

LIVRODEATAS



SI24

# SIMPÓSIO DE ENGENHARIA INFORMÁTICA

Departamento de Engenharia Informática  
Instituto Superior de Engenharia do Porto  
*15 de janeiro de 2025*

## SEI'24 Proceedings Book

---

**TITLE:**

Simpósio de Engenharia Informática 2024: Proceedings Book

**EDITORS:**

Emanuel Silva, João Gil, José Marinho, Marílio Cardoso, Miguel Ferreira, Nuno Morgado, Paulo Proença, Piedade Carvalho, Ricardo Dias, Sérgio Moreira

**PUBLISHER:**

Instituto Superior de Engenharia do Porto (ISEP) – P.Porto

**DESIGN/LAYOUT:**

Gabinete de Design e Multimédia do ISEP

**VENUE/DATE:**

ISEP/January 15th, 2025

**ISBN:**

978-989-36167-0-3



This work is licensed under a Creative Commons License - Attribution-NonCommercial-NoDerivs 4.0 International.

Instituto Superior de Engenharia do Porto (ISEP)  
Rua Dr. António Bernardino de Almeida, 431  
4249-015 Porto  
phone: +351 22 83 40 500

[www.isep.ipp.pt](http://www.isep.ipp.pt)

<http://sei.dei.isep.ipp.pt>

©2025 ISEP – P.Porto

## Acknowledgment

We would like to thank everyone who contributed to the realisation of SEI'24.

First of all, we would like to express our gratitude to both the Presidency of the Engineering School of the Porto Polytechnic (ISEP) and the Management of the Department of Computer Engineering (DEI), for their continuous support and the resources made available that made this event possible.

We also thank the editors, reviewers, and session chairs who dedicated their time and expertise to ensuring the quality of the papers and presentations delivered during the symposium.

Special thanks to the Computer Engineering Center (NEI), students and participants, whose commitment, energy and enthusiasm were fundamental to the dynamics of the symposium.

We thank the guest speakers, who shared their valuable knowledge and experiences, enriching the discussions.

Furthermore, we would like to thank the technical team, the DEI secretariat and the ISEP Communication and Image Office for their essential support and collaboration throughout the event.

Last but not least, we would like to thank the companies and organisations that supported this event, contributing to the promotion of innovation and collaboration in the area of Computer Engineering and related areas.

# Contents

Acknowledgment . . . . .	3
Symposium's Vision . . . . .	6
Committees . . . . .	7
Technical Support and Secretariat . . . . .	10
Invited Speakers . . . . .	11
Papers . . . . .	18
Big Data e Análise de Dados na Tomada de Decisão Empresarial <b>Authors:</b> Cátia Santos, Mário Amorim, Vitor Ventura and Helder Rodrigo Pinto . . . . .	19
Integrated Management Platform for Pet Care Providers <b>Authors:</b> Leandro Fernandes and Emanuel Silva . . . . .	28
Optimizing Military Training Simulations: Development and Security Optimization of SisTrade's Simulation Platform <b>Authors:</b> Nuno Rocha and Emanuel Silva . . . . .	38
Using AI as a Development Accelerator <b>Authors:</b> Rui Soares and F. Jorge Duarte . . . . .	48
Cross-Platform Development Frameworks and Tools: Comparative study <b>Authors:</b> Tiago Machado and Helder Pinto . . . . .	58
NutriScan: Sistema De Monitorização Alimentar baseado em Inteligência Artificial para Reconhecimento de Imagem <b>Authors:</b> Ana Caraméz, Ana Marcelino, Beatriz Sousa, Bruna Cunha, Pedro Guimarães and Frederico Jacob . . . . .	67

Semi-automatic Image Tagging for ML workflows: a GUI for Industrial Applications <b>Authors:</b> Vasco Costa, Lobinho Gomes, Carlos José Campos, Veríssimo Lima, António Sousa and Fernando Carvalho . . . . .	77
Comparação de Arquiteturas Monolíticas e de Microserviços em Sistemas de Informação <b>Authors:</b> Ricardo Moura, Vitor Rocha, Vitor Ventura and Hélder Pinto . . . . .	87
Análise da distribuição geográfica de painéis fotovoltaicos em Portugal continental <b>Authors:</b> Tomás Nogueira, Beatriz Teixeira, Manuel Barbosa, Tatiana Teixeira, Alberto Pereira, Marílio Cardoso, Pedro Guedes and Telmo Matos . . . . .	96
Sistema Interativo de Visualização de Posters <b>Authors:</b> Bruno Sousa Ferreira, Carlos José Campos, António Sousa and Fernando Carvalho . . . . .	106
Accessible Tourism Travel Coach App <b>Authors:</b> João Anjos and Piedade Carvalho . . . . .	116
Deep Learning para a identificação de marcas de implantes dentários <b>Authors:</b> Diana Cardoso . . . . .	126
Simulação de Sinais de Oximetria de Pulso (SpO <sub>2</sub> ) <b>Authors:</b> Ana Gomes, Ângela Carvalho, Inês Ribeiro, Carlos Augusto Ramos and Gina Maria Vilão . . . . .	137
Interfaces Gráficas para Apoio ao Ensino de Séries Numéricas e Séries de Taylor/MacLaurin <b>Authors:</b> Marcos Menezes, André Pinho, Eduardo Almeida, Francisco Rêgo, Fábio Costa, Leonardo Coelho, Nuno Gonçalves, Vítor Cardoso, Helena Brás, Pedro Silva, António Sousa, Jorge Mendonça and Fernando Carvalho . . . . .	147
Deteção de Anomalias em Vagões Ferroviários Baseada em Machine Learning <b>Authors:</b> Ricardo Pereira, Pedro Ribeiro, António Sousa, Carlos José Campos, Veríssimo Lima, Jorge Mendonça and Fernando Carvalho . . . . .	157
<b>Authors Index . . . . .</b>	<b>167</b>

## Symposium's Vision

The Computer Engineering Symposium (SEI) provides a dynamic platform for showcasing the latest innovations, research discoveries, and emerging solutions in Computer Engineering and related fields. This year's edition shone a spotlight on Artificial Intelligence (AI), offering a unique opportunity for students, professionals, and researchers to convene, exchange ideas, and explore how generative AI is poised to transform various aspects of our lives, from education to industry and organizational functions.

SEI'24 emphasized the potential of AI as a tool to facilitate personalized learning and enrich the overall educational experience. Participants delved into a diverse array of topics, ranging from integrating AI technologies for increased accuracy in furniture detection and volume estimation to examining the impact of AI in higher education, including its promises and challenges.

Furthermore, the symposium showcased cutting-edge research and practical applications of AI across multiple domains. Participants shared insights on topics such as:

- Comparisons between monolithic and microservices architectures
- The development of AI-based food monitoring systems
- Machine learning applications in anomaly detection and image labeling
- Innovations in management platforms and military simulations
- Accessibility in tourism
- The use of deep learning to identify dental implant brands

SEI'24 aimed to inspire a new generation of innovators and leaders, empowering them to harness the potential of emerging technologies in a variety of contexts. By fostering knowledge exchange and showcasing real-world applications, the symposium promotes a well-informed generation capable of making decisions that benefit society as a whole.

# Committees

## Organization Committee

---

Emanuel Silva	Instituto Superior de Engenharia do Porto
João Gil	Núcleo de Estudantes de Engenharia Informática do ISEP
José Marinho	Instituto Superior de Engenharia do Porto
Marílio Cardoso	Instituto Superior de Engenharia do Porto
Miguel Ferreira	Núcleo de Estudantes de Engenharia Informática do ISEP
Nuno Morgado	Instituto Superior de Engenharia do Porto
Paulo Proença	Instituto Superior de Engenharia do Porto
Piedade Carvalho	Instituto Superior de Engenharia do Porto
Ricardo Dias	Núcleo de Estudantes de Engenharia Informática do ISEP
Sérgio Moreira	Instituto Superior de Engenharia do Porto

---

## Technical-Scientific Committee

---

A. Augusto Sousa	Faculdade de Engenharia da Universidade do Porto
Alexandre Gouveia	Instituto Superior de Engenharia do Porto
Álvaro Rocha	Instituto Superior de Economia e Gestão
Ana Raquel Faria	Instituto Superior de Engenharia do Porto
Ana Rita Teixeira	Instituto Superior de Engenharia do Porto
Anabela Serrano	Instituto Superior de Contabilidade e Administração do Porto
Bertil Marques	Instituto Superior de Engenharia do Porto
Emanuel Silva	Instituto Superior de Engenharia do Porto
Fátima Rodrigues	Instituto Superior de Engenharia do Porto
Fernando Duarte	Instituto Superior de Engenharia do Porto
João Pascoal Faria	Faculdade de Engenharia da Universidade do Porto
José Marinho	Instituto Superior de Engenharia do Porto
Luís Alves	Instituto Politécnico de Bragança
Luiz Faria	Instituto Superior de Engenharia do Porto
Manuela Raposo-Rivas	Universidade de Vigo
Marílio Cardoso	Instituto Superior de Engenharia do Porto
Mário Cruz	Escola Superior de Educação
Mário Pinto	Escola Superior de Media Artes e Design
Nuno Bettencourt	Instituto Superior de Engenharia do Porto
Olalla García Fuentes	Universidade de Vigo
Orlando Sousa	Instituto Superior de Engenharia do Porto
Paula Flores	Escola Superior de Educação
Paula Tavares	Instituto Superior de Engenharia do Porto
Paulo Baltarejo	Instituto Superior de Engenharia do Porto
Paulo Proença	Instituto Superior de Engenharia do Porto
Piedade Carvalho	Instituto Superior de Engenharia do Porto

---

## SEI'24 Proceedings Book

---

Ricardo Almeida	Instituto Superior de Engenharia do Porto
Ricardo Queirós	Escola Superior de Media Artes e Design
Rosa Reis	Instituto Superior de Engenharia do Porto
Rui Soares	Escola Superior de Tecnologia e Gestão
Sara Paiva	Instituto Politécnico de Viana do Castelo
Sérgio Moreira	Instituto Superior de Engenharia do Porto
Tânia Rocha	Universidade de Trás-os-Montes e Alto Douro
Telmo Matos	Instituto Superior de Engenharia do Porto
Teresa Terroso	Escola Superior de Media Artes e Design

# Technical Support and Secretariat

---

Carla Cunha	Instituto Superior de Engenharia do Porto
-------------	---

---

Lurdes Santos	Instituto Superior de Engenharia do Porto
---------------	---

---

Nuno Fonseca	Instituto Superior de Engenharia do Porto
--------------	---

---

Pedro Rocha	Instituto Superior de Engenharia do Porto
-------------	---

---

## Invited Speakers

---

Ana Almeida	ISEP
-------------	------

---

Dalila Durães	UMinho
---------------	--------

---

Fernando Moreira	UPT
------------------	-----

---

Filipe Prata	Softingal
--------------	-----------

---

Goreti Marreiros	ISEP
------------------	------

---

## Keynote Session

### ENHANCING MOVING SERVICES WITH AI: A DEEP LEARNING APPROACH FOR FURNITURE DETECTION

The moving services industry, traditionally reliant on manual processes and estimations, faces significant challenges in accurately identifying and estimating the volume of furniture for service quotations. This thesis explores the integration of advanced AI technologies to address these challenges, focusing on the development of a mobile application that leverages artificial intelligence for furniture detection and volume estimation for moving services. Machine Learning, particularly Deep Learning, has demonstrated considerable success in tackling complex problems such as object detection and image recognition. This thesis capitalizes on these advancements by initially employing traditional object detection models to classify and estimate the volume of furniture from static images. However, these models, while effective within certain constraints, were limited by their inability to accurately differentiate between overlapping classes, detect sizes, and handle complex furniture configurations. Recognizing these limitations, the research pivoted to integrate GPT-4o, a state-of-the-art multimodal AI model, which brought significant improvements in detection accuracy and contextual understanding. Alongside the development of this application, a thorough study was conducted on the evolution and effectiveness of different machine learning architectures, with a deep focus on Convolutional Neural Networks (CNNs) and their advancements in object detection tasks. This study provided a comprehensive comparison of these architectures, illustrating their strengths and weaknesses in the context of the moving services industry. The integration of GPT-4o into the system allowed for superior performance, particularly in scenarios where traditional models struggled. This enhanced the application's ability to deliver more accurate and reliable service quotations, ultimately improving operational efficiency and customer satisfaction. The thesis concludes by reflecting on the project's achievements, including the successful application of advanced AI models, and suggests avenues for future research, particularly in fine-tuning AI models for specific use cases and exploring new AI technologies as they emerge.



**Filipe Prata**  
Softingal

Filipe Prata é mestre em Inteligência Artificial pelo Instituto Superior de Engenharia do Porto. A sua dissertação, desenvolvida em parceria com a Softingal - Software and Consulting, Lda, focou-se na criação de uma aplicação mobile com integração de IA para o setor de mudanças, utilizando técnicas avançadas de aprendizagem profunda e RAG (Retrieval-Augmented Generation) para a classificação e estimativa de volumes de mobiliário a partir de imagens. Atualmente, é Engenheiro de Software na Softingal, onde aplica a sua experiência em desenvolvimento full-stack com diversas tecnologias. É apaixonado por explorar o potencial da Inteligência Artificial na resolução de desafios do mundo real.

