical care alone is powerless to mitigate T2DM. There is a need for innovative and effective ways to improve diabetes care. Therefore, it is essential to empower diabetic patients with tools and information that will improve their health behaviours and clinical outcomes. Education is vital to support diabetic patients in making informed decisions concerning their health [7].

Digital health interventions are promising in improving health self-management [8]. Digital technologies, particularly mobile wireless technologies, are attractive tools due to their accessibility, low cost, widespread use, and applicability. They can become essential for medical and public health practices [8, 9]. Mobile phones are a viable platform to improve the provision of health services for management and education. These technologies are commonly called Mobile Health (mHealth) [9]. So, mHealth could be a core element in promoting diabetes self-management.

A multidisciplinary team developed the Food Friend app to help the self-management of T2DM for the Portuguese population (https://foodfriend.pt/). Different studies have been designed to guarantee the feasibility and validation of the Food Friend App (papers not yet published). Diabetic patients were recruited to use the Food Friend app over three months and to evaluate and validate the impact of this app on diabetes management. This paper aims to describe the user's interaction with the mobile application developed in the Food Friend project over the first month.

2 Methods

2.1 Participants and Settings

For this study, we recruited patients with T2DM that are independent smartphone users. The recruitment comprised three different contexts: in context 1, we recruited patients in a primary care health centre. These patients were previously referred by the clinicians as potential participants; in context 2, we selected patients with T2DM from endocrinology appointments in secondary healthcare; in context 3, we included voluntary participants that previously manifested interest in participating in our study. The Food Friend study was disseminated on social media, in different news sources, and on the Faculty of Medicine of the University of Porto (FMUP) webpage with a link to register their interest. The potential participants were contacted by a team member to confirm that they have T2DM and to schedule a date for participation in the study. All the participants were invited to install and use the Food Friend app for three months. This study started on January 3, 2023.

2.2 Food Friend App

The Food Friend application was created to help T2DM patients have a positive experience by improving diabetes management. This app aims to assist patients' meal choices by providing an overall picture of the daily meal summary. Moreover, it incorporates reminders for medication, monitoring physical activity, and registration of glycemic levels. All the features of this app were previously described in [10, 11] and in Fig. 1.

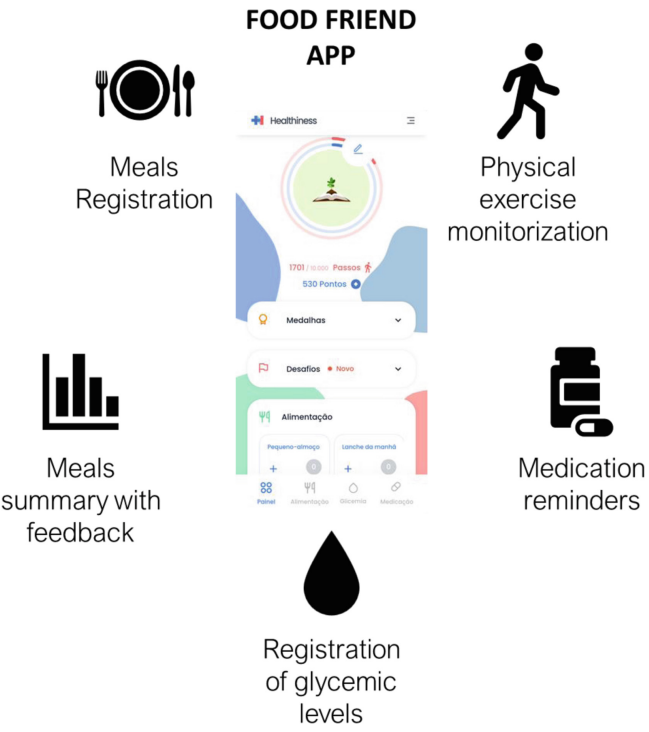


Fig. 1. Food Friend Application features.

2.3 Data Collection

The user's socio-demographic data were obtained from a questionnaire. The engagement data from the application was acquired by the © UXCam software. The UXCam is a mobile app analytics tool that helps to understand and interpret customers' interactions with mobile applications (https://uxcam.com/).

2.4 Data Analysis

Descriptive data on demographic and clinical characteristics of the sample are reported as mean (standard deviation) or median (interquartile range) for continuous variables. Frequencies and their percentages are reported for categorical variables. All analyses were performed using SPSS Statistics for Windows version 28.0 (Armonk, NY: IBM Corp).

2.5 Ethics Approval

This study was approved by the ethics committee of the Faculty of Medicine of the University of Porto, Hospital São João, and Administração Regional de Saúde do Norte in Portugal. Written informed consent was obtained from all participants. Participants were not compensated. The study was carried out in accordance with the ethical standards of the Declaration of Helsinki.

3 Results

3.1 Participants' Characteristics

During the first month of the study, we included 49 participants (mean age of 55.6 ± 10.6 years; 69% males), most recruited from voluntary interest (Context 1). The majority were treated exclusively with oral antidiabetic drugs (69%). Only one-third of our participants need to measure the glycemic level every day. All participants were invited to use the app daily. Participants' characteristics are shown in Table 1.

Table 1. Demographic and clinical characteristics of the participants.

Characteristics	Total (n = 49)
Age, years (mean ± SD)	55.6 ± 10.6
Male sex, n (%)	34 (69)
Less than 12 years of formal education, n (%)	13 (27)
Marital status, n (%)	
Married	38 (78)
Divorced	7 (14)
Single	4 (8)
Employment status, n (%)	
Employed	24 (49)
Student	1 (2)
Housewife	2 (4)
Unemployed	6 (12)
Retired	16 (33)
Years since T2DM diagnosis, median (q1-q3)	10 (4.5 – 15)
Treatment, n (%)	
Oral antidiabetic drugs	34 (69)
Insulin	2 (4)
Insulin + oral antidiabetic drugs	12 (25)
Other	1 (2)
Weekly glycemic measurement regularity, n (%)	
Every day	16 (33)
Few days a week	7 (14)
Once a week	8 (16)
No need	18 (37)

To monitor the activity of our participants, we used the UXCAM tool to verify the engagement of the participants with the application. The users tend to spend more time

in the Food Friend app during the week, as observed in Fig. 2. In all four-day parts, the period of the morning (6 am–12 pm) and the period of the afternoon (12 pm–6 pm) present more visits than the other periods.

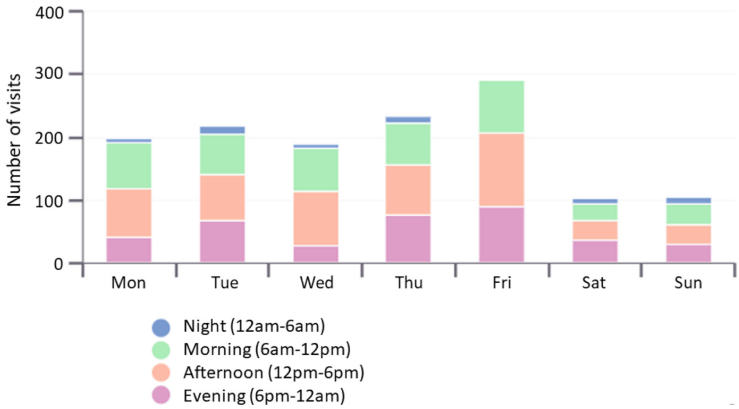


Fig. 2. The total number of visits to the Food Friend app by the day of the week and the time of the respective day.

Figure 3 shows the distribution of new users in the study and the active users observed across the first month. The inclusion of new users during the first month was not accompanied by an increment in the number of daily active users at the end of the month. For this observation, we considered an active user as a user that visited the app at least once.

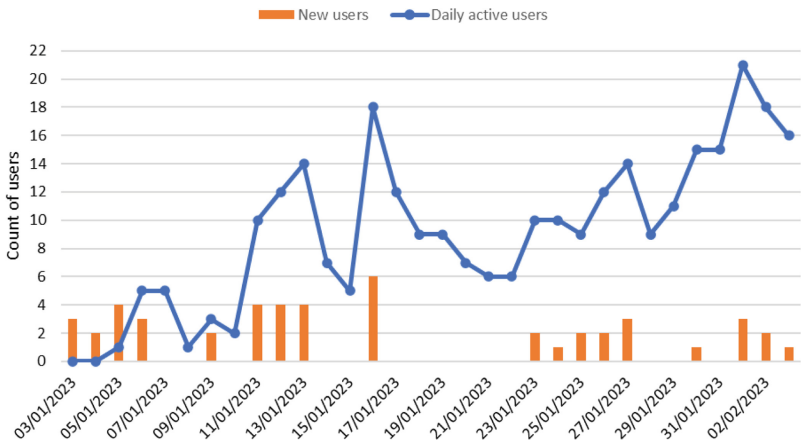


Fig. 3. Trend of daily active users between 03/01/2023 and 03/02/2023.

The interaction with the app results in *Events*, which are actions the users perform in the app that can be measured independently of the screens visited. All the *Events* considered in this paper are related to features of diabetes management, such as *Update*

meal, *Saved meal*, *Search ingredient photo* and others (Fig. 4). We excluded all the events relative to core functionalities, such as login or rage Tap. In total, 2240 events were recorded by the 49 participants included in the first month of this study. The most accomplished event was the *Update meal* (45.1%), followed by the *Saved meal* (22.1%) and *Import steps data* (19.2%).

The Food Friend app allows the recognition of the food by photography (*Search ingredient photo* event) and barcode (*Search ingredient barcode* event). However, these actions only represent 1.0% and 0.9% of the events, respectively, showing that the manual entry of the food is more intuitive for our users.

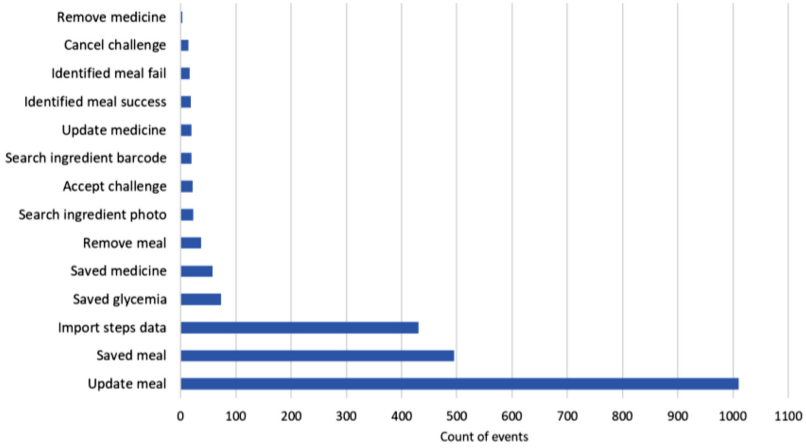


Fig. 4. Count of events by event between 03/01/2023 and 03/02/2023.

Figure 5 demonstrates that *Add meal* screen is the screen with more visits (57.7%), followed by the *Dashboard* screen (23.3%) and the *Register meal* screen (5.9%).

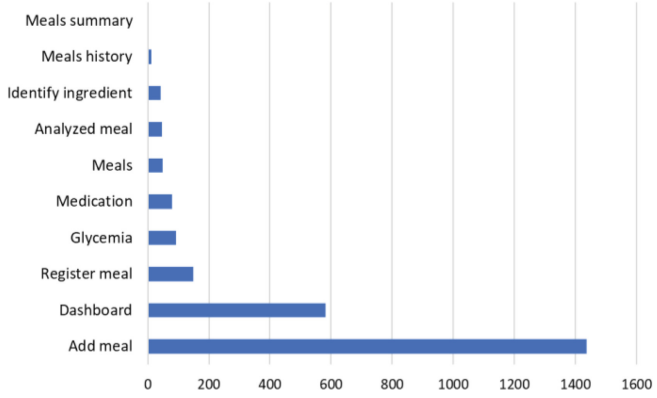


Fig. 5. Count of events grouped by event screen between 03/01/2023 and 03/02/2023.

The Food Friend app contains a recommendation system that provides feedback about the meal quality through nutrient discrimination of the registered food. The analysis of the count of events grouped by event screen reveals that *Meals history* (0.4%) and *Meals summary* (0.2%) are the screens less visited. These screens are represented in Fig. 6.

4 Discussion

Digital Health solutions have been considered important strategies for chronic disease management, such as T2DM [12]. The Food Friend app aims to help diabetic patients through self-management of lifestyle. The Food Friend app version presented in this study was the culmination of several studies.

Looking at our results, we observed that the number of daily active users is lower than the number of users recruited for the study. The use of the mobile app is entirely voluntary and self-motivated. The literature identified key factors influencing users' usability after the initial interaction with an application, such as motivation, patients' goals and their commitment to these goals [13]. One month of usability shows that the utilization of the app is lower during the weekend, and even not all the users that installed the app are active. This inconsistent usage jeopardizes the intended outcomes, since the reduced time spent, and may be related to a tendency to stop using the application [14].

The Food Friend app has several features for diabetes self-management. The events and screens visits show the parts that were used in the app. Diabetic patients play a determinant role in their own treatment, having to take relevant decisions about dietary choices, glucose monitoring, medication intake and physical activity [15]. The glucose monitoring was less observed in our app. Although, the characteristics of our participants highlight that each participant has different necessities for this evaluation, which can explain this observation. Contrarily, meal registrations or updates were the most performed tasks. Notably, their meals have been registered in the app. This result was expected as it was in line with the findings of other articles which demonstrated that this feature was ranked as one of the most useful features or that users would value in similar apps [16, 17]. However, we noted that the meal summary or meals history are the screens less visualized. These screens give automatic feedback based on the food choices, giving patients important information about food nutrients (Fig. 6). Based on this information, a diabetic patient can evaluate their diet and encouraging to make conscious food choices for glycemic control. So, it is challenging to guarantee that the app helps our participants with meal decisions and, consequently, to ensure that the app's purpose has been fulfilled.

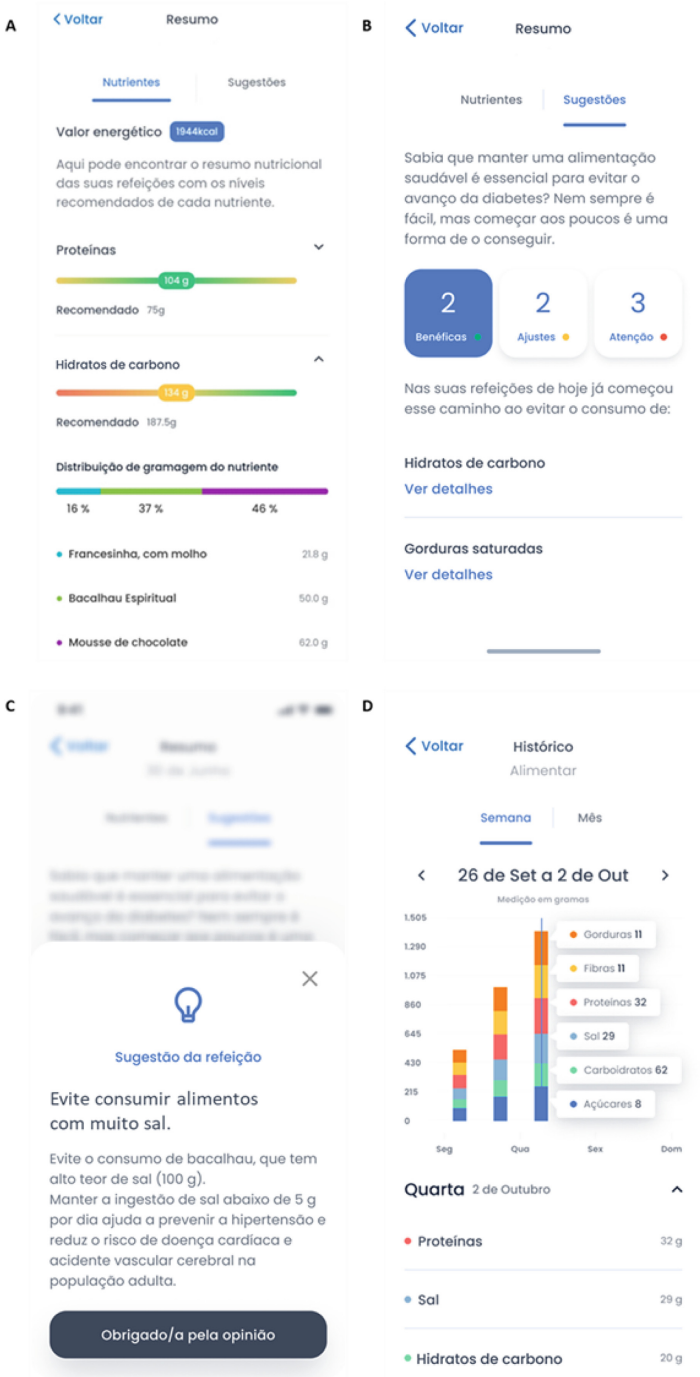


Fig. 6. Representative screens of *Meals summary* (A, B and C) and *Meals history* (D) (in Portuguese).

5 Conclusion

This work seeks to describe the first perception of how users react to the application's use in its first month of use and may give us indications of the following months of use. Furthermore, it highlights the importance of establishing factors to promote long-term commitment to the Food Friend app. These data showed the necessity to develop and implement new strategies for integrating the app and its components for their actual purpose with diabetic patients.

Acknowledgements. This research work was developed under the project Food Friend – an "Autonomous and easy-to-use tool for monitoring of personal food intake and personalized feedback" (ITEA 18032), co-financed by the North Regional Operational Program (NORTE 2020) under the Portugal 2020 and European Regional Development Fund (ERDF), with the reference NORTE-01–0247-FEDER-047381. Andreia Pinto was funded by FCT, Portugal, under PhD grant number UI/BD/151485/2021.

References

1. Cerf, M.E.: Beta cell dysfunction and insulin resistance. Front. Endocrinol. **4**, 37 (2013)
2. Al Kaabi, J., Mustafa, H., Govender, R.D., et al.: Epidemiology of type 2 diabetes – global burden of disease and forecasted trends. J. Epidemiol. Global Health **10**(1), 107 (2019)
3. Guariguata, L., Whiting, D.R., Hambleton, I., et al.: Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res. Clin. Pract. **103**(2), 137–149 (2014)
4. Arrieta, F., Rubio-Terrés, C., Rubio-Rodríguez, D., et al.: Estimación del impacto económico y sanitario de las complicaciones de la diabetes mellitus tipo 2 en la Comunidad de Madrid. Endocrinol. Nutr. **61**(4), 193–201 (2014)
5. Andersson, E., Persson, S., Hallén, N., et al.: Costs of diabetes complications: hospital-based care and absence from work for 392,200 people with type 2 diabetes and matched control participants in Sweden. Diabetologia **63**(12), 2582–2594 (2020)
6. Al-Khaledi, M., Al-Dousari, H., Al-Dhufairi, S., et al.: Diabetes self-management: a key to better health-related quality of life in patients with diabetes. Med. Princ. Pract. **27**(4), 323–331 (2018)
7. Gómez-Velasco, D.V., Almeda-Valdes, P., Martagón, A.J., et al.: Empowerment of patients with type 2 diabetes: current perspectives. Diabetes, Metab. Syndr. Obes. Targets Therapy. **12**, 1311–1321 (2019)
8. Petersen, M., Hempler, N.F.: Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study. BMC Med. Inform. Decis. Making. **17**(1), 1–10 (2017)
9. Executive Board. mHealth: use of appropriate digital technologies for public health: report by the Director-General.5 p. (2017)
10. Pinto, A., Viana, J., Martinho, D., et al.: Improving the lifestyle behavior of type 2 diabetes mellitus patients using a mobile application. In: 2022 IEEE Symposium on Computers and Communications (ISCC), pp. 1–5 (2022)
11. Martinho D, Crista V, Pinto A, et al. An Architecture for a Coaching System to Support Type 2 Diabetic Patients. In: In: Julián, V., Carneiro, J., Alonso, R.S., Chamoso, P., Novais, P. (eds.) Ambient Intelligence—Software and Applications—13th International Symposium on Ambient Intelligence. ISAmI 2022. Lecture Notes in Networks and Systems, vol. 603, pp. 167–78 (2023). https://doi.org/10.1007/978-3-031-22356-3_16

12. Fan, K., Zhao, Y.: Mobile health technology: a novel tool in chronic disease management. Intell. Med. **2**(1), 41–47 (2022)
13. Mustafa, A.S., Ali, N.A., Dhillon, J.S., et al.: User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare. **10**(2), 221 (2022)
14. Vaghefi, I., Tulu, B.: The continued use of mobile health apps: insights from a longitudinal study. JMIR mHealth uHealth **7**(8), e12983 (2019)
15. Borse, S.P., Chhipa, A.S., Sharma, V., et al.: Management of type 2 diabetes: current strategies, unfocussed aspects, challenges, and alternatives. Med. Princ. Pract. **30**(2), 109–121 (2021)
16. Kebede, M.M., Pischke, C.R.: Popular diabetes apps and the impact of diabetes app use on self-care behaviour: a survey among the digital community of persons with diabetes on social media. Front. Endocrinol. **10**, 135 (2019)
17. Borsci, S., Adu, M.D., Malabu, U.H., et al.: Users' preferences and design recommendations to promote engagements with mobile apps for diabetes self-management: multi-national perspectives. Plos One **13**(12), e0208942 (2018)

CERDL: Contextual Emotion Recognition Analysis Using Deep Learning

Aayushi Chaudhari[1], Chintan Bhatt[2(✉)], Achyut Krishna[1], and Juan M. Corchado[3,4,5]

[1] U & P U. Patel Department of Computer Engineering, Chandubhai S Patel Institute of Technology (CSPIT), CHARUSAT Campus, Charotar University of Science and Technology (CHARUSAT), Changa 388421, India
aayushichaudhari.ce@charusat.ac.in, 20ce120@charusat.edu.in

[2] Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar 382007, India
chintan.bhatt@sot.pdpu.ac.in

[3] BISITE Research Group, University of Salamanca, 37007 Salamanca, Spain
corchado@usal.es

[4] Air Institute, IoT Digital Innovation Hub, 37188 Salamanca, Spain

[5] Department of Electronics, Information and Communication, Faculty of Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan

Abstract. This paper delves into the critical importance of understanding emotions from a person's perspective, and the potential for machines to improve human interaction by possessing this ability. While existing research on emotion recognition in computer vision has mainly focused on analyzing facial expressions and categorizing them into six basic emotions, it is important to recognize that contextual factors also play a crucial role in emotion perception. Emotions are not just limited to facial expressions but also include body language, the pitch of voice, and other nonverbal cues. We then trained a convolutional neural network model on this vast dataset and demonstrated the importance of incorporating context to recognize rich information about emotional states in images. Our model surpasses previous benchmarks and confirms the value of contextual information in emotion recognition. We have used the Emotions in Context (EMOTIC) [1] and Body Language Dataset (BoLD) [2] datasets for recognizing emotions by taking their contextual information into account. By incorporating contextual factors, machines can enhance human interaction by accurately recognizing emotional states in various situations. Based on the experiments, we recognized that the emotions of engagement (93.62%), confidence (92.41%), and excitement (95.93%) were predicted accurately. In contrast, the emotions of yearning, disapproval, and pain had low classification accuracy, with less than 40% accuracy. Lastly, this paper highlights the importance of understanding emotions beyond just facial expressions and provides a benchmark for emotion recognition in a contextual setting.

Keywords: Emotion recognition · Scene context · Pattern recognition · Convolutional Neural Network · Fusion Network

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 143–156, 2023.
https://doi.org/10.1007/978-3-031-43461-7_15

1 Introduction

The study of emotions is a critical area of research in fields such as psychology, philosophy, and neuroscience. Emotions play a crucial role in human behavior and can influence our interactions with others, decision-making processes, and overall well-being. However, emotions are complex phenomena, and their expression can be influenced by a variety of factors, including individual differences, social context, and cultural norms.

AER is widely applied in fields like security, where it can be used to identify suspicious behavior or potential threats, and in behavioral science, where it can help researchers understand the underlying mechanisms of emotional processing. AER is also becoming increasingly important in healthcare, where recognizing and addressing a patient's contextual emotions can help healthcare providers develop a deeper sense of empathy and understanding.

In education, AER can also be used in online learning platforms to identify a student's level of motivation or frustration, leading to more effective feedback and a better understanding of how to optimize the learning experience.

In driving, AER can play a crucial role in enhancing road safety. For example, a car that can detect a driver's fatigue or anxiety can intervene or issue an alert, potentially preventing accidents.

The traditional approach to AER has been to analyze facial expressions, body posture, and gestures. Nonetheless, this research suggests that contextual information is also important in comprehending emotional states. Contextual information contains a person's surroundings, such as the category of place, location qualities, objects, or actions that occur around the person.

The proposed approach incorporates body pose, non-targeted subjects, and background information within the contextual part of the image. This approach is said as "Emotions based on visual information with the help of Context," or "EmoViCon" for short. By combining multiple sources of information, including facial expressions, body posture, gestures, and context, EmoViCon aims to improve the accuracy and robustness of AER algorithms. We are also considering two emotional dimensions in our research, called Valence and Arousal.

2 Literature Review

Context-aware emotion recognition has been a relevant topic undertaken several times in recent years. A two-layered Convolutional Neural Network (CNN) includes the scene's structure and semantic information, Kosti et al. [3] in-contextual emotion comprehension job. In order to infer emotion detection, Zhang et al. [4] first extracted scene semantics as node features using the regional proposal network, later they used Graph Convolutional Network (GCN) for building an emotion graph. After the hidden face, Lee et al. [5] researched the scene for pertinent context clues using the attention mechanism. Mittal et al. [6, 7] to combine data from diverse modalities and context analyses, used multiplicative fusion.

Hoang et al. [8] proposed a new line of reasoning for quantifying the interaction between key actors and objects. In a suboptimal manner, the aforementioned methodologies study, the emotional linkages between the agents as well as their impact on

object interactions. In comparison, our method's four contexts are more complimentary and constructive in emotion identification. Ankit et al. [9] have proposed a fusion model based on a Mixture of Experts (MoE) which dynamically mixes data from the audio and video modes to predict any emotion produced by movies. Liu et al. [10] consider the weight of every modality by discriminating them appropriately using the attention network and later fuse all multi-modal features with good incorporation. Gupta et al. [11] have proposed an approach for detecting emotion among online learners in real time using a deep learning approach. Chaudhari et al. [12] have proposed a method using fine-tuned transformers for detecting faces, cropping, and extracting facial features from the FER image dataset. The author has also proposed another work using an inter-modality attention transformer-based fusion approach for using the multi-modal self-supervised technique to recognize emotions [13]. Kothadiya et al. [14] have surveyed various deep learning techniques and challenges in various domains.

Zhao et al. [15] and Ye et al. [16] used CNN and LSTM are combined to form a novel framework for the detection of emotions. They generated spectrograms from audio recordings and input the 3D log-mel spectral values into their model. Gao et al. [17] Emotions should be classified using a kernel function SVM with sequential minimal optimization. They use windows ranging from 20 to 100 ms to compute intensity, pitch, LSP (line spectral pairs), MFCC (Mel Frequency Cepstral Coefficient), and ZCR (Zero crossing rate) as feature points of audio clips. Milton et al. [18] employed RBF kernels, SVM classifiers using linear features, and MFCC features to recognize emotions, They push the envelope by proposing to use a 3-layer SVM. Huang et al. [19] utilized emotionally significant features for emotion recognition in speech using CNN. Lim et al. [20] present an approach based on concatenated CNN and RNN that does not require any obligation to include features, i.e., spectrogram frames. They report higher accuracy than they did using traditional approaches at the time of publishing.

3 Dataset

3.1 EMOTIC Dataset

The EMOTIC dataset is a collection of images of humans in natural settings that have been tagged with their apparent moods. To classify images in the database, the Amazon Mechanical Turk (AMT) framework is used [1]. The resulting collection includes 18,313 images and 23, 788 annotated people. An extensive record of 26 emotion categories, as well as the three dimensionalities namely Arousal, Valence, and Dominance, are labeled on the images. Emotions are defined by the VAD paradigm in three ways: Valence (V) assesses how positive or pleasant an emotion is, ranging from negative to positive; Arousal (A) assesses the person's aggression level, ranging from passive/calm to agitated/ready to act; and Dominance (D) assesses the individual's contextual control, varying from submissive/non-control to dominant/in-control [1]. (see Fig. 1) for the samples emotion categories and images for 26 emotions.

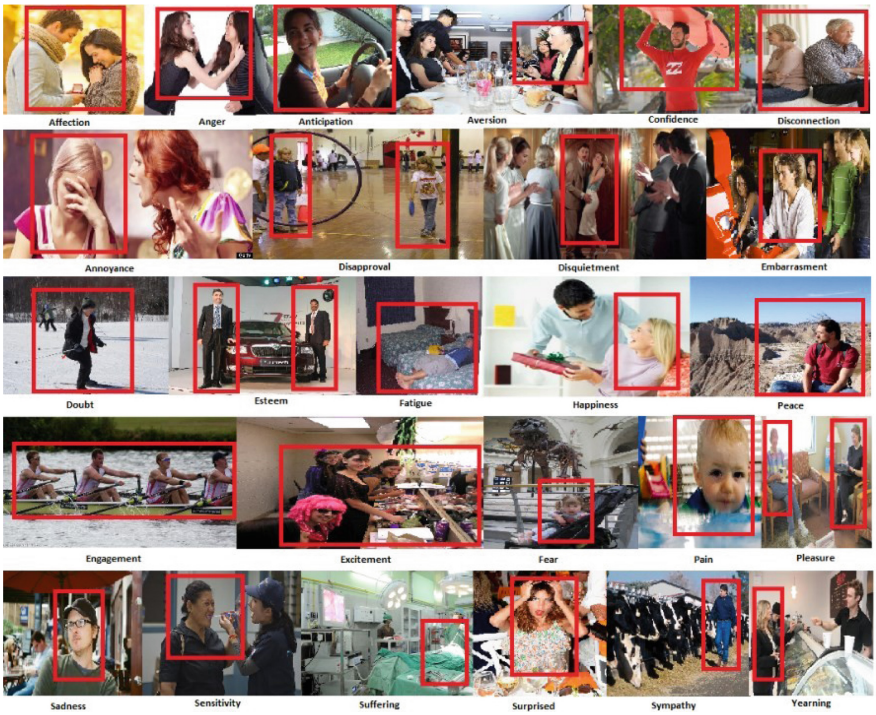


Fig. 1. Image samples of the EMOTIC dataset having 26 emotional categories

3.2 BoLD Dataset

The BoLD stands for Body Language Dataset, which is aimed to understand the emotional states of humans based on the body language of a person. It is a completely annotated and large-scale dataset platform for benchmarking and a model evaluation to compare the effectiveness of models in achieving this goal [2]. A wide range of cutting- edge applications, such as information administration, retrieval, public safety, social media and patient care improvement, may be made possible by computerized bodily emotion expression detection skills. Researchers are challenged to construct a model that can assist computers in recognise human mood by integrating geospatial and human body position data obtained from the BoLD Dataset. Given a brief video sequence of an entity, the model assesses both the continuous emotion values (Valence, Arousal, and Dominance) as well as the 26 discrete emotion categories (Peace, Affection, Esteem, Anticipation, Engagement, Confidence, Happiness, Pleasure, Excitement, Surprise, Sympathy, Doubt/confusion, Disconnection, Fatigue, Embarrassment, Yearning, Disapproval, Aversion, Annoyance) [2]. Every short video in BoLD has been validated for viewers' perceived emotional reactions. (see Fig. 2) for sample images from BoLD dataset for 26 emotions.

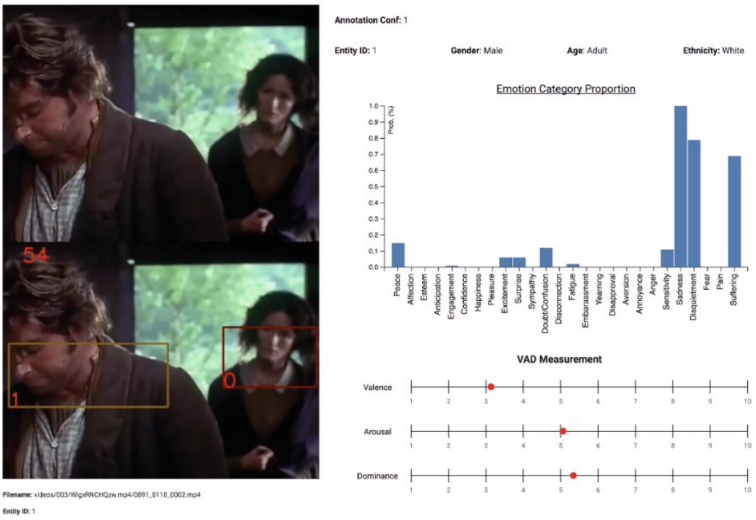


Fig. 2. Image specimen from BoLD Dataset for 26 emotional categories

4 Implementation of Architecture

As a baseline of emotion detection in contextual form, the authors introduced a Convolutional Neural Network (CNN) model. The proposed model is made up of three major components: image (background) extraction of features, extraction of body features, and fusion network. The image feature extraction module processes an entire image to generate scene-context features, which are context-specific characteristics that encapsulate scene classification, attributes, items in the scene, and the interactions between persons in the scene. The authors pre-trained this component with scene-centric information to capture contextual information for the Locations dataset [21]. To retrieve the graphic context features, the authors remove the final Fully connected layers and replace them with an Average Pooling layer, which is followed by a layer called nn that compresses each value in a 1-dimensional vector representation.

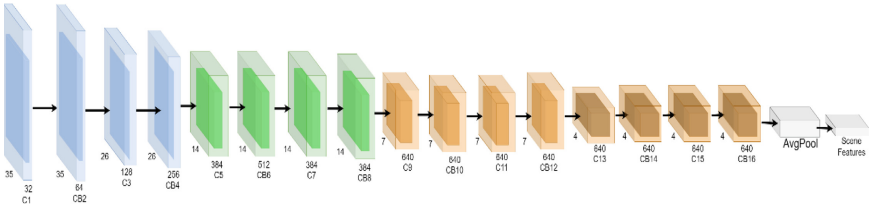


Fig. 3. Basic structure of network for extracting visual contextual features

Figure 3 displays the fundamental network structure used to extract visual context information, where C indicates the Conv layer that follows a ReLU layer then CB indicates the Fully-connected layer followed by a Batch Normalisation and a ReLU

layer. All of the Conv filter utilized by the model is one-dimensional. By using these contextual features, the model can identify the emotions of the person more accurately, as they take into account the visual context of the scene along with the person's body features. The authors claim that this model can improve emotion recognition accuracy by taking into account the context-specific information showcased in the scene.

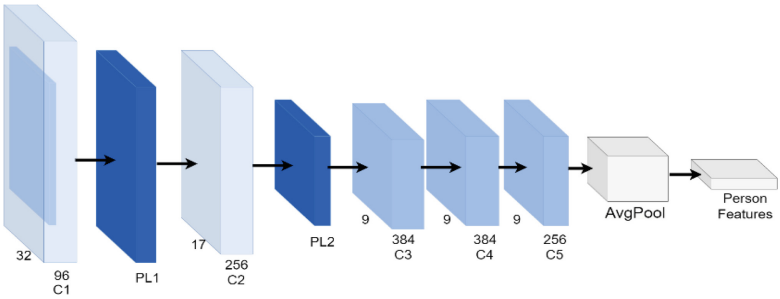


Fig. 4. Person Recognition Component of Architecture

The system's second module is responsible for extracting body-related features from a person's visible body (see Fig. 4). The body feature extraction module, in particular, takes visible portions of the target person's body as input and generates characteristics that encapsulate valuable insights such as face and head facets, as well as body pose and visual appeal. To ensure that these features are extracted correctly, the component is pretrained using ImageNet, a scene-centric set of data that contains the classification of a person. This enables us to extract person-specific features from the image's bounding box that used a Convolutional Neural Network (CNN) pretrained on object recognition. More particularly, we employ a classic object recognition algorithm CNN called Alexnet that has a basic structure comprising five Convolutional layers and three Fully Connected layers, as shown in Figure 4. Note that the middle layers are not trainable, such as ReLU, and are not included in the shown network. Alexnet offers us with characteristics that reflect the human body which includes people. To describe the perception of emotions using EMOTIC, we replace the Fully Connected layers with a Spatial Average Pooled layer, which is followed by a nn layer that compresses all values in a 1-dimensional vector form. View in the Torch7 framework. Following each Convolutional layer is a Rectified Linear Unit (ReLU) layer, which corrects negative values.

The two modules for feature extraction are designed using the 1-dimensional filter of CNN architecture designed in [22]. The CNN networks exhibit excellent results with relatively fewer parameters. Each network is made up of sixteen convolution layers with 1-d kernels that alter between vertical and horizontal orientations, modeling 8 layers with 2-dimensional kernels. We used global average convolution layers to compress the features of the final convolutional layer in order to keep the directional data for various regions of the image. To prevent internal covariant shift, we added a batch normalization layer [23] after every layer of convolution and corrected units in order to speed up our training process.

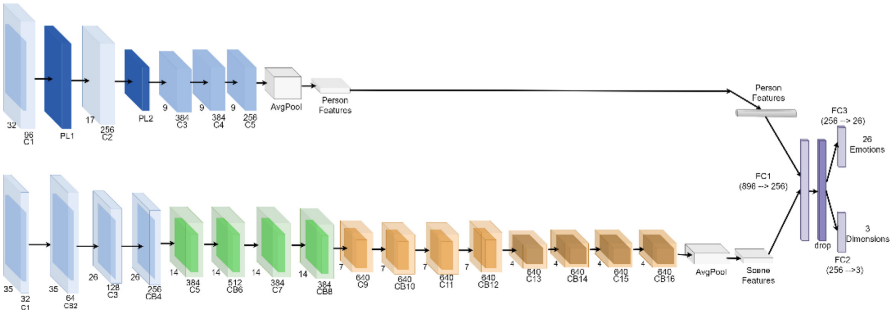


Fig. 5. Prototype for emotion detection considering contextual information.