## Round Table

### *IMPACTO DA IA GENERATIVA NO ENSINO SUPERIOR*

Moderator



**Goretí Marreiros**  
ISEP

Goretí Marreiros is a Coordinator Professor with Aggregation at the Department of Computer Engineering at the Instituto Superior de Engenharia do Porto (ISEP) – Instituto Politécnico do Porto (IPP), Director of GECAD - Research Group on Intelligent Engineering and Computing for Innovation and Development, and President of APPIA (Portuguese Association for Artificial Intelligence). She is the Deputy Director of the Master's in Engineering and Artificial Intelligence (MEIA) and Director of the postgraduate course in BigData & Decision Making at ISEP. Her research focuses on the scientific field of Artificial Intelligence and Intelligent Systems, with applications in different areas: Tourism (e.g. projects GIGADESSEA, EKRUCAmI, TheRoute, GROUPLANNER, SmarTravel, ATT); Industry 4.0 (e.g. projects MANTIS, InVALUE, INVALUE\_PT, NIS, PIANISM, Cyberfactory, SEC-OIIA, Ferrovia 4.0, Produtech\_R3); Healthcare and well-being (e.g. projects such as EKRUCAmI, PHE, AIRDOC, Inno4health, Secur-e-Health, Food-Friend). She is a co-author of more than 200 scientific articles published in international journals and conferences, and books. She has co-organized more than 20 international conferences, workshops, and seminars in the field of Artificial Intelligence. She has served as an expert for the Foundation for Science and Technology, National Innovation Agency, among other entities.

## Participants



**Ana Almeida**  
ISEP

Ana Almeida holds a PhD in Industrial and Systems Engineering (2003) from University of Minho, a MSc in Electronic and Computer Engineering Specialization in Industrial Informatics (1996) from University of Porto, a BSc in Industrial Informatics (1993) and a BSc in Electronic Engineering (1990) both from Polytechnic Institute of Oporto - Institute of Engineering. She is an Associate Teacher at the Institute of Engineering of Polytechnic of Porto (ISEP/IPP), Sub Director of two post-graduation programmes, in BigData & Decision Making and Industry 4.0. She is also the President of the ISEP Scientific Council since 2018, a council she has been a member of since 2007. Ana Almeida is responsible for the diverse accreditation processes of all ISEP Study Programmes, namely: A3ES (Agency for Assessment and Accreditation of Higher Education) and EUR-ACE (a framework and accreditation system that provides a set of standards that identifies high-quality engineering degree programmes in Europe and abroad). She is a member of the A3ES board of programme evaluators, and a member of ANI (Nacional Agency for Innovation) board of I&D projects evaluators. In both these positions, she has already evaluated several study programmes and research projects. She is a Senior Researcher of GECAD (Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, UIDP/00760/2020 and UIDB/00760/2020), a R&D unit ranked as Excellent by the Portuguese Science&Technology Foudation (FCT). She is also member of LASI (Associated Laboratory of Intelligent Systems LA/P/0104/2020), the Portuguese AI Associated Lab recognized by FCT. In the past few years, she has taught and been responsible for several courses on Decision Support Systems, Information Systems and Artificial Intelligence at the graduate and postgraduate levels. Participates and/or participated as Principal investigator (PI) or Co-Principal investigator (Co-PI) in 6 projects and Researcher in more than 20 project(s). She works in the area(s) of ambient Intelligence, decision support systems, and intelligent systems. Her R&D work had applications to different domains, namely Tourism; Industry 4.0; Health Care and Wellbeing. Published several papers in journals, books, and international conferences. Organized several international conferences/workshops. Supervised or co-supervised PhD thesis(es) and MSc dissertation(s). Throughout

her career, she has held various positions, including Director of the CICESI research group at the IPP School of Technology and Management (2003-2008), Head of ISEP's Department of Computer Engineering (2007-2014) and President of ISEP's Pedagogical Council (2014-2016).



## Dalila Durães

UMinho

Dalila Durães is a Assistant Professor of Computer Science at the Department of Informatics, the School of Engineering, the University of Minho (Portugal), and a researcher at the ALGORITMI Centre and LASI Intelligent Systems Associate Laboratory. She holds two international PhD: one in Educational Sciences, in Teachers, Curricula and Educational Institutions, from the University of Granada, Spain, completed in 2012, and another in Artificial Intelligence by the Polytechnic University of Madrid, Spain, completed in 2018. She is vice-chair of the IEEE in Computational Intelligence Society, Portugal Chapter and vice-chair of the Portuguese Association for Artificial Intelligence (APPIA). She began her career in 2002, developing scientific research in nanotechnology until 2004. Subsequently, between 2007 and 2013, scientific research in education. From 2015, she developed research in Computer Engineering, with applications in Human-Computer Interaction, Behavior Analysis, with particular attention to the detection of violence, sentiment analysis, Intelligent Tutoring, Recognition of Human Actions, and LLM. Dalila is the author of more than 80 scientific publications in international peer-reviewed journals, book chapters, books, and conferences. She is also a member of the editorial board of several international journals. During the past few years, she has served as an expert/reviewer for several conferences and journals. She was also a supervisor of master's and doctoral students and was a member of the jury of several doctoral and master's theses.



## Fernando Moreira

UPT

Fernando Moreira fundou o Departamento de Ciência e Tecnologia da Universidade Portucalense, criou os primeiros cursos de Engenharia e foi o seu Diretor entre maio de 2018 e fevereiro de 2022. Licenciado em Informática/Matemáticas Aplicadas (1992), Mestre (1997) e Doutoramento (2003) em Engenharia Eletrónica e de Computadores pela FEUP, obteve a Agregação em 2018. É Professor Catedrático na Universidade Portucalense. Supervisiona estudantes de mestrado e doutoramento, com mais de 275 publicações indexadas na Scopus. É membro de conselhos editoriais de revistas científicas, organiza edições especiais de revistas JCR e atua regularmente como Chair e orador em conferências internacionais. Foi coordenador do Mestrado em Informática durante uma década e integra a Comissão Diretiva do REMIT. Recebeu o prémio Atlas Elsevier em 2019.

# Papers



# Big Data e Análise de Dados na Tomada de Decisão Empresarial

Cátia Santos<sup>1</sup>, Mário Amorim<sup>1</sup>, Vítor Ventura<sup>1</sup>, and Helder Rodrigo Pinto<sup>2</sup>

<sup>1</sup> ISLA Gaia, Vila Nova de Gaia, Portugal

<sup>2</sup> ISEP, Porto, Portugal

r.catia1990@gmail.com

**Resumo.** O artigo enfatiza a importância do Big Data na transformação digital das empresas, destacando a BI4ALL, uma consultora portuguesa especializada em Big Data Analytics. A análise de dados e a inteligência de negócios são apresentadas como fundamentais para capacitar organizações. O texto explora como o Big Data é crucial para decisões informadas, otimização de operações e personalização da experiência do cliente, com aplicações em áreas como marketing, vendas, desenvolvimento de produtos e gestão de riscos.

São discutidos casos práticos da aplicação do Big Data pela BI4ALL em diversos setores, mostrando como soluções personalizadas melhoram operações e estratégias empresariais. O artigo também aborda os desafios éticos e de privacidade na análise de grandes volumes de dados, destacando a importância da conformidade com o RGPD. Esses exemplos ilustram como a BI4ALL auxilia empresas a enfrentar os complexos desafios éticos e legais do Big Data, assegurando práticas responsáveis e em conformidade com as normas vigentes.

O texto explora tendências futuras e inovações em Big Data, como a integração da inteligência artificial, DataOps, Data Mesh e o uso de plataformas na cloud. Destaca-se a necessidade de adaptação das estratégias empresariais para manter a competitividade e inovação num mercado em constante evolução. O artigo oferece uma visão abrangente sobre o impacto do Big Data nas estratégias empresariais e ressalta o papel da BI4ALL como parceiro estratégico essencial, demonstrando que a capacidade de analisar e interpretar grandes volumes de dados é crucial para o sucesso na era digital.

**Palavras-chave:** Big Data, Análise de Dados, Tomada de Decisão, Inteligência de Negócios

## 1 Introdução

Na era digital atual, o Big Data emergiu como um elemento transformador na gestão empresarial. A evolução tecnológica das últimas décadas permitiu às organizações recolher e processar volumes de dados sem precedentes, criando novas oportunidades e desafios. "O Big Data tem-se tornado um recurso inestimável para as empresas, oferecendo insights sobre tendências do mercado e comportamento do consumidor" [4].

A análise de grandes volumes de dados revolucionou a forma como as organizações tomam decisões estratégicas. "A expertise em Big Data permite às empresas não só recolher e armazenar grandes volumes de dados, mas também analisá-los eficazmente para extrair informações valiosas" [8]. Esta capacidade tornou-se crucial num ambiente empresarial cada vez mais competitivo e dinâmico.

No entanto, a implementação eficaz de soluções de Big Data apresenta desafios significativos. As organizações enfrentam questões complexas relacionadas com infraestrutura tecnológica, segurança de dados, privacidade e conformidade regulamentar. "Na era digital, a informação é o novo petróleo, e as empresas que sabem explorar este recurso têm uma vantagem competitiva significativa" [14].

Este artigo analisa o impacto do Big Data na tomada de decisão empresarial, explorando casos práticos do mercado português. Utilizando uma metodologia baseada em estudos de caso, examina-se a implementação de soluções de análise de dados em diferentes setores, incluindo seguros, educação e hotelaria. O estudo foca-se particularmente nas experiências da BI4ALL, uma consultora portuguesa reconhecida no setor de analytics.

A estrutura do artigo compreende: fundamentação teórica sobre Big Data, análise de casos práticos de implementação, discussão dos desafios éticos e de privacidade, e análise das tendências futuras. Esta investigação contribui para a compreensão das melhores práticas na implementação de soluções de Big Data no contexto empresarial português.

## **2 Fundamentos de Big Data nas Empresas**

### **2.1 Big Data e a Revolução na Análise de Dados**

O conceito de "Big Data" é definido por conjuntos de dados de dimensões tão vastas que desafiam os métodos tradicionais de processamento. "Caracteriza-se pelo volume extraordinário, pela diversidade de tipos de dados e pela velocidade com que são gerados e processados" [17].

A evolução tecnológica em termos de capacidade de armazenamento e poder computacional tem sido fundamental para lidar com estes grandes volumes de dados. "O desenvolvimento de tecnologias de armazenamento distribuído e computação paralela revolucionou a nossa capacidade de processar e analisar grandes volumes de dados" [11]. Esta evolução permitiu o surgimento de data centers mais eficientes e sistemas de armazenamento escaláveis, fundamentais para o processamento de Big Data.

O Data Mining surge como uma componente crucial do Big Data, permitindo descobrir padrões, tendências e correlações ocultas nos dados. "As técnicas de Data Mining, como clustering, classificação e regras de associação, são essenciais para extrair conhecimento significativo de grandes volumes de dados" [12]. Esta capacidade de mineração de dados transformou a forma como as empresas compreendem os seus clientes e otimizam as suas operações.

## 2.2 Importância e Mudança de Foco nas Empresas

A ascensão do Big Data marca uma transformação significativa nas empresas modernas. "Transforma-se num recurso estratégico que permite às organizações tomar decisões mais informadas, identificar novas oportunidades de mercado e personalizar as experiências dos clientes" [5]. Esta mudança evidencia-se na transição das empresas, de uma abordagem baseada em intuições, para uma orientada por dados, resultando em estratégias empresariais mais eficientes e sustentáveis.

## 2.3 Obtenção, Uso e Estratégias de Dados

As empresas recolhem dados de uma multiplicidade de fontes, incluindo transações online, redes sociais, dispositivos IoT (Internet of Things), e registos de clientes. "A eficiência no processamento e análise destes dados é possível graças a tecnologias como Apache Hadoop e Spark" [21]. No contexto empresarial, o Big Data é utilizado em diversas estratégias, que vão desde a personalização de ofertas em marketing e vendas até à otimização de processos operacionais e desenvolvimento de produtos inovadores. A análise de riscos financeiros e operacionais também beneficia desta abordagem, permitindo decisões estratégicas mais fundamentadas.

## 2.4 Estratégias Baseadas em Big Data

Na prática, as empresas aplicam o Big Data em várias áreas:

- Marketing e Vendas: Personalização de ofertas e serviços para os clientes, otimização de campanhas de marketing, e previsão de tendências de consumo através de técnicas avançadas de Data Mining [7]. "A análise preditiva em marketing tem demonstrado aumentos de até 15-20% na eficiência das campanhas" [15].
- Operações: Melhoria da eficiência operacional, otimização da cadeia de fornecimento e redução de custos através da análise preditiva. "A implementação de análise de Big Data nas operações pode reduzir custos operacionais em até 25%" [9].
- Desenvolvimento de Produto: Inovação baseada no feedback dos clientes e identificação de oportunidades de mercado emergentes utilizando análise de sentimentos e padrões de comportamento. "O uso de análise de dados no desenvolvimento de produtos reduz o tempo de lançamento em média 20%" [3].
- Gestão de Riscos: Análise de riscos financeiros, operacionais e de segurança através de modelos preditivos e técnicas de machine learning. "Sistemas baseados em Big Data podem identificar fraudes com precisão superior a 90%" [18].

# 3 Análise de Dados e Inteligência de Negócios

## 3.1 Processo de Análise de Dados nas Empresas

A análise de dados nas empresas constitui um processo detalhado e metódico que se inicia com a recolha de dados de fontes variadas, incluindo transações online, interações

nas redes sociais e sistemas internos. Estes dados são posteriormente preparados e purificados para assegurar a sua precisão e relevância. "A utilização de tecnologias como o Hadoop para armazenamento e o Spark para processamento rápido permite às empresas gerir eficientemente grandes volumes de dados" [16]. A fase seguinte centra-se na análise propriamente dita, onde entram em cena ferramentas de Business Intelligence (BI) e análise de dados, tais como o Tableau ou o Power BI. Estas ferramentas facilitam a visualização dos dados, a identificação de tendências, padrões e anomalias, e a sua transformação em insights acionáveis, frequentemente através de dashboards interativos, relatórios e análises, constituindo uma base sólida para a tomada de decisões.

### **3.2 Transformação de Dados em Insights para Decisões de Negócios**

A conversão de dados brutos em insights relevantes é crucial para decisões empresariais informadas. "Por exemplo, a análise de dados de vendas e comportamento do cliente pode revelar tendências de mercado, auxiliando as empresas a ajustar as suas estratégias de marketing e desenvolvimento de produtos" [19]. Similarmente, a análise de dados operacionais pode resultar na otimização de processos e na redução de custos.

### **3.3 Casos Práticos de Decisões Aprimoradas pela Análise de Dados**

- Marketing Direcionado: As empresas podem utilizar análises de dados para compreender melhor os seus clientes e segmentar campanhas de marketing, resultando em taxas de conversão superiores e maior satisfação do cliente.
- Previsão de Necessidades: Através da análise de dados, uma empresa de retalho pode antecipar a procura por produtos, otimizando o seu stock e evitando excessos ou carências.
- Desenvolvimento de Produtos: Os insights obtidos da análise de feedbacks dos clientes permitem às empresas inovar e ajustar os seus produtos às necessidades e preferências dos consumidores.
- Gestão de Riscos: No setor financeiro, a análise de dados é utilizada para avaliar riscos de crédito, otimizar carteiras de investimento e prevenir fraudes.

Estes casos demonstram como a análise de dados e inteligência de negócios podem ter um impacto significativo em diversas áreas das empresas, conduzindo a decisões mais estratégicas e informadas.

## **4 Aplicações Práticas de Big Data na Empresa**

A BI4ALL (<https://bi4allconsulting.com/>), consultora portuguesa especializada em Big Data e análise de dados, tem transformado organizações através de soluções inovadoras de gestão de dados. Através dos seus casos de sucesso, destaca-se a implementação de soluções tecnológicas que permitem às empresas maximizar o valor dos seus dados ([https://bi4allconsulting.com/knowledgecenter\\_categoria/casos-de-sucesso/](https://bi4allconsulting.com/knowledgecenter_categoria/casos-de-sucesso/))

#### 4.1 Caso de Estudo: Mudum Seguros

Fundada em 1996, a Mudum Seguros, que faz parte do grupo Crédit Agricole, dedica-se a fornecer seguros não-vida para o mercado de particulares.

Pauta-se pela orientação ao cliente, excelência, iniciativa individual, honestidade e integridade, e ainda espírito de equipa. Visa contribuir para a confiança e segurança das pessoas, distinguindo-se pela simplicidade, clareza e qualidade de serviço.

O Desafio:

- Inovação em Dados: A Mudum Seguros utiliza dados para operar e inovar há mais de 20 anos, adaptando as tarifas de seguros de acordo com o perfil do cliente.
- Necessidade Estrutural: Com mudanças acionistas, surgiu a necessidade de revisão do parque tecnológico e da estrutura organizacional, com foco reforçado na área de dados.

A Solução pela BI4ALL:

- Estruturação de Dados: A BI4ALL ajudou a criar uma estrutura organizativa dedicada a dados, revendo a arquitetura tecnológica de suporte a dados e participando na criação do catálogo e glossário de dados da companhia.
- Gestão de Dados e Ferramentas: Assistência na escolha de ferramentas de gestão de dados e apoio na execução de provas de conceito, além de suporte na implementação de gestão, boas práticas e regras de utilização de tecnologia, como o Power BI.

Os Resultados:

- Adoção Interna de Ferramentas de BI: Implementação e uso eficaz das ferramentas de trabalho para exploração de dados e produção de informação.
- Foco nos KPIs(Indicadores-Chave de Desempenho) e Satisfação do Cliente: Disponibilização centralizada dos KPIs de cada área para contribuir ativamente para a tomada de decisão e melhoria das ofertas no mercado.
- Alinhamento Estratégico e Novos Insights: Criação de políticas de dados e gestão de informação alinhadas estrategicamente com o negócio, promovendo a criação de novos insights e a melhoria geral do desempenho da organização.

Este caso exemplar da Mudum Seguros serviu como modelo para outras implementações bem-sucedidas da BI4ALL, como na Universidade de Aveiro e Pestana Hotel Group, onde metodologias similares foram adaptadas às necessidades específicas de cada setor.

## **5 Desafios Éticos e de Privacidade na Análise de Grandes Volumes de Dados**

### **5.1 Introdução às Implicações Éticas e Desafios de Privacidade**

No contexto atual, marcado pelo avanço tecnológico e pela dependência crescente da Internet, a recolha e o uso de dados pessoais tornaram-se uma questão premente, desencadeando preocupações significativas com a privacidade online. Com a integração de indivíduos, empresas e dispositivos numa rede global, surge a necessidade de regulamentação ética e legal, particularmente no que diz respeito à proteção de dados pessoais em ambientes privados e públicos. "A necessidade de regulamentação ética e legal é imperativa, especialmente para a proteção de dados pessoais em ambientes privados e públicos" [20].

### **5.2 Desafios na Implementação do RGPD e Compliance**

A implementação do Regulamento Geral sobre a Proteção de Dados (RGPD) exigiu que as organizações revissem as suas práticas de privacidade e ajustassem as suas políticas para estarem em conformidade com as disposições do regulamento. Isto resultou numa maior transparência no uso de dados pessoais e numa crescente consciencialização sobre os direitos individuais de privacidade. O termo "compliance" (conformidade) refere-se ao dever de cumprir regulamentos internos e externos, sendo essencial para as empresas no contexto do RGPD. Implica a adoção de práticas e medidas de segurança mais robustas para proteger os dados dos utilizadores. "O cumprimento do RGPD resultou numa maior transparência no uso de dados pessoais e num aumento da consciencialização sobre os direitos individuais de privacidade" [6].

#### **Implementação do RGPD nos Casos de Estudo da BI4ALL.**

A BI4ALL (<https://bi4allconsulting.com/>), nos seus casos de implementação de Big Data Analytics, incorpora medidas específicas para garantir a conformidade com o RGPD:

- Mudum Seguros: Implementação de catálogo de dados com classificação de dados pessoais e definição de políticas de acesso alinhadas com o RGPD.
- Universidade de Aveiro: Desenvolvimento de protocolos de proteção de dados dos estudantes e funcionários no portal de indicadores.
- Pestana Hotel Group: Implementação de medidas de segurança no repositório único de dados para proteção de dados pessoais dos clientes.

### **5.3 Impacto Global e Adoção de Medidas Semelhantes noutros Países**

Além do RGPD, outros países e regiões também adotaram medidas semelhantes para proteger a privacidade online. Por exemplo, no Brasil, foi implementada a Lei Geral de

Proteção de Dados (LGPD) em agosto de 2020, com objetivos e abordagens semelhantes ao RGPD. Tais regulamentações são fundamentais para consciencializar as pessoas sobre a importância da privacidade online e seus direitos relacionados aos dados pessoais, levando a uma mudança cultural em relação à privacidade e à proteção de informações. O impacto do RGPD estende-se além das fronteiras da União Europeia, influenciando a adoção de regulamentos semelhantes em diferentes países. "Este regulamento serve como um modelo para a proteção de dados num contexto global, realçando a importância da privacidade online e dos direitos relacionados aos dados pessoais" [6].

#### **5.4 Conclusão**

As empresas deparam-se com o desafio de adotar políticas de privacidade rigorosas, garantir a segurança dos dados, realizar auditorias e avaliações de risco regulares, e promover a formação contínua em ética e privacidade de dados. "Essas medidas são cruciais para manter a confiança dos utilizadores e garantir a conformidade com as normativas vigentes, assegurando a proteção efetiva dos dados pessoais no ambiente digital" [1].

## **6 Tendências Futuras e Inovação**

### **6.1 Inteligência Artificial, Análise de Dados e Automatização**

A integração da inteligência artificial na análise de dados está a revolucionar a forma como as empresas interpretam o Big Data. "A IA está a mudar drasticamente a maneira como interagimos e analisamos grandes volumes de dados" [13]. Especialmente em empresas como a BI4ALL, onde "o foco é transformar dados em insights relevantes para o negócio" [2], a IA permite análises mais rápidas, precisas e a automatização de tarefas repetitivas. "Uma solução de IA permite reduzir custos e aumentar a produtividade, automatizando tarefas ou otimizando processos de negócio de forma confiável e precisa" [2].

### **6.2 DataOps, Data Mesh e Colaboração de Dados**

O DataOps está a emergir como uma prática essencial para as empresas que desejam otimizar o uso de dados. Combinando os princípios de DevOps com a Ciência de Dados, "o DataOps permite a agilidade e a eficiência na análise de dados" [10]. Em paralelo, o conceito de Data Mesh está a promover a democratização e partilha de dados dentro das organizações, "promovendo a colaboração de dados em ambientes regulamentados e protegidos como os Data Clean Rooms" [10].

### **6.3 Big Data na Nuvem e Visualização de Dados**

As plataformas baseadas em nuvem estão a tornar-se cada vez mais populares para gerir Big Data. "Estas plataformas oferecem uma solução eficiente para reunir, organizar e

analisar grandes volumes de dados” [10]. Além disso, a visualização de dados é uma ferramenta poderosa para apresentar informações complexas de forma compreensível, "facilitando a tomada de decisão baseada em dados para públicos não técnicos" [10].

#### **6.4 Eficiência Operacional e Melhoria da Experiência do Cliente**

A eficiência operacional e a melhoria da experiência do cliente são focos principais para as empresas que utilizam Big Data. "Soluções de IA como as desenvolvidas pela BI4ALL para otimizar chamadas de clientes e processos de recepção estão a transformar a eficiência operacional das empresas" [2]. Estas inovações permitem atender mais clientes em menos tempo e melhorar significativamente a satisfação do cliente.

Esta secção ressalta a importância de adaptar as estratégias empresariais para aproveitar ao máximo as inovações em Big Data, garantindo que as empresas se mantenham competitivas e inovadoras num mercado em constante evolução.

### **7 Conclusão**

Este artigo apresenta uma perspetiva abrangente sobre o impacto do Big Data nas estratégias empresariais, destacando o papel essencial da BI4ALL como parceiro estratégico nesta área. Conclui-se que, numa era em que a informação é um ativo de elevado valor, a capacidade de analisar e interpretar grandes volumes de dados tornou-se indispensável para o sucesso empresarial.

Foi explorada a importância estratégica do Big Data na transformação digital das empresas, com enfoque na BI4ALL, uma consultora portuguesa de referência em Big Data Analytics. O Big Data é apresentado como um recurso fundamental, permitindo às empresas obter insights valiosos e tomar decisões estratégicas baseadas em dados concretos. A revolução na análise de dados alterou significativamente a forma como as empresas compreendem o mercado, interagem com os clientes e otimizam operações, marcando uma transição para estratégias mais eficientes e sustentáveis.

Os casos práticos analisados demonstram como a BI4ALL, através de soluções personalizadas de Big Data, tem contribuído para melhorar operações e estratégias em diversos setores. Estes exemplos reforçam a relevância de analisar e interpretar grandes volumes de dados para alcançar sucesso no contexto empresarial atual. O artigo também aborda os desafios éticos e de privacidade inerentes à análise de grandes volumes de dados, sublinhando a importância de cumprir regulamentações como o RGPD. As empresas enfrentam desafios significativos na implementação de políticas de privacidade rigorosas, assegurando a segurança dos dados e promovendo formação contínua em ética e privacidade.

Por fim, foram exploradas tendências futuras e inovações relacionadas com o Big Data, incluindo a integração da inteligência artificial, DataOps, Data Mesh e o uso de plataformas na cloud. Estas inovações reforçam a necessidade de as empresas adaptarem as suas estratégias para manterem a competitividade num mercado em constante evolução, sublinhando o papel crucial da BI4ALL como parceiro estratégico para empresas que pretendem prosperar nesta nova era.