The presented model, aims to perform finely tuned regression of 2 variations of emotional representation (see Fig. 5). The third module combines the components extracted from the previous two modules, and the fusion layer concatenates them to estimate both discrete emotional categories and three emotion dimensions. For training, the concatenated features pass through a tuple of fully connected layers, including three fully connected layers in the fusion module. First, the fully connected layer (FC-1) reduces the total amount of dimensions of cascaded features to 256. The reduced features are then passed through two fully connected layers, one for learning independent representations of VAD Dimensions and the rest for 26 Emotional Categories. A dropout layer is applied to the result of fully connected layer 1 to prevent overfitting. The result of fully connected layer-2 provides three continuous dimensions' regressed values, while fully connected layer 3 produces a probability distribution. The threshold values for each category are calculated using this distribution., identifying the activated categories for each test image. The fusion network component is made up of two fully interconnected layers. The very first fully connected layer reduces feature dimensionality to 256, and the another fully interconnected layer is assigned to discover independent depictions for every task. This layer's outcome is split into two distinct representations, including one 26 units representing the distinct emotion classifications and the other with three units representing the three continuous dimensions.

5 Performance Evaluation Metrics

Multi-class classification is used to model the task of learning emotion categories, and regression is used to model the task of learning continuous dimensions. So, we used two different evaluation criteria to evaluate the efficacy of our fusion model.

5.1 Euclidean Loss ($Loss_{disc}$)

We codified 26 different emotion categories in our emotion recognition task. In EMOTIC, each person can be tagged with various emotion categories. We can see that there are various emotion classes to choose from (26 discrete emotion categories) and that each discrete input could be labeled with several labels, giving this a multiclass and multilabel

challenge. Furthermore, inherent class disparity occurs because the amount of training instances for each emotion category and each score of the continuous dimensions is not the same. So, we have used a Euclidean loss that is weighted to overcome this problem. Equation 1 represents the Euclidean loss for our emotion's categories.

$$Loss_{disc} = \frac{1}{Num} \sum_{c=1}^{Num} wei(Y_c^{\wedge disc} - Y_c^{disc}) \tag{1}$$

where Num stands for the number of emotional categories (26), $Y_c^{\wedge disc}$ stands for the prediction and Y_c^{disc} states the ground-truth marking for the cth category. The *wi* states the weight assigned for each category. Values associated with weight are defined as $wei = \frac{1}{\ln(co+pi)}$, where pi stands for probability of the ith category and co for controlling the valid values ranges for *wei*. The values of wei are constrained by this weighting method when the instances of a class approach zero.

5.2 Euclidean Loss on Continuous Dimension ($Loss_{2cont}$, $SLoss_{1cont}$)

There are three aspects whose values are acquired in this learning activity. Each dimension has a value that ranges from 1 to 10. As a result, we represent this work as a regression problem. In the validation and test sets, each annotation has numerous annotators, and because the annotation is a subjective judgment, we contrast its performances using two alternative losses:

1. a marginal Euclidean loss ($Loss_{2cont}$),
2. the Smooth Loss1 Smooth $Loss_{1cont}$.

While computing loss, the former specifies an error margin (V k) that is not taken into account. The border Euclidean loss is defined as follows for VAD dimensions:

$$Loss_{2cont} = \frac{1}{\#D} \sum_{L \in D} V_L (Y_L^{\wedge cont} - Y_L^{cont})^2 \tag{2}$$

where D is for {Valence, Arousal, Dominance} and #*D* has 3, ($Y_L^{\wedge cont}$ and Y_L^{cont}) predicted originally and ground-truth normalization for Lth dimension and *VL* (= 0, 1) is represented as a binary weight for representing the marginal error:

$$\begin{aligned} V_L = 0, \, if \left| Y_L^{\wedge cont} - Y_L^{cont} \right| < \varepsilon \\ = 1, \ \text{rather} \end{aligned} \tag{3}$$

If the error is smaller than a specific number, the Smooth Loss1 losses are related to the error value by calculating the squared error. This loss is widely used for identifying objects, as demonstrated in our research, and is less reliant on outliers. The Smooth Loss1 losses are precisely defined as follows:

$$SmoothLoss_{1cont} = \sum_{L=1}^{3} V_L \begin{cases} 0.5 x_k^2, \ if \, |x_L| < 1 \\ |x_L| - 0.5, \ \ Otherwise \end{cases} \tag{4}$$

where xL = ($Y^{\hat{}cont} - Y^{Cont}$), and V_L stands for weight that is assigned to each continuous dimension and is set to 1.

5.3 Combined Losses ($\mathbf{Loss_{comb1}}$, $\mathbf{Loss_{comb2}}$)

As explained below, the composite loss function is defined as a weighted mixture of two independent losses:

$$Loss_{comb1} = \lambda_{disc} Loss_{disc} + \lambda_{cont} Loss_{2cont} \tag{5}$$

$$Loss_{comb2} = \lambda_{disc} Loss_{disc} + \lambda_{cont} SLoss_{1cont} \tag{6}$$

$\lambda_{(\text{disc, cont})}$, weights the significance of every loss that is set observationally keeping the validation into consideration. After multiple mergers, we realized, $\lambda_{\text{disc}} = 0.5$, $\lambda_{\text{cont}} = 0.5$ delivers the greatest results. Lossdisc and Losscont ($\text{Loss}_{2\text{cont}}$, $\text{SLoss}_{1\text{cont}}$) indicate the losses associated with learning the emotion classifications and the continuous dimensions. This combined loss parameter is used in the model to effectively impose multi-task training. When training, it determines the loss from tasks and inputs it back to the model. By acquiring variables in a common manner, the model is able to generalize its performance.

5.4 Average Precision (AP)

For measuring the performance, we first define a threshold for every emotion category based on the predictions of the validation set. Following framework activation in the assessment step, a forward track of a sample from the validation set produces 26 emotion classification possibilities for 26 emotional expressions. For each validation set sample, we collected all such predicted probability values. Following that, Precision-Recall curves are plotted for each emotion category based on the sample predictions and labels. We rectify the point where Precision is equivalent to Recall for all 26 curves, which is the classification threshold. Every one of the 26 emotional states now has 26 thresholds. These threshold values are later tried on the projected probabilities of the model for testing dataset samples. Average Precision refers to the region bounded by the curves. This score assesses the model's overall performance across all emotion categories. Average Precision ranges from 0 to 100, where 0 indicates no memory and precision and 100 represents perfect recall and precision.

5.5 Average Absolute Errors (AAE)

We utilize Mean Absolute Error (AAE) to assess the efficacy of the continuous dimension task, which is a regression. An error for a given continuous dimension is an absolute measure of the difference between its target value and predicted value. Errors that are fewer than the predetermined error margin are considered in the calculation. These errors are subsequently averaged across all inputs for every continuous dimension. As a result, the AE is determined as follows:

$$\text{AAE} = \frac{1}{\#\text{D}} \sum\nolimits_{L \in D} V_L \left| Y_L^{\wedge cont} - Y_L^{cont} \right| \tag{7}$$

where $D = \{\text{Valence, Arousal, Dominance}\}$, $Y_L^{\wedge Cont} - Y_L^{Cont}$ serve as the Lth dimension's prediction and normalized ground-truth, m stands for the sample numbers and $VL(= 0,1)$ acts as binary weight used to present the error margin. As lower the AAE is, performance is that better is presented (see Fig. 6).

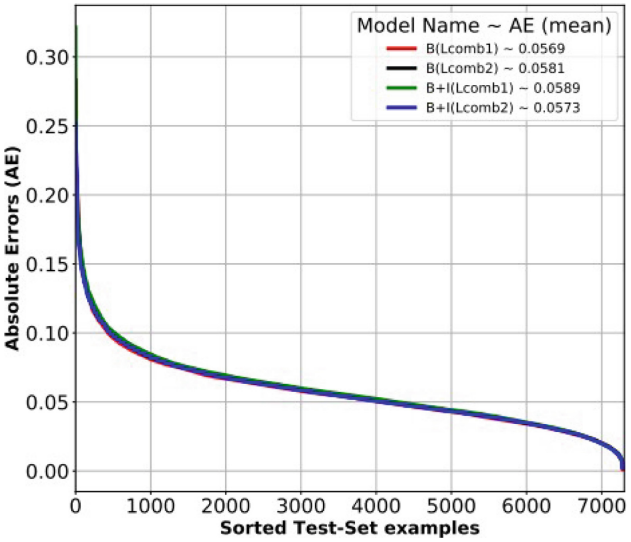


Fig. 6. VAD estimation with AAE

6 Results

We can note that Table 1 represents the average absolute errors calculated from various models for all three dimensions, by comparing it with $\text{Loss}_{2\text{cont}}$ and $\text{Loss}_{\text{comb1}}$. Table 2 represents the cumulative average precision on various models by making the comparison among emotion categories considering $\text{Loss}_{\text{disc}}$ and $\text{Loss}_{\text{comb1}}$. Here 'B' stands for Person's features and 'I' stand for Image Background features. (see Fig. 7) for the confusion matrix indicating covariance and correlation for all 26 emotion categories.

Table 1. Comparative analysis of performance using Average Absolute Errors (AAE) for various models for criterion $\text{Loss}_{2\text{cont}}$ and $\text{Loss}_{\text{comb1}}$ criteria

Continuous Dimensions	Input to model					
	B		I		B + I	
	$\text{Loss}_{2\text{cont}}$	$\text{Loss}_{\text{comb1}}$	$\text{Loss}_{2\text{cont}}$	$\text{Loss}_{\text{comb1}}$	$\text{Loss}_{2\text{cont}}$	$\text{Loss}_{\text{comb1}}$
V-Valence	0.0564	0.0527	0.0510	0.0543	0.0643	0.0596
A-Arousal	0.0681	0.0610	0.0621	0.1162	0.0632	0.0448
D-Dominance	0.0529	0.0571	0.0624	0.0887	0.0624	0.0574
Mean	0.0590	0.0589	0.0618	0.0743	0.0583	0.0689

Table 2. Average Precision (AP) for some models, comparing their performance with the $Loss_{disc}$ and $Loss_{comb1}$ criteria

Emotion Categories	Model Inputs					
	B		I		B + I	
	$Loss_{disc}$	$Loss_{comb1}$	$Loss_{disc}$	$Loss_{comb1}$	$Loss_{disc}$	$Loss_{comb1}$
Affection	22.85	22.9	20.63	22.13	20.56	22.26
Anger	7.46	7.55	7.98	6.74	9.2	7.55
Annoyance	10.64	8.92	6.83	9.59	10.89	12.28
Anticipation	56.74	59.71	54.85	58.22	53.37	59.71
Aversion	6.24	6.18	6.47	6.43	6.68	7.55
Confidence	70.05	74.89	66.86	69.33	61.69	79.07
Disapproval	9.64	8.73	10.11	8.94	9.2	12.1
Disconnection	19.44	21.88	17.99	18.82	21.8	21.47
Disquietment	14.82	15.42	14.85	15.18	15.72	16.64
Doubt	28.37	30.29	29.49	29.21	29.54	29.25
Embarrasment	2.62	3.48	3.7	3.62	3.66	3.54
Engagement	85.28	85.1	82.43	85.82	82.82	87.34
Esteem	17.35	19.46	19	18.36	18.65	18.45
Excitement	69.42	74.83	67.08	69.19	66.3	78.06
Fatigue	8.65	8.95	7.73	8.65	8.97	9.97
Fear	11.65	13.95	14.11	12.64	13.21	13.44
Happiness	57.17	59.81	56.74	57.14	55.65	61.79
Pain	5.76	4.75	3.87	4.98	5.89	5.52
Peace	20.21	18.95	18.45	19.89	18.79	20.53
Pleasure	44.16	43.68	43.27	43.96	43.44	43.22
Sadness	10.11	9.23	12.82	7.24	10.21	11.46
Sensitivity	5.42	5.33	5.86	5.34	5.19	5.92
Suffering	9.34	6	9.44	7.11	8.51	8.75
Surprised	17.43	18.3	17.15	17.36	17.87	17.52
Sympathy	12.3	11.76	12.7	12.39	11.62	12.54
Yearning	7.46	8.92	9.38	9.03	8.74	9.44
Mean	24.25	24.96	23.83	24.13	23.77	25.97

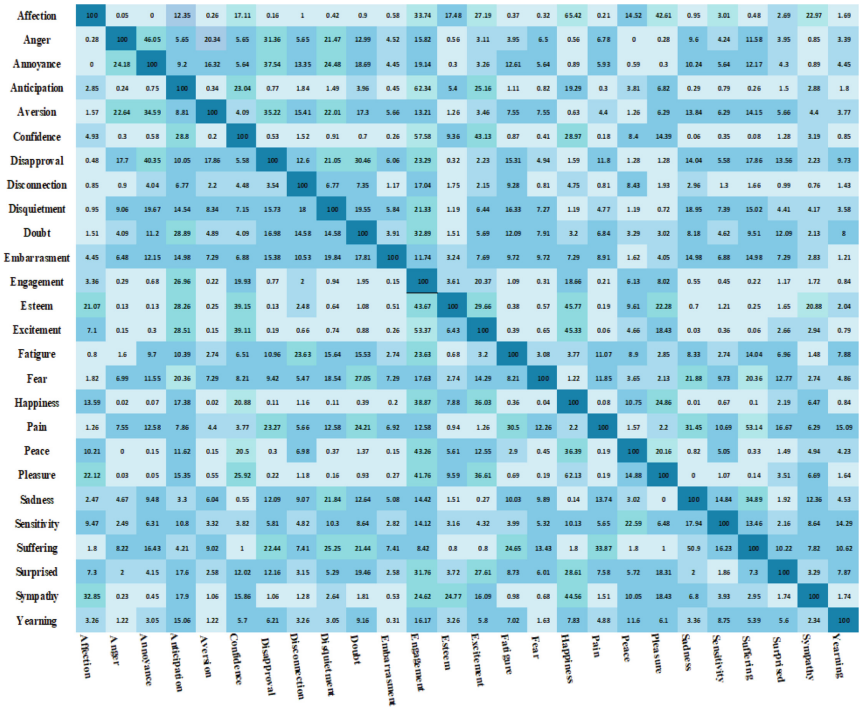


Fig. 7. Correlations and covariance between pairs of 26 emotion categories generated from the EMOTIC and BoLD datasets, where each row shows the probability of occurrence of all other categories

7 Conclusion

The present study highlights the critical importance of incorporating visual scene context for the task of automated emotion recognition. Our investigation begins by introducing the concept of emotion detection and providing an overview of its significance (as discussed in section 1.1). We subsequently delve into the role of contextual cues in emotion perception and explore prior research that has focused on incorporating contextual information to improve emotion recognition. However, our analysis of the existing literature reveals that previous approaches have not fully met the requirements of emotion recognition in context. To address this gap, we have proposed a novel baseline fusion CNN model for identifying emotions in a visual scenic context. Our model has been tested on two datasets, EMOTIC and Bold, and we have performed a series of baseline experiments to evaluate the efficacy of our approach. We have also compared different feature types for encoding contextual information. Our experimental findings confirm that the use of visual scene context significantly enhances the accuracy of emotion recognition, thereby highlighting the relevance of our proposed model. Our study underscores the potential of the EMOTIC and Bold datasets in advancing research in this direction and encourages further exploration of this important topic. Therefore, we believe that our

work will inspire further investigations into the importance of visual scene context in the development of automatic emotion recognition systems.

References

1. Kosti, R., Alvarez, J. M., Recasens, A., Lapedriza, A.: EMOTIC: emotions in context dataset. In: Computer Vision and Pattern Recognition (2017). https://doi.org/10.1109/cvprw.2017.285
2. Luo, Y., Ye, J., Adams, R.B., Li, J., Newman, M.G., Wang, J.Z.: ARBEE: towards automated recognition of bodily expression of emotion in the wild. Int. J. Comput. Vision **128**, 1–25 (2018). https://doi.org/10.1007/s11263-019-01215-y
3. Kosti, R., Alvarez, J.M., Recasens, A., Lapedriza, A.: Context-based emotion recognition using emotic dataset. IEEE Trans. Pattern Anal. Mach. Intell. **42**(11), 2755–2766 (2019)
4. Zhang, M., Liang, Y., Ma, H.: Context-aware affective graph reasoning for emotion recognition. In: 2019 IEEE International Conference on Multimedia and Expo (ICME), pp. 151–156. IEEE (2019)
5. Lee, J., Kim, S., Kim, S., Park, J., Sohn, K.: Context-aware emotion recognition networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10143–10152 (2019)
6. Mittal, T., Bera, A., Manocha, D.: Multimodal and context-aware emotion perception model with multiplicative fusion. IEEE Multimedia **28**, 67–75 (2021)
7. Mittal, T., Guhan, P., Bhattacharya, U., Chandra, R., Bera, A., Manocha, D.: Emoticon: context-aware multimodal emotion recognition using frege's principle. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14234–14243 (2020)
8. Hoang, M., Kim, S., Yang, H., Lee, G.: Context-aware emotion recognition based on visual relationship detection. IEEE Access **9**, 90465–90474 (2021). https://doi.org/10.1109/access.2021.3091169
9. Goyal, A., Kumar, N., Guha, T., Narayanan, S.S.: A multimodal mixture- of-experts model for dynamic emotion prediction in movies. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2822–2826). IEEE (2016)
10. Liu, S., Gao, P., Li, Y., Fu, W., Ding, W.: Multi-modal fusion network with complementarity and importance for emotion recognition. Inf. Sci. **619**, 679–694 (2023)
11. Gupta, S., Kumar, P., Tekchandani, R.K.: Facial emotion recognition based real- time learner engagement detection system in online learning context using deep learning models. Multimed Tools Appl **82**, 11365–11394 (2023). https://doi.org/10.1007/s11042-022-13558-9
12. Chaudhari, A., Bhatt, C., Krishna, A., Mazzeo, P.L.: ViTFER: facial emotion recognition with vision transformers. Appl. Syst. Innovation **5**, 80 (2022). https://doi.org/10.3390/asi5040080
13. Chaudhari, A., Bhatt, C., Krishna, A., Travieso, C.M.: Facial emotion recognition with inter-modality-attention-transformer-based self-supervised learning. Electronics **12**, 288 (2023). https://doi.org/10.3390/electronics12020288
14. Kothadiya, D., Chaudhari, A., Macwan, R., Patel, K., Bhatt, C.: The convergence of deep learning and computer vision: smart city applications and research challenges. In: Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication &Amp; Security (ICIIC 2021) (2021). https://doi.org/10.2991/ahis.k.210913.003
15. Zhao, J., Mao, X., Chen, L.: Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomed. Signal Process. Control **47**, 312–323 (2019). https://doi.org/10.1016/j.bspc.2018.08.035
16. Ye, M., Qian, H., Guangyuan, L.: CNN-LSTM facial expression recognition method fused with two-layer attention mechanism. Comput. Intell. Neurosci. **2022**, 1–9 (2022). https://doi.org/10.1155/2022/7450637

17. Gao, Y., Li, B., Wang, N., Zhu, T.: Speech emotion recognition using local and global features. In: Zeng, Y., He, Y., Kotaleski, J.H., Martone, M., Xu, B., Peng, H., Luo, Q. (eds.) BI 2017. LNCS (LNAI), vol. 10654, pp. 3–13. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70772-3_1
18. Milton, A.H., Roy, S.S., Selvi, S.T.: SVM scheme for speech emotion recognition using MFCC feature Int. J. Comput. Appl. (2013).https://doi.org/10.5120/11872-7667
19. Huang, Z., Dong, M., Dong, M., Zhan, Y.: Speech Emotion Recognition Using CNN. ACM Multimedia (2014).https://doi.org/10.1145/2647868.2654984
20. Lim, W., Jang, D., Lee, T.: Speech emotion recognition using convolutional and Recurrent Neural Networks. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (2016). https://doi.org/10.1109/apsipa.2016.7820699
21. Kalliatakis, G., Ehsan, S., Leonardis, A., Fasli, M., McDonald-Maier, K.D.: Exploring object-centric and scene-centric CNN features and their complementarity for human rights violations recognition in images. IEEE Access **7**, 10045–10056 (2019). https://doi.org/10.1109/access.2019.2891745
22. Sun, G., et al.: Deep fusion of localized spectral features and multi-scale spatial features for effective classification of hyperspectral images. Int. J. Appl. Earth Obs. Geoinf. **91**, 102157 (2020). https://doi.org/10.1016/j.jag.2020.102157
23. Lu, M., Du, G., Li, Z.: Multimode gesture recognition algorithm based on convolutional long short-term memory network. Comput. Intell. Neurosci. **2022**, 1 (2022). https://doi.org/10.1155/2022/4068414

The Impact of Data Selection Strategies on Distributed Model Performance

Miguel Guimarães[1(✉)], Filipe Oliveira[1], Davide Carneiro[1(✉)], and Paulo Novais[2]

[1] CIICESI, Escola Superior de Tecnologia e Gestão, Instituto Politécnico do Porto, Felgueiras, Portugal
{mam,8170164,dcarneiro}@estg.ipp.pt

[2] ALGORITMI Research Centre/LASI, University of Minho, Braga, Portugal
pjon@di.uminho.pt

Abstract. Distributed Machine Learning, in which data and learning tasks are scattered across a cluster of computers, is one of the answers of the field to the challenges posed by Big Data. Still, in an era in which data abounds, decisions must still be made regarding which specific data to use on the training of the model, either because the amount of available data is simply too large, or because the training time or complexity of the model must be kept low. Typical approaches include, for example, selection based on data freshness. However, old data are not necessarily outdated and might still contain relevant patterns. Likewise, relying only on recent data may significantly decrease data diversity and representativity, and decrease model quality. The goal of this paper is to compare different heuristics for selecting data in a distributed Machine Learning scenario. Specifically, we ascertain whether selecting data based on their characteristics (meta-features), and optimizing for maximum diversity, improves model quality while, eventually, allowing to reduce model complexity. This will allow to develop more informed data selection strategies in distributed settings, in which the criteria are not only the location of the data or the state of each node in the cluster, but also include intrinsic and relevant characteristics of the data.

Keywords: Machine Learning · Distributed Computing · Data Selection

1 Introduction

New approaches to Machine Learning (ML), such as Distributed ML [9] or Federated ML [5], emerged in the last years as a response to the challenges of learning from Big Data [11].

The appearance of Distributed ML followed a previous trend, in which storage was moved from large monolithic data architectures to distributed mesh ones [6]. Evidently, once data were distributed, it would not be feasible or possible to move all the data to a single location in order to run learning tasks. Hence, learning

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 157–166, 2023.
https://doi.org/10.1007/978-3-031-43461-7_16

tasks became parallel and distributed, starting to include coordination and/or message passing mechanisms such as ZooKeeper [3] or ZeroMQ [4].

Even so, decisions on which data to use when training a ML model must often still be made, be it because the data are simply too much, or because there are restrictions on training time or model complexity.

In the past we have proposed approaches to deal with this issue [7], namely by selecting the data (and running the distributed learning tasks) in the nodes in which the data are located, and according to the state of each node. Given that modern distributed file systems have some form of replication, this can be explored to select from the best candidate nodes according to their state.

Other typical approaches include selecting data based on their freshness. That is, more recent data will be more likely to be selected. While this is, at first sight, intuitive, it might not always hold the best results in terms of model quality. This happens because old data might describe patterns that are still relevant and useful, as happens in scenarios of streaming data with recurring concept drift (i.e. seasonality) [10], and that would otherwise be lost.

The main goal of this paper is to analyze the influence that different data selection strategies have on the quality and complexity of ML models. Specifically, we aim to ascertain whether selecting data based on a diversity criteria allows to obtain better models, or smaller but equivalent models (in terms of predictive performance) than when the full set of data are used. This will allow to devise new strategies of data selection, based not only on data locality, data availability and cluster state, but that also take into account intrinsic characteristics of the data, in an attempt to select data that are as rich and diverse as possible. This will contribute to the design of overall more efficient ML pipelines.

2 Distributed Machine Learning

The experiments described in this paper, in which a large number of ML Ensembles were trained, were carried out in the CEDEs platform. CEDEs - Continuously Evolving Distributed Ensembles - is a Distributed Learning platform developed in the context of a funded research project.

This section outlines the major modules comprising the architecture of CEDEs. The Storage Layer serves as the central element, implemented on top of an HDFS cluster. In this file system, large datasets are broken down into fixed-size blocks, usually of 128 MB. Various data sources, including files in different formats (e.g., CSV, Parquet, Avro) and databases, can be employed.

Thus, the user of CEDEs initiates a ML project by uploading a data source into the HDFS, which is subsequently split, distributed, and replicated throughout the cluster. Once this process is concluded, the user can create an actual ML project by choosing the desired dataset, and supplying the relevant information such as project name, description and owners, or base model(s) hyperparameters. After configuring the project, the user can begin training a new Ensemble. This task is controlled by the Coordination module, which interacts with the Optimization and Metadata modules.

Training an ensemble occurs in a distributed manner and benefits from two aspects: the replication of blocks within files, enabling them to be accessed concurrently on multiple nodes, and the principle of data locality. To train an ensemble, one base model is trained for each block of a file (excluding replicas). These base models will collectively form the Ensemble. Consequently, it is essential to determine the optimal node for training each base model, based on criteria such as the state of the node where the block is available within the cluster.

Specifically, the Coordination module uses the following inputs:

- Block locations are used to know where in the cluster each individual block exists. Generally, each block is simultaneously available in at least 3 nodes;
- Task cost prediction, in terms of time and computational resources (e.g. RAM). This is given by a previously developed meta-learning module [2];
- The state of nodes, in terms of the makespan of their tasks and a prediction of the necessary computational resources;

The goal of this paper, as elaborated further below, is to determine whether the characteristics of the data blocks can also be used as a relevant optimization criteria when selecting the blocks for training a new Ensemble.

Once the Coordination module defines the allocation of tasks, the process of training the base models initiates. In order to coordinate the worker nodes, a communication protocol was implemented in ZeroMQ: a brokerless asynchronous messaging library that allows the implementation of multiple socket communication patterns, useful for implementing distributed systems.

During the distributed training process, worker nodes regularly report their progress to the coordinator, allowing client applications to access this information in real time. When a worker node finishes training a base model, it notifies the Coordinator, which then moves on to the next task in its queue, if it exists. Additionally, the base model is serialized and stored in the HDFS, and automatically replicated and distributed across the cluster, based on the replication and balancing factors defined in the configuration. This allows to serve the same base model on multiple nodes shortly after the end of its training, significantly increasing model availability and the robustness of the system.

This architecture is illustrated in Fig. 1, in which the following terminology is used: $F_iB_jR_k$ represents a replica k of block j of file i. The names of base models follow the same logic, but start with the letter'M'. The arrow between a specific block replica and a model means that the base model was created based on that specific block, on that node. Replicas of that base model may however exist in other nodes, created through replication.

When a client requests predictions from an ensemble, a similar process to that of training base models takes place. The Coordination module first queries the Optimization module for information about which base models will be used in the ensemble. Indeed, not all base models may be included, depending on the specific project configuration. If the number of intended base models is smaller than the number of blocks in the dataset, the Optimization module selects the best group of base models based on a specific heuristic, that may include models'

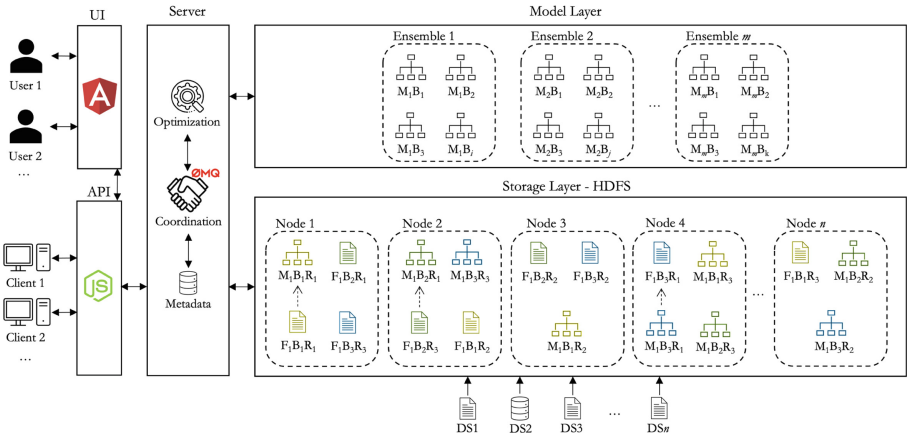


Fig. 1. Depiction of the CEDEs architecture.

performance metrics and the state of the nodes where they are located. The main goal of this work is to check whether including a criteria of data diversity improves the quality and decreases the necessity for data, of the Ensemble.

When the Ensemble is defined, the Coordinator requests all the corresponding worker nodes for predictions for a specific dataset. Once all predictions are received, they are combined to calculate the final predictions by the ensemble. For regression problems, the mean is used, while for classification problems, the mode is used. Ensembles are thus abstract and dynamic constructs, as depicted in Fig. 1 in the Model Layer, built out of a selection of available base models for a given ML problem.

The architecture includes two more modules: the API module, and the User Interface module. The API module provides endpoints for interacting with the system's services, which in turn communicate with the Coordination module, when applicable, via ZeroMQ messages. External client applications can use this API. Finally, the User Interface module is intended for human users and communicates with the API module.

3 Methodology

This work analyzes how different data selection strategies impact the performance of a distributed Machine Learning model, implemented as an Ensemble, with the goal of determining the strategy that produces the best model with the least amount of data and base models.

To this end, 60 different Machine Learning experiments were run. Each experiment, described below, refers to a publicly available dataset (from those described in Table 1), one block size (i.e. 4 MB, 8 MB and 16 MB) and one of five different heuristics for building the Ensemble. The datasets are all traditional

tabular datasets, with the exception of the "covid infections" dataset, which is a time-series one.

The interest in experimenting with different block sizes lies in its potential influence on the quality of the base models: in principle, the smaller the block, the lower and less reliable the quality of the base models (due to significant variations between blocks and the data not being representative of the whole problem). Thus, in each scenario, CEDEs was restarted and configured with different block sizes to implement the experiments described.

Each dataset was first split into training (75%) and testing (25%) sets. The training data was uploaded into CEDEs, as described below, while the test data was held-out to assess the generalization ability of the Ensembles created.

Table 1. Characterization of the datasets used.

Dataset	Rows	Columns	Size
covid infections	4623663	19	585 MB
solar production	2782080	15	399 MB
temperature	2434608	20	351 MB
car prices	1957675	1108	4160 MB

The methodology followed in each experiment, which is also depicted in Fig. 2, was as follows. CEDEs was configured with one of the tested block sizes (i.e. 4 MB, 8 MB and 16 MB) and one of the datasets was uploaded into CEDEs. The dataset was then automatically split into blocks, distributed and replicated. For some datasets, some data preparation tasks were executed (e.g. cleaning, feature engineering, encoding), which were implemented in Apache Spark. However, given that the goal is not at this stage to find the best model possible but only to compare different approaches, no special effort was put on a further improvement of the dataset.

Once the dataset was pre-processed, the meta-features were extracted for each block. This was achieved using the pymfe - Python Meta-Feature Extractor library[8]. Meta-features are features that encode characteristics of an underlying dataset and may represent many different dimensions such as data distribution measures, data quality measures, measures of variability, other statistical measures, etc.

Once meta-features were extracted for each block, a base model was also trained for each block and, as usually happens in CEDEs, their performance metrics (e.g. RMSE, MAE, MSE, Logloss) recorded in a database.

Next, a k-means algorithm was run on the meta-features dataset. The goal was to find clusters of blocks that share similar characteristics (meta-features), thus allowing to select data (and consequently blocks) also based on their similarity. K-means was run with $k = 5$, thus it always finds 5 groups of datasets, based on their meta-features.

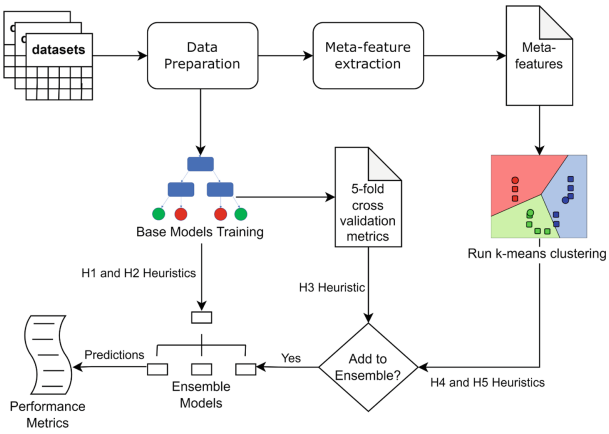


Fig. 2. Methodology followed in each of the experiments.

Then, for each dataset and block size, 5 different Ensembles were trained, with the following heuristics being used for base-model selection:

- H_1 - The Ensemble is created following a Bagging approach, with all the base models being considered. However, each base model has a weight that is inversely proportional to its quality, given by the RMSE calculated through 5-fold cross-validation.
- H_2 - The Ensemble is created following a traditional Bagging approach, with all the base models being used, and each having equal weight.
- H_3 - The Ensemble is created by selecting only the top 50% base models, and with all selected models having the same weight. Models are ranked by their RMSE, calculated through 5-fold cross-validation.
- H_4 - The Ensemble is created by picking one random model from each of the 5 clusters. The goal of this heuristic is to maximize data diversity, by picking models trained on blocks with different characteristics.
- H_5 - The Ensemble is created by picking 5 random models from the cluster that has the highest density, i.e. the datasets have more in common or share similar traits, according to the meta-features.

Although CEDEs allows to build heterogeneous Ensembles, for the work described in this paper, all the Ensembles trained are Random Forests, i.e., collections of Decision Trees. For the sake of fairness, all Decision Trees were trained with the same configuration: max_depth = 25, min_samples_split = 2, max_leaf_nodes = infinite and ccp_alpha = 0.5.

Each Ensemble was then evaluated both through 5-fold cross validation and on the test data that was held-out. The results of this evaluation and of comparing the different heuristics are analyzed in Sect. 4.

4 Results

This section describes the results obtained while comparing the 5 previously mentioned heuristics. The experiments were run in a CEDEs instance, that was configured with 3 different block sizes. The goal is to analyze the impact of the data selection heuristic on the quality of the ensemble, and whether block size affects it.

Specifically, we want to evaluate the proposed heuristic (H_4), and determine whether it produces simpler but equally good ensembles, independently of the number of blocks of the original dataset. This selection is based on the diversity of the data, through the extraction of meta-features [1]. Specifically, meta-features were considered block-wise. As previously mentioned, in this heuristic, data blocks are selected based on their belonging to a different cluster. This means that under H_4, ensembles always have 5 models. The same is true for H_5. However, in this case, the blocks are all obtained from the same cluster.

The analysis carried out was extensive, taking into consideration several factors including block size (x3), dataset (x4) and heuristics (x5). Moreover, we looked at several indicators including the number of base models whose training was spared using H_4 and H_5, the amount of CPU time saved, and model performance metrics (e.g. RMSE). For this reason, this section presents only a summary of the results.

Table 2 presents some results with the 'solar production' dataset. Here we simply compare our proposed approach H_4 with a traditional bagging one H_2. The proposed approach only outperforms the traditional one in one scenario (8 MB block size), in which the ensemble build with 5 diverse data blocks has an RMSE that is 0.58% lower. In the other two scenarios (4 MB and 16 MB), the proposed approach results in Ensembles that are slightly worse in terms of RMSE (0.11% and 0.3% worse, respectively).

While this lower performance is quite marginal, the ensembles build clearly outperform the traditional ones in terms of the number of models trained or the time saved. In this case, the number of models trained is between ≈80% and ≈95% smaller, depending on the block size, which also results in a reduction of ≈80% to ≈95% of the ensemble training time.

Tables 3 and 4 provide the remaining insights in terms of the analysis carried out. Table 3 contains an overall analysis, concerning all the experiments carried

Table 2. Summary of the results for the 'solar production' dataset with ensembles trained with 4, 8 and 16 MB block size. **M** - the number of models whose training was avoided (%); **T** - CPU time saved (%); **R** - difference in RMSE (%).

Block Size	Heuristic	M	T	R
4 MB	$\boldsymbol{H_2}$ vs H_4	−94.79%	−94.81%	+0.11%
8 MB	H_2 vs $\boldsymbol{H_4}$	−89.56%	−89.53%	**−0.58%**
16 MB	$\boldsymbol{H_2}$ vs H_4	−79.17%	79.38%	+0.3%

Table 3. Compiled summary of the total results containing the four datasets with models trained in 4, 8 and 16 MB block size. **TM** - total number of models trained; **TT** - total training time (seconds); **TR** - total time reduction (seconds); **UR** - used RAM (GB); **RR** - RAM reduction (GB).

Heuristic	TM	TT	TR	UR	RR
H_1	618	4561.99	–	2680.82	–
H_2	618	4561.99	–	2680.82	–
H_3	618	4561.99	–	2680.82	–
H_4	60	771.24	3790.74	272	2409
H_5	60	765.37	3796.62	268	2413

out. In what concerns heuristics H_1 to H_3, 618 base models were trained in total, for each one. This includes the 3 different block sizes and the 4 different datasets. The only difference is that the Ensembles built with H_3 only use half the base models: the half with the best RMSE. However, these base models need to be trained to calculate their RMSE, the number is the same.

On the other hand, H_4 and H_5 only need the training of 5 base models per Ensemble, for the 12 experiments made (4 datasets * 3 block sizes), which results in 60 base models trained in total. This means that these heuristics require 558 fewer models each, in comparison with the other 3. Following the same rationale, these heuristics avoided approximately 3790 s of CPU time in the training of models, and approximately 2409 GB of used RAM (measured at half-second interval). In this table, the columns TR and RR represent the reduction in terms of CPU time and RAM, when compared to traditional approaches.