## Referências

1. BENNETT, C. J. [2016], "The Privacy Advocates: Resisting the Spread of Surveillance", MIT Press, 2016.
2. BI4ALL. "AI & Data Analytics eBook", visualizado no dia 03 de dezembro de 2023. Ebook Disponível em: [www.bi4allconsulting.com](http://www.bi4allconsulting.com)
3. BROWN, B. et al. [2011], "Are You Ready for the Era of 'Big Data'?", McKinsey Quarterly, 2011.
4. BRYANT, R. et al. [2017], "The Impact of Big Data on Business Strategy", Journal of Business Research, 2017.
5. BUGHIN, J., CHUI, M., & MANYIKA, J. [2010], "Big Data: The Next Frontier for Innovation, Competition, and Productivity", McKinsey Global Institute, 2010.
6. BUSINESS IT. [2020], 18 de junho de 2020. Visualizado no dia 05 de dezembro de 2023. Disponível em: <https://business-it.pt/2020/06/18/bi4all-reconhecida-como-uma-das-dez-melhores-empresas-de-big-data-analytics-da-europa/>
7. CHEN, H., CHIANG, R. H., & STOREY, V. C. [2012], "Business Intelligence and Analytics: From Big Data to Big Impact", MIS Quarterly, 2012.
8. CHEN, M. et al. [2014], "Big Data: Related Technologies, Challenges and Future Prospects", Springer, 2014.
9. DAVENPORT, T. [2014], "Big Data at Work: Dispelling the Myths, Uncovering the Opportunities", Harvard Business Review Press, 2014.
10. DATA SCIENCE ACADEMY, Visualizado no dia 02 de dezembro de 2023. Disponível em: <https://blog.dsacademy.com.br>
11. DEAN, J. & GHEMAWAT, S. [2008], "MapReduce: Simplified Data Processing on Large Clusters", Communications of the ACM, 2008.
12. HAN, J., KAMBER, M., & PEI, J. [2011], "Data Mining: Concepts and Techniques", Morgan Kaufmann, 2011.
13. INDICIUM TECH BLOG, Visualizado no dia 26 de novembro de 2023. Disponível em: <https://blog.indicium.tech>
14. KHAN, N. et al. [2017], "Big Data: A Paradigm Shift in the Business World", Economic Journal, 2017.
15. KUMAR, V. et al. [2013], "Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know", John Wiley & Sons, 2013.
16. MARZ, N. & WARREN, J. [2015], "Big Data: Principles and Best Practices of Scalable Realtime Data Systems", Manning Publications, 2015.
17. MAYER-SCHÖNBERGER, V. & CUKIER, K. [2013], "Big Data: A Revolution That Will Transform How We Live, Work, and Think", Houghton Mifflin Harcourt, 2013.
18. NGAI, E.W.T. et al. [2011], "The Application of Data Mining Techniques in Financial Fraud Detection", Decision Support Systems, 2011.
19. PROVOST, F. & FAWCETT, T. [2013], "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking", O'Reilly Media, 2013.
20. SOLOVE, D. J. & HARTZOG, W. [2014], "The Future of Privacy", American University Law Review, 2014.
21. WHITE, T. [2012], "Hadoop: The Definitive Guide", O'Reilly Media, 2012.

# Integrated Management Platform for Pet Care Providers

Leandro Fernandes<sup>[0009-0001-4099-1721]</sup>, and Emanuel Silva<sup>[0000-0002-5046-8942]</sup>

Instituto Superior de Engenharia do Porto, Porto, Portugal  
{1211118,ecs}@isep.ipp.pt

**Abstract.** This paper presents the development of an integrated management platform for pet care providers, addressing the challenges faced by professionals in this field due to fragmented tools and inefficient workflows. By adopting a modular architecture and leveraging technologies such as FlutterFlow [1] and Firebase [2], the proposed platform offers comprehensive solutions including scheduling, client and pet management, financial operations, and real-time communication. The platform's features enhance productivity, streamline operations, and improve service quality, while ensuring data security and scalability. By centralizing critical management tasks, the solution enables animal care professionals to focus more on delivering high-quality services and less on administrative complexities. Preliminary user feedback highlights the platform's user-friendly interface and its potential to transform the Portuguese pet care market. The development methodology emphasizes continuous improvement, with planned beta testing phases aimed at refining the system based on real-world user insights. By bridging technological gaps and providing an integrated management solution, this platform represents a significant innovation in animal care service management, promising to elevate professional standards and operational excellence in the industry.

**Keywords:** Pet Care Management, Mobile Application Development, Low-Code Development, FlutterFlow, Firebase.

## 1 Introduction

### 1.1 Context

Fresh Observation Unipessoal Lda. [3] is a Portuguese company based in Alvaiázere, founded in 2023, focused on the pet care market. The company aims to provide innovative solutions to the challenges faced by pet owners, including finding reliable pet sitters and managing the health and well-being of their animals. Recognizing the need for an integrated solution for pet sitters to manage their daily activities effectively, this project was initiated.

### 1.2 Problem Description

Pet sitters currently face difficulties in managing their operations due to reliance on disparate tools such as WhatsApp for communication, phone calendars for scheduling,

and billing software for invoicing. This fragmentation leads to inefficiencies and potential gaps in service delivery and customer satisfaction.

### 1.3 Goals

The goal of this project is to develop a comprehensive management software for pet sitters that integrates functionalities for scheduling, client management, pet management, invoicing and communication.

## 2 State of the art

### 2.1 Related Works

The mobile app market for pet sitters is growing rapidly, with prominent solutions like Rover [4], PetBacker [5], Wag! [6], and TrustedHousesitters [7] offering various features but leaving gaps that the proposed solution aims to address.

Rover is the global leader in pet sitting and dog walking, with features like scheduling, GPS tracking, online payments, and pet sitter ratings. However, it is not available in Portugal. Similarly, Wag! provides scheduling and flexible pricing but lacks availability in Portugal. In contrast, PetBacker and TrustedHousesitters are accessible in Portugal and focus on connecting pet owners with sitters or house caretakers, though both lack comprehensive client and financial management systems.

### 2.2 Identified Gaps

Academic literature also highlights some of these gaps. For instance, Flockr [8] explores how pet care apps can integrate more efficient management functionalities, while Easypet [9] emphasizes the need for systems that manage both pets and their owners comprehensively. Furthermore, Grandpaw [10] discusses the increasing demand for customizable features tailored to specific pet needs, suggesting that a more personalized approach is essential to address these market gaps.

### 2.3 Existing Technologies

Fresh Observation employs FlutterFlow for frontend development and Firebase for backend services. FlutterFlow offers rapid cross-platform development with a user-friendly interface, though its flexibility is limited for complex requirements. Firebase ensures scalability and seamless integration with Google services but incurs significant costs at scale.

Alternative technologies, such as React Native and native development (Java/Kotlin/Swift), provide greater flexibility and performance but require higher development effort and expertise. Fresh Observation prioritizes rapid prototyping and scalability, justifying its technological choices.

### 3 Analysis and Design of the Solution

This section presents the analysis and design of a solution aimed at optimizing pet sitter and pet management, addressing the needs of pet owners, animals, and the professionals involved. Understanding the problem domain is crucial for developing a system that meets these needs, and this work outlines the relevant entities, as well as the proposed architecture.

#### 3.1 Domain Problem

The analysis of the domain problem revealed several interconnected entities that play essential roles in the pet care ecosystem. The main entities identified include:

**Pet Sitters:** Professionals responsible for direct animal care, whose efficiency relies on clear communication with pet owners and understanding individual animal needs;

**Pet Parents:** Individuals providing crucial information about their pets, such as medical history and preferences, with specific expectations regarding services;

**Pets:** Each animal has unique characteristics influencing the type of care required, including breed, age, and health;

**Service Types:** The solution must encompass a variety of services, from basic care to specialized offerings, ensuring flexibility to meet pet owner needs;

**Appointments:** Efficient management of appointments is critical for timely service delivery.

#### 3.2 System Architecture

The proposed solution is based on a microservices architecture [11], which separates the application into independent services, each focused on a specific function. This modular approach allows better code organization and facilitates system maintenance and scalability. The diagram in **Fig. 1** illustrates the system architecture.

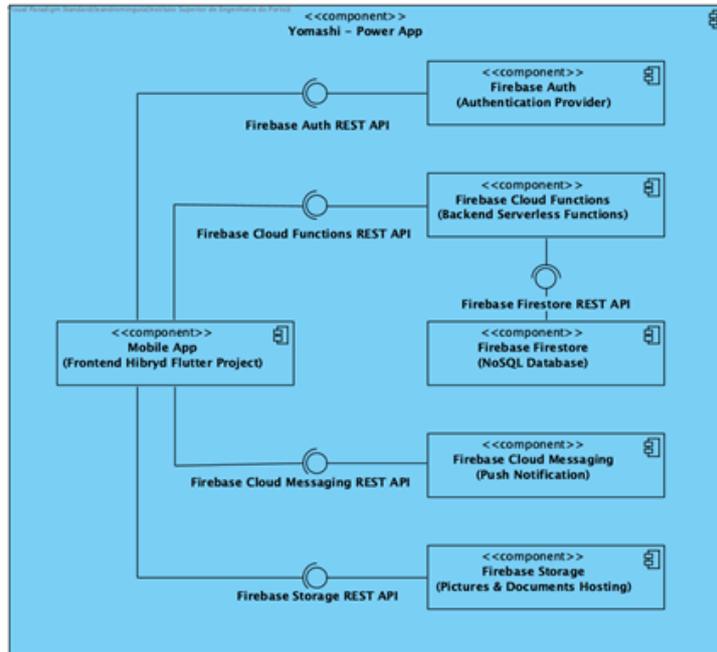


Fig. 1. System Architecture

The choice of a microservices architecture is justified by the need for flexibility, scalability, and security. Each service can be scaled independently, allowing efficient responses to workload spikes. Additionally, data security is ensured through specific access policies for each service, safeguarding sensitive information.

The microservices architecture, combined with a detailed understanding of the problem domain and the functional and non-functional requirements, provides a solid foundation for developing a system that meets user needs and promotes a satisfactory user experience. The next step involves the implementation and testing of the proposed solution to validate its effectiveness and efficiency in a real-world environment.

## 4 Implementation

The platform provides several integrated functionalities aimed at streamlining pet sitters workflows. Below is a detailed exploration of the main modules, supported by visual representations.

### 4.1 Pet Parent Management

This module is responsible for managing pet owners, also referred to as pet parents. It enables registration, updating, and viewing of detailed information about animal caretakers. For example, a pet sitter can create a new pet parent by entering their contact details and address in the system, ensuring smooth communication for service appointments. **Fig. 2** illustrates the user interface for pet parent registration.

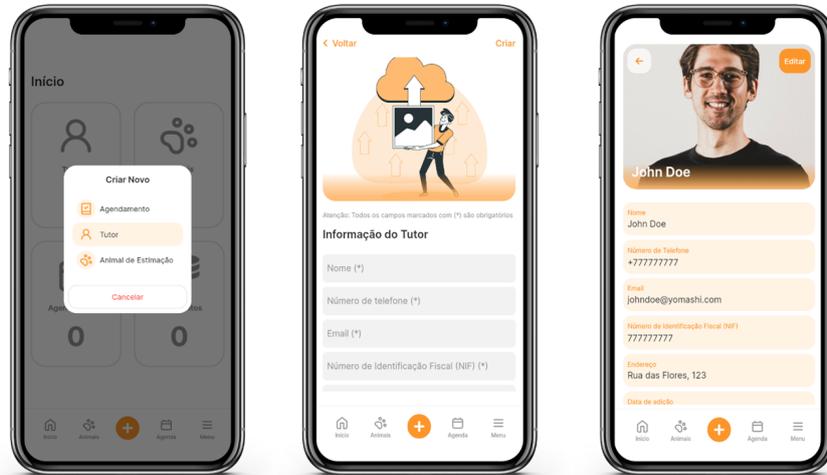


Fig. 2. - Pet Parent Creation Process

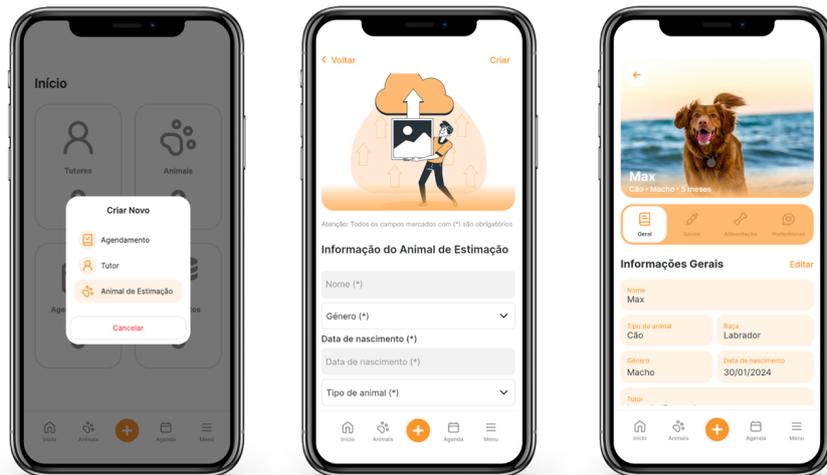


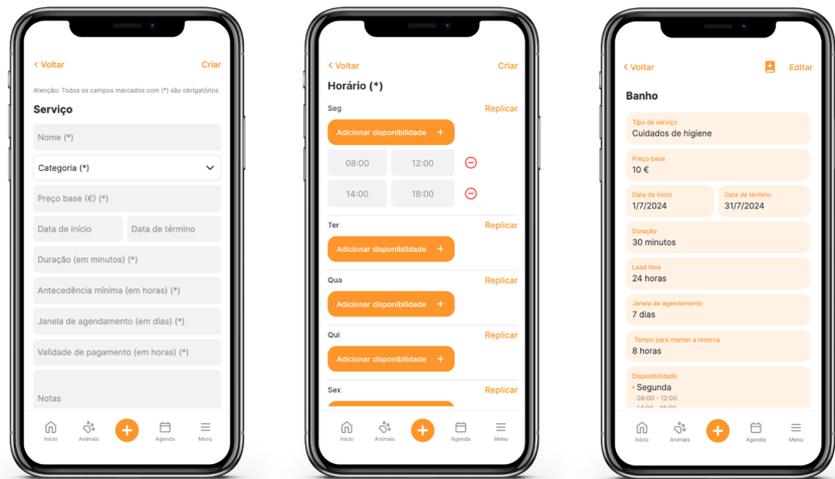
Fig. 3. Pet Creation Process

## 4.2 Pet Management

The Pet Management module handles the registration and management of pet information. It includes functionalities to add, update, and view pet details, such as species, breed, age, and specific needs. Additionally, it allows the recording of medical information, including allergies, chronic conditions, treatments, vaccines, and surgeries, as well as dietary and behavioral preferences. For instance, a pet sitter can create a pet profile by entering details such as the pet's age, breed, and medical history, ensuring proper care during services. **Fig. 3** depicts the pet management interface.

### 4.3 Service Management

This module is dedicated to managing services offered by pet sitters. It allows the creation, editing, and organization of different types of services, such as pet sitting, pet walking, pet boarding, and pet grooming. The system enables the definition of details like duration, price, and specific requirements for each service type. For example, a pet sitter can create a new service offering by specifying its duration, price, and any requirements like special care for certain breeds. **Fig. 4** shows how service management is displayed on the platform.



**Fig. 4.** Service Creation Process

### 4.4 Appointment Management

The Appointment Management module is responsible for scheduling and managing appointments between pet sitters and pet parents. It offers functionalities to create, view, and update appointments, including details such as date, time, duration, service type, and information about the pet and its owner. This module is crucial for the efficient organization of daily activities for pet sitters. For instance, a pet sitter can create an appointment by selecting the pet parent, pet, service type, and scheduling the session at a convenient time. **Fig. 5** demonstrates the appointment creation and management interface.

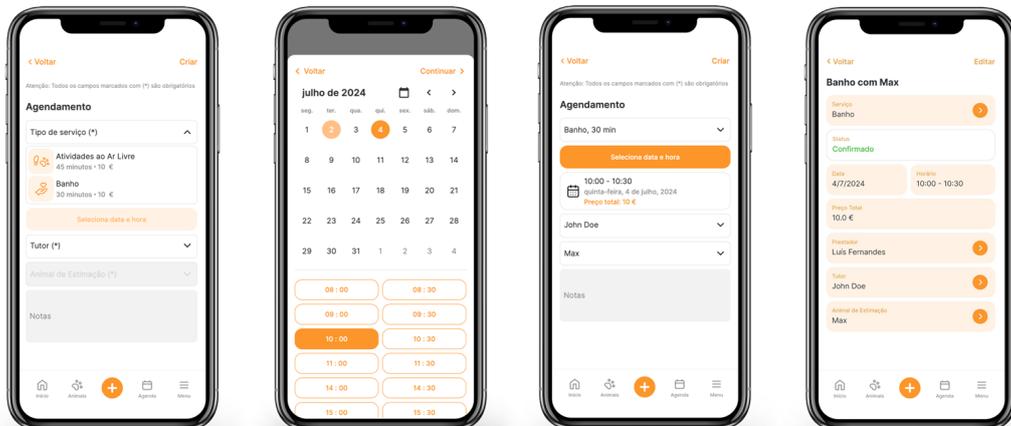


Fig. 5. Appointment Creation Process

### 4.5 Subscription Management

The Subscription Management module handles the management of subscription plans for pet sitters. It includes functionalities such as registering different subscription levels, processing payments, and controlling access to premium platform resources. For instance, a pet sitter can update a subscription plan by selecting a specific tier (e.g., basic, premium) and processing the corresponding payment. Fig. 6 illustrates the user interface for managing subscription plans, including payment options.

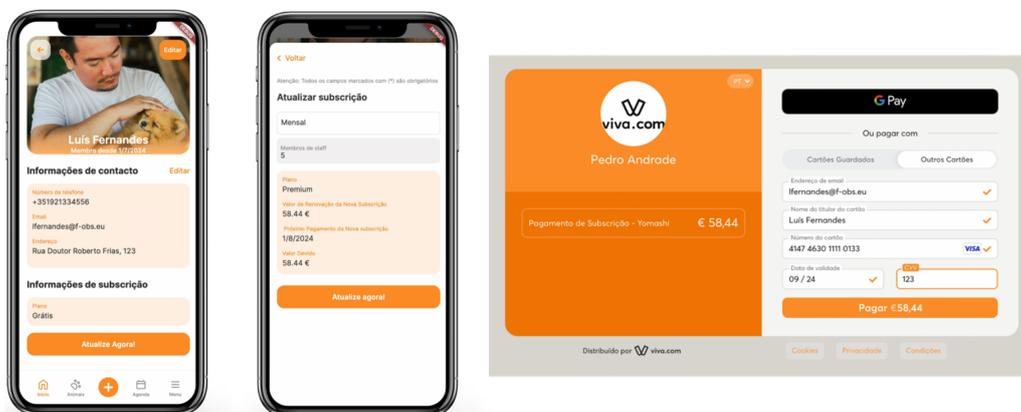


Fig. 6. Update Subscription Process

## 5 Results and Discussion

### 5.1 Preliminary Testing

The feedback collected highlighted several key points of satisfaction. Users expressed positive reactions to the intuitive design of the interface, which facilitated easy navigation between the pet management, scheduling, and payment functionalities. While the

feedback was largely favorable, there were some areas identified for improvement, particularly related to the integration of communication tools and the management of customer inquiries. Overall, this phase confirmed that the core functionalities were effective and that the platform offered tangible benefits to pet sitters, especially in terms of operational efficiency.

## 5.2 Impact Analysis

The impact of the platform on the operational dynamics of pet sitters was significant. One of the most prominent effects was the reduction in operational overhead due to the platform's real-time scheduling system and centralized data repository. Pet sitters no longer had to manage multiple sources of information, as the platform integrated all pet-related data, appointment schedules, and financial records in one place. This allowed for smoother daily operations, reduced the chances of booking errors, and enabled more efficient time management.

Moreover, the platform's ability to provide reminders and notifications for upcoming appointments helped increase reliability and punctuality, further contributing to client satisfaction. With real-time access to pet data, sitters were better prepared to meet the specific needs of each pet, such as dietary restrictions, medical conditions, and behavioral preferences. By offering a user-friendly interface, the platform significantly improved both the customer experience and the pet sitter's workflow, enhancing the overall service quality.

## 5.3 Additional Metrics Supporting Evaluation

Based on Firebase Performance [12] metrics and alpha testing data, a significant improvement in performance was observed, with the startup time reduced by 78%, achieving a current load time of 362ms.

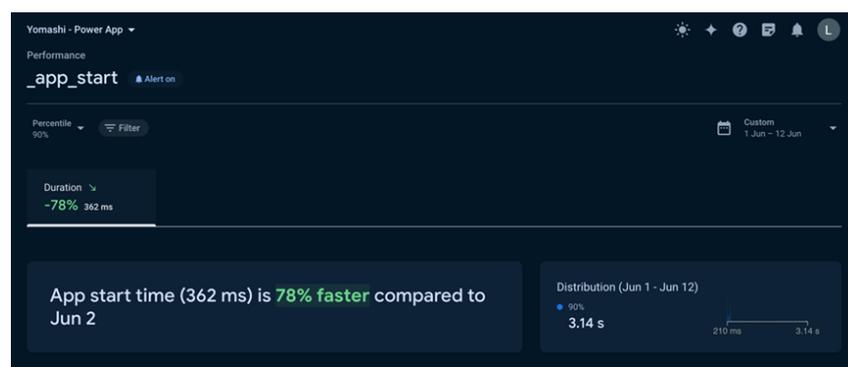


Fig. 7. Application startup time

Fig. 7 illustrates the performance data collected during testing, highlighting the improvements made throughout the development process.

#### 5.4 Limitations in Data Collection

While the tests conducted so far have provided valuable insights into the application's performance and usability, the depth of the evaluation was constrained by the absence of beta testing. During development, tools such as the "Test Mode" in FlutterFlow [1] were used to validate the interface and functionalities in the initial iterations. These tests, combined with alpha testing conducted by the development team, helped identify and resolve specific usability and performance issues.

However, the lack of access to a broader base of external users limited a more comprehensive analysis. Beta testing would have enabled the collection of real-world data from a diverse audience, allowing for a better evaluation of how the application performs under varied conditions, including different devices, network environments, and user profiles.

Thus, the evaluation presented in this paper reflects only a preliminary analysis based on internal feedback and technical data extracted by tools like Firebase Performance [12].

## 6 Conclusion and Future Work

The "Yomashi - Power App" [13] represents a major advancement in the pet care management sector, offering a solution that effectively addresses the challenges faced by pet sitters and pet parents alike. The platform combines key functionalities such as pet and appointment management, subscription handling, and financial tracking, all within an intuitive, user-friendly interface. The positive results from initial testing validate the potential of the platform in reducing administrative overhead, improving service delivery, and enhancing client satisfaction.

However, despite the success of the initial release, there are several areas for improvement and further development.

Future work will focus on expanding the platform's capabilities and refining its features to ensure a more comprehensive and scalable solution. The future plans include a broader beta testing phase involving a more diverse group of users, including different types of pet sitters and pet parents, as well as expanding to various regions. This will provide more detailed feedback, ensuring the platform meets the needs of a wider audience. Additionally, the platform's features will be expanded to include real-time GPS tracking for pet sitters during walks or visits and communication tools to provide both sitters and pet parents with real-time updates, improving trust and transparency. Advanced analytics may also be integrated to offer insights into sitter performance, customer preferences, and service trends, which can inform future business decisions.

Efforts will be made to adapt the platform for international markets, focusing on localization, multi-currency payment systems, and regional regulations. This will allow the platform to scale globally, reaching a broader user base and increasing its market share.

Finally, for a better evaluation of the platform's performance, stress tests will be conducted using simultaneous access from multiple users.

## 7 Project Evolution

Following its initial development and testing phases, the "Yomashi - Power App" [13] has successfully transitioned to a public release. Since October 2024, the application has been available for download on the Google Play Store, and its launch on the Apple App Store is expected soon. Users can also find more information or download links directly from the official website [13].

This advancement underscores the project's successful transition from concept to a market-ready solution.

This milestone highlights the quality and scalability of the work undertaken. Releasing the application on major platforms highlights its potential to address the needs of the pet care industry while providing a reliable, user-friendly solution for professionals.

The deployment validates the effectiveness of the low-code development approach, the modular architecture, and the continuous improvement cycle adopted throughout the project's lifecycle. Besides that, the imminent availability on multiple platforms ensures accessibility to a diverse audience, solidifying the app's role in the modernization of pet care management.

## 8 References

1. FlutterFlow. (n.d.). Retrieved from <https://flutterflow.io>.
2. Firebase. (n.d.). Retrieved from <https://firebase.google.com>.
3. Fresh Observation Unipessoal Lda. (n.d.). Retrieved from <https://www.racius.com/fresh-observation-unipessoal-lda>.
4. Rover. (n.d.). Retrieved from <https://www.rover.com>.
5. Pet Backer. (n.d.). Retrieved from <https://www.petbacker.com>.
6. Wag. (n.d.). Retrieved from <https://www.wagwalking.com>.
7. TrustedHousesitters. (n.d.). Retrieved from <https://www.trustedhousesitters.com>.
8. Lume, U. (2020). Flockr: Flockr: A mobile cross-platform KMM app for pet owners combining social networking and pet health management. Retrieved from <https://lume.ufrgs.br/handle/10183/223537>.
9. UFPE. (2021). Easypet: An app for optimized pet care and management. Retrieved from <https://repositorio.ufpe.br/jspui/handle/123456789/58459>.
10. UFC. (2020). Grandpaw: A proposed mobile educational app for managing the needs of senior dog owners. Retrieved from <https://repositorio.ufc.br/handle/riufc/70355>.
11. Newman, S. (2015). Building Microservices: Designing Fine-Grained Systems. O'Reilly Media.
12. Firebase Performance. (n.d.). Retrieved from <https://firebase.google.com/docs/perf-mon>
13. Yomashi. (n.d.). Retrieved from <https://yomashi.pt>.
14. Firebase Authentication. (n.d.). Retrieved from <https://firebase.google.com/docs/auth?hl=en>
15. Firebase Firestore. (n.d.). Retrieved from <https://firebase.google.com/docs/firestore?hl=en>.
16. Firebase Cloud Functions. (n.d.). Retrieved from <https://firebase.google.com/docs/functions?hl=en>.
17. Firebase Cloud Messaging. (n.d.). Retrieved from <https://firebase.google.com/docs/cloud-messaging?hl=en>.