Table 4, on the other hand, shows the average values for each experiment performed, grouped by heuristic. For each experiment the results were ranked, and analyzed relatively. For instance, for each experiment the best heuristic in terms of RMSE was identified. The remaining were ranked by decreasing RMSE, and their RMSE was analyzed relatively to the lowest one. Thus, column AR depicts how much worse the RMSE of a given heuristic was, in percentage. The results were then averaged.

The table shows that the proposed heuristic H_4 is the one that has the lowest value in this indicator, which means that it was most frequently the best approach in terms of predictor performance.

On average, the proposed approach manages to obtain the best performance in terms of RMSE, and saving training resources, both in terms of CPU time and of memory used, saving on average approximately 85% of models per ensemble: corresponding to ≈5 min of training time avoided and ≈201 GB spared RAM usage. This was the difference of training an Ensemble in only one minute (heuristics H_4 and H_5) instead of 6 min (heuristics H_1 to H_3). However, this difference might be much bigger in real scenarios with the typical 128 MB block size and with even bigger datasets.

Table 4. Compiled summary of the mean results considering the 4 datasets with models trained with 4, 8 and 16 MB block sizes. **AM** - average number of avoided models (%); **AR** - average difference in RMSE compared with the best ensemble (%); **TT** - average training time (s); **TR** - average time reduction (s); **UR** - used RAM (GB); **RR** - average RAM reduction (GB).

Heuristic	AM	AR	TT	AT	UR	RR
H_1	0%	−6.480%	380.17	–	223.40	–
H_2	0%	−6.680%	380.17	–	223.40	–
H_3	0%	−3.310%	380.17	–	223.40	–
H_4	85.16%	**−2.710%**	64.27	315.90	22.64	200.76
H_5	85.16%	−3.280%	63.78	316.39	22.31	201.09

Thus, heuristic H_4 results in ensembles that train quicker and achieve better performance most of the times, although not always. Nonetheless, even when the performance is not the best, it is close to it. In opposition, the amount of necessary resources and the complexity of the resulting ensemble is far smaller, which is a significant advantage, especially in Big Data and streaming scenarios.

The proposed heuristic for data selection is thus one that might sometimes result in slightly worse model performance, but one that significantly reduces training time and computational resources.

5 Conclusions and Future Work

As is widely accepted, data and their quality are a key determinant of model quality. While in some cases all the available data is used, in others a subset of the data must be selected. This happens when the available data is too much, or when there are strict requirements concerning model training time or model complexity. In such cases, the strategies used for selecting data must be carefully considered, to maximize the quality of the model.

In this paper we analyze 5 different heuristics for selecting data. Two of them are novel approaches, and are based on meta-features, that is, characteristics of the data. Our goal was to ascertain whether selecting a smaller number of subsets of data but while maximizing diversity, would result in better ensembles. The results show not only this is generally true for the cases tested, but also that the resulting Ensembles are much smaller. Moreover, we avoid the training of a significant amount of models, which results in significant savings in terms of computational resources and time.

There are, however, limitations to this work. One of the major ones is that each experiment was only run once. Given that a methodology based on a holdout dataset was used, the split of data and the results are dependent on luck. In future work we will run the same experiments with cross-validation, to have more reliable results (cross-validation was used but only at the level of the base models). Moreover, in order to reduce variability in the results, a single algorithm

was used (Decision Trees). In future work we will repeat the experiment with other algorithms, in far more extensive experiments. Also, the number of clusters (5) was arbitrarily selected and is, most likely, not the best one for each problem. In the future, a prior heuristic for selecting the optimum value of k (number of clusters) will be implemented.

Finally, the findings of this work will be included in the optimization module of CEDEs, so that it can also take into account the characteristics of the data when deciding which blocks to use.

Acknowledgments. This work has been supported by national funds through FCT - Fundação para a Ciência e Tecnologia through projects UIDB/04728/2020, EXPL/CCI-COM/0706/2021 and CPCA-IAC/AV/475278/2022.

References

1. Alcobaça, E., Siqueira, F., Rivolli, A., Garcia, L.P.F., Oliva, J.T., de Carvalho, A.C.P.L.F.: MFE: towards reproducible meta-feature extraction. J. Machine Learn. Res. **21**(111), 1–5 (2020). http://jmlr.org/papers/v21/19-348.html
2. Guimarães, M., et al.: Predicting model training time to optimize distributed machine learning applications. Electronics **12**(4), 871 (2023)
3. Hunt, P., Konar, M., Junqueira, F.P., Reed, B.: ZooKeeper: wait-free coordination for internet-scale systems. In: USENIX Annual Technical Conference, vol. 8 (2010)
4. Lauener, J., Sliwinski, W., Cern, G.: How to design & implement a modern communication middleware based on ZeroMQ. In: Proceedings of ICALEPCS, vol. 17, pp. 45–51 (2017)
5. Liu, J., et al.: From distributed machine learning to federated learning: a survey. Knowl. Inf. Syst. **64**(4), 885–917 (2022)
6. Machado, I.A., Costa, C., Santos, M.Y.: Data mesh: concepts and principles of a paradigm shift in data architectures. Procedia Comput. Sci. **196**, 263–271 (2022)
7. Monteiro, J., Oliveira, O., Carneiro, D.: Task scheduling with makespan minimization for distributed machine learning ensembles. In: 2022 IEEE 4th Eurasia Conference on IOT, Communication and Engineering (ECICE), pp. 435–438 (2022). https://doi.org/10.1109/ECICE55674.2022.10042894
8. Rivolli, A., Garcia, L.P., Soares, C., Vanschoren, J., de Carvalho, A.C.: Characterizing classification datasets: a study of meta-features for meta-learning. arXiv preprint arXiv:1808.10406 (2018)
9. Verbraeken, J., Wolting, M., Katzy, J., Kloppenburg, J., Verbelen, T., Rellermeyer, J.S.: A survey on distributed machine learning. ACM Comput. Surv. (CSUR) **53**(2), 1–33 (2020)
10. Wang, H., Abraham, Z.: Concept drift detection for streaming data. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–9. IEEE (2015)
11. Zhou, L., Pan, S., Wang, J., Vasilakos, A.V.: Machine learning on big data: opportunities and challenges. Neurocomputing **237**, 350–361 (2017)



AI-Powered Smart Book: Enhancing Arabic Education in Palestine with Augmented Reality

Murad Al Rajab[1], Suhail Odeh[2(✉)], Sherin Hazboun[2], and Emran Alheeh[2]

[1] College of Engineering, Abu Dhabi University, Abu Dhabi, UAE
murad.al-rajab@adu.ac.ae

[2] Department of Software Engineering, Bethlehem University Bethlehem, Bethlehem, Palestine
{sodeh,shazboun}@bethlehem.edu

Abstract. The global outbreak of COVID-19 has had a major impact on the education of millions of children, hindering their involvement in educational settings and negatively affecting their academic progress. This situation highlights the need to reassess teaching methods and find ways to encourage student involvement in different learning scenarios. Specifically, elementary-level students studying the Arabic language have faced significant challenges in comprehending academic subjects. To tackle this issue, we propose a cutting-edge and creative proposed application called SBAR (Smart Book Augmented Reality), which leverages artificial intelligence (AI) to support students in their interaction with stories from their Arabic textbook curriculum in the State of Palestine. Our proposed application utilizes augmented reality to create a 3D environment featuring book characters that interact with students, inspiring them to learn in a unique and stimulating manner. The app can be run on smartphones and tablets and has been designed to maintain student engagement. It provides two key features: a 3D augmented reality representation of the story that presents the tale through voice and animation, and an interactive component that allows students to respond to questions through natural speech. The interactive component relies on AI and deep learning algorithms to analyze student answers, evaluate them, and provide feedback through interactive voice. Our proposed application is user-friendly, efficient, resource-friendly, and innovative compared to other interactive educational apps. It has been trialed in local schools in Palestine and received positive responses from both students and teachers.

Keywords: Smart Book · Arabic · Augmented Reality · Voice Recognition · Interactive Learning · COVID-19

1 Introduction

The integration of Artificial Intelligence (AI) in the education sector has been slow to develop, despite the availability of AI-powered solutions. However, the recent COVID-19 pandemic has significantly altered this situation, with a shift toward virtual learning technology. A recent survey found that 86% of teachers believe technology should be an integral part of education [1]. According to the eLearning industry, up to 47% of

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 167–178, 2023.
https://doi.org/10.1007/978-3-031-43461-7_17

learning management products are expected to be equipped with AI capabilities in the next three years, with the potential to improve both learning and teaching [1].

In recent years, there has been an increasing interest in leveraging emerging technologies to enhance educational experiences. This study focuses specifically on Arabic education in Palestine and explores the integration of AI and AR technologies to create adaptive learning environments that align with the principles of ambient intelligence. By combining these technologies, we aim to provide Arabic language learners with interactive and personalized learning experiences that adapt to their individual needs and preferences.

Traditional face-to-face education mixed with traditional teaching methods has proven to be a failure during the COVID-19 pandemic, where online education was utilized as a solution to overcome the challenges of physical distancing. However, it was observed that the quality of education was negatively impacted. As a result, there is a pressing need for new technological innovations to overcome this problem.

Research work has been done in the area of using AR in education. A paper published in 2017 reports a review of literature on AR in educational settings considering 55 studies published between 2011 and 2016 and found most of the studies reported that AR in educational settings lead to better learning performance and promotes learning motivation, which because AR supplies the authenticity of graphical content and interaction [8]. Another study published in 2019 examined 58 articles between 2006 and 2016 containing AR to determine the trends of using it in education and found that mobile technologies such as tablets and smartphones are the most used. Mobile AR systems allow the user to take and use the AR wherever the user desires by removing the limitations of the user being in a specific location [9]. A 2020 study presented a quantitative meta-analysis of 46 empirical studies to identify how pedagogical approaches affect the impact of AR on education and found that an adequate combination of AR technology and specific pedagogical approaches enrich educational settings. It encouraged researchers to explore different pedagogical approaches that will potentially benefit AR interventions [10]. Another 2022 study's findings revealed an increase in the number of AR studies during the last four years where the most reported advantage of AR is that it promotes enhanced learning achievement and is also of interest to the younger generation but one of the limitations is the lack of educational applications with augmented reality [11].

In this research paper, we propose an innovative AI-powered augmented reality application called "SBAR", which is a smart book augmented reality assistant application that aims to help students in the Arabic language learning process in the State of Palestine. Our proposed application allows students to learn interactively and independently under the supervision of teachers and also helps teachers to explain lessons. SBAR utilizes 3D characters that explain lessons with an interactive sound system between the characters in the application and the user, which helps in increasing student engagement and interaction with the book. The application also enables students to target their device camera towards the story title or any of the story images, and then displays its characters in an animated 3D environment that tells the story through voice and animation. Additionally, the application allows the user to answer related questions through natural voice. It uses AI and deep learning algorithms to analyze the student's answers and

provide instantaneous interactive voice feedback. Our proposed application is designed for students who want to learn a specific topic on their own, and for teachers who are open to utilizing technology to interact more with their students. In the context of AR mobile applications, the proposed AI-based augmented reality application SBAR is a highly innovative solution. By combining artificial intelligence, augmented reality and interactive features, SBAR introduces a new way of learning Arabic in Palestine. While existing research has explored the use of AR in education, the current landscape lacks augmented reality education applications. This study aims to address this limitation and presents SBAR as a revolutionary solution to improving learning experience.

The SBAR incorporates advanced technologies such as artificial intelligence and deep learning algorithms to provide a unique set of features:

1. Interactive Learning: SBAR enables students to learn interactively and independently while receiving guidance from teachers. Through its 3D characters and interactive sound system, the application enhances student engagement and interaction with the book content. By targeting specific story titles or images with their device camera, students are transferred into an animated 3D environment that tells the story and allows for natural voice interactions.
2. AI-Powered Feedback: SBAR utilizes AI and deep learning algorithms to analyze students' answers and provide instantaneous interactive voice feedback. By incorporating natural language processing techniques, the application creates intelligent interactions with students, going beyond traditional books. This personalized feedback raises a dynamic and interactive learning experience, tailored to individual students' needs.
3. Accessibility: An important contribution of SBAR is its commitment accessibility. The application incorporates features such as text-to-speech capabilities, making Arabic books accessible to students with visual impairments. Voice interaction functionality accommodates diverse learning styles and abilities, ensuring that all students can actively participate in the learning process.

These unique features highlight the novelty of SBAR within the AR mobile applications landscape. By leveraging AI and AR technologies, SBAR revolutionizes the traditional approach to learning, providing a personalized, interactive, and inclusive experience for Arabic language learners in Palestine. The subsequent sections of the research paper will delve into the methodology, a demonstration of the AI-powered AR book, and conclude with the overall findings and implications of this innovative application.

These unique features highlight SBAR's uniqueness and novelty in AR's mobile application domain. By incorporating AI and AR technologies, SBAR revolutionized traditional learning methods, providing personalized, interactive and student friendly experiences for Palestinian Arabic students.

The main contributions of this research paper are:

- Personalized learning experience: The AI Smart Book revolutionizes the learning process by offering students a personalized approach to comprehending and engaging with Arabic stories from their school books. Through its smart, simple, and interactive

features, it tailors the learning experience to individual students, allowing them to grasp the material more effectively.

- Enhanced interactivity: By incorporating artificial intelligence components, such as augmented reality, into traditional educational books, the AI Smart Book adds a new level of interactivity to the learning journey. Students are actively engaged in the educational process, fostering greater interest and enjoyment while exploring the content.
- Adding natural language processing: Leveraging cutting-edge artificial intelligence techniques like natural language processing, the AI Smart Book enables intelligent interactions with students. It analyzes and responds to their answers, offering a dynamic and interactive learning experience that goes beyond traditional books.
- Accessibility for all: An essential contribution of the AI Smart Book is its commitment to making Arabic books accessible to students with disabilities or unique learning needs. With features like text-to-speech capabilities, students with visual impairments can access the content audibly, while voice interaction functionality accommodates diverse learning styles and abilities. This inclusivity ensures that all students can actively participate in the learning process.

The subsequent sections of this paper are structured as follows: The methodology of the proposed research application is outlined in Sect. 2. An exemplary demonstration of the AI-powered AR book is presented in Sect. 3. Finally, the paper concludes in Sect. 4.

2 Methodology of the Proposed Application

Our research methodology involved developing an innovative application that utilizes AI and AR to deliver immersive Arabic language learning experiences. The application incorporates natural language processing algorithms to analyze learners' performance and provide tailored feedback and recommendations. Through AR technology, learners can interact with virtual objects and environments that enhance their understanding and engagement with the Arabic language. The application also utilizes ambient intelligence principles to seamlessly integrate with the educational context, promoting a harmonious coexistence between technology and the physical learning environment.

2.1 Application System Analysis

The questionnaire was developed to suit the target group and facilitate the extraction of necessary project requirements. It addresses the issues faced by teachers and parents, providing solutions and suggestions to resolve these problems. This study was conducted in the following cities: Hebron, Bethlehem, Jerusalem, Tulkarem, and Nablus. A total of 268 questionnaires were collected from teachers, while 253 were obtained from parents. The age range of participants was between 25 and 50 years, with 42% being male and 58% female. Their educational background varied, ranging from high school diplomas to bachelor's degrees, and some even held postgraduate degrees such as PhDs and Masters.

The results of the questionnaires were as we expected, as 85% percent of teachers and parents faced difficulties during the e-learning period, due to the lack of assistance applications that suit the capabilities of teachers, parents, and students, and the lack of

training for teachers to deal with e-learning applications such as the Zoom application. And there was a high percentage of support for the presence of educational applications to help students that increase the level of student absorption of the curricula, as there were many suggestions by teachers and parents for the features of the application such as the presence of interactive activities, an interactive audio system, a simplified explanation of academic materials and also that the application be easy to use.

The main requirements that resulted from the questionnaire:

- Existence of activities that enhance the understanding of the student.
- The interactive audio system between the student and virtual characters within the application.
- Quiz exercises and questions.
- Send a report to the teacher and parents, including the student's assessment and the extent of his interaction with the application.

2.2 Application System Design

The application architecture comprises several components, including an augmented reality component and a natural language processing (NLP) component, which includes the analysis of sounds and images. Figure 1 illustrates the application's workflow and main features, which consist of two primary parts: (1) augmented reality technology and (2) NLP.

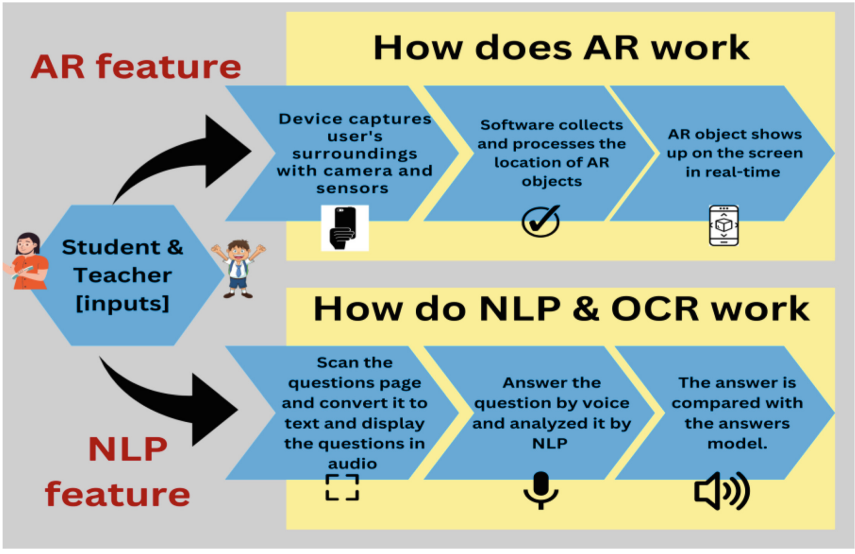


Fig. 1. SBAR system workflow.

2.3 System Architecture

The architecture of SBAR consists of several components, including an augmented reality component developed using Unity and Vuforia, and a natural language processing (NLP) component that encompasses two sub-features: optical character recognition (OCR) and natural language processing (NLP). The augmented reality component of the architecture employs Vuforia to store target images for augmented reality. These images serve as the targets that the Vuforia engine can detect and track by comparing natural characteristics extracted from camera images to known target resources in the database. Both features work simultaneously and integrate with each other. The OCR component is utilized to convert questions from images to text. The user captures the question page from the book using their device's camera, and the system sends it to the OCR script. After that, the system reads these questions, and the user can start answering them one by one using voice, with each voiced answer sent to the Google Cloud Speech-to-Text API for recognition, and then to our algorithm, which decides whether the answer given is correct or not. The system architecture diagram is depicted in Fig. 2.

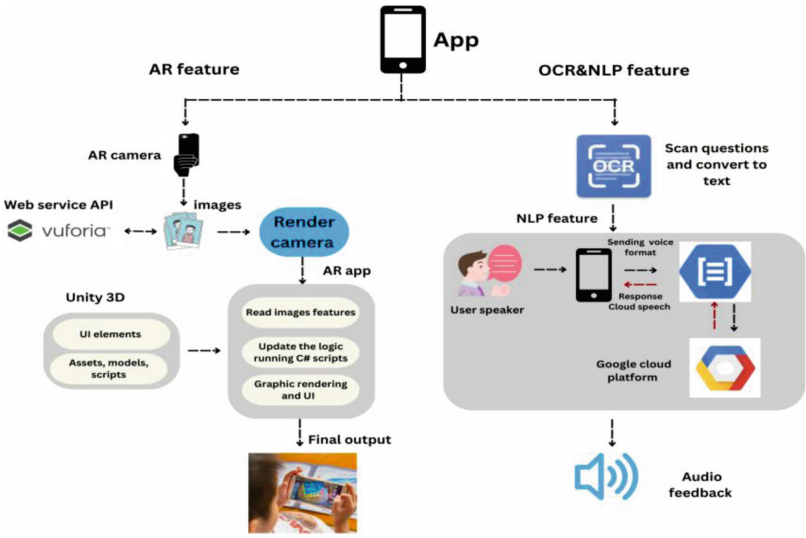


Fig. 2. System architecture.

2.4 Application Main Features

Augmented Reality. The augmented reality feature of the application was developed using Unity and Vuforia. When the user opens the AR camera, the application detects the image target that is saved in the Vuforia database. These images correspond to the one presenting all the stories written in the Arabic language book. The system then analyzes this image and compares it with the saved image target, and displays the corresponding 3D characters as depicted in Fig. 3.

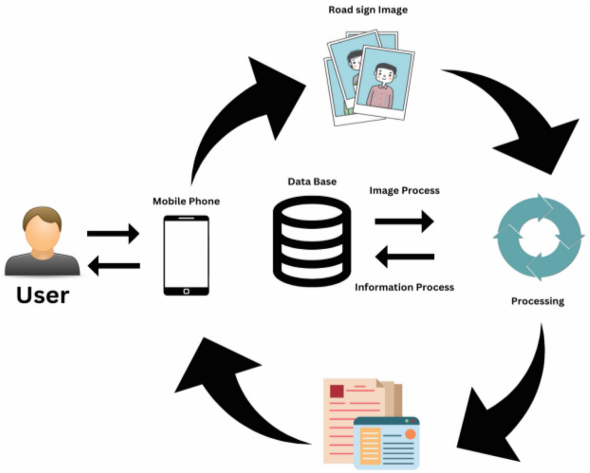


Fig. 3. AR architecture.

Speech-to-Text: Natural language processing (NLP) is a technology that enables computers to understand human languages as they are spoken and written. We used it to convert the voice of a student's answer to the related questions into text, so the application can analyze whether the answer is correct or not. When the student inputs their answer through voice, the system sends it to the NLP Google API to convert speech to text and then sends it to the system for comparison with the correct answer to process feedback. Figure 4 illustrates how the NLP functions work.

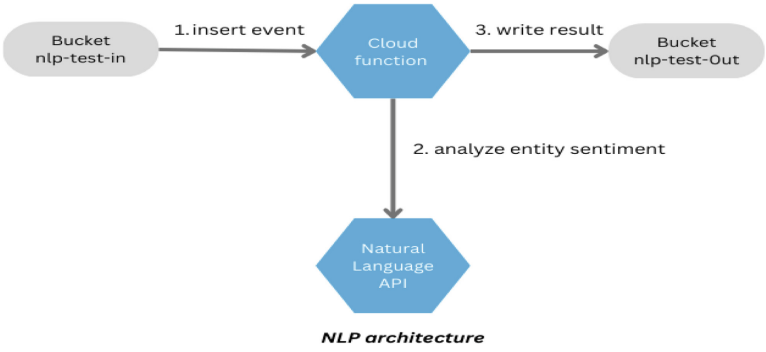


Fig. 4. NLP architecture

The algorithm for converting students' voice input to text to answer questions and provide feedback can be outlined as follows:

1. Initialize the NLP system and establish a connection to the Google Cloud Natural Language API.
2. Begin listening for students' voice input.

3. Once input is detected, send the audio to the NLP API for conversion to text. Let the converted text be represented as T.
4. Compare T to the correct answer, A.
5. Compute the similarity score between T and A using a similarity measure such as cosine similarity (S), where S = (T * A) / (||T|| * ||A||).
6. If S > threshold (set by administrator), then the answer is considered correct and feedback "Correct" is provided.
7. If S ≤ threshold, then the answer is considered incorrect and feedback "Incorrect" is provided.
8. Repeat steps 2–7 for each new input received.

Optical Character Recognition (OCR). The deep learning-based system used in our application was constructed utilizing the Long-Short Term Memory algorithm LSTM, is a type of a Recurrent Neural Network (RNN), the LSTM consists from memory blocks which is contains from one or more self-connected memory cells associated to each other. RNN-LSTM has accuracy ratings ranging from 70.2% to 92.9%, depending on the picture resolution, text type, text size, and image clarity. The model receives an input image to be processed and converted into binary text. Then it is processed to identify the beginning of each line, line by line. The model locates the text lines and then divides the image into characters that are classified and combined to form words. These words are then compared with a list of words learned from the dataset and used to predict the output (Fig. 5).



Fig. 5. OCR text recognition.

Finally, to verify the accuracy of answers, an API algorithm was developed that uses a Universal Sentence Encoder (USE) to encode the answer into vectors of real numbers. These vectors can be used to determine sentence similarities between the answers provided by the user and the answers saved in the application. For instance, if the saved answer is " ذهب الفيل إلى الغابة ليتنزه" In English; "*The elephant went to the forest for a walk*" and the answer provided by the user is " ذهب الفيل ليمشي في الغابة" In English "*The elephant went for a walk in the forest*", both answers are correct but in different forms. This problem had to be addressed as the fuzzy string algorithm only finds similarities between answers, not between the meaning of the answer as illustrated in Fig. 3.

The focus of the proposed application is on modern standard Arabic, as it is widely used across all Arabic-speaking countries and is the academic standard. Additionally, it is the language used in social media applications and is integrated with Unity.

3 An Illustrative Example with User Interfaces

3.1 Illustrative Example

The user interface for the proposed SBAR application has been meticulously crafted based on previous research. It was determined that children have a preference for cool and simple colors such as light pink and light blue. The design of the application's user interface aims to maximize ease of use, with an estimated training time of just three minutes to learn how to navigate it. Additionally, the application has been designed to use the Arabic language in order to provide simple and clear responses and requests, ensuring that the child can easily understand and utilize the app. Overall, the SBAR user interface has been created to provide a user-friendly experience. It consists of a limited number of screens, as illustrated in Fig. 6.

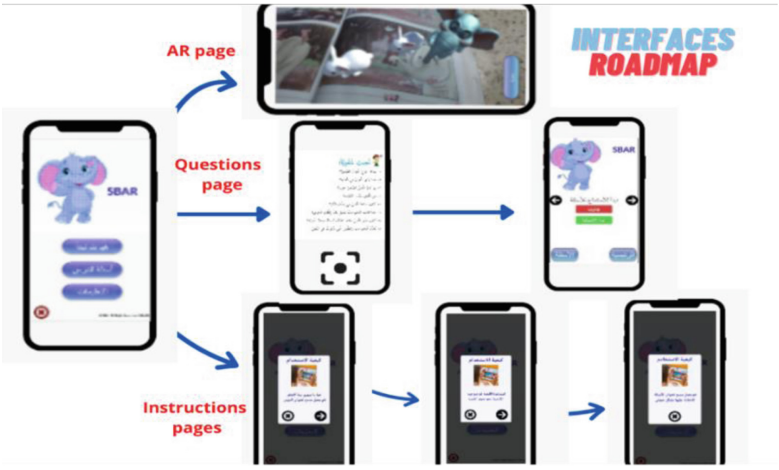


Fig. 6. SBAR journey interfaces roadmap

3.2 Evaluation

The study involved a sample of almost 150 participants, including students and teachers from various primary schools located in Hebron, Palestine, such as Sourif Primary School for Boys, Amro Ibn Al'ase Primary School for Boys, and Sourif Primary School for Girls. A random selection of students from the primary level (grade four and year five) was chosen to use the application on an Arabic language textbook. The evaluation process consisted of several stages. Initially, the participants were provided with access to the SBAR application and given time to familiarize themselves with its features and functionalities. They were then assigned specific tasks or learning activities within the application, designed to assess their comprehension and engagement with the Arabic language content.

Various measures were employed to collect data during the evaluation, including quantitative assessments where we had measure the achievement of the learning outcomes through SBAR, and how it improved the students' Arabic language skills and their understanding to the book content, surveys in which we gathered subjective verbal feedback from both students and teachers regarding their experiences with the SBAR application, and observations were made during the evaluation process to gather first hand insights into how students and teachers interacted with the SBAR application. These measures aimed to capture information on factors such as user satisfaction, ease of use, learning outcomes, and overall educational value.

To analyze the collected data, descriptive statistics were employed. This allowed for a demanding examination of the results and the identification of any significant findings. The results of these tests were overwhelmingly positive, with both students and teachers expressing satisfaction with the application's ease of use. They did not encounter any difficulties when using the application. The positive reactions from these users indicate that the application has effectively added value to their education and has been successful in capturing the attention of students toward learning. Many students were able to answer all questions related to the lesson correctly, after being explained by the 3D characters within a short period, not exceeding several minutes.

3.3 Impact and Benefit of the Proposed Application

We have designed the application with ease of use in mind. The user interface is visually appealing and intuitive to navigate. Additionally, we have incorporated an instruction button that provides step-by-step guidance on how to use the SBAR application. The goal of SBAR is to promote self-directed learning by presenting school subjects engagingly and innovatively, different from traditional classroom instruction. We anticipate that the application will bring about significant educational advancements by leveraging technology to enhance student engagement.

An augmented reality application for Arabic book stories can have a significant impact and provide numerous benefits when used to teach students in elementary schools in Palestine. The use of augmented reality technology can enhance the student's learning experience by bringing the stories to life and making the content more interactive and engaging. This can lead to a deeper understanding of the material and increase the student's motivation to learn. Furthermore, the use of an augmented reality application

can help to bridge the gap between traditional classroom instruction and the use of technology, which is becoming increasingly prevalent in education. Additionally, an augmented reality application for teaching Arabic book stories in Palestine can help improve language proficiency and cultural understanding among students. Overall, an augmented reality application for Arabic book stories is a valuable tool for enhancing the education of students in elementary schools in Palestine.

4 Conclusion and Future Work

As technology has rapidly progressed in recent decades, it has become integrated into various aspects of economic, educational, industrial, and social life. Artificial intelligence has replaced human labor in many fields, resulting in significant advancement in human life. In this research, we have explored various ways to implement artificial intelligence in education for Arabic books to make them more personalized and interactive. Our proposed application has demonstrated a positive impact on students learning the Arabic language in Palestine at the preparatory level. By applying various artificial intelligence techniques such as augmented reality and natural language processing, we aim to personalize the learning process and assist students in learning effectively. Additionally, by making the tool interactive, we have been able to increase student engagement with Arabic stories, making the learning experience more immersive. Our proposed idea emphasizes the importance of technology in the field of education, especially during periods of pandemics like Covid-19, and the challenges faced by teachers. It is important to use technology to support the education process.

In plans, the proposed idea can be extended to:

- Make the application compatible with all textbooks of the Palestinian curriculum and for all levels.
- Test the application with more AI-based models to improve detection and responsive accuracy.
- Add more features, such as: generating and sending feedback reports to teachers and parents, and adding more interactive exercises.

References

1. Karandish, D.: 7 benefits of AI in education. Journal (2021). https://thejournal.com/articles/2021/06/23/7-benefits-of-ai-in-education.aspx. Accessed 23 July 2022
2. Wrike: What is Agile Methodology in Project Management?, Wrike (2019). https://www.wrike.com/project-management-guide/faq/what-is-agile-methodology-in-project-manacement/
3. IBM Cloud Education. "What is Natural Language Processing?," IBM (2020). https://www.ibm.com/cloud/learn/natural-language-processing
4. IBM Cloud Education. "What is optical character recognition (OCR)?," IBM (2022). https://www.ibm.com/cloud/blog/optical-character-recognition. Accessed 23 July 2022
5. Chen, Y., Wang, Q., Chen, H., Song, X., Tang, H., Tian, M.: An overview of augmented reality technology. In: Journal of Physics: Conference Series, vol. 1237, no. 2, p. 022082 (2019)

6. Alkhaldi, N.: Evolution of education: how interactive technologies reshaped learning. eLearning Industry (2021). https://elearningindustry.com/evolution-of-education-how-interactive-technologies-reshaped-learning
7. Sinha, S.: Augmented reality in education: a staggering insight into the future. eLearning Industry (2021). https://elearningindustry.com/augmented-reality-in-education-staggering-insight-into-future
8. Chen, P., Liu, X., Cheng, W., Huang, R.: A review of using Augmented Reality in Education from 2011 to 2016. In: Popescu, E., et al. (eds.) Innovations in Smart Learning. LNET, pp. 13–18. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-2419-1_2
9. Altinpulluk, H.: Determining the trends of using augmented reality in education between 2006–2016. Educ. Inf. Technol. **24**(2), 1089–1114 (2019)
10. Garzón, J., Baldiris, S., Gutiérrez, J., Pavón, J.: How do pedagogical approaches affect the impact of augmented reality on education? A meta-analysis and research synthesis. Educ. Res. Rev. **31**, 100334 (2020)
11. Расулова, H.: Prospects for the use of Augmented Reality in Education. Embed. Selforganising Syst. **9**(4), 26–29 (2022)
12. Google. "Dinosaur 4D+ - apps on Google Play" Google (n.d.). https://play.google.com/store/apps/details?id=com.OctagonStudio.ARDinoVR&hl=en&gl=US
13. "By AI Initiative: Interactive Educational Books and Active Learning". IBAL (n.d.). https://ibal.ps/
14. QuiverVision. "QuiverVision," QuiverVision. http://www.quivervision.com. Accessed 28 Apr 2020
15. Arloon Plants. "Arloon Plants," Arloon. http://www.arloon.com/en/apps/plants/



How to Handle Fuzzy Aspects in African Sociocultural Ontology

Abdoul Azize Kindo[1(✉)], Sadouanouan Malo[1], Gaoussou Camara[2], Guidedi Kaladzavi[3], Théodore Marie Yves Tapsoba[1], and Kolyang[3]

[1] Ecole Supérieure d'Informatique, University of Nazi Boni, Bobo-Dioulasso, Burkina Faso
azize.kindo@univ-bobo.gov.bf
[2] Department of Mathematics, University of Alioune Diop of Bambey, Bambey, Senegal
[3] Department of Computer Science and Telecommunications, University of Maroua, Maroua, Cameroon

Abstract. In this article, we propose an innovative methodology for creating a fuzzy sociocultural ontology based on the CriSO ontology, aiming to comprehensively represent the rich and diverse aspects of African culture. We address the limitations of the CriSO ontology and emphasize the necessity of incorporating fuzzy logic in order to effectively capture the nuances of African sociocultural knowledge. Our methodology begins with the extension of the CriSO ontology by integrating fuzzy concepts and properties. We subsequently present the formalization process of the fuzzy ontology using advanced fuzzy logic techniques and showcase its practical application in a restaurant recommendation scenario. In conclusion, our research demonstrates that employing a fuzzy ontology offers a more flexible and accurate approach to representing sociocultural knowledge. This methodology holds the potential to serve as a solid foundation for the development of fuzzy ontologies in various other knowledge domains, contributing to a better understanding and preservation of cultural heritage.

Keywords: knowledge representation · Fuzzy logic · Fuzzy ontology · Fuzzy OWL · reasoning

1 Introduction

African culture is a diverse and rich heritage that remains largely underrepresented in global knowledge systems. While efforts have been made to promote African sociocultural knowledge, these efforts have not fully considered the fuzzy logic aspect of ontology. Ontologies can be instrumental in representing and organizing knowledge about African culture, but there is a need to incorporate fuzzy logic for more flexible reasoning.

In this paper, we propose a methodology for creating a fuzzy sociocultural ontology based on the classic CriSO [1] ontology, which provides a spatial representation of sociocultural infrastructure for Africa. However, CriSO has limitations in representing the fuzzy aspects of African sociocultural knowledge, specifically the lack of consideration for fuzzy logic.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 179–188, 2023.
https://doi.org/10.1007/978-3-031-43461-7_18

This study proposes an innovative method for creating a fuzzy ontology based on the CriSO ontology. While efforts have been made to promote African sociocultural knowledge, few have fully considered the fuzzy logic aspect of ontology. Therefore, we strive to bring a new perspective to this field, utilizing fuzzy logic to enhance the representation of sociocultural knowledge.

We present an approach to extend the CriSO ontology by incorporating fuzzy concepts and properties. This extension aims to bring a new perspective to this field, utilizing fuzzy logic to enhance the representation of sociocultural knowledge. We discuss the need to consider fuzzy aspects in the representation of African sociocultural knowledge and provide a methodology for creating a fuzzy ontology. Our methodology stands out for its ability to handle uncertainty and flexibility in knowledge representation, unlike traditional methods that fall short of capturing the nuanced and non-linear aspects of sociocultural knowledge. We then detail the formalization and implementation of the fuzzy ontology, demonstrating the potential for its application in representing complex knowledge domains.

The paper is structured as follows: In Sect. 2, we provide the main definitions, including sociocultural knowledge and fuzzy ontology, as well as an introduction to fuzzy logic and its components, such as membership functions. In Sect. 3, we provide background information on ontologies and their significance in organizing knowledge about African culture, discussing the limitations of classical ontologies in representing the fuzzy aspects of African sociocultural knowledge. Section 4 focuses on the CriSO ontology, its description, and the need to consider fuzzy aspects. In Sect. 5, we detail the methodology for creating a fuzzy ontology based on CriSO. Sections 6 and 7 present the formalization and implementation of the fuzzy ontology, respectively. Finally, in Sect. 9, we conclude the paper and discuss future work.

2 Main Definitions

2.1 Sociocultural Knowledge

Sociocultural [1] knowledge encompasses a wide range of concepts, including infrastructure, social structures, art, culture, and more. These concepts are essential to understanding the beliefs, values, and practices of a community and can be represented using ontologies.

2.2 Fuzzy Ontology

Fuzzy logic. Fuzzy logic [2] is a mathematical logic that enables the representation of vague information. It was initially developed by Lotfi Zadeh in the 1960s to handle problems that are too complex or too difficult to model using traditional binary (true/false) logic. Fuzzy logic is based on the concept of degrees of truth, which allows for the representation of information that is not entirely true or false, but rather falls somewhere in between. This means that fuzzy logic can handle data that is uncertain, vague, or ambiguous. In fuzzy logic, variables can take on values that range from completely true (1) to false (0), as well as any value in between. This approach provides a more nuanced understanding of complex systems and allows for a more accurate modeling of real-world phenomena.

Membership Functions. The membership function [3] is a mathematical function used in fuzzy logic that maps a range of values to a degree of membership in a fuzzy set. It is used to represent the degree of truth or membership of an element in a set and is typically defined by a curve or other mathematical function. The most function used are triangular membership function, trapezoidal membership function, left, right function gaussian and sigmoidal membership function. It used to represent fuzzy properties.