# Optimizing Military Training Simulations: Development and Security Optimization of Sistrade's Simulation Platform

Nuno Rocha<sup>1,2</sup> [0009-0004-2405-3251], and Emanuel Silva<sup>1</sup> [0000-0002-5046-8942]

<sup>1</sup> Polytechnic Institute of Porto, Porto, Portugal

<sup>2</sup> Sistrade - Software Consulting, S.A., Porto, Portugal  
{1190912,ecs}@isep.ipp.pt

**Abstract.** In modern military training, immersive simulations offer essential preparation across a wide range of combat scenarios, enabling the creation of more interactive and engaging training scenarios, in which participants can experience and react to complex situations safely, without the risks associated with training in real conditions. This paper presents advancements in Sistrade's military simulation platform, focusing on two primary areas: enhanced interactivity through Collaborative Virtual Environments (CVE) and robust data security with Attribute-Based Encryption. The integration of Augmented Reality (AR) within CVEs allows personnel to engage in more realistic, interactive environments. This work lays down the groundwork for future developments into multi-user editing, enabling users to work collaboratively and simultaneously on Sistrade's products.

**Keywords:** Military Simulation, Collaborative Environments, Attribute-Based Encryption.

## 1 Introduction

Due to the inherent dangers of modern military training, soldiers and expensive equipment are frequently at risk during training exercises. Military operations require extensive planning for a variety of battle scenarios, from urban warfare to isolated locations, each of which has its own set of difficulties and perils.

Immersive simulations are increasingly vital in military training, allowing personnel to experience realistic environments without real-world risks. Military simulations facilitate complex scenario preparation and provide safe, adaptable environments for practising tactical responses. Considering technological advancements, simulations can now incorporate collaborative and augmented reality features, expanding their utility and realism.

Sistrade's military simulation platform offered basic scenario editing and playback but lacked interactive and secure multi-user functionality. This study addresses these limitations by enhancing collaborative capabilities and incorporating data security protocols. Specifically, it integrates CVEs to support multi-user interactivity and Attribute-

Based Encryption (ABE) to protect sensitive information. This paper explores the design, implementation, and evaluation of these improvements, demonstrating the potential of enhanced, secure military simulations.

## 2 Literature Review

### 2.1 Military Training Simulations

According to Hill and Miller [1], simulations serve as a means to deploy a model over an extended duration, being a model a simplified representation of an entity. Additionally, they function as a technique for testing, analyzing, or training, whereas the model replicates the real world and its conceptual systems.

As such military training simulations have evolved to include virtual and constructive approaches, enabling real-time scenario simulation and strategic planning exercises. Virtual simulations allow users to interact with computer-generated environments, while constructive simulations run self-contained models for high-level scenario analysis.

### 2.2 Collaborative Virtual Environments

Collaborative environments are digital spaces where individuals, often geographically dispersed, can interact, share information, and work collectively, facilitating real-time communication and cooperation [2]. There are several types of environments such as Online Collaboration Platforms, Augmented Reality Collaborative Environments, Collaborative Virtual Environments, among others.

Collaborative Virtual Environments (CVEs) enable users to interact within a shared digital space, promoting collective problem-solving and real-time communication. In a military simulation context, CVEs can replicate the coordination required in multi-person missions, enhancing teamwork and adaptability.

### 2.3 Collaboration Frameworks

Some of the main frameworks that enable collaboration are Y.js, Derby.js, and Firebase real-time database.

Y.js is a real-time shared editing library designed to facilitate collaborative editing of shared documents and data structures within web applications, in peer-to-peer settings [3]. It enables multiple users to concurrently edit and synchronize changes in real time, making it suitable for collaborative environments and applications.

DerbyJS [4] is a real-time, full-stack JavaScript framework that facilitates the development of dynamic and collaborative web applications. This framework provides automatic two-way data binding, ensuring that changes to the data model are instantly reflected in the user interface and vice versa.

Firebase Realtime Database [5] is a NoSQL database that enables real-time data synchronization between clients. It offers a flexible JSON-like data structure, along with,

support for offline usage meaning that changes made offline are synchronized when the device is online again.

## 2.4 Data Security in Military Simulations

With the increasing usage of cloud computing technology, it became paramount to ensure data protection, confidentiality, and data immutability. However safeguarding data protection has added several steps that have to be done to allow a user to send or access this encrypted data, creating a lack of flexibility and scalability.

There are three types of encryption processes:

- Symmetric Encryption, involves using a single shared key for both encryption and decryption processes. As such the key must be kept confidential between the communicating parties;
- Traditional Public-Key Encryption, employs a pair of keys, a public key, and a private key. The public key is distributed openly, while the private key remains secret. This type of encryption is the most common and used encryption method, due to its ease of implementation;
- Attribute-Based Encryption, employs encryption and decryption methods that operate based on the user's specified property set.

Attribute-Based Encryption (ABE) provides fine-grained access control by assigning encryption policies based on user attributes. ABE's compatibility with military environments, such as NATO's Object-Level Protection protocols, makes it suitable for sensitive data protection in this application.

## 3 Analysis and Design

As previously stated, the objectives of this work are to complement and improve the existing solution, and as such, the need to first analyse the solution is necessary.

Starting with the problem domain, the existing software was implemented using Domain Driven Design. As such, the classes are configured in aggregates, allowing for ease of modularity and ease of change. In a simplified manner, 3 major groups of models exist, the asset group, the scenario group, and the session group.

1. The Assets group, include the type of vehicle and its specific properties, speed in each terrain, and other assets it possesses, such as sensors, weapons, etc;
2. The Scenario group includes the map, the tiles contained between the boundaries of the map, and the scenario information, such as the teams involved, assets being used, their location, their allegiance, etc;
3. The Session group is responsible for the lobby, the communication between the users, the game itself and the summary at the end of each session.

Regarding the architecture, the system presents itself as a standard web browser formation, having a frontend, a backend, a database, and finally, another database containing the information of the tiles of the entire world.

With the interactivity of the user being the major point of this work, it was necessary to analyze the construction, functionalities and existing pages of Sistrade's simulator. The present prototype has 9 major pages, ranging from the splash screen and the home screen to the edition of scenarios, scripting of assets and the execution of those same scenarios. It was found that it allowed single-user scenario design and playback but lacked collaborative functions and data security. To address these gaps, it was decided to incorporate multi-user support and encrypted data handling into the design.

After analyzing the existing pages, potential changes to be made were identified, namely, the addition of Y.js to simultaneously allow changing the scenario information, placing assets and scripts, among more options and also adding AR functions. Table 1 summarizes these potential changes.

**Table 1.** Functionalities and possible alterations to the screens

Screen	Functionalities	Possible Alteration
<b>Splash Screen</b>	Login into system	N/A
<b>Home Screen</b>	View existing lobbies	N/A
<b>Asset Management Screen</b>	View, create and management of assets	Addition of Y.js
<b>Creation of Scenarios Screen</b>	Creation of new Scenarios	N/A
<b>Edition of Scenarios Screen</b>	View and manage existing scenarios	Addition of Y.js
<b>Placement of Assets Screen</b>	Place units on the scenario map and assign their allegiance	Addition of Y.js Addition of AR functions
<b>Scripting of Assets Screen</b>	Manage the script of each of the assets	Addition of Y.js
<b>Lobbies Screen</b>	Select a scenario to play, and assign to each player a team	Addition of Y.js
<b>Game Screen</b>	Execute the simulation,	Addition of Y.js Addition of AR functions

## 4 Solution Implementation

### 4.1 Implementation

As the development process could easily escalate in difficulty, it was decided to divide it into 3 phases for a systematic implementation. The first phase regards implementing collaborative work through the usage of Y.js. The second phase concerns the usage of WebXR in Babylon.js. The last phase, relating to the implementation of ABE has not yet been completed, but will therefore be considered for future work.

### Phase. 1 – Collaborative Work

As previously stated, one of the main objectives of this work was the implementation of collaborative work into the simulator, which was achieved through the usage of Y.js. This was decided to be the first phase of development, as it was the less disruptive and easiest to implement.

A few problems arose during the development. First was the way of how Y.js documents work. The contents of these documents are normally objects, meaning that they should be easily integrated with the existing schemas of the implemented software. However, due to the way the existing software was developed, there was a need to identify what type of asset the document was referring to.

To solve this, one of the possible solutions, and determined to be the easiest, was to add in the object the type of the asset, and then use the constructor to transform the object into that class, as seen in the following code:

```
this.ymap.observeDeep(() => {
  var temp = this.ymap.get(0) as any
  switch (temp.Type) {
    case "Sensor": temp = new Sensor(temp);
                  break;
  }
})
```

Although it is not the ideal solution, as it adds technical debt, due to the additional work added if any new type of asset is added.

Another issue arose due to how the different copies were transmitted between users. Among several possible technologies, the usage of “y-websocket”<sup>1</sup>, a WebSocket provider was decided. This required the creation of a new server in the backend of the application, as well as the connection with it through the front end.

With several users now connected and working on the same information, discrepancies and overlaps of information started to occur.

The first and easiest to resolve was the phantom work, that is, the work that was not saved but still overlapped. This correction was as simple as commenting on a few lines regarding the persistence of data, as seen in the following code:

```
//before
if (doc.conns.size === 0 && persistence !== null) {
  persistence.writeState(doc.name, doc).then(() => {
    doc.destroy()
  })
  docs.delete(doc.name)
}
//after
if (doc.conns.size === 0) {
  doc.destroy()
}
```

---

<sup>1</sup> <https://github.com/yjs/y-websocket>

Another problem was when a new user joined an editing session, the object reverted to its last saved state, causing a loss of alterations. As such, it was decided to prioritize edited objects. This means that when a new user starts editing, the system must load two times, the first being the original object, while the second is the altered object.

## Phase 2 – AR Implementation

This phase regards the second major objective of the work, the implementation of AR technology. As previously asserted, the existing application already uses Babylon.js which possesses the ability to easily add VR and AR functionalities.

The first major roadblock of this phase was how to test the AR experience since the availability of such capable equipment was scarce. Several ideas were tough off, with the two most promising being:

- The use of a smartphone, for the most ideal and realistic;
- The use of Google Chrome's extension "Immersive Web Emulator"<sup>2</sup>, for an easier time debugging.

Since the development of this work was in a developer branch not deployed in the company's servers, the necessity of the smartphone to connect to the computer running the software was paramount. To achieve this, Internet Information Services<sup>3</sup> (IIS) was used, as it allowed to host a local web server, granting smartphones and other devices access to the software through the local IP of the hosting server. After allowing other devices to access the program, development with XR started, and a problem was immediately detected - WebXR only works with HTTPS. The solution was the creation of a local Secure Sockets Layer certificate.

Although WebXR now worked on other devices connected, all the collaborative functions previously developed would not work, no matter how many different solutions and fixes were tried, due to the inability to connect with WSS. As such, it was decided to abandon the idea of using other devices, and the implementations continued with the Immersive Web Emulator extension.

Having now a more stable development and debugging platform, the real implementation of WebXR started to gain momentum.

Starting with the first objective, originally the maps were scaled with a fixed value, allowing them to have a size closer to their real counterparts, in which their coordinates correspond to the ones in the real world. This brought a major problem, because when a user enters the AR environment, he is always located at the coordinate (0,0,0), and the map would be outside of the play area.

To resolve this, it was decided to remake the creation of the map, which involved working out the new map mesh sizes, scaling values and positioning. This alteration also involved creating new converters from the real-world coordinates to coordinates in the Babylon.js environment and vice versa.

---

<sup>2</sup> <https://chromewebstore.google.com/detail/immersive-web-emulator/cgffilbpci-bhmcfbgggfhfolhkfbbmik>

<sup>3</sup> <https://www.iis.net/>

The next major blockade was the usage of the 3D objects in the game. Currently, in the Asset Management screen, the assets have a slot to add a 3D model, however changing it does not change the appearance of the asset in the map view. Currently, the assets are represented by a cube with the image and colour of the team to which they are assigned.

The first problem that appeared was the orientation of the loaded models. Due to how an asset is represented by several meshes, it was decided to use the original representing cube as the parent to the models. This allowed all the several meshes to stay together with the caveat that the orientation of the asset could be incorrect depending on the model used.

Another issue encountered was the scaling of each asset. In a fully fledged simulator, all the assets would be checked beforehand, so that they be scaled correctly.

As this project is a proof of concept, several different open-source assets were used, causing the assets to not correlate correctly. This was fixed by obtaining the biggest bounding box of the meshes and scaling it in relation to it allowing for more uniform scaling.

## 4.2 Testing and Validation

In this section, the two testing procedures that are essential to the software development process and that were used during the solution development process will be discussed, namely performance and acceptance tests.

During meetings between the responsible organization and the author of this paper, a proposal for validating the solution was defined. The proposal encompasses the implementation of performance and acceptance tests, ensuring sound testing methods and valid results.

### Performance and Stress Tests

Performance tests aim to evaluate a system's efficiency and responsiveness by measuring its ability to handle varying workloads and stress levels [6].

Apache JMeter<sup>4</sup> was used for load, performance and functional testing of web applications. It simulates variable loads to measure and analyze performance under different conditions, indicating how well the system meets stakeholder needs.

With this tool, it was possible to compare the capabilities of the WebSockets with the HTTP posts. This was important as the ability of the real-time collaborative work was dependent on them.

Two configurations were used:

- Configuration 1:
  - Number of Threads (users): 100
  - Ramp-Up Time: 50 seconds
  - Test Duration: 60 seconds
- Configuration 2:
  - Number of Threads (users): 400
  - Ramp-Up Time: 100 seconds
  - Test Duration: 60 seconds

---

<sup>4</sup> <https://jmeter.apache.org>

With these configurations, in total, there were 4 group threads, the first 2 threads were configured to simulate 100 users, and the last 2 threads were configured to simulate 400 users, with each run done three times and the average was presented.