Ontology. According to Studer [4] ontology is defined as a formal, explicit specification of a shared conceptualization. Ontology is a specific vocabulary for a domain and gives a good description of things within a domain. Ontology structure [3] is quadruple O = (C, P, R, A) and C, P, R, A stand respectively for a set of concepts within a domain, set of concept properties, set of the relation between concept and axiom which is a real fact. It can be defined as a description language knowledge base a tuple K = (T, A, R) where T is a set of terminological axioms (TBox), A is a set of assertional axioms (ABox), R is a role (RBox).

Fuzzy Ontology. Fuzzy ontology is an ontology that uses fuzzy logic to handle vagueness. According to [3] the structure of the fuzzy ontology is a tuple OF = (C, P, CF, PF, R, RF, AS, ASF, A) where C is a set of crisp concepts defined for the domain. P is a set of crisp concept properties. CF is a set of fuzzy concepts, PF is a set of fuzzy concept properties, R is a set of crisp binary semantic relations defined, between concepts in C, RF is a set of fuzzy binary semantic relations defined between concepts in C or fuzzy concepts in CF, AS is a set of crisp binary associations defined between concepts in C or fuzzy concepts in C, ASF is a set of fuzzy binary associations defined between crisp concepts in C or fuzzy concepts in CF, A is a set of axioms. An axiom is a real fact or reasoning rule.

A fuzzy concept CF is a tuple <***Ncf***, ***P***, ***PF***> Where ***Ncf*** is the name of a fuzzy concept, e.g. "Near Restaurant", or "Good Restaurant". ***P*** is a set of crisp properties.

PF is a nonempty set of fuzzy properties. A fuzzy property ***PF*** is a tuple <**Npf**, **T**, **M**> **Npf** is a name of the fuzzy property, e.g., "Distance". **T** is a set of terms, which are the values of the fuzzy property, e.g., **T** = {near, middle, far}. **M** is a mapping rule, which maps every term of T to a fuzzy set.

3 Background

Many projects in the sociocultural domain have focused on knowledge representation, resulting in the creation of various ontologies to capture different aspects of this knowledge. These include the High-level sociocultural ontology USCO [5], the CriSO ontology which focuses on the spatial representation of sociocultural infrastructure [1], the Human Time Ontology Huto [6], and the high-level sociocultural ontology OntoSOC [7], which is an improvement over USCO. However, all these ontologies are crisp and do not address fuzzy aspects. Africa has a rich sociocultural heritage that is often underrepresented in global knowledge systems. However, there is a growing interest in documenting and preserving this knowledge, both for the benefit of Africans and for those interested in learning more about African culture. Ontologies can play an important role in this effort, as they provide a structured way of representing and organizing knowledge. By creating

ontologies that capture different aspects of African sociocultural knowledge, we can help ensure that this knowledge is preserved and accessible to a wider audience. For example, ontologies could be created to capture information about African languages, traditional practices, beliefs, art, music, and literature. By creating these ontologies, we can help to ensure that this knowledge is not lost and that it can be used to benefit both Africans and tourists who are interested in learning more about African culture.

Fuzzy aspect. The current projects in African sociocultural knowledge representation have neglected to address the vague aspects of this knowledge. Fuzzy concepts, relations, axioms, instances, and other fuzzy elements were not taken into account during the design process of existing ontologies. Furthermore, the reasoning tools built for African sociocultural knowledge representation do not allow for fuzzy reasoning. As a result, it can be difficult to find sociocultural elements using natural language containing vague terms, such as restaurants that are "close" to the market or pharmacies that are "not far" from the stadium.

In this work, we propose a solution for considering the fuzzy logic aspect in sociocultural ontology, which will make African sociocultural ontology more expressive. By incorporating fuzzy logic, we can better capture the uncertain and vague aspects of African sociocultural knowledge, allowing for more accurate and comprehensive representation of this knowledge. This, in turn, will make it easier for people to access and utilize this knowledge, benefiting both Africans and those interested in learning more about African culture.

4 The CriSO Ontology

4.1 Description of the CriSO Ontology and Its BasicConcepts

CriSO ontology is a spatial representation of sociocultural infrastructure. It provides a structured way of representing and organizing knowledge about diverse types of infrastructure such as education, religion, health, tourism, transportation, and sports facilities. The ontology also includes concepts related to public spaces and non-built areas. The CriSO ontology is organized into two main categories: infrastructure and locality. While we focus on CriSO ontology as a case study in this paper, the methodology we propose is intended to be generalizable and can be applied to any ontology that could benefit from the incorporation of fuzzy logic.

4.2 Description of the Need to Consider Fuzzy Aspects

The CriSO ontology has limitations in representing sociocultural knowledge, specifically with regards to fuzzy concepts. These limitations are not unique to CriSO, and indeed can be found in many classical ontologies, particularly when dealing with imprecise and vague knowledge. To address this limitation, fuzzy logic must be considered to provide a more flexible representation of sociocultural knowledge. By incorporating fuzzy concepts, nuances that may be missed in a crisp ontology can be captured. The methodology for creating a fuzzy ontology based on the CriSO ontology will be presented in the next section. By incorporating fuzzy logic into an existing ontology, we can enhance its functionality, enabling it to better capture and represent the intricacies of the knowledge domain it models.

5 Methodology for Creating a Fuzzy Ontology

To create a fuzzy ontology, we start with an existing ontology and add fuzzy concepts and properties to it. This extension involves defining fuzzy classes, properties, and axioms using the fuzzy logic framework. This methodology provides an excellent starting point for the creation of a fuzzy ontology.

1. **Analysis of the classical ontology**
   In this step, the ontologist analyzes the classical ontology to identify concepts that could benefit from a fuzzy representation. This involves reviewing the concepts and relationships within the ontology to identify where a more flexible and nuanced representation could be beneficial.
2. **Definition of linguistic variables and terms**
   For each identified concept, the ontologist defines an appropriate linguistic variable that captures the fuzzy notion in question. This involves selecting linguistic terms that capture the nuances of the concept being represented. The choice of linguistic terms depends on the context and domain being represented and should be chosen to accurately represent the concept being modeled.
3. **Attribution of membership degrees or membership functions**
   For each defined linguistic term, the ontologist assigns membership degrees or membership functions based on the values associated with the concept. Membership degrees or functions determine the degree to which an individual belongs to a particular linguistic term. For example, if the concept being modeled is "age," the membership function might assign a higher degree of membership to the term "old" for buildings that are more than 50 years old.
4. **Creation of simple fuzzy concepts**
   For each identified concept, the ontologist creates a simple fuzzy concept using the defined linguistic variables and associated membership degrees or functions. This involves combining the linguistic terms and membership functions with a crisp concept to create a fuzzy concept. For example, a fuzzy concept for the age of a building might be "OldCityHall," using the linguistic term "old" and the membership function defined for the "age" property.
5. **Creation of complex fuzzy concepts**
   For more complex concepts, the ontologist can use fuzzy rules to combine the linguistic variables defined for a concept. For example, for the Restaurant concept, considering its properties of service quality and cleanliness, we can combine the concept name "restaurant" with linguistic terms "clean" and "good" to create the fuzzy concept "CleanAndGoodRestaurant."

   Fuzzy rules can be used to define how linguistic variables (such as "good quality" and "clean") are combined to create a fuzzy concept (such as "GoodQualityAndCleanRestaurant"). Here are some example fuzzy rules for this concept:

   If the service quality is "high" and the cleanliness is "very clean," then the restaurant belongs to the "GoodQualityAndCleanRestaurant" fuzzy concept.

   If the service quality is "medium" and the cleanliness is "clean," then the restaurant partially belongs to the "GoodQualityAndCleanRestaurant" fuzzy concept.

   If the service quality is "low" and the cleanliness is "not clean," then the restaurant does not belong to the "GoodQualityAndCleanRestaurant" fuzzy concept.

These fuzzy rules define how different values for the linguistic variables of service quality and cleanliness can be combined to create a membership value for the "GoodQualityAndCleanRestaurant" fuzzy concept.

6. **Creation of fuzzy relations**
   If the classical ontology contains relations between concepts, the ontologist can use fuzzy rules to create fuzzy relations by combining the defined linguistic variables. For example, to create a fuzzy relation between two concepts using distance, the ontologist can use terms like "close," "medium," and "far" with a target concept. This approach allows for more flexible representations of relations between concepts and enables the creation of fuzzy concepts such as "close restaurant" or "restaurant medium distance."

6 Formalization of Fuzzy Ontology

To formalize the CriSO ontology with fuzzy concepts and classes, we can follow the methodology outlined in Sect. 5 of this article. The steps of the methodology are applied for each of the concepts presented below. While we demonstrate this formalization process using the CriSO ontology, it is worth emphasizing that this process can be similarly applied to any other ontology in a different domain, enhancing its ability to represent and handle imprecise and vague knowledge.

1. Classical ontology analysis: Identification of concepts for which a fuzzy representation could be beneficial.
2. Definition of linguistic variables and terms: For each identified concept, we define an appropriate linguistic variable and select linguistic terms to capture the nuances of the represented concept.

   The following table shows the linguistic variables and associated terms for different concepts (Table 1):

Table 1. Selected Concepts and their Corresponding Linguistic Variables

Concept	Linguistic Variable	Linguistic Terms
Education	Teaching Quality	Good, Average, Poor
Health	Medical Equipment	Good, Average, Poor
Health	Location Proximity	Close, Moderately Close, Far
Health	Specialization Type	General, Specialized, Highly Specialized
Health	Quality	Good, Average, Poor
Infrastructure	Building Age	Old, Historic, Recent
Restaurant	Cleanliness	Clean, Moderately Clean, Dirty
Restaurant	Service Quality	Good, Fair, Poor

(*continued*)

Table 1. (*continued*)

Concept	Linguistic Variable	Linguistic Terms
Restaurant	Distance	Close, Moderately Close, Far
Supermarket	Product Variety	Diverse, Moderately Diverse, Non-Diverse
Transport	Distance	Close, Moderately Close, Far
Infrastructure	RoadState(rural zone)	Good, Average, Poor
District	PopularNeighborhood	High, Medium, Low

3. Attribution of membership degrees or membership functions: For each defined linguistic term, we assign membership degrees or membership functions based on the values associated with the concept.
4. Creation of simple fuzzy concepts: For each identified concept, we create a simple fuzzy concept using the defined linguistic variables and associated degrees or membership functions. To represent the fuzzy concepts, we use fuzzy data properties that capture the fuzziness of the concept.

   Examples of simple fuzzy concepts:

   Fuzzy class: GoodQualityEducation $\sqsubseteq$ Education $\sqcap$ hasQuality some High
   Fuzzy class: GoodQualityRestaurant $\sqsubseteq$ Restaurant $\sqcap$ hasQuality some Good
   Fuzzy class: WellEquippedHealthFacility $\sqsubseteq$ HealthFacility $\sqcap$ hasEquipmentQuality some High
   Fuzzy class: CleanPlace $\sqsubseteq$ Place $\sqcap$ hasCleanliness some Clean

5. Creation of complex fuzzy concepts: For more complex concepts, associate multiple simple fuzzy concepts to create complex fuzzy concepts, such as the intersection of two or three fuzzy data properties.

   Example of complex fuzzy concept:

   Fuzzy class: GoodQualityCleanRestaurant $\sqsubseteq$ Restaurant $\sqcap$ hasQuality some Good $\sqcap$ hasCleanliness some Clean

   In this example, the complex fuzzy concept "GoodQualityCleanRestaurant" represents restaurants that have both good service quality and a high cleanliness level.
6. Fuzzy relationships: Fuzzy relationships, such as distance, can be used to associate fuzzy concepts. For example, the distance between a user and a restaurant could be represented using a fuzzy membership function to describe whether the restaurant is "close", "moderately close", or "far". Additionally, combinations of fuzzy relationships and membership functions can be used to create complex fuzzy concepts.
7. Fuzzy reasoning and querying: Fuzzy reasoning allows for more expressive queries and more precise results based on user preferences. Using the fuzzy concepts and relationships defined earlier, queries can be formulated to include fuzzy aspects. Example of a fuzzy query:

   "Find all restaurants that are both GoodQuality and CleanRestaurant and close to the user's location."

   This query can be translated into the following fuzzy logic expression:

(GoodQualityRestaurant ⊑ Restaurant ⊓ hasQuality some Good) AND (CleanRestaurant ⊑ Restaurant ⊓ hasCleanliness some Clean) AND (LocationProximity some Close)

The query will return all restaurants that have a high membership degree to the fuzzy class "GoodQualityRestaurant" (defined by the "hasQuality some Good" property) and a high membership degree to the fuzzy class "CleanRestaurant" (defined by the "hasCleanliness some Clean" property), and have a high membership degree to the "Close" linguistic term of the "LocationProximity" linguistic variable. This allows for more tailored recommendations based on user preferences, taking into account both service quality and cleanliness in addition to proximity.

7 Implementation

```
prefix = "PREFIX : <http://www.semanticweb.org/hp/ontologies/naziboni/criso#>"
entity = entity
qr_str1 = " select * WHERE { ?x    a  " + ":"+entity+";"
qr_str2 = """ :Infras_Nom ?n; :Infras_Latitude ?Lat; :Infras_Longitude ?Lon;
:isLocatedIn   ?Loc; :QualityOfService ?Quality} """
q_requets = prefix + "" + qr_str1+ "" + qr_str2 # requête  finale
```

The Python code snippet provided at the beginning of this section shows an example of constructing a SPARQL query string to retrieve information from the fuzzy ontology. The variable "entity" represents the type of entity being queried (e.g., restaurant), and the query string is constructed by concatenating the appropriate prefixes, variables, and criteria related to the entity.

To implement our fuzzy ontology, we used the Protégé tool with the fuzzyOWL plugin. We created fuzzy classes for each fuzzy concept identified in the modeling, as well as fuzzy properties to represent membership functions. We also created fuzzy axioms to define the relationships between the different classes.

For example, for the fuzzy class "CloseRestaurant," we created a fuzzy data property "Close" with a triangular membership function defined previously for the linguistic variable "Distance." We then defined the fuzzy axiom that relates the "CloseRestaurant" class to the "Restaurant" class using the fuzzy data property "Close." This same process was followed for the other fuzzy classes identified in the modeling.

Here is an example of a fuzzy class using the fuzzy data property "Close":

Fuzzy class: CloseRestaurant ⊑ Restaurant ⊓ hasDistance some Close.

In this example, the fuzzy class "CloseRestaurant" represents restaurants that are considered close to the user's location, as captured by the fuzzy data property "Close" associated with the "hasDistance" property.

8 Evaluation

In our evaluation, we focused on the enhanced query capabilities enabled by the integration of fuzzy logic into our ontology. Specifically, our fuzzy ontology allows for more sophisticated, nuanced queries that incorporate subjective and contextual factors.

For instance, in the original CriSO ontology, users could list infrastructures. However, with the integration of fuzzy logic, users can now perform queries to identify nearby infrastructures or historic infrastructures. This ability to handle and reason with imprecise and vague information significantly improves the utility and flexibility of the ontology, allowing for more personalized and relevant recommendations in the African sociocultural domain.

We further demonstrated this with a use case in which the system received a request, retrieved relevant sociocultural data from the fuzzy ontology, calculated the degree of relevance for each restaurant using fuzzy logic, and returned a list of recommended restaurants with their corresponding relevance degrees. See Fig. 1

These results demonstrate the effectiveness of our fuzzy ontology in modeling and capturing the complexities of the African socio-cultural domain, thereby fulfilling the goals of our research.

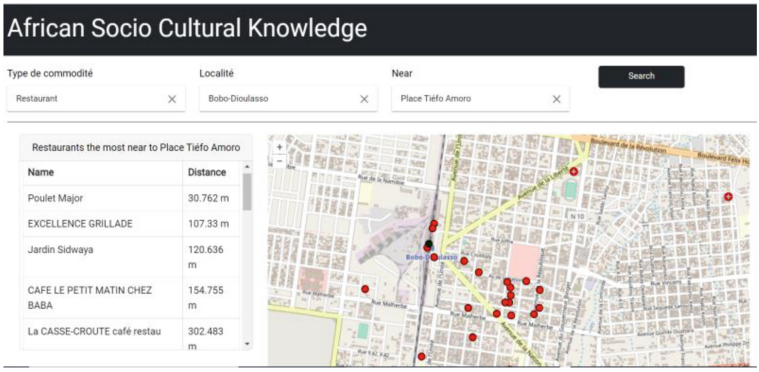


Fig. 1. Use case

9 Conclusion and Future Work

Our methodology addresses the challenges of representing imprecise and vague sociocultural knowledge by considering subjective and contextual factors that can influence their definition. This approach offers a more precise representation that is better suited to the complexity of certain concepts such as quality or proximity, which may vary depending on individual preferences and situations.

For future work, it would be interesting to explore other theories for representing uncertainty in ontologies, such as the possibility theory. Integrating the possibility theory into the CriSO ontology could enable the handling and reasoning with incomplete knowledge, thus enhancing the quality of recommendations. This approach could be extended to other domains to improve inform decision-making processes.

References

1. Kaladzavi, G., Traore, Y., Ari, A.A.A.: Geo-spatial domain ontology: the case of the socio-cultural infrastructures. Int. J. Comput. Appl. **178**(1), 42–48 (2017)
2. Bobillo, F., Straccia, U.: Fuzzy ontology representation using OWL 2. Int. J. Approximate Reasoning **52**(7), 1073–1094 (2011). https://doi.org/10.1016/j.ijar.2011.05.003
3. Ghorbel, H., Bahri, A., Bouaziz, R.: Fuzzy ontologies building method: fuzzy OntoMethodology. In: 2010 Annual Meeting of the North American Fuzzy Information Processing Society, Toronto, ON, Canada, pp. 1-8. IEEE (2010). https://doi.org/10.1109/NAFIPS.2010.5548211
4. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. Data Knowl. Eng. **25**(1), 161–197 (1998). https://doi.org/10.1016/S0169-023X(97)00056-6
5. Diallo, P.F., Corby, O., Mirbel, I., Lo, M., Ndiaye, S.: Ontologies-based platform for sociocultural knowledge management. J. Data Semant. **5**(3), 117–139 (2016). https://doi.org/10.1007/s13740-016-0065-4
6. Diallo, P.F., Corby, O., Mirbel, I., Lo, M., Ndiaye, S.M.: HuTO: an human time ontology for semantic web applications. Cs (2015). Consulté le: 9 novembre 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1506.05969
7. Kaladzavi, G., Diallo, P.F., Kolyang, Lo, M.: OntoSOC: Sociocultural Knowledge Ontology (2015)
8. Lamy, J.-B.: Owlready: ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif. Intell. Med. **80**, 11–28 (2017). https://doi.org/10.1016/j.artmed.2017.07.002

VRPrOE Toolbox for Virtual Pre-occupancy Evaluation: Proof of Concept on a BIM Model of a Conservatory Classroom

José L. Gómez-Sirvent[1,2(✉)], Desirée Fernández-Sotos[1], Francisco López de la Rosa[1,3], Alicia Fernández-Sotos[4], and Antonio Fernández-Caballero[1,2,5]

[1] Neurocognition and Emotion Unit, Instituto de Investigación en Informática de Albacete (I3A), Universidad de Castilla-La Mancha, Avenida de España s/n, 02071 Albacete, Spain
jose.gomez@uclm.es
[2] Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Avenida de España s/n, 02071 Albacete, Spain
[3] Department of Electrical, Electronic, Automatic and Communications Engineering, Universidad de Castilla-La Mancha, Avenida de España s/n, 02071 Albacete, Spain
[4] Conservatorio de Música de Murcia, Calle Cartagena 74, 30002 Murcia, Spain
[5] Biomedical Research Networking Center in Mental Health (CIBERSAM-ISCIII), 28016 Madrid, Spain

Abstract. Feedback on how people perceive the characteristics of the built environment is very useful in identifying design flaws. Using virtual environment tools, it is possible to show building designs to potential future users before construction begins, allowing for design refinement. Creating experiments to evaluate the built environment is often a laborious and building-specific process. In this paper, we present a toolbox designed to simplify this task. Starting from a 3D model of a building that can be created using Building Information Modeling software, the toolbox allows the creation of different design alternatives with a photorealistic appearance that can be used to perform a variety of pre-occupancy evaluation experiments. Through our tool, it is possible to define the parameters of the experiment through a simple user interface and generate questionnaires that allow the user to evaluate the design of the building in a virtual environment. To demonstrate the usefulness of the toolbox, we successfully conducted a proof-of-concept experiment to evaluate the effect of color and lighting on the perception of the environment in a conservatory classroom. The results of the experiment indicate that there is a significant relationship between the type of lighting (artificial, natural and mixed) and the perception of pleasantness, openness and brightness of the environment. The combined effect of color and lighting also shows a strong relationship with openness and brightness.

Keywords: Pre-occupancy evaluation · Building Information Modeling · Virtual reality · Neuroarchitecture · Proof of concept

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 189–198, 2023.
https://doi.org/10.1007/978-3-031-43461-7_19

1 Introduction

For thousands of years, humans have lived in nature, in an environment we could not control, where the sun marked the beginning and end of the day, and where there were few straight lines. Today, most of us spend most of our lives in a built environment. Modern buildings are designed to efficiently maintain a comfortable temperature and good indoor air quality. However, there are still many aspects that are not considered in the design of a building that can affect the well-being of its occupants. Numerous studies suggest that there is a relationship between the environment around us and our brain activity [2,5]. Visual characteristics of building components such as color, shape, size or texture can act as stimuli capable of eliciting an emotional response in people [4,14,20].

There is a growing interest in the scientific community to study the relationship between the human nervous system and the built environment. This emerging field of study, called neuroarchitecture, seeks to explain these relationships by applying neuroscience principles to architectural design. Subjective measures, such as self-reports or neurophysiological response measurement techniques inherent to neuroscience, can be used to identify environmental factors that produce positive emotional responses in architectural spaces [8,9,12]. In recent years, several experiments have been conducted in this area, analyzing various neurofeedback techniques and built environment variables. However, the results are rather limited, mainly due to the small sample size and the variety of methods used [4,15]. In addition, these experiments evaluate a small number of variables in a specific environment. A building is the result of the combination of thousands of elements; to reliably know whether its design is comfortable for its users, it is necessary to analyze the building as a whole.

Pre-occupancy evaluation (PrOE) is the process of obtaining feedback on how a building performs and interacts with its future users [1]. This information allows architects to identify design flaws before construction of the building begins. In a PrOE, virtual reality (VR) is a tool with great potential [21]; through this technology, different design alternatives can be shown to potential building occupants, with the aim of identifying which design variant generates the best response from users. However, the validity of the data obtained with this method is conditioned by the user's sense of presence in the virtual environment. When users are shown a VR scene, they need to feel as if they are there so that their responses are comparable to those in a real environment. [11,18].

Traditionally, creating realistic and immersive VR scenes has been a long and tedious process. Today, it has been greatly simplified thanks to Building Information Modeling (BIM) software [16]. BIM can be defined as a digital representation of the physical and functional characteristics of a building [4]; using BIM software, it is possible to quickly model a building in 3D and then export it to a VR visualization tool. The use of these dimensionally accurate models in conjunction with game engines allows the creation of frameworks for PrOE experiments in VR environments [17]. However, the appearance of BIM models is not very realistic due to the use of low quality textures to reduce the size of the model. Giving the BIM model a realistic appearance is still a fairly manual

process. In this study, we present a toolbox for quick and easy creation of VR occupancy evaluation experiments. Starting from a BIM model, a set of scripts were developed to (1) give the model a photorealistic appearance, (2) create multiple variants of the same model, (3) automatically generate evaluation questionnaires, and (4) manage experiment data. To the best of our knowledge this is the first toolbox for realising of PrOE in VR from BIM models that allows to automate the process of visual enhancement and the creation of different design variants from a single BIM model.

2 Toolbox Description

This section describes the VRPrOE toolbox and its various modules. VRPrOE was created with the goal of simplifying the process of creating PrOE experiments, allowing architects with no experience in creating VR environments to test their designs with potential end users (see Fig. 1). The toolbox was developed using the Unreal Engine 4 (UE4) (https://www.unrealengine.com/) and the Datasmith plugin (https://www.unrealengine.com/en-US/datasmith), which allows building geometry to be imported from various BIM software.

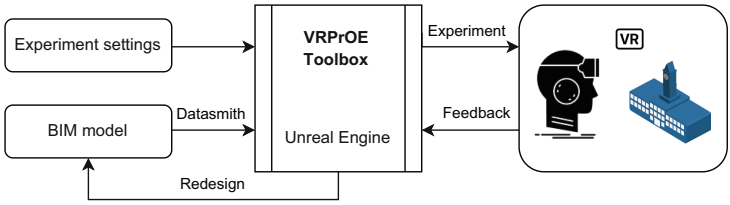


Fig. 1. VRPrOE toolbox

2.1 Visual Enhancement Module

BIM models typically use low-resolution textures, and non-architectural elements such as chairs or tables are modeled with low-density mesh. This makes BIM models lightweight, which facilitates collaboration among the various professionals using the model, but significantly reduces visual realism. This module is responsible for replacing unrealistic textures and geometries, adding elements not included in the BIM model, and adjusting lighting parameters such as intensity, color, or attenuation radius. Figure 2 shows a room of a building before and after applying this module.

The visual enhancement module uses three databases to list the changes to be made. First, the program reads a database containing 4 fields: *existing object identifier*, *new object identifier*, *transformation* and *replace*. Then it finds the existing object and inserts the new object in the place of the first one,



(a) (b)

Fig. 2. Sample room created with BIM software. (a) Before and (b) after running the visual enhancement module

applying the transformation (translation, rotation and scaling) specified in the *transformation* field. Finally, if the *replace* field is true, the existing object is deleted. If there is more than one existing object with the same identifier, this sequence is repeated for all objects with that identifier. In the example in Fig. 2, the landscape seen through the windows, the light simulating the sun, and other objects used to adjust reflections, rendering, and the VR experience have been added. The computer has also been replaced.

Using a database containing the object identifier, the current material identifier, and the new material identifier, the materials of the objects are replaced after the new objects are added. As before, if there is more than one object with the same identifier, the process is repeated for all objects with that identifier. Finally, the light identifier, intensity and color are stored in the third database.

2.2 Experiment Creation Module

A graphical user interface (GUI) allows setting the configuration of each experiment as shown in Fig. 3.

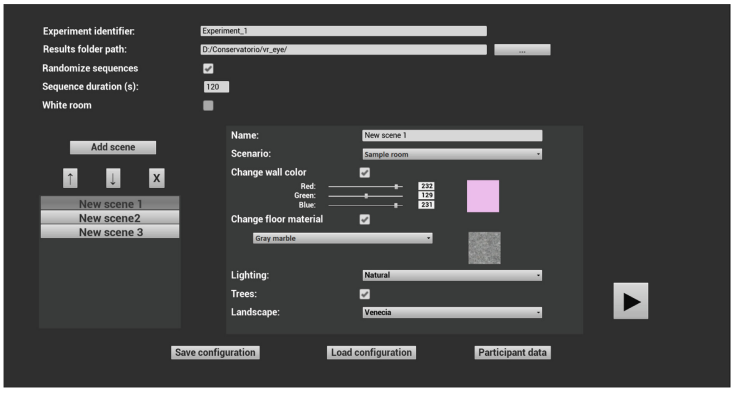


Fig. 3. Experiment setup window

The interface allows scenes to be added to the experiment to be displayed to the user for the specified time. The different scenes are displayed in the order in which they appear at the bottom left of the window, unless the *randomize sequences* checkbox is checked, in which case they appear in random order. Also included is the *white room* checkbox which, when checked, is used to insert an initial sequence in which a neutral colored room is displayed to allow the user to get used to the VR environment. In each scene it is possible to change (or not) the scenario, the color of the walls, the material of the floor, the type of lighting, the presence of trees and the exterior landscape. The exterior landscape is a 360° image that can be added with the previous module. By using images instead of 3D models, the cost of rendering is significantly reduced and it is possible to have countless landscapes without having to model them.

2.3 Questionnaire Creation Module

Questionnaires are the easiest way to get feedback from your users. The VRPrOE toolbox includes three types of questionnaires by default:

- Sociodemographic data: Participants' personal information, such as age, gender, and occupation, will be collected.
- Sense of presence: To assess the degree to which participants in the experiment feel they are in a real building, the Slater-Usoh-Steed (SUS) questionnaire [19] was used.
- Semantic differential scales (SDS) [13]: Pairs of opposite adjectives are used for experiment participants to rate their subjective perception of various architectural features on a 7-point Likert scale.

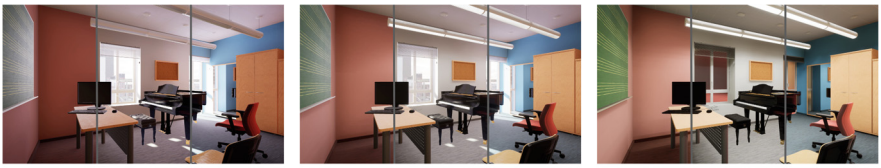
Questions are stored in a plain text file, which can be modified to change, add or delete questions without modifying the program. Prior to the start of the experiment, the participant reports his or her sociodemographic data; at the end of each scene, the participant is asked to fill out the SDS questionnaire. Finally, at the end of the experiment, the participant is asked to complete the SUS questionnaire. Questions can be responded to inside the immersive VR environment using the VR controller or outside the environment using a computer mouse.

2.4 Data Management Module

This module is responsible for managing the experiment data. Self-reported data from participants in previous questionnaires are stored in comma-separated value (CSV) format. In addition to this information, the person's location within the virtual environment is also recorded and stored along with a timestamp. The module also allows to send and receive data externally via web socket or http request. This feature can be used to receive data from biosensors and/or send the collected data to an external server. Data management is done in separate threads from the main thread to avoid interfering with game execution.

3 Proof of Concept

This section describes the experiment carried out as a proof of concept with the purpose of verifying that the toolbox can be used for the purpose for which it was designed. Several studies suggest that wall color and exposure to natural light have positive effects on human subjective well-being [3,6,10]. We used the proposed toolbox to evaluate the effect of color and lighting type on the perception of the environment in a conservatory classroom. For this purpose, a BIM model of the classroom was created using Autodesk Revit software and exported to UE4. Nine different scenes were created from this classroom using the experiment creation module of the toolbox (see Fig. 4). Each scene shows a variation of the classroom that can have 3 different color tones (warm, cold, and neutral) and 3 types of lighting (natural, artificial, or mixed).



(a) (b) (c)

Fig. 4. Variations of the conservatory classroom. On the left side of each image the classroom is shown with warm colors, in the middle with neutral colors and on the right side with cold colors. (a) Natural lighting. (b) Mixed lighting. (c) Artificial lighting

The colors have been applied to the floor and walls, leaving the rest of the elements intact. However, the light reflected on the floor and walls slightly changes the color of all the elements in the room, just as it would in a real environment. As can be seen in Fig. 4, the type of lighting affects the perception of color. For example, the left wall appears darker in natural light than in artificial light because the wall receives much less direct light when only natural light is used. Color perception is also affected by camera exposure; in our experiment, automatic exposure was used to simulate the human eye's response.

3.1 Experiment Design

Figure 5 shows the workflow of the experiment. First, the participant has to fill out a socio-demographic questionnaire, then he/she is given the VR controller and the head-mounted display (HMD) and is helped to fit it on the head. Once the HMD is properly fitted, the participant is shown a white room with a panel of buttons to interact with the controller. When the participant indicates that he/she is comfortable with the VR environment and the use of the controller, the core experiment begins. First, one of the 9 possible scenes, chosen at random, is shown.

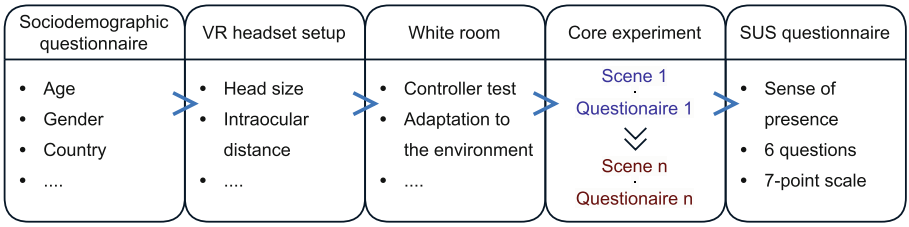


Fig. 5. Experiment workflow

After one minute of viewing, the participant is shown a SDS questionnaire to rate different aspects of the environment, which can be answered using the VR controller. After answering the questionnaire, the participant is moved to another scene and after one minute answers the same questionnaire again. This process is repeated until the participant has visualized a total of 3 scenes. Although scene selection is random, it is conditioned so that all participants see the three colors and three possible lights. Finally, the participant must answer the SUS questionnaire to assess his/her sense of presence in VR. During the experiment, participants are seated in an office chair that allows them to rotate freely. The HMD chosen is the HTC Vive Pro eye. A convenience sample of 13 members of the educational community who were not involved in creating the Toolbox was used to conduct the study. All gave their informed consent to participate.

In the SDS questionnaire we used 4 pairs of opposite adjectives related to different architectural features and human experiences [7] (see Table 1).

Table 1. Opposite adjective pairs from the differential semantic scales questionnaire

Adjectives	Human experience	Architectural feature
Pleasant/unpleasant	Restorativeness	Presence/absence of natural light
Open/closed	Restorativeness, stress and anxiety	Openness of spaces & Presence/absence of windows
Warm/cold	Stress and anxiety	Color of surfaces
Light/dark	Stress and anxiety	Luminance

3.2 Results and Discussion

The experiment was created by the VRPrOE developers, so it was not possible to get independent feedback on the toolbox. The levels of presence experienced by the participants in the experiment were determined using the SUS questionnaire. The mean of the 6 questionnaire items was 5.46 out of 7, with a standard deviation of 1.30, which can be considered a high sense of presence. The use of VR may be unpleasant for some users, the level of sense of presence is correlated with the

degree of comfort with the virtual environment and the hardware. Therefore, the result obtained indicates that in general the participants feel comfortable with the technology used. Table 2 shows the mean and standard deviation of the scores of the opposite adjective pairs for the different design variables. A one-way ANOVA analysis was performed to determine whether the differences between the means of the variables studied were statistically significant (see Table 3). Differences were considered statistically significant when p-value was <0.05.

Table 2. Opposite adjective pair scores with respect to color and lighting

Variable	Pleasant/unpleasant		Open/closed		Light/dark		Warm/cold	
	mean	std	mean	std	mean	std	mean	std
warm color	2.85	1.46	3.00	1.78	2.15	1.28	2.46	1.66
cold color	2.18	1.08	2.45	1.81	1.91	1.04	3.36	1.63
neutral color	3.08	1.83	3.58	2.19	3.25	1.91	3.42	1.56
artificial light	3.58	1.68	4.75	1.86	3.58	1.78	3.67	1.87
natural light	2.50	1.31	2.50	1.31	2.08	1.24	3.00	1.48
mixed light	2.08	1.16	1.83	1.27	1.67	0.78	2.50	1.45

Table 3. One-way ANOVA p-values of the variables studied

	Pleasant/ unpleasant	Open/closed	Light/dark	Warm/cold
Color	0.3428	0.3867	0.0751	0.2683
Lighting	0.0366	0.0001	0.0032	0.2204
Color & Lighting	0.0919	0.0087	0.0193	0.1148

In the case of color alone, the differences between the means are not large enough to be considered statistically significant for any of the pairs of adjectives. In contrast, in the case of lighting, a relationship between the type of lighting and the perception of the environment was found for 3 of the 4 pairs of adjectives (pleasant/unpleasant; open/closed; light/dark). With a combination of natural and artificial lighting, the room is perceived as more pleasant, open and light. With only artificial lighting, they are perceived as more unpleasant, closed and cold. The exclusive use of natural light would be in the middle.

4 Conclusions

The built environment can affect brain activity, and assessing how different design variables impact it can guide architects in creating buildings that maximize well-being. This paper presents a toolbox for creating PrOE experiments in VR environments from BIM models. This first version of the toolbox has some

limitations, only a small number of environmental variables can be manipulated, it has been tested with only one HMD model and the only tool included to assess the environment are questionnaires. Nevertheless, we believe it is sufficient to demonstrate the potential of this tool. In our proof-of-concept experiment, we started with an unrealistic-looking BIM model, and thanks to the toolbox, we were able to instantly create different versions of the classroom with a highly realistic look and feel, evoking a strong sense of presence for the participants.

The results of the experiment reveal a strong relationship between the type of lighting and the subjective perception of the environment. However, the sample size used in this experiment is quite small, and an experiment with a larger number of participants would be necessary to confirm these results. Future versions of the software could also increase the number of environmental parameters, other VR devices could be integrated and it could handle and integrate biosensor data processing, which can be very useful in understanding how participants in experiments feel without requiring them to spend time answering surveys.

Acknowledgments. Grants PID2020-115220RB-C21 and EQC2019-006063-P funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way to make Europe". Grant BES-2021-097834 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future". Grant 2023-PRED-21291 funded by Universidad de Castilla-La Mancha and by "ESF Investing in your future". Grant 2022-GRIN-34436 funded by Universidad de Castilla-La Mancha and by "ERDF A way of making Europe". This research was also partially funded by CIBERSAM, Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación and co-funded by "ERDF A way to make Europe".

References

1. Alzoubi, H., Al-Rqaibat, S., Bataineh, R.F.: Pre-versus post-occupancy evaluation of daylight quality in hospitals. Build. Environ. **45**(12), 2652–2665 (2010). https://doi.org/10.1016/j.buildenv.2010.05.027
2. Azzazy, S., Ghaffarianhoseini, A., GhaffarianHoseini, A., Naismith, N., Doborjeh, Z.: A critical review on the impact of built environment on users' measured brain activity. Archit. Sci. Rev. **64**(4), 319–335 (2020). https://doi.org/10.1080/00038628.2020.1749980
3. Beute, F., de Kort, Y.A.: The natural context of wellbeing: ecological momentary assessment of the influence of nature and daylight on affect and stress for individuals with depression levels varying from none to clinical. Health & Place **49**, 7–18 (2018). https://doi.org/10.1016/j.healthplace.2017.11.005
4. Bower, I., Tucker, R., Enticott, P.G.: Impact of built environment design on emotion measured via neurophysiological correlates and subjective indicators: A systematic review. J. Environ. Psychol. **66**, 101344 (2019). https://doi.org/10.1016/j.jenvp.2019.101344
5. Coburn, A., Vartanian, O., Chatterjee, A.: Buildings, beauty, and the brain: a neuroscience of architectural experience. J. Cogn. Neurosci. **29**(9), 1521–1531 (2017). https://doi.org/10.1162/jocn_a_01146
6. Edwards, L., Torcellini, P.: Literature review of the effects of natural light on building occupants. Technical report, National Renewable Energy Laboratory (2002). https://doi.org/10.2172/15000841

7. Ergan, S., Shi, Z., Yu, X.: Towards quantifying human experience in the built environment: a crowdsourcing based experiment to identify influential architectural design features. J. Build. Eng. **20**, 51–59 (2018). https://doi.org/10.1016/j.jobe.2018.07.004
8. Fernández-Sotos, A., Martínez-Rodrigo, A., Moncho-Bogani, J., Latorre, J.M., Fernández-Caballero, A.: Neural correlates of phrase quadrature perception in harmonic rhythm: an EEG study using a brain-computer interface. Int. J. Neural Syst. **28**(05), 1750054 (2018). https://doi.org/10.1142/S012906571750054X
9. Kim, J., Kim, N.: Quantifying emotions in architectural environments using biometrics. Appl. Sci. **12**(19), 9998 (2022). https://doi.org/10.3390/app12199998
10. Kwallek, N., Lewis, C.M., Robbins, A.S.: Effects of office interior color on workers' mood and productivity. Percept. Mot. Skills **66**(1), 123–128 (1988). https://doi.org/10.2466/pms.1988.66.1.123
11. Latini, A., Giuseppe, E.D., D'Orazio, M.: Immersive virtual vs real office environments: a validation study for productivity, comfort and behavioural research. Build. Environ. **230**, 109996 (2023). https://doi.org/10.1016/j.buildenv.2023.109996
12. Martínez-Rodrigo, A., Fernández-Aguilar, L., Zangróniz, R., Latorre, J.M., Pastor, J.M., Fernández-Caballero, A.: Film mood induction and emotion classification using physiological signals for health and wellness promotion in older adults living alone. Expert. Syst. **37**(2), e12425 (2020). https://doi.org/10.1111/exsy.12425
13. Messick, S.J.: Metric properties of the semantic differential. Educ. Psychol. Measur. **17**(2), 200–206 (1957). https://doi.org/10.1177/001316445701700203
14. Peralta, A., Fernández-Caballero, A., Latorre, J.M.: Determining the ambient influences and configuration of optimised environments for emotional wellbeing of older adults. Ergonomics **64**(9), 1146–1159 (2021). https://doi.org/10.1080/00140139.2021.1909756
15. Rad, P.N., Behzadi, F., Yazdanfar, S.A., Ghamari, H., Zabeh, E., Lashgari, R.: Exploring methodological approaches of experimental studies in the field of neuroarchitecture: a systematic review. HERD: Health Environ. Res. Des. J. 193758672211331 (2023). https://doi.org/10.1177/19375867221133135
16. Safikhani, S., Keller, S., Schweiger, G., Pirker, J.: Immersive virtual reality for extending the potential of building information modeling in architecture, engineering, and construction sector: systematic review. Int. J. Digit. Earth **15**(1), 503–526 (2022). https://doi.org/10.1080/17538947.2022.2038291
17. Schneider, S., Kuliga, S., Fuchkina, E., Weiser, R.: VREVAL - a virtual reality based framework for user-centered evaluation of complex buildings. In: Conference on Education and Research in Computer Aided Architectural Design (2018)
18. Slater, M., Khanna, P., Mortensen, J., Yu, I.: Visual realism enhances realistic response in an immersive virtual environment. IEEE Comput. Graph. Appl. **29**(3), 76–84 (2009). https://doi.org/10.1109/mcg.2009.55
19. Slater, M., Usoh, M., Steed, A.: Depth of presence in virtual environments. Presence: Teleoperators Virtual Environ. **3**(2), 130–144 (1994). https://doi.org/10.1162/pres.1994.3.2.130
20. Sokolova, M.V., Fernández-Caballero, A.: A review on the role of color and light in affective computing. Appl. Sci. **5**(3), 275–293 (2015). https://doi.org/10.3390/app5030275
21. Tseng, K.C., Giau, D.T.N., Huang, P.-H.: A systematic review of the potential application of virtual reality within a user pre-occupancy evaluation. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2017. LNCS, vol. 10279, pp. 612–620. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58700-4_50

Consumption Performance, of a Multi Agent System Used to Achieve Comfort Preferences

Pedro Filipe Oliveira[1,2(✉)], Paulo Novais[2], and Paulo Matos[1]

[1] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
poliveira@ipb.pt, pmatos@ipb.pt

[2] Algoritmi Centre, Department of Informatics, University of Minho, Braga, Portugal
pjon@di.uminho.pt

Abstract. In this work, a multi-agent system was developed to manage comfort preferences, in an autonomous and completely automatic and non-invasive way for the user. This system was based on the use of an architecture supported by low cost hardware, namely Raspberrys, to support the different actuators present in the different spaces. To carry out its validation, a methodology was created to analyze consumption performance, and the results obtained in two scenarios are demonstrated here, a domestic housing scenario, and a professional environment scenario. The results obtained were quite positive for the prototype developed, and validate the option of using low cost hardware.

Keywords: multi-agent · consumption · preferences

1 Introduction

Currently, users are increasingly looking for automatism's that make their daily lives easier. And thus allow them to have more free time for themselves, not having to worry about routine tasks.

In which most of the time they are not able to optimize them in the same way, as is done by any automatic system. Still achieving autonomy, which is perhaps the characteristic most sought after by users of this type of system.

With this work, it has been validated a multi agent system to achieve the best comfort preferences using low cost hardware like Raspberry's, and at the same time improve consumption performance. This work aims to give continuity and finalize the doctoral work presented in previous editions [1–6].

This project demonstrates the complete development of a multi-agent system, and its effective performance related to energy consumption.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 199–208, 2023.
https://doi.org/10.1007/978-3-031-43461-7_20

2 Materials and Methods

In this section, the different used actuators are detailed, as well the multi-agent system developed.

2.1 System Actuators

For the different actuators operation, in the different scenarios, different valences were used, and are following detailed:

Temperature/Relative Humidity
Namely in terms of heating, this was achieved through a hydraulic underfloor heating that is divided into different circuits to cover the different house areas, as well for its control, six thermostats from the *Tado* brand were used that allow in real time to send, using an API, the desired temperature. The thermostat and its operation mode, can be seen at Figs. 1 and 2.



Fig. 1. Actuator desired temperature.



Fig. 2. Thermostat current temperature.

For cooling and relative humidity control, also six fan coils were used, one for each area, and controlled by individual thermostats, which also allow the desired temperature definition through an API.

Luminance/Brightness
For luminance and brightness, *Shelly* bulbs that have WIFI connection are used, that allow to control different luminance and brightness present at each individual environment, in the same way they have an API to integrate with other smart home systems, and that allow its direct control. This device can be seen at Fig. 3.

Sound
For sound, were used *Echo* speakers from the brand *Amazon* which have WIFI connection, and allow to control the sound volume and also the played music (sound source, playlist or gender) present at each individual environment, in the same way they have an API to integrate with other smart home systems, and that allow its direct control. This device can be seen at Fig. 4.



Fig. 3. Smart bulb.



Fig. 4. Smart speaker.

Security Systems
Also, was tested the possibility to use some security systems, and enable/disable this according to the user detection at the environment. The used device is from the brand *Ring* and it can be seen at Fig. 5.



Fig. 5. Smart security system.

2.2 Multi-agent System Architecture

The multi agent system was developed using JASON and ARGO, and the Fig. 6 represent the different layers architecture separation, to easily identify the purpose of each, and agents containing it.

There will be one principal agent who will represent local system, namely each individual environment, where it was a need to ensure individualized comfort conditions, such as a room in a house, or a office in a building. This agent will take into account any directives that may exist for this environment, such as lower or upper limits to different comfort conditions, or also safety parameters that may be critical for a given space. This agent will have a obviously prevalence relative to others, since it will be the dominant for a given environment.

With users respect, each one in the space, will also be represented by an agent, this will receive user preferences from main system, for the place where it is, as well for the time in which it is. Also in this situation there will be a prioritization that identifies which user will have environment supremacy, so it also has an increase in the negotiation process.

In decision-making process, all users agents and agents representing the environment will be taken into account. With the different priorities that each of them has, and with this information will begin the negotiation process.

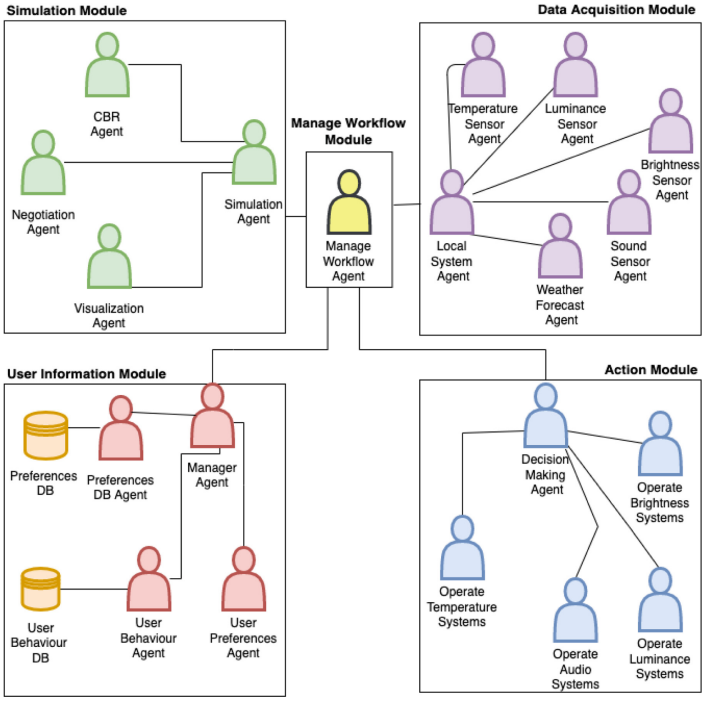


Fig. 6. Multi-agent system architecture

2.3 Evaluation Scenarios

For the proposed framework analysis and evaluation, different scenarios were formulated. Initially it was applied in a two floors house.

In this way, it was possible to validate the domestic space concept, with a family composed of two adult users and a child, characterized before.

Their individual preferences were defined, and the MAS system analysis was carried out during a six months period.

The workspace concept was also defined, with different local systems being installed in the partner company's offices.

It was also planned to install some local systems, in partnership with the partner higher education institution, as well in a local health unit. But due to budget constraints, and costs associated with acquiring the high number of equipment's (*Raspberry's*) necessary for data acquisition, this was not possible.

This fact was also aggravated, due to the constraints introduced by the pandemic, having been completely impossible to access the health unit at that period, as well the higher education institution.

In Sect. 3 the two defined scenarios results, are detailed, and explained for each of the aspects analyzed.

Home Scenario

Table 1 characterizes the different users that compose the home scenario.

Table 1. Home Scenario - Users characterization.

Username	Type	Proportion
User1	Adult	1
User2	Adult	1
User3	Child	0,75

Table 2 was developed, where all the entry records (samples) considered for analysis are represented, and they are divided by the six months under analysis (October 2021, November 2021, December 2021, January 2022, February 2022 and March 2022).

Totalizing *15420* log records for the six months in question. Each of these samples represents one user entrance/presence, recorded by the local system. We can see an average of *84,45* samples registered for each day.

Table 2. Home Scenario - Total registered samples.

	Oct	Nov	Dec	Jan	Feb	Mar	Total/Average
Nr. of Days	31	30	31	31	28	31	182
Nr. of Periods	992	960	992	992	896	992	5824
Total samples	3219	3033	1548	2988	1737	2895	15420
Average/Day	103,84	101,1	49,94	96,39	62,04	93,39	84,45

Work Scenario

Table 3 characterizes the six users that compose the work scenario.

Table 3. Work Scenario - Users characterization.

Username	Type	Proportion
User10	Hierarchy_1	(100-1)
User20	Hierarchy_2	(100-2)
User30	Hierarchy_2	(100-2)
User40	Hierarchy_2	(100-2)
User50	Hierarchy_2	(100-2)
User60	Hierarchy_3	(100-3)

Table 4 was developed, where all the entry records (samples) considered for analysis are represented, and they are divided by the six months under analysis (October 2021, November 2021, December 2021, January 2022, February 2022 and March 2022).

Totalizing *36578* log records for the six months in question. Each of these samples represents one user entrance/presence, recorded by the local system. We can see an average of *200,98* samples registered for each day.

Table 4. Work Scenario - Total registered samples.

	Oct	Nov	Dec	Jan	Feb	Mar	Total/Average
Nr. of Days	31	30	31	31	28	31	182
Nr. of Periods	682	660	682	682	616	682	4004
Total samples	6024	8170	3676	6420	4968	7320	36578
Average/Day	194,32	272,33	118,58	207,1	177,43	236,13	200,98

3 Results

To assess the results, the scenarios identified in Sect. 2.3 were defined, and implemented. Thus, a six-month period was defined for the identified scenarios analysis, as well the users present.

For the spaces characterized in Sect. 2.3, information was then collected over a six months period. Thus, it was possible to carry out all the statistical analysis, in order to execute the results compilation presented below at Sect. 3.1 and 3.2 and at Tables 6 and 8.

As previously mentioned, the results presented are preliminary and subject to industrial secrecy by the partner company. Therefore, all possible information is presented, considering the company's intention to commercialize the developed product, there are thus several restrictions on more data availability.

3.1 Home Scenario

Thus, all manual changes made during the testing phase were analyzed, and the satisfaction metric was calculated, by period of time/place.

The average satisfaction was also measured, for the different periods: morning (8am–1pm), afternoon (1pm–7pm) and night (7pm–12pm).

Also regarding energy savings, and knowing that it is currently a factor that isn't and cannot be neglected by any individual user or any business entity.

Considering the costs increase with different energy types, as well the ecological footprint that its production represents, the savings metric was also calculated, always considering that the purpose of this solution would not have this as prime factor, but indeed the maximum user comfort.

But knowing from the start that with all the introduced automatism's (detection of users present at the space, adjustment to minimum reference values in empty spaces, etc.) by the proposed solution, a decrease in consumption would be expected by itself.

Compared to solutions that only implement pre-programmed fixed adjustments and which most of the time don't include any automatism, such as simply allowing to detect absence periods, for example in the workspace, such as vacations, holidays or others, in this scenario, savings are expected to be even more significant.

To check exact values, the month global consumption was been verified for each analyzed space, and compared with the same month global consumption, after applying the solution.

At Table 5 we can see the mean value for the baseline day consumption, and the day consumption for the analyzed period, and also the difference in kWh, and the savings in percentage value.

Table 5. Home Scenario - Day Energy consumption (mean value).

Scenario	Baseline (kWh)	Period analyzed (kWh)	Difference (kWh)	Savings (%)
Home	35,2	32,05	3,15	9,84

At Table 6 we can see the total consumption value for the baseline, and for the 6 months period analyzed for the home scenario, and also the difference in kWh, and the savings in percentage value. At Fig. 7 we can see the plot of this information.

Table 6. Home Scenario - Energy consumption - 6 Months.

	Oct	Nov	Dec	Jan	Feb	Mar	Total
Nr. of Days	31	30	31	31	28	31	182
Baseline (kWh)	806	960	1240	1426	952	930	6314
Period analyzed (kWh)	682	870	1209	1209	868	868	5706
Difference (kWh)	124	90	31	217	84	62	608
Savings (%)	18,18	10,34	2,56	17,95	9,68	7,14	10,66

3.2 Work Scenario

At Table 7 we can see the mean value for the baseline day consumption, the day consumption for the analyzed period, and also the difference in kWh, and the savings in percentage value.

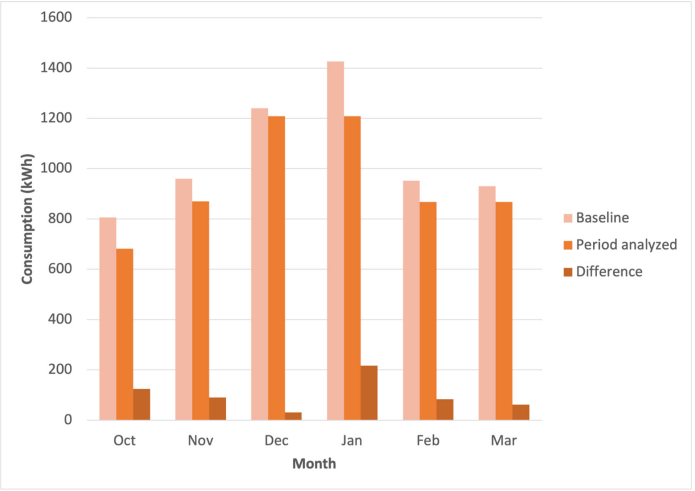


Fig. 7. Home Scenario - Energy consumption - 6 Months.

Table 7. Work Scenario - Day Energy consumption (mean value).

Scenario	Baseline (kWh)	Period analyzed (kWh)	Difference (kWh)	Savings (%)
Work	42,5	36,4	6,1	16,76

At Table 8 we can see the total consumption value for the baseline, and for the 6 months period analyzed for the work scenario, and also the difference in kWh, and the savings in percentage value. At Fig. 8 we can see the plot of this information.

Table 8. Work Scenario - Energy consumption - 6 Months.

	Oct	Nov	Dec	Jan	Feb	Mar	Total
Nr. of Days	31	30	31	31	28	31	182
Baseline (kWh)	992	1050	1519	1643	1036	992	7232
Period analyzed (kWh)	899	840	1364	1612	868	868	6451
Difference (kWh)	93	210	155	31	168	124	781
Savings (%)	10,34	25	11,36	1,92	19,35	14,29	12,11

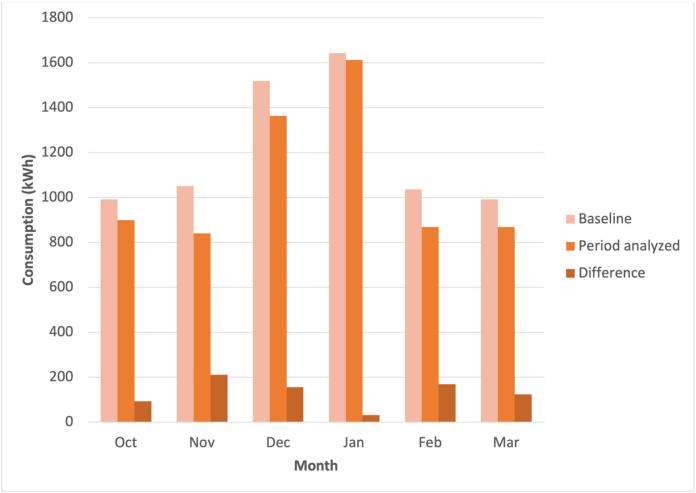


Fig. 8. Work Scenario - Energy consumption - 6 Months.

4 Conclusions

With this project, was achieved the complete development of a multi-agent system, an effective reduction in consumption.

The possibility of using low cost hardware (40€) to control this type of system was also validated. Therefore, it is effectively possible to use this type of equipment for the development of this type of project.

The 6 months analyzed period is not very extensive, but it can be seen as sufficient for this kind of spaces (domestic, small company) analysis, because users remain in some way very constant, and where there is thus no significant variance in their preferences.

For future work, some more scenarios have to be tested, to reinforce the full effectiveness of this system, and the values of his overall performance.

Acknowledgments. The authors are grateful to the Foundation for Science and Technology (FCT, Portugal) for financial support through national funds FCT/MCTES (PIDDAC) to CeDRI (UIDB/05757/2020 and UIDP/05757/2020) and SusTEC (LA/P/0007/2021).

References

1. Oliveira, P., Matos, P., Novais, P.: Behaviour analysis in smart spaces. In: 2016 International IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), pp. 880–887. IEEE (2016)
2. Oliveira, P., Novais, P., Matos, P.: Challenges in smart spaces: aware of users, preferences, behaviours and habits. In: De la Prieta, F., et al. (eds.) PAAMS 2017. AISC, vol. 619, pp. 268–271. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61578-3_34
3. Oliveira, P., Pedrosa, T., Novais, P., Matos, P.: Towards to secure an IoT adaptive environment system. In: Rodríguez, S., et al. (eds.) DCAI 2018. AISC, vol. 801, pp. 349–352. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-99608-0_43
4. Oliveira, P.F., Novais, P., Matos, P.: A multi-agent system to manage users and spaces in a adaptive environment system. In: De La Prieta, F., et al. (eds.) PAAMS 2019. CCIS, vol. 1047, pp. 330–333. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24299-2_31
5. Oliveira, P.F., Novais, P., Matos, P.: Using Jason framework to develop a multi-agent system to manage users and spaces in an adaptive environment system. In: Novais, P., Vercelli, G., Larriba-Pey, J.L., Herrera, F., Chamoso, P. (eds.) ISAmI 2020. AISC, vol. 1239, pp. 137–145. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58356-9_14
6. Oliveira, P.F., Novais, P., Matos, P.: Security constraints on an multi-agent system to manage users and spaces in an adaptive environment system. In: Novais, P., Carneiro, J., Chamoso, P. (eds.) ISAmI 2021. LNNS, vol. 483, pp. 171–176. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-06894-2_16

Pleasure and Displeasure Identification from fNIRS Signals

Daniel Sánchez-Reolid[1(✉)], Roberto Sánchez-Reolid[1,2], Antonio Fernández-Caballero[1,2,3], and Alejandro L. Borja[4]

[1] Neurocognition and Emotion Unit, Instituto de Investigación en Informática de Albacete, 02071 Albacete, Spain
{Daniel.Sanchez,Roberto.Sanchez,Antonio.Fdez}@uclm.es
[2] Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, 02071 Albacete, Spain
[3] CIBERSAM (Biomedical Research Networking Center in Mental Health), 28016 Madrid, Spain
[4] Departamento de Ingeniería Eléctrica, Electrónica, Automática y Comunicaciones, Universidad de Castilla-La Mancha, 02071 Albacete, Spain
Alejandro.Lucas@uclm.es

Abstract. This paper presents two feature extraction methods for training different classification models for the detection of pleasure and displeasure defined by high and low valence levels using functional near-infrared spectroscopy (fNIRS). The study involved fifty-four volunteers who were presented with emotion-inducing image blocks while their prefrontal cortex brain activity was recorded by an fNIRS device. Results from the participants' responses to a questionnaire showed no significant differences in valence-related scores. The study used statistical and ROCKET methods to extract features and trained six models to discriminate high and low valence. The results showed that ROCKET features performed best with kNN, MLP and SVM classifiers. The research highlighted the potential of different feature extraction methods to improve the accuracy of biosignal analysis from fNIRS devices.

Keywords: Emotion induction · Functional near-infrared spectroscopy · Machine learning · Signal processing

1 Introduction

Emotions are a crucial aspect of human interaction and have recently become a topic of growing interest in the field of human-computer interaction. The branch of affective science known as emotion classification aims to distinguish one emotion from another. This classification has traditionally been studied from two main perspectives. The discrete approach, proposed by Ekman in 1972, suggests that there are six basic emotions that are universal across cultures (happiness, disgust, anger, fear, sadness, and surprise) [5]. On the other hand, the dimensional approach, proposed by Russell in 1980, argues that different emotional

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 209–219, 2023.
https://doi.org/10.1007/978-3-031-43461-7_21

states can be represented by two or more dimensions, with the most commonly used dimensions being activation and valence [17]. The dimensional approach has seen a significant increase in publications in recent years [7,11,18].

Functional near-infrared spectroscopy (fNIRS) is a non-invasive neuroimaging technique that has been increasingly used in recent decades to study cognitive processes. It measures changes in blood oxygenation in the brain by emitting infrared light through the skull and detecting the amount of light absorbed or scattered by brain tissue. The technique is based on the haemodynamic response, i.e. the increase in oxygen consumption and blood flow in activated brain regions. fNIRS has several advantages over other neuroimaging techniques, such as its ability to be used in natural environments, its cost-effectiveness, its portability, its high temporal resolution and its high ecological validity. It can also be used to study linguistic, auditory and visual topics and to investigate areas that are difficult to study with other neuroimaging techniques. It is also advantageous for the study of clinical populations, as there is less physical and psychological burden on patients and less need to control their actions.

The aim of this paper is to compare high and low valence classification using different methods of feature extraction from fNIRS biosignals. The paper presents a study in which participants were presented with a series of images from the International Affective Picture System (IAPS) database to elicit emotional responses. The images were organised into blocks, with each block containing in random order images related to a particular emotion based on their arousal (only high arousal) and valence (pleasant/displeasure). The signals corresponding to each block of images were segmented and obtained from each participant. The signals were then pre-processed and features extracted using two different methods. These features were used to train different machine learning models for further comparison [19].

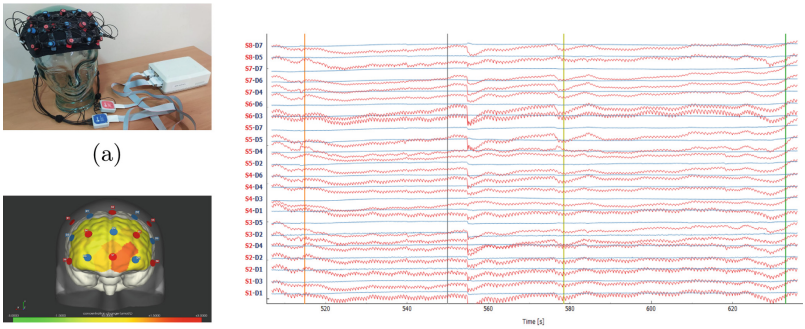
2 Materials and Methods

2.1 The fNIRS Device

The measurement of haemodynamic responses by oxygenated and deoxygenated haemoglobin was performed using the NIRSport2 fNIRS https://nirx.net/nirsport. The assembly used in this work consisted of a headband into which 8 emitting optodes and 7 detecting optodes are inserted, resulting in a 22-channel configuration covering the prefrontal cortex (PFC), as shown in Fig. 1a. This configuration allows for rapid placement of the optodes on the subject's head, thus reducing the time required for device placement.

2.2 The Software Suite

The software PsychoPy [13] was used to regulate the conditions of the experiment. The open source software PsychoPy provides an easy-to-use interface for designing and analysing experiments in psychology and neuroscience.



(a) (b) (c)

Fig. 1. (a) Prefrontal cortex fNIRS optode assembly. (b) fNIRS 3D plot sample. (c) Raw fNIRS signal fragment from a study participant.

Researchers can create designs with text, images, video and custom questionnaires on slides and the software is compatible with a variety of devices.

The software used to record the data was Aurora fNIRS (https://nirx.net/software) as it is the software compatible with the NIRSport2 device. The use of Aurora fNIRS together with PsychoPy allows easy generation of event markers in our signals when using the integrated Lab Streaming Layer protocol (https://labstreaminglayer.readthedocs.io/).

2.3 Stimuli: The IAPS Database

The International Affective Picture System (IAPS) database [9], known for its validation and reliability, was chosen as the source of colour pictures to evoke emotions in participants, as pictures are the primary medium used to elicit. The aim of showing these images was to evoke the same emotion in the participant as the one depicted in the photograph. Two different experimental conditions were chosen: (1) the HV-HA conditions for high valence and high arousal, and (2) the LV-HA conditions for low and high arousal. Table 1 shows the mean values for each group of selected images.

Participants were asked to rate their level of pleasantness/unpleasantness, calmness/arousal and control when viewing each image, allowing us to obtain the mean and standard deviation of the valence, arousal and dominance of each

Table 1. Mean and standard deviation for valence, arousal and dominance for each group of images selected from the IAPS.

Experimental Condition	Valence	Arousal	Dominance
HV-HA	7.33 (0.40)	6.59 (0.61)	5.55 (0.44)
LV-HA	1.68 (0.28)	6.86 (0.32)	2.69 (0.37)

image. Their responses were collected using a Self-Assessment Manikin (SAM) graphic scale questionnaire [10].

2.4 Participants

A young group of 54 participants took part in the experiment, of which 31 were female and 23 were male. Their age ranged from 18 to 32 years (M = 23.32, SD = 2.67). All participants were in good health and did not have any neurological or psychiatric disorders. It should be noted that 92.59% of the participants were right-handed and had normal (40.74%) or corrected (59.26%) vision.

This research was carried out in accordance with the Declaration of Helsinki and did not involve real patients, so the approval of a clinical research ethics committee was not mandatory under Spanish and European law. All participants volunteered without any financial incentive and accepted the informed consent form after being informed of the nature of the images they would be exposed to. In addition, participants had the right to withdraw from the experiment at any time without prejudice.

2.5 Experimental Design

The study was conducted in a controlled and comfortable environment to remove any external factors that might influence the results. Participants were seated in a low-light room in front of a computer monitor with a 24-inch screen at a distance of 60 cm. The stimuli were delivered using the PsychoPy software [13] and the procedure is illustrated in the flowchart shown in Fig. 2. During the recording of the fNIRS measurement, the subjects were required to look attentively at each set of stimuli and to maintain their focus on the images for the entire duration of the exposure.

The experiment began with the participant being informed of the study procedure and completing consent, sociodemographic and clinical forms. Prior to the presentation of the stimuli, a baseline recording of 120 s of rest was made at the beginning of the experiment. The images were presented in random order within each group of images. These images were presented sequentially in the

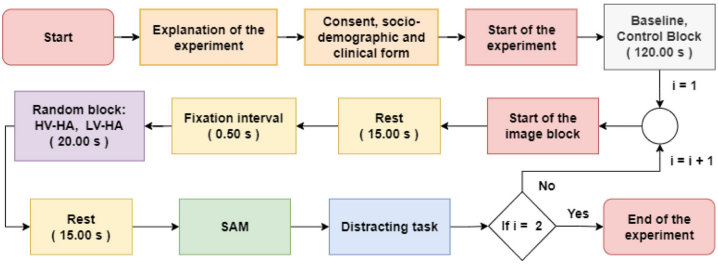


Fig. 2. Flowchart of the experimental design used in PsychoPy.

centre of the computer screen for 2 s each, for a total of 10 images per block, for a total of 20 s per block. Before and after the presentation of each block, the participant had 15 s to initialise and terminate the display in a calm state.

For all participants, the order of appearance of the other blocks was randomised and each block appeared only once during the experiment. For each block of images, participants were given time to rate their emotional experience on the SAM scale [10] by rating valence, arousal and dominance on a bipolar scale applied to each group of images. This was followed by a distraction task aimed at eliminating the emotion evoked by the pictures. Once the final distraction task was completed, the test was terminated and the participant was thanked for their time.

2.6 Preprocessing

First, the raw signals underwent a transformation process to become optical density signals, and during this stage the quality of the coupling between the scalp and the optodes was assessed using the scalp coupling index (SCI) [14]. Any defective channels were identified and removed, and any gaps were filled using interpolation based on nearby channels.

The temporal derivative distribution repair technique was used [6] to correct for any motion artefacts present in the optical density phase. In addition, the modified Beer-Lambert law was used to convert the optical density data to haemoglobin concentration, specifically oxygenated and deoxygenated haemoglobin (HbO_2 and HbR, respectively). Finally, to remove artefacts due to the heartbeat ∼1 Hz, a bandpass filter was applied within a specific frequency range of 0.05–0.4 Hz and the signals were segmented.

2.7 Statistical Feature Extraction

12 statistical features were calculated for both oxygenated and deoxygenated haemoglobin for each of the 22 fNIRS channels. These 12 features were calculated per trial and include mean (M), median (Md), standard deviation (SD), variance (Va), skewness (SKU), kurtosis (KU), fourth moment (4Mo), fifth moment (5Mo), root mean square (RMS), maximum (Max), minimum (Min) and zero crossing (ZC).

2.8 Time Series Feature Extraction

The ROCKET method [4] was used to extract time series features. The study defined a collection of convolutional kernels with parameters such as length, weights, bias, dilation and padding. However, the only hyperparameter in the ROCKET implementation that requires manual configuration is the number of kernels, and the method operates in an unsupervised manner.

All other ROCKET hyperparameters, including length, weights, bias, dilation and padding, are randomly sampled. This means that each kernel in ROCKET

assigns random values to these parameters. The study used 1000 kernels and two operators: (1) the maximum value operator, and (2) the positive value ratio operator [4].

2.9 Machine Learning Algorithm Selection and Evaluation

The study used six different supervised classifiers from different categories, each with separate statistical and ROCKET features. The classifiers and their respective hyperparameters are presented below. Identical hyperparameters were used for both feature extraction approaches, indicating that any discrepancies in classification metrics between the two for a given classifier are due to differences in how each feature extraction method works.

- Gaussian Naive Bayes (GNB) is a classifier that applies Bayes' theorem under the assumption of feature independence. It estimates the probability of a new instance's class membership from the probabilities of its features in that class, assuming no relationship between them [15]. The default parameters for GNB were used in this study.
- k-Nearest Neighbours (kNN) is based on the idea that data points belonging to the same class are more similar to each other than to those belonging to different classes. Various similarity metrics such as distance, proximity or closeness can be used to compute the degree of similarity [1]. In this study, similarity was calculated using a neighbour count of 3.
- Multi-Layer Perceptron (MLP) is a popular type of Artificial Neural Network (ANN) that uses interconnected nodes in multiple layers to make predictions and learn from input data. Nonlinear activation functions are applied to the weighted sum of inputs in each node, allowing the network to learn complicated patterns in the data [16]. The current study used a fully connected layer with a hidden layer consisting of 100 neurons. The hidden layer used rectified linear unit (ReLU) activation and an alpha value of 0.0001, and the optimiser used was ADAM. In addition, the maximum number of iterations was set to 1000.
- Random Forest (RF) is a machine learning technique that uses multiple decision tree classifiers on different subsets of the input data. It improves predictive performance and prevents overfitting by taking the average decision of these decision trees [8]. This study used 100 decision tree classifiers with a maximum depth of 5 and 1 feature to consider when searching for the optimal split.
- Support Vector Machines (SVM) are a type of supervised learning algorithm used for classification and regression analysis. SVM constructs a hyperplane in a multidimensional space that separates the data into different classes. The algorithm maximises the distance between the hyperplane and the nearest data point of each class to optimise the hyperplane definition [3]. This work used a non-linear radial basis function kernel with a gamma value of 2 and a C-value of 1.

- XGradient Boost (XGB) is a machine learning algorithm that creates a strong model by using decision trees as base models. The algorithm uses a gradient boosting framework to improve the performance of weaker models through iterative processing [2]. In this study, the XGBoost algorithm was used with a multiclass objective and specific hyperparameters such as a subsample fraction of 0.3, a learning rate of 0.1, a maximum depth of 3, an alpha of 5, a number of estimators of 15, and a number of classes of 3.

The Scikit-learn library [12] was used to implement various classifiers in the research. The default parameters were used for all models except those where parameters were specified.

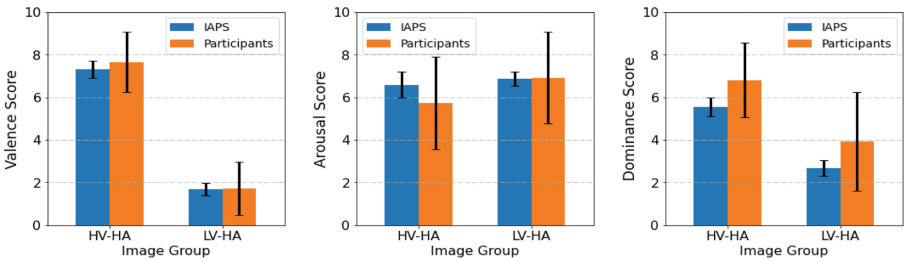
3 Result and Discussion

3.1 Comparing SAM Responses to IAPS Values

After analysing the mean scores obtained from the SAM scale (see Table 2), it can be seen that the results obtained for both image groups are similar to those provided by the IAPS (see Fig. 3). In order to check whether there are significant differences between the values obtained and those provided by the IAPS, a one-sample t-test is carried out for each condition with p-value < 0.05.

Table 2. Mean and standard deviation for valence, arousal and dominance for each group of images from the SAM scale scores of each participant.

Experimental condition	Valence	Arousal	Dominance
HV-HA	7.65 (1.42)	5.72 (2.17)	6.81 (1.75)
LV-HA	1.72 (1.27)	6.93 (2.14)	3.94 (2.32)



(a) Valence Scores (b) Arousal Scores (c) Dominance Scores

Fig. 3. IAPS and participants' SAM scale scores for valence, arousal and dominance.

No significant differences were found for valence (p-value = 0.104 for the HV-HA image group and p-value = 0.807 for the LV-HA image group). However, for the HV-HA image group, significant differences were observed for arousal (t = 2.940, p-value = 0.005, η = 0.044, $1 - \beta$ = 0.830), with participants perceiving a lower level of activation compared to the values presented in the IAPS. In terms of dominance levels, significant differences were found for both the HV-HA (t = 5.326, p-value < 0.001, η = 0.732, $1 - \beta$ = 1.000) and LV-HA (t = 3.989, p-value = 0.001, η = 0.548, $1 - \beta$ = 0.977) image groups. In both cases the dominance values were significantly higher than those provided by IAPS.

3.2 Different Classifier Results

Table 3 shows various statistical measures such as accuracy, precision, recall, F1-score and AUC for 6 classifiers. The results were obtained by running 10 rounds of stratified k-fold cross-validation with 5 splits. The data set was divided into 80% for training and validation and 20% for testing. Means and standard deviations are presented for each statistical measure.