Key metrics included Average Bytes, Standard Deviation, and Error Rate, which represent the average data transmitted, standard deviation, and the percentage of failed requests or interactions.

**Table 2.** Performance Tests Results

<b>Technology</b>	<b>Configuration</b>	<b>Average (ms)</b>	<b>Std. Dev. (ms)</b>	<b>Avg. Bytes (ms)</b>
<b>WebSocket</b>	1	83	24.66	131.0
<b>HTTP</b>	1	116	21.01	1348282.0
<b>WebSocket</b>	2	79	26.75	131.0
<b>HTTP</b>	2	278	695.12	1348282.0

As expected, the WebSocket implementation is more efficient and has better performance than the HTTP implementation. This comes down to various factors, such as the size of the packages, the time needed to access the database, and the overall latency of the communication. Table 2 depicts the results of the performance tests.

A stress test group was also implemented, with the objective of knowing how many threads the system could handle until having a ~10% error. It was concluded that the HTTP implementation with a 3000 set of samples obtained an error rate of 16.77% while the WebSockets obtained an error rate of 85.73% on a set of 3000 samples. Table 3 depicts the results of the stress test.

**Table 3.** Stress Test Results

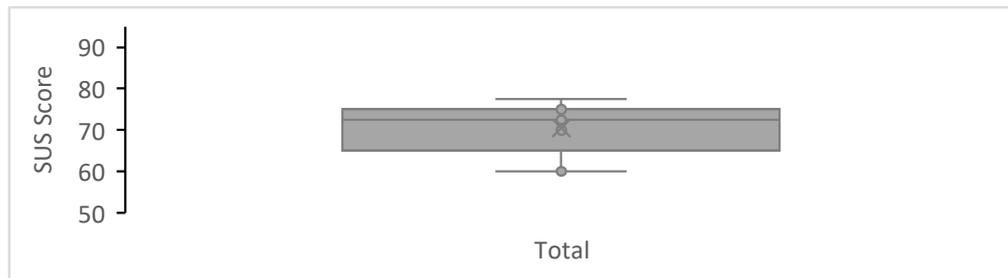
<b>Technology</b>	<b>Average (ms)</b>	<b>Error (%)</b>	<b>Avg. Bytes (ms)</b>
<b>WebSocket</b>	5891.99	85.73%	131.0
<b>HTTP</b>	7946.52	16.77 %	1348282.0

### Acceptance Tests

Alongside the performance tests, acceptance tests are used to ensure that the results of a solution align with the requirements established during the analysis process, in which there is the involvement of the party interested in the solution (client, product owner, etc) [7]. These tests involve a sample of 10 collaborators using the application simultaneously, particularly with editing assets and scenario functionalities. This realistic approach allows for comprehensive validation of collaborative features. The participants are then asked to fill out two questionnaires, a System Usability Scale (SUS) and an Interface Validation.

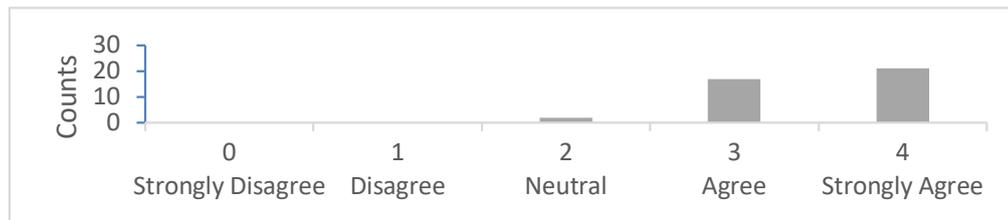
The SUS [8] questionnaire is a ten-item scale used to assess the usability of a system. It is used after the respondent has had the opportunity to use the system but before debriefing or discussion. Simple ten-item scale giving a global view of subjective assessments of usability.

The mean score of 71.25 indicates the application's usability, which is converted to a percentile rank of 70%. This score is interpreted as a grade of a C [9], corresponding to the level Good of the System Usability Scale [10]. The normality of the data was assessed using the Shapiro-Wilk test, which yielded a p-value of 0.02142. Since this p-value is less than the threshold of 0.05, the assumption of normality was rejected. Fig. 1 depicts the results of the SUS questionnaire.



**Fig. 1.** Boxplot results of SUS questionnaire

The second questionnaire, an Interface Validation, consisted of 13 Likert categorical variables. The mean user intention detection accuracy was 88.6%, with a scale accuracy of 3.544. The Shapiro-Wilk test assessed data normality, yielding a p-value of 0.2496, which was below the significance level of 0.05, rejecting the assumption of normality. Fig. 2 depicts users' opinions.



**Fig. 2.** Overall accuracy of Users' intention

### 4.3 Solution Evaluation

As shown through the performance and acceptance tests, the system is well formulated and functional, corresponding greatly to the expectations of the company owners and project managers. Although still a prototype, and with only some knowledge in the area, the results indicate a promising direction for specification and further development. The different approaches adopted and experimented with throughout the course of this project facilitated the identification of potential issues related to the usage of concurrent editing, synchronization, and data consistency.

Below expectations was the lack of implementation of ABE and an approach plan for this technology in a timely manner. This was largely due to the first phase of development, where significant trial and error in implementing the IIS consumed a major portion of the development time. Even so, it is concluded that the alterations made to the software were a success and are considered a relevant improvement.

## 5 Conclusions

This work achieved its primary objectives by integrating collaborative features into Sistrade's military simulation platform. The improvements enabled multi-user interaction within an AR-enhanced environment, increasing training realism and user engagement. Improvements to the implementation of collaborative work in the simulator were achieved using Y.js and y-websocket, and the implementation of AR technology was achieved using Babylon.js, which facilitates the inclusion of VR and AR features. Future work will involve extending AR capabilities and refining encryption protocols for more advanced data protection.

## References

1. R. R. Hill., J. O. Miller: A history of United States military simulation. In: Winter Simulation Conference (WSC), pp. 346–364. IEEE, Las Vegas (2017)
2. D. T. Ahmed, S. Shirmohammadi, J. C. de Oliveira, J. Bonney: Supporting Large-Scale Networked Virtual Environments. In: 2007 IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, pp. 150–154. IEEE (2007)
3. P. Nicolaescu, K. Jahns, M. Derntl, R. Klamma: Yjs: A Framework for Near Real-Time P2P Shared Editing on Arbitrary Data Types. In: Engineering the Web in the Big Data Era. Cham, pp. 675-678. Springer International Publishing (2015)
4. A. Mardan: Real-Time Apps with WebSocket, Socket.IO, and DerbyJS. In: Practical Node.js, pp. 307-330. (2018)
5. L. Moroney: The Definitive Guide to Firebase: Build Android Apps on Google's Mobile Platform. (2017)
6. Isabel Azevedo: QESOF - Qualidade na Engenharia de Software Mestrado em Engenharia Informática Lectures 04.1 and 04.2 How to measure quality? Measuring performance, ISEP (2023)
7. ARQSI, [https://moodle.isep.ipp.pt/pluginfile.php/147512/mod\\_resource/content/2/Introduction%20to%20Software%20Testing%20v2.pdf](https://moodle.isep.ipp.pt/pluginfile.php/147512/mod_resource/content/2/Introduction%20to%20Software%20Testing%20v2.pdf), last accessed 2024/11/10
8. J. Brooke: SUS: A 'Quick and Dirty' Usability Scale. In: Usability Evaluation In Industry. CRC Press, 1996
9. J. Brooke: SUS: A Retrospective. In: Journal of User Experience. vol 8, pp. 29-40, 2013
10. A. Bangor, P. Kortum, J. Miller: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. In: Journal of User Experience. vol 4, pp. 114-123, 2009

# Using AI as a Development Accelerator

Rui Soares<sup>1</sup> e F. Jorge Duarte<sup>2</sup>

<sup>1,2</sup> ISEP, Porto, Portugal

<sup>1</sup> 1191045@isep.ipp.pt; <sup>2</sup> fjd@isep.ipp.pt

**Abstract.** This article explores the transformative potential of Artificial Intelligence (AI) in accelerating and automating the development of Java microservices, focusing on the Spring Boot framework. As software development becomes increasingly complex, AI integration promises to simplify coding tasks while meeting technical standards and business requirements. The study begins with a comprehensive literature review on AI-driven code generation, identifying recent advancements and current knowledge gaps. Practical experimentation follows, involving tools like GitHub Copilot, TabNine, and various versions of ChatGPT to assess their effectiveness in generating functional Java microservices code, integrating open APIs, and maintaining quality standards. The primary goal is to develop an AI-based methodology for generating specific code for a Java microservice within the Spring Boot environment. This goal involves creating open APIs, integrating databases, and connecting to external REST services. The thesis also examines the broader implications of AI in software development, such as data privacy, security, and employment impacts. It considers ethical issues to ensure fair and unbiased code generation. The adopted methodology is reflective and iterative, adjusting to emerging findings and challenges, with regular project team discussions refining AI tool application and evaluation. In conclusion, this research takes a comprehensive approach to AI integration in software development, highlighting its benefits and identifying areas for future research. The findings are expected to significantly contribute to the understanding of how AI technologies can revolutionize software development practices, enhancing efficiency, reducing development time, and ensuring code quality.

**Keywords:** Artificial Intelligence, AI, Machine Learning, Software Development, Code Generation, Automation, Java, Microservices, Spring Boot, AI Tools.

## 1 Introduction

### 1.1 Background and Context

In recent years, software development has undergone transformative changes driven by evolving technological patterns and dynamic market demands. A key driver of this transformation is the emergence of artificial intelligence (AI) technologies, which act as catalysts for expediting development processes. The increasing complexity and scalability demand of modern software systems have outpaced traditional development

methodologies, creating a need for more efficient and sophisticated approaches to code generation.

AI-based methodologies offer a promising solution to the demand for faster software development without compromising quality. However, integrating AI into software development introduces significant challenges, particularly in generating functional code that adheres to rigorous technical and business specifications. Research highlights these limitations, such as the difficulty of AI systems in interpreting and applying complex business logic [1]. Moreover, substantial human involvement is still required to refine AI-generated code and ensure it meets specific requirements [2].

This integration is further complicated by the need for conformity with industry standards, alignment with user and business needs, and seamless incorporation into existing workflows. Financial and resource constraints also present obstacles, particularly for smaller organizations, due to the high costs associated with AI implementation and the limited availability of skilled professionals in AI and software engineering [3].

While large enterprises are leading advancements in AI-driven software development, smaller firms face challenges stemming from resource constraints and a lack of specialized expertise [4]. This thesis focuses on the use of AI for Java microservices development, specifically in the context of Spring Boot, to evaluate the opportunities and constraints of AI-driven code generation and its broader implications for software development.

### 1.2 Objectives

The primary objective of this research is to develop and critically evaluate an AI-driven methodology for generating code specifically tailored for Java microservices within the Spring Boot framework. This methodology includes creating Open APIs, integrating databases, and connecting to external REST services. To achieve this, the study undertakes a comprehensive review of existing literature to identify advancements, effective techniques, and gaps in the field of AI-driven code generation.

A detailed comparison of various AI tools is performed, evaluating their advantages and limitations to select the most suitable option for the research context. The performance of the selected methodologies is rigorously assessed using objective metrics and practical testing scenarios. This evaluation is further complemented by the development of a functional case study involving a microservice of medium-to-high complexity. The research also compares the AI-based approach to traditional code generation methods, highlighting strengths and weaknesses, while examining the level of autonomy AI tools can achieve across different development stages, such as generating user stories, producing code, and integrating observability components, API documentation, and CI/CD processes. Key outcomes of the research are analyzed to extract meaningful insights, which contribute to refining the AI code generation methodology and identifying areas for future exploration. Additionally, the study addresses security and privacy concerns, recognizing potential vulnerabilities in handling sensitive data during the AI-driven development process. Comprehensive documentation of all phases ensures a coherent and thorough record of the research. This structured approach facilitates a holistic

understanding of the potential, limitations, and practical applications of AI in Java microservice development, while addressing ethical and security aspects.