Table 3. Mean and standard deviation for the different classifiers.

Classifier	Features	Accuracy	Precision	Recall	F1-Score	AUC
GNB	ROCKET	67.30 (5.37)	69.68 (4.58)	67.30 (5.37)	68.17 (4.60)	67.70 (7.80)
	Statistical	39.19 (16.98)	58.36 (7.32)	39.19 (16.98)	42.26 (15.66)	51.49 (7.01)
kNN	ROCKET	**76.49 (2.93)**	**71.09 (5.97)**	**76.49 (2.93)**	**71.40 (3.53)**	71.65 (7.23)
	Statistical	73.80 (1.50)	57.15 (2.85)	73.80 (1.50)	64.07 (1.57)	51.16 (5.45)
MLP	ROCKET	74.14 (3.18)	68.64 (4.47)	74.14 (3.18)	70.59 (3.34)	77.02 (5.78)
	Statistical	69.59 (3.22)	60.70 (4.63)	69.59 (3.22)	64.17 (3.12)	57.69 (6.37)
RF	ROCKET	76.23 (0.71)	58.12 (1.08)	76.23 (0.71)	65.95 (0.96)	68.80 (6.03)
	Statistical	74.78 (0.71)	55.93 (1.06)	74.78 (0.71)	63.99 (0.95)	59.15 (7.75)
SVM	ROCKET	76.23 (0.71)	58.12 (1.08)	76.23 (0.71)	65.95 (0.96)	**78.42 (5.78)**
	Statistical	74.78 (0.71)	55.93 (1.06)	74.78 (0.71)	63.99 (0.95)	59.37 (7.73)
XGB	ROCKET	76.35 (0.84)	59.25 (3.74)	76.35 (0.84)	66.27 (1.35)	70.39 (5.66)
	Statistical	**75.01 (1.44)**	**61.98 (6.84)**	**75.01 (1.44)**	**65.55 (2.18)**	**63.52 (7.19)**

Based on the results, the kNN classifier achieved the highest accuracy and F1 score mean values of 76.49% and 71.40% respectively. The MLP classifier had the second highest accuracy (74.14%) and F1 score (70.59%). The GNB classifier had the lowest accuracy and F1 score of all classifiers with a mean value of 67.30% and 68.17% respectively. For the AUC metric, the SVM classifier achieved the highest value with a mean of 78.42%. The MLP classifier achieved the second highest AUC value with a mean of 77.02%. The kNN classifier achieved AUC values of 71.65%. These percentages give a clear picture of the performance of each classifier. It is evident that the kNN classifier achieved the highest accuracy and F1-score values, indicating its effectiveness in accurately predicting the classification labels. The MLP classifier performed slightly worse than the kNN

classifier, but still achieved relatively high accuracy and F1-score values. The GNB classifier, on the other hand, had the lowest accuracy, F1-score and AUC values, indicating its limitations in accurately classifying the data points.

The AUC metric also provided valuable insight into the performance of each classifier. The SVM classifier achieved the highest AUC, indicating its ability to effectively discriminate between classes. The MLP classifier had the second highest AUC, indicating its usefulness in classifying the data. The GNB classifier had the lowest AUC, indicating its limitations in discriminating between classes.

Table 4 shows the percentage improvement for the classifiers when using ROCKET features instead of statistical features. Overall, the results suggest that the use of ROCKET features can lead to significant improvements in the performance of some classifiers, in particular GNB in terms of accuracy (71.75%) and F1-score (61.30%), and kNN and MLP in terms of AUC, 40.05% and 33.49% respectively. However, for other classifiers, such as RF and XGB, the improvements are relatively small or negative, indicating that the use of ROCKET features does not always lead to better performance.

Table 4. Improvement (%) in the use of ROCKET over statistical features.

Classifier	Accuracy	Precision	Recall	F1-Score	AUC
GNB	**71.75**	19.40	**71.75**	**61.30**	31.50
kNN	3.65	**24.40**	3.65	11.44	**40.05**
MLP	6.54	13.07	6.54	10.00	33.49
RF	1.94	3.91	1.94	3.06	16.31
SVM	1.94	3.91	1.94	3.06	32.09
XGB	1.78	−4.41	1.78	1.10	10.82

4 Conclusions

This paper has compared different methods for classifying pleasure and displeasure (high and low valence) using statistical and ROCKET feature extraction methods on fNIRS biosignals obtained from participants viewing emotional response images. The SAM scale was chosen to quantify the levels of arousal, valence and dominance as perceived by the participants and compared to the levels reported in the IAPS database. Results from the SAM scale indicate differences in arousal and dominance ratings for some groups of images, while no significant differences are reported for valence.

Based on the results obtained for the metrics presented, the study concluded that the ROCKET features were the most effective with the kNN, MLP and SVM classifiers in identifying pleasure and displeasure in fNIRS signals. The results highlight the importance of selecting the appropriate feature extraction

method and classifier to achieve accurate classification results. In addition, the study highlights the potential of different feature extraction methods to improve the accuracy of fNIRS biosignal analysis.

Acknowledgements. Grants PID2020-115220RB-C21 and EQC2019-006063-P funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way to make Europe". Grant 2022-GRIN-34436 funded by Universidad de Castilla-La Mancha and by "ERDF A way of making Europe". Grant PTA2019-016876-I funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future". This research was also supported by CIBERSAM, Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación.

References

1. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. **46**(3), 175–185 (1992)
2. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM (2016)
3. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
4. Dempster, A., Petitjean, F., Webb, G.I.: Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. arXiv:1910.13051 (2019)
5. Ekman, P.: Universal and cultural differences in facial expression of emotions, pp. 207–283. University of Nebraska Press (1972)
6. Fishburn, F.A., Ludlum, R.S., Vaidya, C.J., Medvedev, A.V.: Temporal derivative distribution repair (TDDR): a motion correction method for fNIRS. Neuroimage **184**, 171–179 (2019)
7. García-Pérez, E., Sánchez-Reolid, D., Sánchez-Reolid, R., Fernández-Caballero, A., Latorre, J.M., Borja, A.L.: Electroencephalographic signal processing from brain-computer-interface following image-based emotion induction. In: Julián, V., Carneiro, J., Alonso, R.S., Chamoso, P., Novais, P. (eds.) ISAmI 2022. LNCS, pp. 239–248. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-22356-3_23
8. Ho, T.K.: Random decision forests. In: 3rd International Conference on Document Analysis and Recognition, vol. 1, pp. 278–282. IEEE (1995)
9. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical report. A-8, University of Florida (2008)
10. Lang, P.J.: Behavioral Treatment and Bio-behavioral Assessment: Computer Applications. Ablex Publishing (1980)
11. Martínez-Rodrigo, A., García-Martínez, B., Alcaraz, R., González, P., Fernández-Caballero, A.: Multiscale entropy analysis for recognition of visually elicited negative stress from EEG recordings. Int. J. Neural Syst. **29**(2), 1850038 (2019)
12. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
13. Peirce, J., et al.: PsychoPy2: experiments in behavior made easy. Behav. Res. Methods **51**(1), 195–203 (2019)

14. Pollonini, L., Olds, C., Abaya, L., Bortfeld, H., Beauchamp, M.S., Oghalai, J.S.: Phoebe: a method for real time mapping of optodes-scalp coupling in functional near-infrared spectroscopy. Biomed. Opt. Express **7**(12), 5104–5119 (2016)
15. Rish, I.: An empirical study of the Naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3, pp. 41–46 (2001)
16. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986)
17. Russell, J.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
18. Sánchez-Reolid, R., et al.: Emotion classification from EEG with a low-cost BCI versus a high-end equipment. Int. J. Neural Syst. **32**(10), 2250041 (2022)
19. Sánchez-Reolid, R., López de la Rosa, F., Sánchez-Reolid, D., López, M.T., Fernández-Caballero, A.: Machine learning techniques for arousal classification from electrodermal activity: a systematic review. Sensors **22**(22) (2022)

Environmental Sound Recognition in Social Robotics

Sara Marques-Villarroya, Aythami Sosa-Aleman, Jose Carlos Castillo(✉), Marcos Maroto-Gómez, and Miguel Angel Salichs

University Carlos III, Madrid, Spain
{smarques,jocastil,salichs}@ing.uc3m.es

Abstract. An important task for a social robot is to understand the world around it. For this, sensory information is a key factor that allows identifying objects and people or listening to users and/or relevant sounds in the environment. This work focuses on the latter: the detection of everyday environmental sounds. We present a system able to recognise common sounds (e.g., air conditioning, car horn, water dripping) and how the integration in the social robot Mini allows for enhanced interaction with the users. We propose using deep learning techniques for sound identification, with the Mel-frequency spectrogram to represent the sound as an image, which allows using Convolutional Neural Networks to distinguish between sounds. The development was integrated into a real social robot to ensure the system's proper operation. The resulting functional system depends on specific settings implemented in the robot to distinguish between actual sounds and background noise.

Keywords: Social Robotics · Sound Recognition · Mel-frequency Spectrogram · Machine Learning

1 Introduction

Automatic environmental sound recognition has become an interesting research area in recent years and has demonstrated great potential in practical applications such as assistance for people with hearing disabilities [3] and safety in environmental surveillance and monitoring [2]. Social robots are intended to collaborate with humans in everyday environments. Equipping a robot with the ability to detect different sounds can therefore serve to improve interaction with people [1].

Mel-frecuency spectrogram and neural networks have become increasingly popular in this context due to their effectiveness in sound feature extraction and identification of complex sound patterns. In this line, using cochleagram images and Convolutional Neural Networks (CNN), Sharan et al. [10] proposed a novel method for recognising acoustic events. The authors extract features from the cochleagram using a CNN and classify the data using a softmax layer. Finally, the authors evaluate their method on two datasets and obtain competitive results

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 220–229, 2023.
https://doi.org/10.1007/978-3-031-43461-7_22

compared to state-of-the-art methods. Another example is the work developed by Pawar et al. [6]. In this case, the authors propose a CNN-based method, with Mel-frequency Cepstrum coefficients as input, for automatic emotion recognition from speech signals. The authors evaluated the proposed method on the German Emotional Speech Database (Emo-DB) dataset. The proposed method achieved an overall accuracy of 82.1%, significantly higher than the SVM- and DBN-based approaches. More recently, Khurana et al. [4] propose a tri-integrated CNN (Tri-CNN) for audio image classification using Mel-frequency spectrograms. They propose a three-stage CNN architecture that integrates three CNN types: spectrogram, Mel-frequency spectrogram, and joint CNNs. Using the three CNNs, the authors extract features from Mel-frequency spectrograms. They evaluate their method on two audio datasets and achieve cutting-edge results compared to other methods. The authors conclude that their Tri-CNN technique efficiently extracts discriminative features from Mel-frequency spectrograms for audio classification tasks.

Following the studies in the literature, this work aims at developing an environmental sound identification system that relies on computer vision and deep learning techniques (convolutional neural networks) to enhance the perceptual capabilities of a social robot. Thus, the detection system will be coupled with a novel skill to endow the robot with the ability of assisting in sound detection and identification tasks, with the potential of supporting people, mainly the elderly, who have lost hearing or speech abilities.

The rest of this paper is organised as follows. First, Sect. 2 explains how the CNN is trained using different publicly-available sound datasets. Next, Sect. 3 describes the main steps of the proposed pipeline: sound detection and identification. Then, Sect. 4 shows the evaluation of the detection system with different datasets, and Sect. 5 presents how the system is integrated into the social robot Mini. Finally, Sect. 6 highlights this work's main contributions.

2 Offline Training

Before developing the online system, integrated with the robot, we trained the selected neural network using three different datasets to assess its performance. In this work, we use the temporal-frequency attention-based convolutional neural network (TF-CNN) presented by Mu et al. [5]. This network consists of four $2D$ convolutional layers, each with 32 4×10 filters, followed by a Batch Normalization layer, a ReLu activation function, and a $2D$ Max Pooling layer with 2×2 filters. It continues with a Flatten layer, another layer of 64 neurons with dense connections, a Batch Normalisation layer, and the ReLu activation function. Finally, the output layer has the same number of neurons as the number of classes in each model, and its activation function is softmax.

We trained three TF-CNN models using three datasets. The first model used UrbanSound8K dataset [8]. This dataset includes ten categories: air conditioner, car horn, children playing, dog bark, drilling, engine idling, siren, street music, jackhammer, and gunshot. The dataset contains 8732 short audio clips (no longer



Fig. 1. Offline training pipeline.

than four seconds). All classes in the dataset have 1000 instances except for "car horn" (429) and "gunshot" (1327). Since the original audio was recorded at different sampling rates, it was necessary to establish a uniform sampling rate, downsampling the sounds captured at a high rate to match the ones with lower rates. The second model was trained using the ESC-50 dataset [7], which retrieves audio from YouTube videos directly. This dataset contains 50 distinct categories of audio data, primarily animals, natural soundscapes, water sounds, human non-speech sounds, internal/home sounds, and external/urban sounds. Each category contains forty-five-second-long audio files, with 40 examples per class, for 2, 8 hours. As this work is intended to be integrated into a social robot for the elderly at home, the latter model was trained using a reduced version of the ESC-50 dataset containing sounds considered common in a domestic environment. After class filtering, we used 24 classes to train the model.

Since in this work we use techniques originally proposed to classify images, the steps to follow will be similar, with the difference that at the beginning we have to transform sound information into visual information. Therefore, the system creates first an image of Mel's spectrogram based on the received sound data. This information is fed into the machine learning model (Fig. 1).

3 Real-Time Sound Detection and Identification

This section describes the detection and identification system proposed in this work when acquiring sounds in real-time. Figure 2 depicts the pipeline followed by the system. First, the sound detection process filters the input information to distinguish between background noise and actual sounds. Then, the system uses the Mel-spectrogram to characterise the sound. Finally, it sends the spectrogram to the classifier for recognition.

3.1 Sound Detection

Sound detection is one of the system's main functions since ambient noise has an impact on the performance of the classifier. The system establishes a window size of 512 samples every half a second and concatenates two of these windows to create an instance. Next, with the instance information, the system calculates the signal-to-noise ratio (SNR), which will serve as a baseline to distinguish between background noise and actual sounds. The SNR indicates the signal-to-noise ratio and informs the system when to begin collecting sound information. The SNR could be defined in terms of the Root Mean Square (RMS) power of the signal (see Eq. 1).

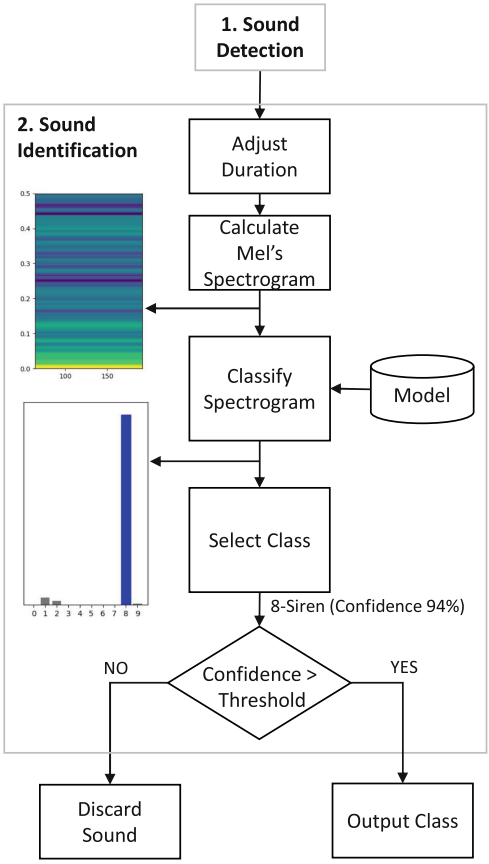


Fig. 2. Complete sound detection and identification real-time pipeline.

$$SNR = \frac{Power_{signal}}{Power_{noise}} = (\frac{Amplitude_{Signal(RMS)}}{Amplitude_{Noise(RMS)}})^2 \quad (1)$$

With this value, the system uses a threshold (SNR_τ) to determine when the sound begins to avoid false positives. In this line, if the ambient SNR ratio is high, it indicates loud ambient noise, so the threshold would also be high. The threshold can have three values to simplify the operation (high, medium and low), and these threshold values have been determined through experimentation in controlled environments.

The first step is to calculate the SNR_{window} by measuring the RMS of the signal and of the ambient noise (see Algorithm 1). As this value can change, the values are recalculated and adjusted on each window time. Next, we have to determine when the system considers the sound complete. Once the SNR threshold (SNR_τ) is exceeded, the system will store 512 samples every half second. Once the samples have been collected, the system determines if they

Algorithm 1. Sound detection process with SNR calculation

$silence_frames \leftarrow 1$
$silence_frames_{\max} \equiv 1$ seconds
$collecting_sound \leftarrow$ False
$sound_finished \leftarrow$ False
while True **do**
  $window \leftarrow microphone_{signal}$
  $A_{window} \leftarrow RMS(window)$
  $SNR_{window} \leftarrow (A_{window}/A_{background})^2$
  $SNR_\tau = \begin{cases} High & if\ SNR_{window} \leftarrow high \\ Medium & if\ SNR_{window} \leftarrow medium \\ Low & if\ SNR_{window} \leftarrow low \end{cases}$
  while $SNR_{window} > SNR_\tau$ **do**
    $collecting_sound \leftarrow$ True
    $A_{window} \leftarrow RMS(window)$
    $SNR_{window} \leftarrow (A_{window}/A_{background})^2$
  end while
  if low noise **then**
    $silence_frames \leftarrow silence_frames + 1$
    if $silence_frames = silence_frames_{max}$ **then**
      $silence_frames \leftarrow 1$
      $sound_finished \leftarrow$ True
    end if
  end if
end while

surpass the SNR_τ. If they do, the samples are stored, and acquisition continues. When the sound intensity drops below the threshold, the system initiates a $silence_frame$, some extra time before closing the instance to allow detection of sounds with low-intensity parts or silences. If the system does not detect any sound above SNR_τ before the $silence_frame$ expires ($silence_frames_{max}$), the sound is considered to be finished. After testing silent times ranging from 0.5 to 2 s, we settled on a 1-s timeout, which allows us to neither cut off sounds due to brief silence nor store irrelevant information.

3.2 Sound Identification

After the sound has been acquired, the system proceeds to the recognition stage. For sound identification, we use the three machine learning models trained as explained in Sect. 2.

In the real-time version of the system, first, we need to adjust the sound duration to the three-seconds interval used in the training phase to improve recognition. If the sound is longer than that interval, it is trimmed. If the sound lasts less than three seconds, it is filled with a null wave (silence).

After resizing the sound, the system calculates Mel's spectrogram to obtain the recognition features and starts the classification process using the pre-trained models. This classification returns the probability of recognition for each of the

Table 1. F1-score value for the recognition of each class of Urban-Sound8k dataset.

Class	F1-score	Class	F1-score
Car horn	0.94	Drill	0.86
Air conditioning	0.93	Jack-hammer	0.81
Gunshot	0.91	Children playing	0.76
Engine	0.91	Dog	0.76
Siren	0.88	Street music	0.71

model's classes. The system chooses the class with the highest probability. If the confidence value for that class is greater than 75% (*Threshold* value in Fig. 2), the system sends the output class. Otherwise, it will be deemed a false positive.

4 Evaluation of the Sound Detection System

This section will discuss the training results and compare the three trained models using the same neural network. Table 1 displays the F1-score calculated from the UrbanSound8k dataset's data. With this model, we aimed to examine the correct operation of the system and confirm that the selected neural network functioned adequately. As can be seen, F1-score ranges between 0.71 and 0.94, which, considering that humans recognise approximately 81% of the sounds in their environment [7], seems to indicate that the proposed system can achieve human-like performance on this reduced dataset.

Next, we tested the system with a more challenging dataset, the ESC-50. Table 2 displays the obtained results that, in this case, offer lower F1-scores for certain sounds, such as *washing machine*, *engine*, and *breathing*. However, other sounds, such as sneezing, digging, and cricket, obtain values greater than 81%. It is worth mentioning that this dataset had 40 instances per class, which affected the generalisation of the CNN in the training process.

Finally, Table 3 presents the results of the reduced version of the ESC-50 dataset. Again, as a result of reducing the number of classes, we achieved better training outcomes than with the last model trained. In this case, we obtain favourable results for most classes; however, for classes such as *laughter* and *siren*, additional samples are required to increase the obtained accuracy value.

4.1 Comparison Between Models

After analysing the training results, they provide valuable insights into the effectiveness of the trained models for sound classification. The first model, trained on the UrbanSound8K dataset, showed good results for all sound classes, with the exception of children playing and street music, which had a higher probability of failure. The second model, trained on a complete dataset of everyday sounds, also showed promising accuracy, but some sound classes had lower precision and

Table 2. F1-score value for the recognition of each class of the complete ESC-50 dataset.

Class	F1-Score	Class	F1-Score	Class	F1-Score
Crackling fire	0.92	Wind	0.64	Helicopter	0.48
Cow	0.90	Hand saw	0.63	Snoring	0.46
Crickets	0.86	Clock tick	0.59	Drinking sipping	0.46
Crow	0.85	Airplane	0.58	Brushing teeth	0.43
Mouse click	0.83	Laughing	0.57	Sheep	0.40
Sneezing	0.80	Hen	0.56	Crying baby	0.40
Toilet Flush	0.80	Siren	0.55	Sea waves	0.40
Cat	0.75	Footsteps	0.53	Snapping	0.40
Train	0.72	Chirping birds	0.52	Pouring water	0.38
Can opening	0.72	Dog	0.50	Car horn	0.36
Insects	0.71	Coughing	0.50	Rain	0.36
Door wood knock	0.71	Keyboard tipping	0.50	Washing machine	0.33
Glass breaking	0.71	Fireworks	0.50	Pig	0.33
Frog	0.67	Thunderstorm	0.50	Chainsaw	0.29
Rooster	0.67	Clock alarm	0.50	Breathing	0.22
Door wood creaks	0.67	Church bells	0.50	Engine	0.21
Vacuum cleaner	0.66	Water drops	0.50		

Table 3. F1-score value for the recognition of each class of the reduced ESC-50 dataset.

Class	F1-Score	Class	F1-Score	Class	F1-Score
Door wood creaks	1.00	Rain	0.83	Can opening	0.67
Insects	1.00	Water drops	0.80	Fireworks	0.67
Clock alarm	0.92	Church bells	0.80	Washing machine	0.63
Pouring water	0.90	Snoring	0.76	Crying baby	0.62
Dog	0.88	Toilet flush	0.75	Siren	0.59
Thunderstorm	0.88	Cat	0.75	Chirping birds	0.50
Snapping	0.87	Coughing	0.70	Laughing	0.46
Footsteps	0.86	Mouse click	0.67		

sensitivity, indicating that further refinement may be necessary. Despite the limitations of the dataset, the models were able to accurately classify many of the sound classes, with some classes having no errors at all. However, there were some sound classes where the models were less certain, such as *mouse clicks* and *laughter*. These results suggest that the models may need further refinement to improve their accuracy for these classes. Overall, the results of this study demonstrate the potential of using machine learning models for sound classification in robotics applications. However, further research is needed to optimize the models for specific sound classes and to evaluate their performance in real-world scenarios.



Fig. 3. Social Robot Mini

Comparing the results of the various models, we can conclude that the recognition rate increases as the number of classes the system must differentiate decreases. It indicates that the system achieves better results with the UrbanSound8K sound model than the ESC-50 model. Furthermore, the distinction between sounds supports this result: as there are fewer distinct classes, the difference between them is more significant, and as a result, we obtain more accurate recognition data.

Finally, for integration with the robot, we selected the third model as the most balanced because the hit rates are high (over 59% in all classes), enabling the robot to respond to different home sounds. Therefore, the social robot was tested using the latter model.

5 Integration in the Social Robot

This section explains how the social robot Mini responds to the stimuli detected by the identification pipeline described in the previous section. The system attempts to provide a human-like response to achieve a natural interaction.

5.1 The Social Robot Mini

In this work, we used the social robot Mini [9] developed at the Roboticslab of the Universidad Carlos III de Madrid (see Fig. 3). Mini is designed to interact with elderly people with cognitive impairment. For cognitive stimulation, it has skills such as dancing, conversing, and playing games. Furthermore, Mini integrates LEDs on its cheeks, heart, and mouth and 5 degrees of freedom to move its head, neck, base, and arms to enhance its expressiveness. In addition, this robot includes touch sensors in its shoulders and belly, a speaker for voice interaction with the user and an external tablet for displaying multimedia content such as videos, images, and menus with which the user can interact during games. Lastly, the robot Mini is equipped with an air microphone to pick up sounds in front

of the robot, particularly voice interactions with the user, and a head-mounted omnidirectional microphone to pick up sounds from any direction and locate them.

5.2 Detector's Integration in Social Robot Mini

Once the robot detects an external sound that exceeds the SNR threshold, the omnidirectional microphone captures it for recognition using the acoustic detection system presented in Sect. 3.1.

Next, the system performs the sound identification process using the trained model. If any of the probabilities obtained by the recognition exceeds the threshold (classification score above 75%), the system sends the detected sound information to the highest level of the robot. Based on this data, the robot responds to the detected sound. In this version, the robot responds to the detected sounds with simple movements (movement of the robot's arms and base) and output utterances referring to the sounds. For instance, if the robot detects a knock at the door, it will raise its arms and announce: *It appears that someone has arrived*. We have compiled a video[1]demonstrating the actual operation of the social robot with the integrated detector during user interaction.

6 Conclusions

The paper describes a sound detector that can capture and transmit the entire sound for classification. The system uses the Mel-frecuency spectrogram as a sound feature, to test and validate a machine learning model using three datasets. The F1-score of the model trained with the UrbanSound8K dataset is 0.85 on average for all sound classes. In the same line, in the model trained with the entire dataset of everyday sounds, sound classes have an average F1-Score of 0.57, which, considering the reduced number of instances in the datases, shows that with more training instances, the system could achieve competitive results. Finally, in the model trained with the reduced everyday sounds dataset has an F1-Score of 0.75 for sound classes. Although some sounds may be more difficult to identify than others, these results indicate that the sound identification system performs well overall. Moreover, it is observed that the model trained with a smaller dataset achieves better results than the model trained with the entire dataset, suggesting that a smaller number of sounds can improve the accuracy of the system. In addition, a robot skill was developed to test system integration, in which the robot speaks or acts based on the detected sound. Multiple tests with the robot were conducted, and although sound detection accuracy was not always accurate, the system avoided reactions to unwanted sounds. Finally, from the tests, it was noted that the SNR and class certainty thresholds are crucial to the system's operation and that the training model impacts the system's accuracy.

[1] Working example video: https://youtu.be/2mO5241CKMU.

In future research, we will create a more complex neural network to improve the system's accuracy in sound identification. This work will also enable the development of new skills for the robot that make use of the sound perception capabilities presented in this paper. For example, new games with sound interaction could be integrated, as well as the possibility of assisting people with hearing impairments.

Acknowledgements. The research leading to these results has received funding from the projects: Robots sociales para mitigar la soledad y el aislamiento en mayores (SOROLI), PID2021-123941OA-I00, funded by Agencia Estatal de Investigación (AEI), Spanish Ministerio de Ciencia e Innovación. Robots sociales para reducir la brecha digital de las personas mayores (SoRoGap), TED2021-132079B-I00, funded by Agencia Estatal de Investigación (AEI), Spanish Ministerio de Ciencia e Innovación. This publication is part of the R&D&I project PDC2022-133518-I00, funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

References

1. Alonso-Martin, F., Gamboa-Montero, J.J., Castillo, J.C., Castro-Gonzalez, A., Salichs, M.A.: Detecting and classifying human touches in a social robot through acoustic sensing and machine learning. Sensors **17**(5) (2017). https://doi.org/10.3390/s17051138, https://www.mdpi.com/1424-8220/17/5/1138
2. Chandrakala, S., Jayalakshmi, S.: Environmental audio scene and sound event recognition for autonomous surveillance: a survey and comparative studies. ACM Comput. Surv. (CSUR) **52**(3), 1–34 (2019)
3. Jain, D., et al.: Homesound: an iterative field deployment of an in-home sound awareness system for deaf or hard of hearing users. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2020)
4. Khurana, A., Mittal, S., Kumar, D., Gupta, S., Gupta, A.: Tri-integrated convolutional neural network for audio image classification using Mel-frequency spectrograms. Multimed. Tools Appl. **82**(4), 5521–5546 (2023)
5. Mu, W., Yin, B., Huang, X., Xu, J., Du, Z.: Environmental sound classification using temporal-frequency attention based convolutional neural network. Sci. Rep. **11**(1), 21552 (2021)
6. Pawar, M.D., Kokate, R.D.: Convolution neural network based automatic speech emotion recognition using Mel-frequency Cepstrum coefficients. Multimed. Tools Appl. **80**, 15563–15587 (2021)
7. Piczak, K.J.: ESC: dataset for environmental sound classification. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 1015–1018 (2015)
8. Salamon, J., Jacoby, C., Bello, J.P.: A dataset and taxonomy for urban sound research. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 1041–1044 (2014)
9. Salichs, M.A., et al.: Mini: a new social robot for the elderly. Int. J. Soc. Robot. **12**, 1231–1249 (2020)
10. Sharan, R.V., Moir, T.J.: Acoustic event recognition using cochleagram image and convolutional neural networks. Appl. Acoust. **148**, 62–66 (2019)



Intelligent Systems in Healthcare: An Architecture Proposal

António Chaves[1], Larissa Montenegro[1], Hugo Peixoto[1], António Abelha[1], Luís Gomes[2], and José Machado[1(✉)]

[1] ALGORITMI/LASI, University of Minho, Braga, Portugal
{antonio.chaves,larissa.montenegro}@algoritmi.uminho.pt, {hpeixoto,abelha,jmac}@di.uminho.pt
[2] ALGORITMI/LASI, University of Azores, Ponta Delgada, Portugal
luis.mp.gomes@uac.pt

Abstract. Multi-Agent Systems has existed for decades and has focused on principles such as loose coupling, distribution, reactivity, and local state. Despite substantial tool and programming language research and development, industry adoption of these systems has been restricted, particularly in the healthcare arena. Artificial intelligence, on the other hand, entails developing computer systems that can execute tasks that normally require human intelligence, such as decision-making, problem-solving, and learning. The goal of this article is to develop and implement an architecture that includes multi-agent systems with microservices, leveraging the capabilities of both methodologies in order to harness the power of Artificial Intelligence in the healthcare industry. It assesses the proposed architecture's merits and downsides, as well as its relevance to various healthcare use cases and the influence on system scalability, adaptability, and maintainability. Indeed, the proposed architecture is capable of meeting the objectives while maintaining scalability, flexibility, and adaptability.

Keywords: Artificial Intelligence · Multi-Agent Systems · Microservices · Intelligent Healthcare

1 Introduction

Artificial Intelligence (AI) has been making significant contributions to the healthcare sector, especially in hospitals, where the complexity of healthcare systems necessitates advanced technologies to enhance patient outcomes. Multi-Agent Systems (MAS) are one of the AI technologies that are being utilized to support architecture in the instantiation of AI applications in hospitals. The inclusion of AI applications in this environment requires a solid structural backbone [1]. In this research, we aim to explore two of the main frameworks for the deployment of such applications, their comparison and how they can work in together to achieve scalability, modularity and robustness.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 230–238, 2023.
https://doi.org/10.1007/978-3-031-43461-7_23

As a research field, MAS, has existed since the late 1980s. Since its inception, notions such as loose-coupling, distribution, reactivity and isolated (local) state have been core concepts [2]. While the notion of an agent has been widely researched and a large range of tools and programming languages developed, there has been little real adoption of these languages and tools within industry. AI refers to the development of computer systems that can perform tasks that usually require human intelligence, such as problem-solving, decision-making, and learning. These systems can be trained using machine learning algorithms that enable them to recognize patterns in data and make predictions or decisions based on that information.

Microservice architecture, on the other hand, is a software development approach that focuses on breaking down complex applications into smaller, modular services that can be developed, deployed, and scaled independently [3]. Each service runs its own processes and communicates with other services via lightweight protocols, allowing for greater flexibility, agility, and scalability [4]. Overall, these three concepts - AI, microservice architecture, and multi-agent systems - are all essential components of modern software development and have the potential to transform the healthcare industry by improving patient outcomes, reducing costs, and increasing efficiency.

In this paper, we will explore the use of microservice architecture and multi-agent systems for implementing AI in the healthcare field. Specifically, we will examine the benefits and drawbacks of the combined approach, its applicability to different healthcare use cases, and their impact on system scalability, flexibility, and maintainability. By analyzing these factors, we aim to provide insights into the usability of the now called Multi-Agent Microservices and whether its application in the healthcare sector can be of use.

2 Background

The goal of the background material presented here is to illustrate the fundamental ideas behind AI, Microservice architecture, and MAS, as well as how contemporary research has been able to use these ideas in both theoretical and practical contexts. AI has received a lot of interest recently in the healthcare sector due to its potential to enhance patient outcomes, save costs, and boost productivity. There are numerous methods for putting AI into practice as it develops, including multi-agent systems and microservice architecture.

2.1 Microservice Architecture

Smaller, modular services that can be created, deployed, and scaled independently are what is meant by the term “microservice architecture” in the context of software development. Greater flexibility, agility, and scalability are possible thanks to the fact that each microservice manages its own processes and communicates with other microservices using simple protocols. Building an application

as a group of loosely coupled, independently deployable services is the fundamental concept behind microservices. This makes it simpler to develop, test, and deploy new features because developers can work on each service independently without affecting the rest of the application [5].

Greater flexibility and scalability are two of microservices' main advantages. It is simpler to add new functionality to the application without affecting the rest of the system because each microservice can be developed and deployed independently. Greater fault-tolerance is also made possible because a failure in one microservice does not always have an impact on the entire system. Additionally, microservices can be scaled horizontally by adding more instances of a specific service, making it easier to handle high traffic loads [6].

Microservice architecture does, however, present some difficulties. More coordination and communication between microservices are needed, as well as more infrastructure for fault tolerance, load balancing and service discovery. Additionally, it calls for a higher level of technical proficiency from developers, who must be knowledgeable of related technologies such as distributed systems, message queues, and others.

Generally speaking, microservice architecture offers a more modular and adaptable method of developing software, enabling greater agility and scalability. Its suitability depends on the particular use case and project requirements, but it has become increasingly well-liked in recent years as a means of developing large-scale, complex applications [7].

2.2 Multi-Agent Systems

MAS are a class of AI systems that employs a number of intelligent agents working together to accomplish a single objective [8]. Each agent is a software entity with the ability to perceive its surroundings, think through its actions, and interact with other agents to plan out their actions [9]. The fundamental idea behind MAS is to represent complex systems as a group of autonomous agents that communicate with one another and their surroundings. In the creation of a multi-agent system, a certain group of high level abstractions should be successfully utilized. For example, agents are autonomous entities with independent spheres of control that are positioned in environments and engage in social interaction. These agents are capable of observation and utilisation of their surroundings and resources. The interactions between agents and other resources are defined functionalities which can be limited and further defined by organizational rules [10].

There are, however, some difficulties with MAS as well. When there are competing objectives or worldviews, it can be complicated and challenging to coordinate the actions of several actors. Additionally, developers need to have a greater degree of competence because MAS demand them to be knowledgeable about technologies like distributed systems, agent-based modeling, and other relevant fields [11].

Overall, MAS provide an effective method for simulating complex systems and enhancing their behavior. Its appropriateness depends on the particular use

case and project objectives, but it has drawn a lot of attention recently as a way to boost coordination and efficiency across a number of industries [12].

2.3 Artificial Intelligence in Healthcare

Artificial intelligence is a sub-field of computer science that focuses on the creation of intelligent robots capable of conducting tasks that normally require human intellect, such as sensing, reasoning, learning, decision making, and natural language processing [13]. Large volumes of data inputs and outputs are used by these algorithms to identify patterns and effectively "learn" in order to teach the computer to make autonomous suggestions or conclusions [14]. Overall, artificial intelligence has the potential to alter numerous sectors, including healthcare, by giving intelligent assistance with decision making, diagnosis, and treatment.

Examples of intelligent AI applications in healthcare span across different fields. [15] reviewed different Machine Learning algorithms applied for the diagnosis of chronic diseases from 2015 until 2019. While the issue does not target a specific type or group of chronic diseases, researchers were able to deduce the best models for prediction over an entire subset of disorders.

Another interest of AI in healthcare is its application in Medical Imaging processing and symptom detection. Both [16] and [17] proposed Deep Learning frameworks for the prediction of and analysis of different medical images.

3 Proposed Architecture

Multi-agent systems involve the development of intelligent software agents that can interact with each other to achieve a common goal [18]. Each agent is capable of autonomous decision-making and can communicate with other agents to exchange information and coordinate actions [19]. Microservices, on the other hand, involve breaking down a large application into smaller, modular services that can be developed and deployed independently. Each microservice focuses on a specific functionality, and communicates with other microservices through standardized interfaces. Modern large-scale software development practices use microservices. They have clearly proven their potential as a tool for developing systems at scale, despite not being an all-encompassing answer [20]. The case study presented hereby, relies on two different proposed applications which employ AI models in order to achieve their goals. Figure 1 aims to explain how different context specific applications can make use of their own structure of a group of microservices which activate intelligent agents in order to preform their specific tasks.

By combining these two approaches, it is possible to develop intelligent software systems that consist of multiple autonomous agents that interact with each other through well-defined APIs. Each agent can represent a specific functionality or service, and can communicate with other agents to exchange information and coordinate actions [6]. This architecture allows for the exposure of Intelligent Agents' capabilities through RESTful APIs, with an impact on how MAS

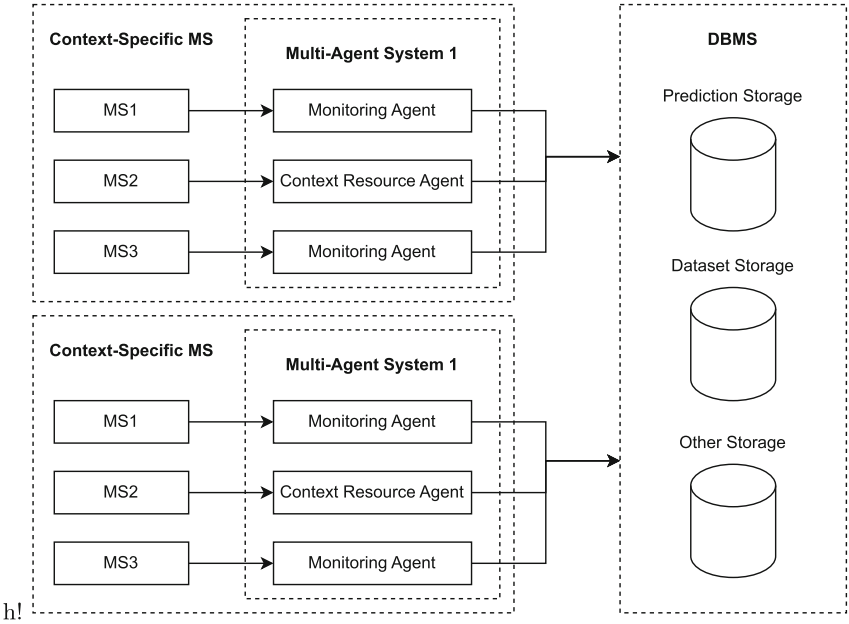


Fig. 1. Proposed architecture comprised of Microservices and Multi-Agent Systems

are viewed and implemented, making it easier to incorporate agile development techniques and modify how agents are accessed and utilized, resulting in systems that are quicker and more intelligent [21].

The sequence diagram in Fig. 2, an explanation is drawn on how the exposure of the intelligent agents' capabilities through the microservice architecture can aid in scabalility, responsiveness and overall better organization in the services provided. In the diagram, the User, through a provided Graphical User Interface can request different actions in the same application context through different microservices, thus reducing server overload and obtaining better performance. On another note, this architecture can help with better load-balancing on the more utlized microservices without the need for the entire system to be replicated and run in parallel.

Overall, the combination of multi-agent systems and microservices can lead to more intelligent and modular software applications that are better suited to handle complex and dynamic environments.

4 Discussion

The fusion between MAS and Microservices has been growing in interest among the academic community, with different applications in mind [22]. MAMS is a term that arises from the belief that organizations combine outward-facing

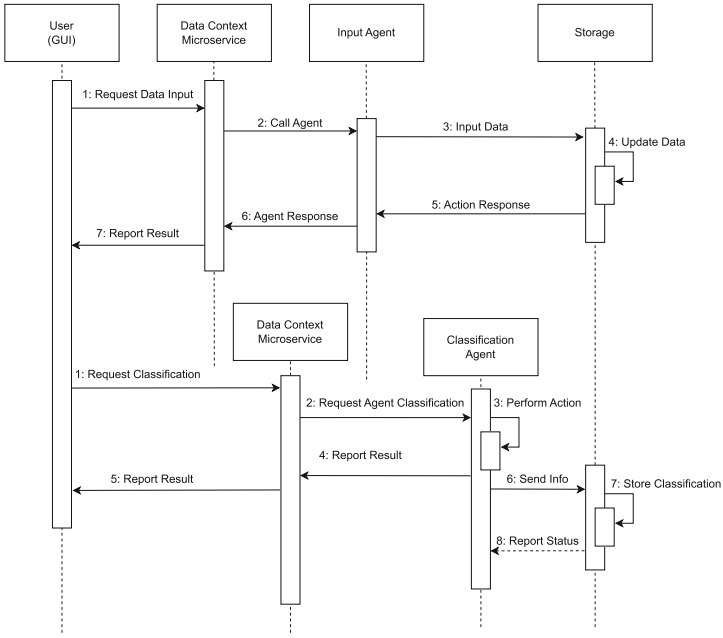


Fig. 2. Sequence Diagram for the Architecture

customer service agents while concealing internal worker agents. The most significant benefit of this technology is that it is now feasible to create packaged components that use agent-level reasoning as common practice [23]. The architecture allows for the deployment of decentralized Multi-Agent Systems, while also enabling easier exposure to the Intelligent Agents' actions.

In order to analyze the prospected architecture, some pros and cons of its utilization need to be taken into consideration:

- In terms of scalability, the microservice architecture offers the possibility to independently boost a desired service's responsiveness, through replication or changes in memory allocation.
- The architecture can leverage both MAS and MS for flexibility and modularity. Agent-Service combinations can easily be altered without affecting the rest of the system. Also, MAS stands out for its capacity in modeling complex systems and the integration or modification of Microservices and Agents can be performed seamlessly [24].
- Decentralizing services facilitates the fault tolerance and high availability. With the autonomous nature of MAS, the coupling of MS and IA allows its independent work without the need for centralized control in distribution.

It is also important to review the drawbacks of implementing the proposed architecture. These are more development centered but still impose some difficulties to deployment.

- Generally speaking, the MS architecture relies on additional resources to run and this may affect the need for scalability options if/when resources are limited.
- Both MAS and MS are applied in complex systems and the modularity should require additional security constraints to ensure system access control.
- It is crucial to give the Healthcare Information System architecture the ability to support communication across various, otherwise isolated systems, avoiding their centralization and dispersion of End-to-End connections [25]. The integration of MS relies on stability of its communication and interoperation protocols.

5 Conclusions

In this paper, we discuss the importance and applicability of the Microservice Architecture, and the possibility of combination with Multi Agent Systems for the deployment of intelligent applications in the healthcare scope.

A Multi-Agent Microservice architecture proves to be an appropriate methodology for developing intelligent healthcare systems. We may achieve great scalability, flexibility, and modularity by breaking down the system into separate agents, each accountable for a certain job. Furthermore, by combining this architecture with various AI models, we can create a system that can learn from data, reason, and make intelligent decisions. However, there are some drawbacks to using this architecture, such as complexity, integration, overhead and security. As a result, before using this design in a healthcare system, it is critical to carefully weigh the benefits and cons. Overall, we think that a Multi-Agent Microservices architecture has the potential to transform healthcare by delivering tailored and intelligent healthcare services. Additionally, the importance of a solid architecture when combined with rapidly growing AI models will allow for faster and more reliable versioning of production ready applications, without compromise in uptime. The proposed Multi-Agent Microservice architecture will also benefit from intrinsic decentralized architecture advantages, especially in terms of application access in downtime situations.

Acknowledgements. 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. Cristani, M., Pasetto, L., Tomazzoli, C.: Protecting the environment: a multi-agent approach to environmental monitoring. Proc. Des. Soc. **1**, 161–170 (2021)
2. Benhajji, N., Roy, D., Anciaux, D.: Patient-centered multi agent system for health care. IFAC-PapersOnLine **48**(3) (2015)
3. Munaf, R.M., Ahmed, J., Khakwani, F., Rana, T.: Microservices architecture: challenges and proposed conceptual design. In: 2019 International Conference on Communication Technologies (ComTech), Rawalpindi, Pakistan, 2019, pp. 82–87 (2019). https://doi.org/10.1109/COMTECH.2019.8737831

4. Francesco, P.D., Malavolta, I., Lago, P.: Research on architecting microservices: trends, focus, and potential for industrial adoption. In: 2017 IEEE International Conference on Software Architecture (ICSA), Gothenburg, Sweden, 2017, pp. 21–30 (2017). https://doi.org/10.1109/ICSA.2017.24
5. Villamizar, M., et al.: Evaluating the monolithic and the microservice architecture pattern to deploy web applications in the cloud. In: 2015 10th Computing Colombian Conference (10CCC), Bogota, Colombia, 2015, pp. 583–590 (2015). https://doi.org/10.1109/ColumbianCC.2015.7333476
6. Heorhiadi, V., Rajagopalan, S., Jamjoom, H., Reiter, M.K., Sekar, V.: Gremlin: systematic resilience testing of microservices. In: 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS), Nara, Japan, 2016, pp. 57–66 (2016). https://doi.org/10.1109/ICDCS.2016.11
7. Kang, H., Le, M., Tao, S.: Container and microservice driven design for cloud infrastructure DevOps. In: 2016 IEEE International Conference on Cloud Engineering (IC2E), Berlin, Germany, 2016, pp. 202–211 (2016). https://doi.org/10.1109/IC2E.2016.26
8. Agrawal, A., Won, S., Sharma, T., Deshpande, M., Mccomb, C.: A multi-agent reinforcement learning framework for intelligent manufacturing with autonomous mobile robots
9. Radisic-aberger, O., Weisser, T., Sabmannshausen, T., Wagner, J., Burggraf, P.: Concept of a multi-agent system for optimised and automated engineering change implementation. Proc. Des. Soc. **2**, 1689–1698 (2022)
10. Cocho-bermejo, A., Navarro-mateu, D.: User-centered responsive sunlight reorientation system based on multiagent decision-making, UDaMaS. In: Proceedings of the International Conference on Education and Research in Computer Aided Architectural Design in Europe, vol. 2, pp. 695–704 (2019)
11. Calvaresi, D., Albanese, G., Calbimonte, J.-P., Schumacher, M.: SEAMLESS: simulation and analysis for multi-agent system in time-constrained environments. In: Demazeau, Y., Holvoet, T., Corchado, J.M., Costantini, S. (eds.) PAAMS 2020. LNCS (LNAI), vol. 12092, pp. 392–397. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49778-1_30
12. Rehman, H.U., et al.: Cloud based decision making for multi-agent production systems. In: Marreiros, G., Melo, F.S., Lau, N., Lopes Cardoso, H., Reis, L.P. (eds.) EPIA 2021. LNCS (LNAI), vol. 12981, pp. 673–686. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86230-5_53
13. Simmons, A.B., Chappell, S.G.: Artificial intelligence-definition and practice. IEEE J. Oceanic Eng. **13**(2), 14–42 (1988). https://doi.org/10.1109/48.551
14. Helm, J.M., Swiergosz, A.M., Haeberle, H.S., et al.: Machine learning and artificial intelligence: definitions, applications, and future directions. Curr. Rev. Musculoskelet. Med. **13**, 69–76 (2020). https://doi.org/10.1007/s12178-020-09600-8
15. Battineni, G., Sagaro, G.G., Chinatalapudi, N., Amenta, F.: Applications of machine learning predictive models in the chronic disease diagnosis. J. Pers. Med. **10**, 21 (2020). https://doi.org/10.3390/jpm10020021
16. Kowsari, K., et al.: HMIC: hierarchical medical image classification, a deep learning approach. Inf. **11**, 318 (2020). https://doi.org/10.3390/info11060318
17. Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., Soufi, G.: Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. Med. Image Anal. **65**, 101794 (2020). ISSN 1361-8415, https://doi.org/10.1016/j.media.2020.101794, https://www.sciencedirect.com/science/article/pii/S1361841520301584
18. Shakshuki, E., Reid, M.: Multi-agent system applications in healthcare: current technology and future roadmap. Procedia Comput. Sci. **52** (2015)

19. Cardoso, L., Marins, F., Portela, F., Santos, M., Abelha, A., Machado, J.: A multi-agent platform for hospital interoperability. In: Ramos, C., Novais, P., Nihan, C.E., Corchado Rodríguez, J.M. (eds.) Ambient Intelligence - Software and Applications. AISC, vol. 291, pp. 127–134. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07596-9_14
20. Collier, R., O'Neill, E., Lillis, D., O'Hare, G.: MAMS: Multi-Agent MicroServices, pp. 655–662 (2019). https://doi.org/10.1145/3308560.3316509
21. Carneiro, J., Alves, P., Marreiros, G., Novais, P.: A multi-agent system framework for dialogue games in the group decision-making context. Adv. Intell. Syst. Comput. **930**, 437–447 (2019)
22. Zouad, S., Boufaida, M.: Using multi-agent microservices for a better dynamic composition of semantic web services. In: Proceedings of the 4th International Conference on Advances in Artificial Intelligence (ICAAI 20), pp. 47–52. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3441417.3441423
23. O'Neill, E., Lillis, D., O'Hare, G.M.P., Collier, R.W.: Explicit modelling of resources for multi-agent microservices using the cartago framework. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '20). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp. 1957–1959 (2020)
24. Pathirana, S., Asirvatham, D., M'djohar, M.: Applicability of multi-agent systems for electroencephalographic data classification. Procedia Comput. Sci. **152**, 36–43 (2019)
25. Miranda, M., et al.: Multi-agent systems for HL7 interoperability services. Procedia Technol. **5**, 725–733 (2012). https://doi.org/10.1016/j.protcy.2012.09.080

ISAMI'23 - DOCTORAL CONSORTIUM

Teaching Innovation in Higher Education Through Virtual Reality

Fernando Lara-Lara[1], Juan-Carlos de la Cruz-Campos[2(✉)], Magdalena Ramos-Navas-Parejo[2], and Blanca Berral-Ortiz[1]

[1] Department of Didactics and School Organization, Faculty of Education Sciences, Campus Universitario de Cartuja, University of Granada, 18071 Granada, Spain
blancaberral@ugr.es

[2] Department of Didactics and School Organization, Faculty of Education and Sport Sciences of Melilla, Campus Universitario de Melilla, University of Granada, 52071 Melilla, Spain
juancarlosdelacruz@ugr.es

Abstract. In the knowledge society in which we find ourselves, traditional expository classes in which students played a passive role have been shown not to meet students' expectations and do not respond to educational demands. Technology has gradually been incorporated into all areas of life to the point where it has become the central axis of life. The education sector has also been strongly influenced by this growing trend in recent years. E-learning and immersive technologies, such as au-mentary and virtual reality, offer a series of benefits and potentialities in training practice at any educational stage and in any area of knowledge, making learning meaningful and contextualised, facilitating the understanding of abstract content, enriching teaching materials, avoiding risks, overcoming spatial and temporal obstacles and arousing the interest of students due to the motivating and innovative content they present, as the main features. Therefore, its inclusion in university classrooms of any discipline should be encouraged, as it represents a significant increase in the quality of teaching.

Keywords: Augmented Reality · Virtual Reality · Higher Education · e-learning

1 Introduction

E-learning, also known as e-learning, is a type of learning that uses electronic technologies to deliver educational content and training. E-learning can take many forms, such as online courses, virtual classrooms, webinars and multimedia presentations [1].

E-learning allows learners to access educational content anytime, anywhere and at their own pace, making it a flexible and convenient way to learn. E-learning can be accessed through a variety of different E-learning can be accessed through a variety of electronic devices, such as computers, tablets and smartphones, and can include interactive and multimedia elements, such as videos, animations and simulations [2].

E-learning can be used for a wide range of educational and training purposes, such as corporate training, professional development, academic education and personal learning.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 241–249, 2023.
https://doi.org/10.1007/978-3-031-43461-7_24

E-learning is especially beneficial for those who have a busy schedule, cannot attend face-to-face classes or need to study at a distance [1]. It also allows for personalised learning experiences, as learners can choose the content and pace that suits them best.

Augmented reality (AR) is a technology that enhances the real-world environment with digital information, such as images, videos and 3D models, in real time. Unlike virtual reality (VR), which creates a completely artificial environment, AR uses a camera or other sensors to capture the user's environment and then uses software to overlay digital content on top of the real-world view. The digital content can be interactive and responsive to the user's actions, allowing the user to manipulate and interact with virtual objects as if they were real [3].

AR is a technology that superimposes digital information, such as images, videos and 3D models, on the real world. AR allows users to interact with digital objects in the physical world, enhancing their perception and experience of the real world. AR can be used in a variety of applications, such as gaming, entertainment, education, marketing and healthcare [4]. For example, in education, AR can be used to create interactive and immersive learning experiences, such as virtual field trips, simulations and experiments. In healthcare, augmented reality can support medical training, surgery and patient care.

VR is a technology that allows users to experience immersive three-dimensional digital environments in real time. Unlike AR, which superimposes digital elements over the real world, VR creates a completely new, virtual world for users to explore and interact with. VR uses specialised hardware and software to simulate the physical presence of users in the virtual environment [5].

Hardware typically includes VR headsets or headsets, motion sensors and controllers, which allow users to see, hear and feel the virtual world and interact with it in a natural way. VR has numerous applications, including gaming, entertainment, education, design and training in industry [6]. For example, in the entertainment industry, VR is used to create immersive and realistic gaming experiences. In the education industry, it is used to create interactive learning experiences and educational simulations. It is also used in the training of professionals, such as pilots and doctors, to practice and hone skills in a safe and controlled environment.

E-learning and VR are technologies that are increasingly being used in Higher Education. These technologies offer new opportunities to improve the quality of education by making learning more interactive, personalised and engaging for students. In Higher Education, e-learning is used to deliver online courses, which allow students to access educational materials anytime, anywhere, through electronic devices such as computers, tablets and smartphones [7]. Online courses can include different types of resources, such as videos, podcasts, online readings, discussion forums, educational games and online quizzes. Teachers can interact with students through online chats, emails and video conferencing, allowing them to provide support and guidance in real time.

In Higher Education, VR can be used to create virtual learning environments that allow students to explore and experience different situations in a safe and controlled way [5]. It can simulate realistic scenarios, such as exploring archaeological sites, performing surgery or simulating natural disasters, allowing students to develop skills and knowledge more effectively. Students can interact with virtual objects, perform tasks and

experiments, and receive immediate feedback, allowing them to learn autonomously and hands-on.

E-learning and AR are technologies that are increasingly used in Higher Education to enhance the quality of learning and the student experience. E-learning is used to deliver online courses that students can access from anywhere and at any time, using electronic devices such as computers, tablets and smartphones [3]. Online courses can include different types of resources, such as videos, podcasts, online readings, discussion forums, educational games and online quizzes. Lecturers can interact with students through online chats, emails and video conferencing, allowing them to provide support and guidance in real time. AR is used in Higher Education to create immersive, hands-on learning experiences, allowing students to interact with digital objects in the real world. AR uses devices such as smartphones and tablets to overlay digital objects on top of the real world [4].

Ultimately, e-learning and AR are technologies that are increasingly being used in Higher Education to improve the quality of learning and the student experience. These technologies offer new opportunities for personalised, interactive and hands-on learning, which can significantly improve the quality of education and better prepare students to face the challenges of the working world.

2 E-Learning and Augmented Reality

The use of AR in e-learning projects opens up a wealth of possibilities for education. This resource can extend online training or complement face-to-face training. AR in education contributes to innovation, offers other possible realities, enables the understanding of theory through experimentation and simulation of contextualised scenarios and allows simultaneous continuous assessment [8].

This technology, which combines virtual and real or physical information at the same time, is generally used through mobile devices, which means that its inclusion in the classroom does not entail extra economic costs, as most students today have mobile devices and are positively disposed towards its use [9–12]. AR has the following characteristics: it works in real time, it is recorded in 3 dimensions to preserve the illusion of reality, and it uses computer-generated information and images superimposed on the users' field of view [13].

There are several levels of AR, depending on the technological devices used, complexity, parameters and techniques used [14, 15]:

- Level 0: this is a basic level of AR, which relates to information links through shapes, such as QR codes.
- Level 1: refers to virtual figures created in 3D. This level is the most commonly used.
- Level 2: covers everything related to image recognition or geolocation. It depends on parameters such as orientation, position and tilt of the device.
- Level 3: at this level, special screens are needed to obtain an augmented view, which allows the virtual environment to be added to the real one.
- Level 4: corresponds to global positioning, as in the case of GPS.
- Level 5: this relates to thermal fingerprinting and contact lenses.

This technological resource offers a series of advantages in educational practice and in the development of fundamental student competences. Fundamental aspects that facilitate the work of teaching and learning from many educational dimensions and that are in line with the demands of the students and with the active learning style that the current educational system demands [16, 17] (Table 1).

Table 1. Benefits of Augmented Reality in education

Advantages AR in education	
Advantages in educational practice	- It allows only the most important content to be shown, eliminating information that could hinder learning
	- It enriches the real information to facilitate its assimilation. It is especially effective in the understanding of abstract concepts
	-The student can observe the object of study from any 3-dimensional perspective
	- Allows for adjustment of the level of teaching
	- Facilitates ubiquitous learning
	- Develops active learning
	- Generates highly motivating educational scenarios for students
	Enables practice in safe contextualised scenarios -Enriches print materials
	- Enriches printed materials with additional information displayed in different media
	- It is a versatile resource that can be used at all stages of education and in any subject
Aspects of AR that contribute to the development of competences	- Participation in learning
	- Immersion and presence in the content
	- Placement of learning within the appropriate context
	- Authentication of content
	- Socialisation of the learner

Own elaboration based on Akçayır [16], Barroso and Gallego [17], Madanipour and Cohrs-sen [18] and Gómez-García [19].

In light of these benefits, it demonstrates the endless possibilities that AR offers to education. [14, 15]. On the other hand, it also has a number of limitations related to the lack of practice in the use of AR in the classroom. Its implementation is still a novelty in education to this day. However, these are aspects that time and experience will make up for (Table 2).

Table 2. Limitations of AR used in the educational setting

Limitations of AR in education	- Lack of teacher training - There are few educational practices and limited learning objects available - There is a need for more in-depth theoretical reflection -There is a lack of educational models for their incorporation - Students have little experience with the interaction of these virtual objects - Some contexts are difficult to handle

Own elaboration based on Cabero et al., [9].

In recent years, the use of mobile technologies combined with AR to create applications has increased, as they benefit from the portability and immediate access to information offered by these devices [20]. In non-face-to-face education, AR has a number of very interesting specific benefits such as [8, 21]:

- In the learning of medicine, it allows the exploration of an organ in 3D, the observation of the functioning of the human body, the analysis of the action of a medicine or to intervene in any type of surgery [22].
- It develops cognitive, spatial-perceptual-motor and temporal skills.
- It strengthens attention, concentration, short- and long-term memory in its visual, auditory and reasoning forms.
- It improves the effectiveness of training, which depends on the active work of students. Because AR interacts actively and consciously on this process, allowing to confirm, refute or expand knowledge, generate new ideas, feelings or opinions about any aspect.
- The transformative nature of the pedagogical and didactic activity supported by AR favours attitudes of reflection, by explaining the observed phenomena, and offers solutions to specific problems.
- Regardless of the academic discipline, it provides an ideal communication environment for educational work, because it contributes to the complete knowledge of an object.
- It greatly increases students' positive attitude towards learning, reinforcing autonomy and initiative [23].

3 E-learning and Virtual Reality in Higher Education

Today's university education combines the traditional method with e-learning systems. Today's modernity is progressively orienting educational innovation towards the possibilities offered by VR and AR as van-guardian technologies [24]. This is further boosted by the rise of distance learning. Thus, these educational technologies are increasingly present in university classrooms [25]. According to Díaz-López, learning through visual, auditory and tactile experience is making homo sapiens more and more homo videns [26]. Thus, there are disciplines that recognise and accept VR as an essential tool in

their teaching and learning processes. In the teaching of medicine, construction, geography, physical education [25], language learning, and nursing [27] among others, VR is positioning itself as an essential technology [28].

This technology enables higher education institutions to break down linguistic, physical, cultural, border, bureaucratic and even economic barriers [24]. The recreation of real or fictional environments is one of the important advantages of VR. Especially in the part of the curriculum that requires the formation of professional competences linked to experiential learning. Immersion in this sense provides students and teachers with the opportunity to use their knowledge in VR and to avoid dangerous and costly situations [27].

An outstanding example is the "Valdecilla Virtual Hospital", recognised and accredited by the Center for Medical Simulation in Boston. This VR initiative is improving the pace of learning for university students at 48 universities on different continents [27]. However, academic performance has been a concern that has accompanied its use in the classroom. In the imagination of many actors in the educational community, VR is still associated with video games. However, it is interesting how this has been overcome by the academic results of its use when connected to its pedagogical component [24]. In another study, it was found that both in the recreation of VR simulation experiences in a traffic emergency, and in the simulation of a laboratory accident, positive academic results were obtained in the students [27].

This immersion experience requires careful preparation guided by its didactic purpose. It also requires a process in which an application is designed and developed that is coherent with the curriculum. Secondly, an efficient application can be created with the proposed learning outcomes and goals. Subsequently, the interactive application can be experienced by the students and learning can take place. Finally, a space for evaluation and feedback is required. In this way, this educational technology will fulfil its function as long as quality didactic materials are developed [24].

The Unity platform, used for teaching the construction of dwellings in different spaces, is interesting in this respect. This tool, in addition to including the essential learning outcomes for the use of resources for the work, includes an important ethical content associated with construction. Thus, for example, it includes concern for water sources in the development of the constructions that are carried out. On the other hand, the Opem Sim and Sloodle [28] platforms stand out in language teaching. These tools promote essential skills such as creativity, collaboration and conflict resolution linked to digital competence. Thus, there are digital painters and sculptors who question this artistic tradition by proposing new ways of creating art from the virtual world. Among the tools that offer these possibilities are Google's Tilt Brush and HTC Vive [26].

The use of VR in university classrooms has led to a debate regarding its contribution to scientific its contribution to the scientific culture of society [28]. It is an opportunity to bring science and technology closer to society, both in formal and informal educational environments [25]. Its power to democratise science by making it more accessible to traditionally excluded sectors of the population is also advocated. VR as a tool facilitates the democratisation of knowledge and makes it possible to bring educational proposals to excluded groups, favouring the development of scientific vocations. This situation has been studied in the Mexican context [26]. They found that university students who studied

in public universities had better academic results than those in private universities. This was also the case when comparing university locations in urban and rural areas, as they varied in their achievement, impact and motivation.

4 Conclusions

In the knowledge society in which we live, technology has become a fundamental tool in our daily lives. In the field of education, the incorporation of technology has been a gradual process that has allowed the forms of teaching and learning to evolve. Traditional expository classes, in which students played a passive role, have proven not to meet students' expectations or to respond to current educational demands. For this reason, the inclusion of technologies in the educational sector has been a growing trend in recent years.

Among the technologies that have been incorporated in the educational field, e-learning and immersive technologies, such as augmented reality and virtual reality, stand out. These technologies offer a series of benefits and potentialities in training practice at any educational stage and in any area of knowledge. Immersive technologies allow students to experience learning situations or scenarios that would be difficult or impossible to recreate in any other way. For example, they can simulate practice in virtual laboratories or virtual trips to historical places.

The inclusion of these technologies in university classrooms in any discipline can significantly improve the quality of teaching. These technologies make learning more meaningful and contextualised, facilitate the understanding of abstract content, enrich teaching materials, avoid risks, overcome spatial and temporal obstacles, and awaken students' interest in the motivating and innovative content they present.

However, the inclusion of these technologies in university education also poses challenges that need to be addressed. One of the most important challenges is to ensure the quality of the educational content and materials offered through these technologies. It is essential that the content is accurate, up-to-date and relevant to the area of knowledge being studied.

Another important challenge is to ensure adequate education and training for teachers who must teach using these technologies. It is important that teachers have the necessary education and training to use these technologies effectively and to design and develop quality educational content that is adapted to the needs and characteristics of the students.

Finally, another important challenge is to ensure accessibility and equity of access to these technologies for all students. Measures need to be taken to avoid the digital divide and to ensure that all students have the same opportunities to access and use these technologies, regardless of their socio-economic background.

In conclusion, the inclusion of technologies in university education can significantly improve the quality of teaching and learning. However, it is important to address the challenges in implementing these technologies to guarantee their effectiveness and to ensure quality and equity in access and use by all students. Ultimately, the inclusion of technology in university education can be a very useful and effective tool to improve the training and preparation of students in an increasingly technological and digitised world.

Acknowledgements. This work is linked to the Advanced Teaching Innovation Project entitled "Aula invertida y recursos tecnológicos inmersivos (xr) para el desarrollo de la competencia digital docente en los futuros profesionales de la educación" (Inmer). Reference: 22-06. FIDO Plan. Unidad de Calidad, Innovación Docente y Prospectiva de la Universidad de Granada.

References

1. Firwana, A., Abu, M., Aqel, M.: Effectiveness of e-learning environments in developing skills for designing E-tivities based on Gamification for teachers of technology in Gaza. Educ. Knowl. Soc. EKS **22**, 15–21 (2021)
2. Segovia-García, N., Said-Hung, E.: Factores de satisfacción de los alumnos en e-learning en Colombia. Rev. Mex. Investig. Educ. **26**(89), 595–621 (2021)
3. Tallón, S., Berral, B., Campos, M.N., Martínez, J.A.: La realidad aumentada y sus efectos en el aprendizaje de los escolares. In: Marín-Marín, J.A., Boffo, A., Ramos-Navas-Parejo, M., y De La Cruz-Campos, J.C. (eds.), Retos de la investigación y la innovación en la sociedad del conocimiento, pp. 45–52. Dykinson (2022)
4. Carballo, L., Fernández, Y.F.: La Realidad Aumentada en el enfrentamiento a la COVID-19. Serie Científica de la Universidad de las Ciencias Informáticas **13**(11), 1–16 (2020)
5. Chirinos, Y.: La realidad virtual como mediadora de aprendizajes: desarrollo de una aplicación móvil de realidad virtual orientada a niños. Revista Iberoamericana de Tecnología en educación y Educación en Tecnología **27**, 98–99 (2020)
6. Londoño, L.M., Rojas, M.D.: De los juegos a la gamificación: propuesta de un modelo integrado. Educación y educadores **23**(3), 493–512 (2020)
7. Oliveros-Castro, S., Núñez-Chaufleur, C.: Posibilidades educativas de la realidad virtual y la realidad combinada: una mirada desde el conectivismo y la bibliotecología. Revista Saberes Educativos **5**, 46–62 (2020)
8. Korowajczenko, K.T.: Realidad Aumentada Sus Desafíos Y Aplicaciones Para El e-learning. No. June (2012)
9. Cabero, J., Barroso, J., y Llorente, C.: La realidad aumentada en la enseñanza universitaria. REDU. Revista de docencia universitaria **17**(1), 105–118 (2019)
10. Seifert, T., Hervás, C., y Toledo, P.: Diseño y validación del cuestionario sobre percepciones y actitudes hacia el aprendizaje por dispositivos móviles. PixelBit. Revista de Medios y Educación **54**, 45–64 (2019)
11. Aznar, I., Cáceres, M.P., Trujillo, J.M., y Romero, J.M.: Impacto de las apps móviles en la actividad física: un meta-análisis. Retos: nuevas tendencias en educación física, deporte y recreación (36), 52–57 (2019)
12. Cabero, J., y Barroso, J.: Los escenarios tecnológicos en Realidad Aumentada (RA): posibilidades educativas. Aula Abierta **47**(3), 327–336 (2018)
13. Ferguson, R., et al.: Innovating Pedagogy 2017: Open University Innovation Report 6. Milton Keynes: The Open University, UK (2017)
14. Cabero, J., Llorente, M.C., y Marín, V.: Comunidades virtuales de aprendizaje. El Caso del proyecto de realidad aumentada: RAFODIUM. Perspectiva Educacional. Formación de Profesores **56**(2), 117–138 (2017)
15. Marín, V., y Muñoz, V.P.: Trabajar el cuerpo humano con realidad aumentada en educación infantil. Revista Tecnología, Ciencia y Educación **9**, 148–158 (2018)
16. Akçayır, M., y Akçayır, G.: Advantages and challenges associated with augmented reality for education: a systematic review of the literature. Educ. Res. Rev. **20**, 1–11 (2017)

17. Barroso, J., y Gallego, O.: Producción de recursos de aprendizaje apoyados en Reali-dad Aumentada por parte de los estudiantes de magisterio. Edmetic. Revista de Edu-cación Mediática y TIC. **6**(1), 23–38 (2017)
18. Madanipour, P., y Cohrssen, C.: Augmented reality as a form of digital technology in early childhood education. Australas. J. Early Child. **45**(1), 5–13 (2020)
19. Gómez-García, G., Rodríguez-Jiménez, C., y Ramos, M.: Virtual reality in physical education area. J. Sport Health Res. **11**, 177–186 (2019)
20. Fabregat-Gesa, R.: Combinando la realidad aumentada con las plataformas de e-elearning adaptativas. Enl@ce: Revista Venezolana de Información, tecnología y conocimiento **9**(2), 69–78 (2012)
21. Almenara, J.C., y Osuna, J.B.: Posibilidades educativas de la Realidad Aumentada. Revista de nuevos enfoques en investigación educativa (NAER Journal) **5**(1), 44–50 (2016)
22. Vidal Ledo, M., Lío Alonso, B., Santiago Garrido, A., Muñoz Hernández, A., Morales Suárez, I.D.R., y Toledo Fernández, AM.: Realidad aumentada. Educación Médica Superior **31**(2) (2017)
23. Cabero, J., Barroso, J. y Gallego, O.: La producción de objetos de aprendizaje en realidad aumentada por los estudiantes. Los estudiantes como prosumidores de información. Tecnología, Ciencia y Educación **11**, 15–46 (2018)
24. Gros-Salvat, B.: La evolución del e-learning: del aula virtual a la red. RIED. Revista Iberoamericana de Educación a Distancia **21**(2), 69–82 (2018)
25. Gómez-García, G., Rodríguez-Jiménez, C., Ramos-Navas-Parejo, M.: Virtual reality in physical education area. J. Sport Health Res. **11**(Supl 1), 177–186 (2019)
26. Díaz-López, L., Tarango, J., Romo-González, J.: Realidad Virtual en procesos de aprendizaje en estudiantes universitarios: motivación e interés para despertar vocaciones científicas. Cuadernos de Documentación Multimedia **31**, e68692 (2020)
27. Mariscal, G., Jiménez, E., Vivas-Urias, M.D., Redondo-Duarte, S., Moreno-Pérez, S.: Aprendizaje basado en simulación con realidad virtual. Educ. Knowl. Soc. (EKS) **21**, 15 (2020)
28. Anacona, J., Millán, E., Gómez, C.: Aplicación de los metaversos y la realidad virtual en la enseñanza. Entre ciencia e ingeniería **13**(25), 59–67 (2019)



Machine Learning Based Ambient Analysis of Railway Steel Bridges for Damage Detection

Muhammad Asad(✉)

Universoty of L'Aquila, 67100 L'Aquila, Italy
muhammad.asad@graduate.univaq.it

Abstract. Structural Health monitoring is becoming important research direction mainly with focus in transportation industry which includes civil engineering structures. Ambient analysis of these structures is evident as their lifespan depends on the damages caused to them due to natural disasters as well as due to ambient factors like vibrations, acceleration and/or temperature. Fuzzy Rule based and Machine Learning based ambient analysis for anomalies detection, damage detection and localization and finally alert intimation based on the level of damage is the main focus of this study. In this study , proposed solution includes Fuzzy rules and machine learning based model for achieving promising results. As part of a project, these proposed solution can then also be deployed to other civil engineering structures and not limited to the steel bridges. As in suggested solution and observed in the experimental FEM simulations, it is expected to get good results that will definitely be helpful for in-time alarm and alert generation. This proposed system can also have some limitations including correct working of sensors, effects of weather conditions and/or other environmental factors but if implemented with great care, these factors can be reduced to some extent.

Keywords: Anomaly detection · Damage detection · Fuzzy rules · Machine learning · Reasoning · Real-time multi-agent systems

1 Problem Statement

In modern transportation, Railways is still playing its role. Railway networks are interlinked and connected through bridges and tunnels that keeps these mean of transportation intact. Such civil engineering structures have long design life but inevitably encounter wear & tear and aging due to environmental factors including i. extreme weather conditions, natural disasters. Ambient analysis of railway bridges (steel or concrete) is necessary for their long life. Using Structural Health Monitoring (SHM) technology, early structural anomalies detection, occurrence of damage with alert intimation and in-time maintenance become possible. This

Supported by University of L'Aquila, AQ Italy.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 250–255, 2023.
https://doi.org/10.1007/978-3-031-43461-7_25

lead researchers to systematic structural research about dynamic responses of vibrations, natural frequencies, accelerations of vehicles on the bridges, temperature variations (when train passes and after train passes) and mode shapes. These ambient metrics can be used to predict the anomalies in bridge structures, damage levels and alert intimation based on those damage indexes.

2 Related Work

SHMs provide many viable solutions to the damage detection but most of these are only limited to anomalies or statistical model comparisons. Svendsen et al. [2]investigated statistical comparison models with supervised and unsupervised learning using Mahalanobis Squared Distance only. This study only focus

Table 1. Damage detection model details.

Study	Model Used	Dataset Availability	Feature extraction method	Data Normalization method	Performance matrices
Damage Detection [2]	KNN, SVM, Gaussian NB, RF	Supervised ML	Auto-regressive	Mahalanobis Squared Distance	ROC, Confusion matrix
Steel railway bridge fatigue damage detection [3]	ANN	Supervised ML	Singular Value Decomposition SVD	RMSs	Damage Intensity accuracy
Numerical FE Model Evaluation [1]	Dynamic response change, static response change	Unsupervised	-	Least Squared method	-
Robust bridge damage detection [4]	Threshold based SVD Classification	Unsupervised	SVD, LSVs, ICA	Euclidean Norm	Novelty index comparison, Outlier analysis, Novelty detection accuracy
Online Structural health monitoring [5]	Model order reduction, Fully CNN	Supervised	PCA	-	Accuracy, Signal to Noise ratio
Bridge health monitoring [6]	Linear regression, Robust PCA	Supervised	PCA	Mahalanobis distance, Chi-square distribution	ROC

of ROC metrics using SVM, RF, kNN; which don't give any insights of the dataset used and the damage levels. Ahmed et al. [3] explains on steel railway

bridge fatigue damage detection using experimental model. This model is base on simple Neural Network using Singular Value Decomposition for measuring damage intensity accuracy. In [1], dynamic response change and static response change in the Finite Element Model (FEM) is studied. FEM is evaluated with ambient vibration analysis, dynamic analysis using traffic measurements and static analysis based on the temperature. Another research by Luca et. al [5] explains the CNN based bridge damage classification and accuracy metrics based on SNR. Although these studies are covering some scenarios of railway damage classification but none of these are providing any information about data acquisition, detailed damage levels and insights about evaluation metrics.

Furthermore, Thanh et al. [7] and Maes et al. [9] also worked extensively on machine learning based model for damage detection. But this work in specifically inspired by the work by Micro et.al. [8] in which an IoT based sensory system is designed for structural damage indication. Accelerometer data is acquired using triaxial sensors and then the data is used generate alarms when the threshold value is exceeded captured from the sensors. Although there exists few system but the IoT based system proposed in their study has few advantages including sensors with high bandwidth, high data rate with high synchronization between nodes and thus making it a low-cost system. In Table 1 summary of literature studies is also shown.

Ambient analysis provide viable solution for damage localization while the accelerometer sensors can be used for data accumulation as accurate ambient analysis requires huge amount of responsive data in less time.

3 Hypothesis

Monitoring the response of a structure during its lifetime has therefore been proposed as a tool to detect damage at the earliest possible stage. Following some predefined question from [10], following structural health monitoring stages are suggested to be studied: **Detection**- Presence of damage in the structure? **Localization** - Location of the damage? **Assessment** - Type and level of damage caused? **Prediction** - Remaining lifetime of the structure?

In this doctoral study, the goal is to propose, fuzzy rule based and machine learning based railway bridge damage anomalies calculation, railway bridge damage levels classification and localization of damage occur and alert intimation based on the level of damage. Increasing importance of transportation means and advancement in technologies, it is becoming effective to address such structural damages and generating in-time alerts to lessen the losses. This leads to the idea to develop a system that is based on the reference FEM for simulations results, perception layer and reasoning layer for complex system events processing (includes damage detection, anomalies detection, damage classification) and finally a response UI to signal alarms.

Currently, reasoning layer is considered for study and in the Reasoning Layer, anomalies detection based on the raw sensory data using Monte Carlo method and Mahalanobis distance is suggested. For classification, fuzzy rules

and machine learning models are being implemented. As a novelty, Radial Basis Function based neural network is considered.

In short, main focus is on building of cost-effective Neural Network framework and fuzzy rule based ambient AI system for estimation of damages to the bridges, anomalies detection and alerts generation based on ambient data acquired by different sensors.

4 Proposal

Starting form the system in Fig. 1 which contains Reference FEM for simulation results, a data bus for acquisition of data from the sensors, a perception layer for raw data handling, reasoning layer for complex events processing and dashboard for alarm alerts.

As main focus of this PhD study is on the reasoning layer, Proposed system architecture is elaborated in the Fig. 2.

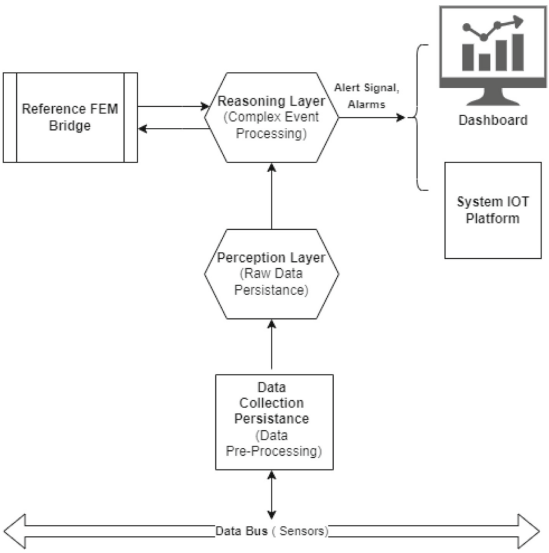


Fig. 1. Overall System Architecture

5 Evaluation Plan

As discussed in the hypothesis section about structural health monitoring system sub-modules, our focus for system evaluation consists of Anomalies detection in data coming from the different sensors. Experts will provide the limitations on the rules. Using these rule a classification model will be generated using fuzzy

rules and machine learning algorithms. As far now, no alert, medium alert and high alert rules are in focus based on the anomalies detected. For anomalies, data pre-processing and feature extraction based on Principal Component Analysis, Single Value Decomposition etc. are proposed. Experimental Model is identified from this anomaly module. This data is passed as a Vector Matrix to the reasoning module using Hankel Matrix H and Residual Matrix R. Experimental results using simulations are almost ready and are in process of validation so far.

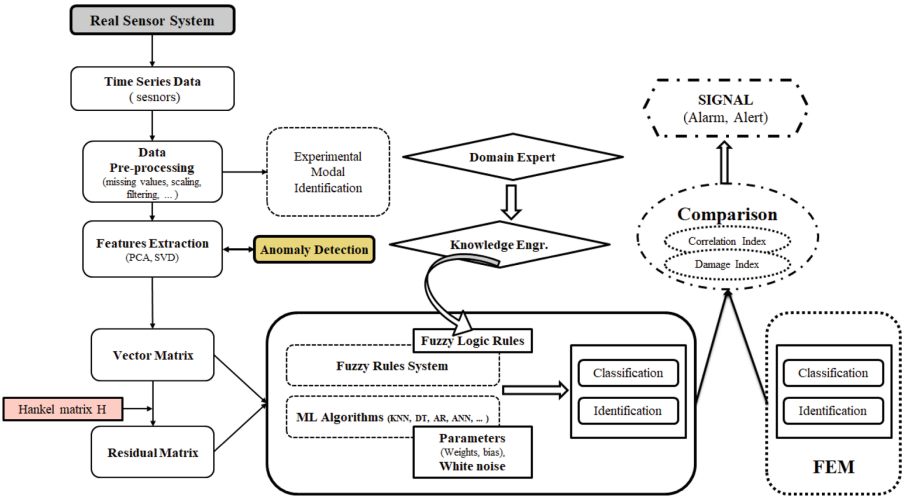


Fig. 2. Proposed Reasoning Layer Architecture

6 Reflections

As far now only very few system are available which provide structural health monitoring especially for steel bridges. This study will be an extensive system which can provide solutions for anomalies detection in data, based on these anomalies, localization of damage can be easy and from these result it will be evident to damage detection and classification which will intimate an alert alarm for in-time damage detection so that possible measures can be taken to minimizes the losses and also helpful for proper maintenance of the bridges.

References

1. Meixedo, A., Ribeiro, D., Santos, J., et al.: Progressive numerical model validation of a bowstring-arch railway bridge based on a structural health monitoring system. J. Civil Struct. Health Monit. **11**, 421–449 (2021). https://doi.org/10.1007/s13349-020-00461-w

2. Svendsen, B.T., Frøseth, G.T., Øiseth, O., et al.: A data-based structural health monitoring approach for damage detection in steel bridges using experimental data. J. Civil Struct. Health Monit. **12**, 101–115 (2022). https://doi.org/10.1007/s13349-021-00530-8
3. Rageh, A., Azam, S.E., Linzell, D.G.: Steel railway bridge fatigue damage detection using numerical models and machine learning: mitigating influence of modeling uncertainty. Int. J. Fatigue, **134**, 105458 (2020). ISSN 0142–1123, https://doi.org/10.1016/j.ijfatigue.2019.105458
4. Akintunde, E., Azam, S.E., Rageh, A., Linzell, D.G.: Unsupervised machine learning for robust bridge damage detection: full-scale experimental validation. Eng. Struct. **249**, 113250 (2021). ISSN 0141–0296, https://doi.org/10.1016/j.engstruct.2021.113250
5. Rosafalco, L., Torzoni, M., Manzoni, A., Mariani, S., Corigliano, A.: Online structural health monitoring by model order reduction and deep learning algorithms. Comput. Struct. **255**, 106607 (2021). ISSN 0045–7949, https://doi.org/10.1016/j.compstruc.2021.106604
6. Mousa, M.A., et al.: Application of digital image correlation in structural health monitoring of bridge infrastructures: a review. Infrastructures **6**(12), 176 (2021). https://doi.org/10.3390/infrastructures6120176
7. Cuong-Le, T., Nghia-Nguyen, T., Khatir, S., et al.: An efficient approach for damage identification based on improved machine learning using PSO-SVM. Eng. Comput. **38**, 3069–3084 (2022). https://doi.org/10.1007/s00366-021-01299-6
8. Muttillo, M., et al.: Structural health monitoring: an IoT sensor system for structural damage indicator evaluation. Sensors **20**, 4908 (2020). https://doi.org/10.3390/s20174908
9. Maes, K., Van Meerbeeck, L., Reynders, E.P.B., Lombaert, G.: Validation of vibration based structural health monitoring on retrofitted railway bridge KW51. Mech. Syst. Signal Process. **165**, 108380 (2022). ISSN 0888–3270, https://doi.org/10.1016/j.ymssp.2021.108380
10. Rytter, A.: Vibrational Based Inspection of Civil Engineering Structures. Dept. of Building Technology and Structural Engineering, Aalborg University. Fracture and Dynamics, Vol. R9314 No. 44 (1993)

Human Action Recognition in Uncontrolled Environments: Application from Artificial Intelligence to Contactless Interfaces

Vanessa Alvear[1,2], César Domínguez[1], and Gadea Mata[1(✉)]

[1] Departamento de Matemáticas y Computación, Universidad de La Rioja, Logroño, Spain
mayra-vanessa.alvear@alum.unirioja.es,
{cesar.dominguez,gadea.mata}@unirioja.es
[2] Innovación Riojana de Soluciones IT SL, Logroño, La Rioja, Spain
vanessa.alvear@irsoluciones.com

Abstract. The recognition of human gestures and actions in images and videos is an active area of research in computer vision. This field has made great advances in the last decade thanks to the use of deep learning techniques. In addition, the recent diffusion of low-cost video camera systems, including depth cameras has enhanced the development of observation systems in a variety of application areas, such as video surveillance, home security, healthcare, etc. However, most of these developments are in closed and controlled environments. The recognition of movements and gestures in real-time through a camera that acquires its images in an uncontrolled environment (such as a shopping mall, a university lobby or a museum hall) and that allows interaction with passers-by in these spaces, contains challenges in various areas that include at least technological, social and legal challenges that need a careful approach. Within this framework, we set as objectives of this project the design of a web interface adapted to interaction without physical contact (i.e., through video images captured with a camera in real-time) and the construction of artificial intelligence models (based on deep learning) that guarantee, in an uncontrolled environment, an interaction with this web interface. As a first step in this work, we propose a literature review in this area. In addition, we include model recognition for a given set of gestures to explore the possibilities of different approaches.

Keywords: Gesture recognition · uncontrolled environment · contactless interface · computer vision

This work was partially supported by Doctorado industrial, Comunidad Autónoma de La Rioja, and Grant PID2020-115225RB-I00 funded by MCIN/AEI/ 10.13039/ 501100011033.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 256–261, 2023.
https://doi.org/10.1007/978-3-031-43461-7_26

1 Introduction

Computer vision is a branch of artificial intelligence that aims to provide computers with the ability to obtain information from images and videos. In recent years there has been a great revolution in this field due to the use of deep learning techniques. An active research area within computer vision is the recognition of human gestures and actions and extensive research work has been carried out in the last decade, as can be seen in the existing literature [9], [5]. In addition, the recent spread of low-cost video camera systems, including depth cameras [12], has reinforced the development of observation systems in a variety of application areas, such as video surveillance, home security, healthcare, etc [9]. However, most of these developments are carried out in closed and controlled environments. The recognition of movements and gestures in real-time through a camera that acquires its images in an uncontrolled environment (such as a shopping mall, a university lobby, a museum hall, a railway platform, or an air terminal) and that allows interaction with passers-by in these spaces, contains challenges in several areas that include at least technological, social, and legal challenges that need to be carefully addressed. In the technological field, it is necessary to develop useful models in different uncontrolled spaces. Different methods can be used, such as data augmentation [2], model transfer [13], or style transfer [4] that can complement the acquisition and annotation of sufficiently general datasets to be able to train and validate these models. Moreover, within these technical aspects, there are difficulties arising from the noise in an uncontrolled space (people passing behind the one in front of the camera focus). The cameras must take into account different depth levels so that they can effectively focus on the aspect to be studied [12]. In social factors, the interface should be carefully designed to avoid feelings of frustration (or even rejection) that could be generated if the movements in front of the camera are not interpreted as the user expects. Finally, legal aspects must be taken into account to analyze how these cameras should be installed in uncontrolled spaces and the request of the necessary permissions to the users when capturing the images.

2 Literature Review

We consider five papers focusing on Human Action or Activity Recognition (HAR) to gather a general view of this field of study. First of all, we analyze a review focused on the concepts, methods, and datasets used in this area [9], followed by a paper explaining two fundamental HAR approaches and their characteristics, as well as describing the difference between recognition and detection problems for solving this task [7]. The third studied article [11] addresses real-time recognition using raw depth video. To find out how to deal with skeleton joints for HAR, we explore Skepxels [6], which are built-in to obtain temporal-spatial information in image sequences designed for HAR. Finally, we are interested in creating a dataset. How2Sign dataset for American Sign Language [3] is a good starting point to be studied.

Human action recognition approaches with video datasets-A survey [9] is an overview covering various HAR methods. This is the starting point of our study. Therefore, it is very useful to understand the existing approaches, identify their performance and technical background, and the details of the main HAR datasets. This paper analyzes fifteen comparative surveys between 2008 and 2020 of different techniques in which all methods have been re-run. Also, twenty-five datasets were examined with detailed comparisons. This survey states that recognition operations are performed using methods such as feature extraction, action learning, action classification, action recognition, and action segmentation. In addition, it is mentioned that is crucial to define the task as a recognition or prediction problem, the type of input data (e.g., images, sensors, videos, etc.), and to define the problem as a single or multimodal.

Vision-Based Human Activity Recognition [7] is the first chapter of the book Contactless Human Activity Analysis [1]. In this paper, we can find an explicit way to address HAR tasks from two broad categories: Handcrafted and Deep Feature-Based Human Activity Recognition. In the former, it is found that colour information can provide important clues for activity recognition. In the latter, CNNs are combined with methods such as LSTM (Long Short-Term Memory) and supervised/unsupervised models to obtain a better performance.

Real-time human action recognition using raw depth video-based recurrent neural networks [11] explain that dealing with people implies the consideration of their privacy, especially in uncontrolled environments. This article proposes two approaches for real-time HAR considering raw depth video sequences as input data. Both different architectures (*stateless model* and *stateful model*) are based on ConvLSTM (Convolutional LSTM). The two approaches differ in the training procedure. *Stateless* ConvLSTM focuses on how the frames to be processed from the video are selected depending on their length and follows a more conventional training; on the other hand, the training of *stateful* ConvLSTM needs special treatment due to the existence of a state that can be updated and reset throughout the processing of each video. Working with depth information in HAR helps preserve people's privacy. Moreover, it does not require any prior calculation, such as skeleton positions, optical flow, etc., and allows estimating human joint positions.

Skepxels: Spatio-temporal Image Representation of Human Skeleton Joints for Action Recognition [6] works on human action recognition through a strategy using Skepxels (Skeleton Picture Elements). We choose this article to center on the benefits of working with skeleton joints and avoiding problems with clothing, illumination, and background condition. The potential of CNNs is made possible by arranging the skeleton joints in images and allowing CNNs to be directly trained on such images. Skepxels are maps of the skeleton data from a fixed-length sequence to an image with the help of an essential building block (similar to a pixel). Their purpose is to make a spatio-temporal representation of skeleton sequences to fully exploit the correlations between joints using the kernels of CNNs. They showed that CNNs using appropriate skeletal time series data representations could lead the HAR task.

How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language [3] discusses how the How2Sign dataset for American English Sign Language (ASL) was created, what considerations were taken into account, the cameras used, the number of cameras, and so on. This multimodal dataset consists of a parallel speech corpus, and English video transcripts corresponding to the ASL translation and its annotations. It is a collection of 83 h of instructional videos that were translated into ASL. It was recorded from multiple viewpoints, including deaf information and clause annotation. Created through the collaboration of eleven ASL users, they employed different studios like a green screen studio containing two high definition cameras and the Panoptic Studio enabling detailed 3D pose estimation built on top of a geodesic dome with more than five hundred cameras, which provides high-quality 3D key points.

3 Proposal

The starting hypothesis in this thesis is that it is possible to develop artificial intelligence models that allow gesture recognition in uncontrolled environments through a contactless interface. We believe that the research required by the study of this hypothesis involves research challenges and that the obtained results can be transferred to society.

We intend to explore this hypothesis through the following two general objectives of the project, which are closely related:

- GO1. Design of a web interface adapted to interaction without physical contact (i.e., through video images captured with a camera in real-time).
- GO2. Construction of artificial intelligence models (based on deep learning) that guarantee, in an uncontrolled environment, interaction with the interface whose design is included in the first general objective.

4 Preliminary Results

As a proof of concept and to implement technical practices of the papers mentioned above, a project has been carried out that, in real-time, is capable of predicting the vowel of the Spanish Sign Language (Lengua de Signos Española - LSE) when a person is signing. For this purpose, we create a dataset corresponding to the LSE vowels named *vowelsLSE dataset*, see Fig. 1. The *vowelsLSE* dataset was created on March 2023, and it consists of RGB images in JPG format with a size of 300 × 300 for a total of 6078 images (see Table 1). For data collection, we took pictures, in a controlled environment to minimize variability, using a conventional web camera of the same person doing the LSE vowel signs using the right and left hand. The recognition was based on a hand detection system and image classification models. Using the libraries *MediaPipe* and *OpenCV*, we defined a hand-tracking module for skeleton detection in real-time. For the image classification, we built and used three models with *Keras*, *Tensorflow*, and *fastAI* libraries. The first one consists of a simple neural network

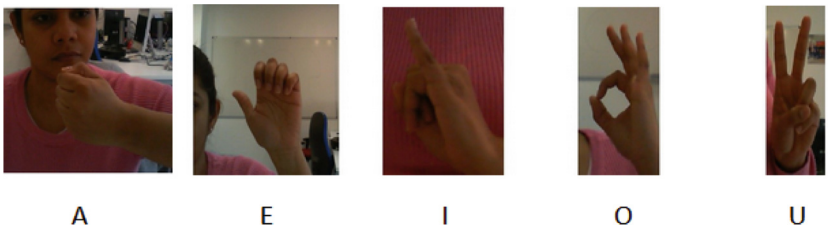


Fig. 1. Samples of the vowels we consider for *vowelsLSE* dataset.

Table 1. Information of the *vowelsLSE* dataset.

Vowel	Right hand	Left hand	Total
A	600	615	1215
E	600	619	1219
I	601	611	1212
O	601	619	1220
U	612	600	1212
Total	3014	3064	6078

(NN1) with three layers, the activation softmax function, the Adam optimizer, and the Sparse Categorical Crossentropy as the loss function. The second model is a more complex neural network (NN2) based on [10] where we implemented a hyperparameter searching using *KerasTuner* [8] based on a *Bayesian Optimization* looking for values for 39 hyperparameters. Finally, the third model was built with the *fastAI* library using the fine-tuning technique on a pre-trained neural network *ResNet50*. For the three models, the dataset was split using 70% of the images for training, 10% for validation, and 20% for testing. The evaluation results are included in Table 2. The best prediction was obtained with the ResNet50 model. NN2 model also obtains good results. It is important to notice that both models are more complex than the simple NN1 model. The architecture of our proposal will integrate the detection system with the best classification model obtained to create full recognition in real-time. The dataset and the code created in this project is available at the project webpage https://github.com/AlvearVanessa/vowelssigndataset.

Table 2. Results of the evaluation according to the accuracy.

Modelo	Accuracy test set (%)
ResNet50	96.9
NN2	92.3
NN1	20

5 Conclusions

In this work, we have undertaken an exploratory study of existing HAR technologies. Also, we have performed a proof of concept (using Spanish Sign Language vowels) to create a dataset and a recognition system that allows us to become familiar with these methods. This test has been conducted in a controlled space, and established the basis to be improved in the future. It also remains to study uncontrolled environments, privacy issues, and add more variability to the dataset, among others.

In future work, we need to address at least the following aspects. Firstly, to extend the study to uncontrolled environments, which will also raise important issues such as privacy. Secondly, to extend the range of human actions to be treated and the variability of the people in the dataset.

References

1. Ahad, M.A.R., Mahbub, U., Rahman, T.: Contactless Human Activity Analysis, vol. 200. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68590-4
2. Casado-García, Á., et al.: CLoDSA: a tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks. BMC Bioinform. **20**(1), 1–14 (2019)
3. Duarte, A., et al.: How2Sign: a large-scale multimodal dataset for continuous American sign language. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2735–2744 (2021)
4. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
5. Jegham, I., Khalifa, A.B., Alouani, I., Mahjoub, M.A.: Vision-based human action recognition: an overview and real world challenges. Forensic Sci. Int. Digit. Investig. **32**, 200901 (2020)
6. Liu, J., Akhtar, N., Mian, A.: Skepxels: spatio-temporal image representation of human skeleton joints for action recognition. In: CVPR Workshops, pp. 10–19 (2019)
7. Mahmud, T., Hasan, M.: Vision-based human activity recognition. Contactless Hum. Act. Anal. 1–42 (2021)
8. O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al.: Kerastuner (2019). https://github.com/keras-team/keras-tuner
9. Özyer, T., Ak, D.S., Alhajj, R.: Human action recognition approaches with video datasets-a survey. Knowl.-Based Syst. **222**, 106995 (2021)
10. Poulinakis, K.: Complete Practical Tutorial on Keras Tuner GitHub. https://github.com/Poulinakis-Konstantinos/Blogging-Journey/blob/main/Keras-Tuner-Complete-Tutorial/keras-tuner.ipynb. Accessed 25 Apr 2023
11. Sánchez-Caballero, A., Fuentes-Jiménez, D., Losada-Gutiérrez, C.: Real-time human action recognition using raw depth video-based recurrent neural networks. Multimed. Tools Appl. 1–23 (2022)
12. Sarkar, A., Banerjee, A., Singh, P.K., Sarkar, R.: 3D human action recognition: through the eyes of researchers. Expert Syst. Appl. 116424 (2022)
13. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)



Analysis of Techniques for the Augmentation and Correction of Unbalanced Data in Binary Cybersecurity Datasets for Machine Learning

Alberto Sánchez del Monte(✉)

Programme in Computer Engineering, Line of Cryptography and Information Security, University of Salamanca, Salamanca, Spain
idu030355@usal.es

Abstract. This document presents an analysis of different methods for correcting an unbalanced binary dataset with a limited number of observations associated with cybersecurity, specifically with Mitre Att&ck. To this end, the application of data manipulation techniques, oversampling, undersampling and synthetic generation of observations is considered. This research contributes to having a balanced dataset with sufficient observations for the subsequent application of machine learning algorithms.

Keywords: Cybersecurity · Mitre Att&ck · Tactics · SMOTE · Neighborhood Cleaning Rule · Oversampling

1 Problem Statement

Binary datasets have certain advantages for the subsequent application of machine learning algorithms, such as easy interpretation of the information, computational efficiency when working with ones and zeros, and easy translation to practical real-life problems. However, it should be borne in mind that these data are generally of high dimensionality and are often unbalanced. The latter phenomenon can generate biases that lead to an erroneous execution of the algorithm that is subsequently applied [1]. Likewise, it must be taken into account that the correction of the balance in the data will not always result in a better performance of the model [2]. It is therefore essential to analyze the results obtained with and without the application of techniques.

Mitre Att&ck [3] is a cyber-intelligence knowledge base based on a matrix that shows the attack techniques used to breach a network or system and the tactics or steps used to achieve this as well as the specific techniques. It has many advantages, but one of its limitations is the scarcity of data on some of the attack vectors. This is due to the fact that the framework has only been applied for a relatively short time and mainly because the collection of cyberattack data can sometimes be complex, especially if it involves data from real attacks where the affected organizations are reluctant to show the shortcomings that led to the success of a given attack. Consequently, the scarcity of data can make a machine learning model based on the Mitre Att&ck matrix less effective in predicting

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 262–267, 2023.
https://doi.org/10.1007/978-3-031-43461-7_27

and defending against future attacks, since, if attack patterns are not observed frequently enough, it is less likely that effective tools and defense measures will be developed in response to the results shown by the algorithm. Thus, lesser-known attack techniques may become more effective, as defense teams are not prepared for them, creating a kind of invisibility to the benefit of cybercriminals.

The data analyzed have been obtained after manipulation of the information contained in the official repository "*Working with Mitre Att&ck*" [4], so that variables have been created based on the Tactics detected in an attack, with "1" indicating their presence and "0" indicating their absence. These data, once they have been manipulated for processing, show a strong disparity in terms of the frequency of positivity in the variables selected for the study, with variability between about 20% and 80%. Also, the number of observations in the initial dataset is only 750.

The objective set in this work is to obtain adequate data, in quantity, quality, and with sufficient fidelity to reality, for the subsequent application of machine learning algorithms to a binary dataset that shows the actions of attacks in Mitre Att&ck.

2 Literature Review

To correct the problem, there are several resampling techniques, both from the point of view of actions on the data, such as oversampling (generally employing techniques such as SMOTE (*Synthetic Minority Over-sampling Technique*) [5], which creates new synthetic examples for the minority class of the dataset by interpolations instead of duplicating the data, as well as its variant Borderline SMOTE [6]; sub-sampling, usually employing random deletion, based on Near Miss [7] or Tomek links [8, 9]; as well as the combination of the two previous techniques, among other options, there is the combination of SMOTE and Tomek links, known as SMOTE-Tomek [10], which combines these two techniques to address class imbalance in a dataset. Finally, Neighborhood Cleaning Rule (NCR) [11], is an instance-based learning algorithm that reduces the size of the training set by removing noisy instances and majority-class instances that overlap with minority-class instances, improving classification performance. It is important to note that if a technique works well on one unbalanced dataset, it does not necessarily follow that it will work on a different dataset [12].

Although for the particular case of a Mitre Att&ck related dataset no specific solutions have been found, to address the generic problem of a sparse or unbalanced dataset, several solutions have been developed in other fields. Among them is ROSE (*Random Over-Sampling Examples*) [13], which is based on *Bootstrap* and focuses on oversampling random data to equalise the number of instances of the two classes in the training dataset and also includes other features, such as the ability to perform cross-validation and feature selection. On the other hand, [14] approaches the problem from a holistic point of view, acting on data preprocessing and algorithm assembly in the context of intrusions in networks and information systems. In [15] two of the best-known methods (SMOTE and subsampling) are adapted to regression tasks and in [16] discusses in detail the possible application of sub-sampling techniques.

3 Hypothesis

To correct the imbalance and the small number of observations in the binary dataset drawn from "*Working with Mitre Att&ck*", this study deems it appropriate to explore the application of several resampling and synthetic generation techniques that can address potential imbalances between classes, as well as increase the volume of observations. Based on this, the study hypothesizes that there is a significant difference in the outcomes when correcting and augmenting a binary dataset using different techniques, and that one of the techniques (SMOTE, Borderline SMOTE, or NCR) yields a better performance in terms of precision and the quantity of newly generated observations, compared to the others.

Derived from this hypothesis, the research objective is to compare the effectiveness of the oversampling techniques SMOTE, Borderline SMOTE and the NCR cleaning strategy in improving the performance of classification models on unbalanced binary datasets. To this end, the accuracy of the classification models, after the application of each method, will be compared to the accuracy of models trained on the original, uncorrected data, with the aim to increase the number of observations and correct the class imbalance.

To test the hypothesis, an architecture based on the following points is proposed:

1. Data pre-processing
2. Apply proposed method
3. Model selection
4. Model training and evaluation
5. Comparison of techniques
6. Interpretation of results and conclusions

4 Proposed Method

The proposed method consists of the progressive execution of two phases that would allow the binary data to be corrected and augmented. For the development and analysis of the proposed method, the Python programming language and the Scikit learn library [17] will be used on a dataset (previously manipulated and adapted) obtained from the official repository "*Working with Mitre Att&ck*." [4].

PHASE I. Feature Selection (Preliminary Phase): In order to predict the minority class, certain features may have a greater impact on the algorithm than others. Therefore, by means of this method, certain relevant features of the dataset can be selected to improve performance. A first filtering will be used to aggregate variables and then statistical measures will be used - such as correlation as done by [18] or variance.

PHASE II. Application of Techniques: The aim will be to carry out subsampling of the majority classes and oversampling of the minority classes by applying three possible options: variant a) SMOTE; variant b) Borderline SMOTE; and variant c) Neighborhood cleaning rule.

5 Preliminary Results

Through the application of the proposal in its different variants, a progressive adjustment in the class values is observed. Thus, in all the variants of Phase II it is observed that the data balancing techniques (SMOTE, Borderline SMOTE and Neighborhood cleaning rule) improve the average accuracy compared to the original data, and in particular, the application of variant c) Neighborhood cleaning rule. Regarding the number of observations, it can be seen that in all three variants it is very significant, exceeding well over a thousand observations in all cases, approaching tens of thousands of observations in the first variant (Fig. 1).

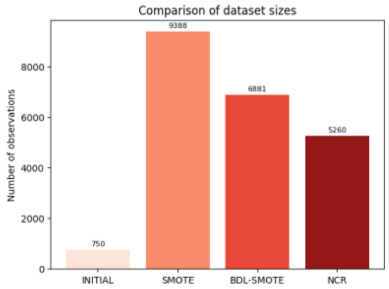


Fig. 1. Number of observations of each dataset after treatment

With respect to the accuracy of each variant, it can be seen that variant c) is the one with the best results, with a percentage of 78% in the average accuracy prediction of variables. In relation to the positivity of the variables observed, the first two variants reduce the positivity in those variables with a greater presence of ones, while variant c) maintains and even slightly increases the positivity in the variables with a greater presence of ones, but on the other hand increases the positivity in those variables with a lower presence of ones (Fig. 2).

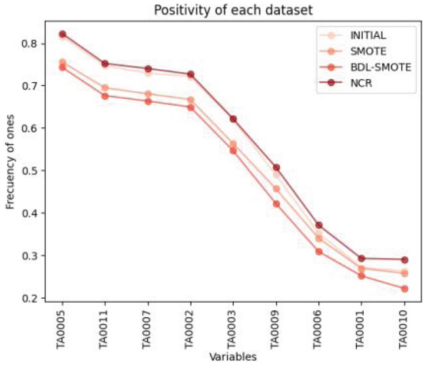


Fig. 2. Positivity of each dataset

In this sense, given that variant c), Neighborhood cleaning rule, offers the best accuracy results, and bearing in mind that the increase in the volume of observations is more than acceptable, it is considered the best option.

6 Conclusions

This document presents an analysis of various options for the augmentation of a binary dataset in the Mitre Att&ck framework. It is observed that the application of Neighborhood cleaning rule is the method that offers the best results, as it allows the balancing and augmentation of data while maintaining homogeneity in the different variables analyzed. On the other hand, the results obtained have been analyzed using a single algorithm, and without the adjustment of the hyperparameters themselves, which may allow future work to increase the precision reflected in this analysis.

References

1. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**, 1263–1284 (2009). https://doi.org/10.1109/TKDE.2008.239
2. Goorbergh, R., van Smeden, M., Timmerman, D., Van Calster, B.: The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. J. Am. Med. Inform. Assoc. JAMIA. **29** (2022). https://doi.org/10.1093/jamia/ocac093
3. Strom, B.E., et al.: Mitre att&ck: design and philosophy. In: Technical report. The MITRE Corporation (2018)
4. Working with ATT&CK | MITRE ATT&CK®. https://attack.mitre.org/resources/working-with-attack/. Accessed 11 Apr 2023
5. 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)
6. Han, H., Wang, W.-Y., Mao, B.-H.: Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: Huang, D.S., Zhang, X.P., Huang, G.B. (eds.) Advances in Intelligent Computing. ICIC 2005. LNCS, vol. 3644, pp. 878–887. Springer, Berlin, Heidelberg (2005). https://doi.org/10.1007/11538059_91
7. Tanimoto, A., Yamada, S., Takenouchi, T., Sugiyama, M., Kashima, H.: Improving imbalanced classification using near-miss instances. Expert Syst. Appl. **201**, 117130 (2022)
8. Two modifications of CNN. IEEE Trans. Syst. Man Cybern. **SMC-6**, 769–772 (1976). https://doi.org/10.1109/TSMC.1976.4309452
9. An experiment with the edited nearest-neighbor rule. IEEE Trans. Syst. Man Cybern. **SMC-6**, 448–452 (1976). https://doi.org/10.1109/TSMC.1976.4309523
10. Batista, G.E., Bazzan, A.L., Monard, M.C.: Balancing training data for automated annotation of keywords: a case study. In: WOB, pp. 10–18 (2003)
11. Laurikkala, J.: Improving identification of difficult small classes by balancing class distribution. In: Quaglini, S., Barahona, P., Andreassen, S. (eds.) Artificial Intelligence in Medicine. AIME 2001. LNCS, vol. 2101, pp. 63–66. Springer, Berlin, Heidelberg (2001). https://doi.org/10.1007/3-540-48229-6_9
12. Rout, N., Mishra, D., Mallick, M.K.: Handling imbalanced data: a survey. In: Reddy, M., Viswanath, K.K.M.S. (eds.) International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications. AISC, vol. 628, pp. 431–443. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-5272-9_39

13. Lunardon, N., Menardi, G., Torelli, N.: ROSE: a package for binary imbalanced learning. R J. **6**, 79–89 (2014). https://doi.org/10.32614/RJ-2014-008
14. Al, S., Dener, M.: STL-HDL: a new hybrid network intrusion detection system for imbalanced dataset on big data environment. Comput. Secur. **110**, 102435 (2021). https://doi.org/10.1016/j.cose.2021.102435
15. Torgo, L., Branco, P., Ribeiro, R.P., Pfahringer, B.: Resampling strategies for regression. Expert Syst. **32**, 465–476 (2015)
16. Dal Pozzolo, A., Caelen, O., Johnson, R.A., Bontempi, G.: Calibrating probability with undersampling for unbalanced classification. In: 2015 IEEE Symposium Series on Computational Intelligence, pp. 159–166. IEEE (2015)
17. scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation. https://scikit-learn.org/stable/. Accessed 11 Apr 2023
18. Yu, L., Liu, H.: Feature selection for high-dimensional data: a fast correlation-based filter solution. In: Proceedings, Twentieth International Conference on Machine Learning, pp. 856–863 (2003)

Author Index

A
Abdulahi, AbdulRahman Tosho 63
Abelha, António 230
Al Rajab, Murad 167
Alheeh, Emran 167
Alonso, Ricardo S. 53
Alonso-García, María 22
Alvear, Vanessa 256
Arrieta, Angélica González 53
Asad, Muhammad 250

B
Babatunde, Akinbowale Nathaniel 63
Babatunde, Ronke Seyi 63
Balderas-Díaz, Sara 103
Berral-Ortiz, Blanca 241
Bhatt, Chintan 143
Borja, Alejandro L. 209

C
Camara, Gaoussou 179
Carneiro, Davide 157
Carrasco-Martínez, Sara 43
Castillo, José Carlos 43, 220
Chamoso, Pablo 22
Chaudhari, Aayushi 143
Chaves, António 230
Chessa, Stefano 113
Conceição, Glória 133
Corchado, Juan M. 53, 143
Costa, Angelo 43
Curto, Javier 22, 33, 84

D
de la Cruz-Campos, Juan-Carlos 241
de la Rosa, Francisco López 189
Dimitri, Giovanna Maria 113
Domínguez, César 256

F
Fernández-Caballero, Antonio 74, 189, 209
Fernández-Sotos, Alicia 189
Fernández-Sotos, Desirée 189
Freitas, Alberto 133

G
Gomes, Luís 230
Gómez-Sirvent, José L. 189
Gómez-Villar, Alesandro 33
González-Briones, Alfonso 63
González-Ramos, Juan Antonio 33, 84
Gori, Marco 113
Guerrero, Graciela 74
Guerrero-Contreras, Gabriel 103
Guimarães, Miguel 157

H
Hazboun, Sherin 167
Hernández, Guillermo 22
Hornos, Miguel J. 13

K
Kaladzavi, Guidedi 179
Kimura, Risa 3
Kindo, Abdoul Azize 179
Kocian, Alexander 113
Kolyang 179
Krishna, Achyut 143

L
Lara-Lara, Fernando 241
López-Blanco, Raúl 84

M
Machado, José 230
Malo, Sadouanouan 179
Maroto-Gómez, Marcos 220
Marques-Villarroya, Sara 220

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Novais et al. (Eds.): ISAmI 2023, LNNS 770, pp. 269–270, 2023.
https://doi.org/10.1007/978-3-031-43461-7

Martinez-Rodrigo, Arturo 43
Mata, Gadea 256
Matos, Paulo 199
Mendes, Carla 93, 123
Miranda, Gonçalo 133
Montenegro, Larissa 230
Muñoz, Andrés 103

N
Nakajima, Tatsuo 3
Negre, Pablo 53
Novais, Paulo 157, 199

O
Odeh, Suhail 167
Ogundokun, Roseline Oluwaseun 63
Oliveira, Filipe 157
Oliveira, Pedro Filipe 199

P
Padovano, Daniele 43
Parra-Domínguez, Javier 33
Peixoto, Hugo 230
Pereira, António 74, 93, 123
Pereira, Rafael 93, 123
Pinto, Andreia 133
Pinto-Santos, Francisco 33, 84
Prieto, Javier 53

R
Ramos-Navas-Parejo, Magdalena 241
Ribeiro, José 93, 123
Ribeiro, Roberto 123
Rodrigues, Nuno 93
Rodríguez-Domínguez, Carlos 13

S
Salichs, Miguel Angel 220
Sánchez del Monte, Alberto 262
Sánchez-Reolid, Daniel 209
Sánchez-Reolid, Roberto 209
Santos, Cristina 133
Santos, Paulo 133
Serrano-Ortega, Alicia 22
Shaheen, Yaqin 13
Silva, Fernando José Mateus da 74
Sosa-Aleman, Aythami 220

T
Tapsoba, Théodore Marie Yves 179
Toledo-Garrote, Alberto 22

V
Viana, João 133