## 2 Literature Review

### 2.1 State of the Art

AI technologies have significantly transformed software development, particularly through tools like ChatGPT, GitHub Copilot, and Amazon CodeWhisperer. These tools offer innovative approaches to code generation, but they also raise challenges in education, security, intellectual property, and software engineering practices. In programming education, tools like ChatGPT and GitHub Copilot have introduced new opportunities and challenges. Lau and Guo (2023) explored how instructors are adapting to these tools, revealing a divide between those seeking to ban them to preserve fundamental programming skills and those integrating them to prepare students for AI-driven industries. This adaptation highlights the need for balanced educational strategies to address AI's potential while maintaining robust learning outcomes [5]. Similarly, Takerngsaksiri et al. (2023) demonstrated that while tools like GitHub Copilot improve productivity, they risk fostering over-reliance among students, potentially undermining problem-solving skills [6]. The quality of AI-generated code has been a focus of several studies. Yang et al. (2023) developed RADAR to evaluate and enhance the robustness of pre-trained code generation models, emphasizing the critical role of method names in producing accurate code [7]. Further, Yetistiren et al. (2022) assessed GitHub Copilot, noting its high success rate in generating valid code but identifying limitations in handling complex programming tasks [8]. A comparative analysis of ChatGPT, Copilot, and Amazon CodeWhisperer by Yetistiren et al. (2023) highlighted ChatGPT's superior performance in generating maintainable code, although all tools require improvements in security and reliability [9]. Addressing intellectual property concerns, Li et al. (2023) proposed ToSyn, a watermarking technique to safeguard the IP of large language model-based code generation. While effective for Python-based systems, broader applicability remains limited, underscoring the need for ongoing innovation in IP protection strategies [10]. Security is another critical concern in AI-driven development. Siddiq and Santos (2022) introduced the SecurityEval dataset to evaluate vulnerabilities in AI-generated code, revealing significant weaknesses that necessitate robust security frameworks [11]. Extending this research, Li et al. (2024) demonstrated ChatGPT's ability to identify and resolve vulnerabilities but noted limitations in cross-file analysis due to token constraints [12]. AI has also reshaped software engineering practices. Lozano et al. (2013) introduced a constraint-based approach for code generation, showcasing flexibility and efficiency but facing scalability challenges in complex systems [13]. Xue (2023) leveraged machine learning to automate code generation in Model-Driven Engineering, offering practical benefits despite limitations in language diversity and optimization [14]. Peng et al. (2024) highlighted GitHub Copilot's potential to enhance developer productivity, particularly for less experienced programmers, suggesting AI's ability to democratize software engineering [15].

While AI tools promise to accelerate software development, challenges in robustness, security, and ethical integration remain. Continued innovation and careful oversight are essential to harness AI's full potential while addressing its limitations.

## 2.2 Ethical Considerations

The ethical implications of using AI as a development accelerator are critical to ensuring responsible integration into various domains. This research adheres to frameworks like the IEEE Code of Ethics [16] and the GDPR [17], emphasizing fairness, transparency, and accountability. AI systems must avoid perpetuating bias or social inequalities and ensure decision-making processes are transparent and accountable. Proper management of intellectual property rights, including compliance with open-source and proprietary licensing, is essential. Furthermore, safeguarding sensitive data requires adherence to international security standards such as ISO/IEC 27001:2022 [18]. Celfocus exemplifies ethical AI practices through principles that prioritize human dignity, non-discrimination, and privacy. Their systems are designed to prevent bias and misuse, while policies prohibit uploading confidential data into generative AI tools. Transparency is emphasized, with users expected to critically evaluate AI outputs, and accountability is central, ensuring responsibility for outcomes and adherence to company values. By addressing these considerations, AI can responsibly accelerate development while mitigating risks. Ensuring fairness, transparency, and security helps maximize its potential for positive impact.

## 3 Methodology

### 3.1 Research Framework

This research investigates the role of AI as a development accelerator through the PRISMA methodology, a systematic framework that ensures transparency and rigor in literature reviews. By following this structured approach, the study aims to comprehensively assess AI's impact on programming education, code quality and security, intellectual property, and software engineering practices. The research begins with a focus on four key questions: how AI transforms programming education and code generation, the implications of AI-generated code on quality and security, its influence on intellectual property rights, and its role in reshaping traditional software engineering methodologies. Data sources include reputable databases such as ScienceDirect, ACM Digital Library, IEEE Xplore, and ArXiv, selected for their relevance to peer-reviewed academic work in AI and software development. Search terms like "Artificial Intelligence," "Code Generation," "ChatGPT," and "GitHub Copilot" were used to construct queries, ensuring a broad yet focused exploration of the topic. Eligibility criteria were applied to refine the dataset, including only peer-reviewed articles in English that provided practical applications of AI in software development. Articles that lacked relevant scenarios or were not published in English were excluded. The PRISMA process was conducted in three stages: first, identifying 270 articles that matched the initial search criteria; second, screening abstracts to narrow the selection to 24 articles; and finally,

analyzing seven articles deemed most relevant to the research questions. This systematic process provided a rigorous foundation for exploring AI's multifaceted impact on software development, addressing technical, ethical, and educational challenges while highlighting opportunities for innovation. By ensuring a methodical and transparent approach, the study delivers meaningful insights into AI's role as a transformative tool in software engineering.

### **3.2 Tools and Processes**

This study evaluates several AI-driven code-generation tools relevant to enhancing software development workflows. The analysis focuses on operational functionalities, integration capabilities, and their potential to improve productivity and code quality within Celfocus's development environment. The primary tool under evaluation is GitHub Copilot, developed by GitHub and OpenAI, which integrates seamlessly with IDEs like Visual Studio Code. Leveraging advanced GPT models, it offers real-time code suggestions and features like Copilot Chat and automated pull request reviews, which enhance coding efficiency and streamline testing workflows [19]. Its widespread adoption and integration capabilities make it particularly relevant for Celfocus. TabNine, a competitor to Copilot, supports multiple programming languages and provides predictive coding assistance. Its efficiency and compatibility with various IDEs will be assessed to determine its viability compared to GitHub Copilot [20]. Although tools like CodeWhisperer and Snyk Code offer benefits in cloud integration and security analysis, they are excluded from this study, as the focus remains on direct coding assistance tools. ChatGPT 3.5 and ChatGPT 4.0 are also under evaluation, with particular attention to the advancements in ChatGPT 4.0. This model's enhanced natural language understanding and cost-effectiveness for specific applications are critical considerations [21]. Additionally, Custom GPTs, tailored for specialized tasks, are assessed for their potential advantages in performance over general models. The evaluation framework uses IntelliJ IDEA and Visual Studio Code as primary testing platforms, reflecting their widespread use within Celfocus and their robust feature sets. Browsers like Brave and Chrome support web-based tests for ChatGPT models, ensuring a holistic assessment across different coding environments. The evaluation combines these tools' real-world performance in generating code and their integration into workflows. This comprehensive approach aligns with Celfocus's priorities, ensuring the tools selected meet both current needs and future scalability.

## **4 Case Studies, Findings and Results**

### **4.1 Behavioral Patterns and Predictions in Companies**

Integrating AI tools like GitHub Copilot, TabNine, and ChatGPT into corporate workflows has the potential to significantly enhance productivity by automating routine tasks, freeing employees to focus on more complex problem-solving. Studies have shown that tools like Copilot increase developer efficiency and satisfaction, making them valuable accelerators in software development processes [22]. These tools also

improve over time through machine learning, refining their suggestions and keeping developers at the forefront of programming techniques [23]. The concept of "pair programming" with AI, where tools assist but humans refine the outputs, has proven effective in enhancing productivity and maintaining developer engagement [24]. To maximize these benefits, companies are investing in upskilling programs that teach employees to work synergistically with AI, ensuring that automation enhances rather than replaces human expertise [25]. While AI tools can uncover patterns and drive innovation, success depends on thoughtful implementation and balancing automation with human skills. Companies that prioritize strategic integration and training are more likely to leverage AI effectively, avoiding over-reliance and fostering sustainable productivity [26].

## 4.2 Prompt Engineering Analysis

Prompt engineering is a vital aspect of interacting with AI systems, shaping the quality and relevance of outputs across diverse applications. It involves crafting precise and effective inputs to guide AI models, particularly those using transformer architectures like GPT, which rely heavily on learned patterns from training data. By translating complex or vague user requirements into structured queries, prompt engineering optimizes the AI's ability to deliver accurate and contextually appropriate responses [31]. The method requires both technical knowledge of AI model capabilities and creative skills to frame prompts effectively. Its influence spans various domains, including natural language processing (NLP), computer vision, and software development. In NLP, prompts direct AI systems to extract specific data or resolve ambiguities in translation. For computer vision, structured prompts guide models to recognize patterns or anomalies, such as in medical imaging diagnostics. In software development, tools like GitHub Copilot and TabNine rely on well-crafted comments or partial code inputs to generate precise code suggestions, reducing cognitive load and accelerating development processes [32][33]. Applications like ChatGPT highlight the importance of prompt clarity and detail, as specific inputs significantly impact the quality of generated responses. Similarly, tools like GitHub Copilot and TabNine depend on precise prompts to align code generation with developer intent. However, challenges remain, such as the complexity of crafting effective prompts, potential biases in model outputs, and the need for AI systems to learn and adapt to user interactions over time. As AI systems become more pervasive, prompt engineering is increasingly essential for optimizing their functionality. Its continued evolution promises more intuitive and efficient interactions, enabling broader adoption of AI technologies in domains ranging from healthcare to software development. By addressing challenges and refining techniques, prompt engineering will unlock new possibilities for accessible and equitable AI applications.

## 4.3 Results

The analysis of collected data underscores the transformative potential of AI-driven code-generation tools in modern software development. Key findings reveal their ability to improve productivity, accelerate development processes, and enhance code

quality when effectively integrated into workflows. The evaluation combined quantitative metrics and qualitative insights. Quantitative data focused on code accuracy rates, efficiency, and performance compared to human-generated baselines. AI tools like GitHub Copilot and ChatGPT 4.0 demonstrated notable advantages in accuracy, speed, and reducing manual interventions. Scalability was also a strength, with these tools handling large-scale projects efficiently. Integration capability was assessed through compatibility with IDEs, operating systems, and CI/CD pipelines, with tools showing strong adaptability in these areas. Qualitative feedback from Celfocus developers emphasized the tools' ease of use, highlighting GitHub Copilot and Custom GPTs as particularly effective in supporting diverse programming tasks. Detailed case studies further validated these insights, demonstrating the tools' practical applications and identifying areas for improvement, such as security and reliability challenges. Prompt engineering emerged as a critical factor in maximizing the utility of these tools. Specific, clear, and contextual prompts were found to significantly enhance the accuracy and relevance of generated code. Iterative refinement of prompts also improved output quality, highlighting the importance of a strategic approach to prompt design. While the findings affirm the potential of AI tools to transform software development, they also underscore the need for thoughtful implementation. Addressing challenges like seamless integration, maintaining security, and optimizing prompt engineering techniques will be essential to fully realize their benefits. By leveraging best practices, organizations can drive innovation and efficiency, making AI an integral part of modern development workflows.

## **5 Discussion**

### **5.1 Integration of AI in Programming**

The integration of AI tools in programming represents a transformative shift, redefining workflows and evolving the role of developers. AI-driven pair programming fosters collaboration by suggesting improvements and introducing new techniques. Rather than replacing human effort, these tools augment it, enabling developers to focus on higher-level tasks such as designing solutions, ensuring software reliability, and maintaining quality assurance. However, challenges accompany these advancements, including potential over-reliance on AI tools and associated security and ethical considerations. Nonetheless, AI-driven debugging, automated documentation, and intelligent code refactoring promise innovation and efficiency across the software development lifecycle.

### **5.2 Limitations of the Study**

While this research provides valuable insights into the integration and effectiveness of AI-driven code-generation tools, it is essential to recognize certain limitations that may affect the generalizability and practical application of the findings. First, the study focused on a select group of tools (GitHub Copilot, TabNine, ChatGPT 3.5/4.0/4o, and Custom GPTs), excluding other options such as Amazon CodeWhisperer and Snyk

Code due to organizational priorities. Although these chosen tools represent some of the most advanced solutions available, their selection narrows the overall scope and may leave out features or capabilities found in other platforms. Future studies should include a broader range of tools to enable more comprehensive comparisons.

Second, the sample size and diversity of participants were limited to developers within Celfocus. This relatively small and homogeneous group may not reflect the varied experiences, programming practices, and cultural contexts of the wider developer community. Extending the analysis to include a larger, more diverse set of respondents from different industries and organizational backgrounds would enrich the validity of the results and provide a more robust basis for generalization.

Third, the research concentrated on generating Java microservices within the Spring Boot framework. While this focus allowed for a detailed assessment of how these tools function in a specific environment, it does not capture their performance in other programming languages, frameworks, or paradigms (such as front-end development or data-focused workflows). Conducting further studies that explore multiple contexts will better illustrate the adaptability and limitations of AI-driven code-generation tools in various domains.

Another concern is the rapid pace at which AI technologies evolve. Frequent updates and new features can significantly alter a tool's capabilities, performance, and integration potential. The conclusions drawn here reflect the state of the tools at the time of the research, underscoring the need for continuous monitoring and re-evaluation to maintain relevance.

Finally, the context-specific nature of these findings should be emphasized. Organizational structures, development workflows, and project-specific needs can greatly influence the success of AI integration. From a practical standpoint, businesses must assess factors such as training requirements, total cost of ownership, and potential security or ethical risks when adopting AI tools. Engineers, meanwhile, may face a learning curve in adapting to AI-driven workflows and must stay vigilant about the transparency and reliability of code suggestions. Addressing these practical implications will be critical for organizations seeking to optimize efficiency, maintain high-quality standards, and cultivate a sustainable competitive edge as AI technologies continue to advance.

## **6 Conclusion and Future Work**

### **6.1 Conclusion**

This study explored the integration of AI-driven tools in software development, focusing on Java microservices using the Spring Boot framework. Through comprehensive literature reviews, comparative analyses, and case studies, the research identified key strengths and limitations of tools like GitHub Copilot, ChatGPT, and TabNine. These tools demonstrated significant potential to improve productivity by automating routine coding tasks, enhancing code quality, and reducing development time. GitHub Copilot excelled in tasks like configuring Jenkins pipelines and integrating database connections, while Custom GPTs delivered the most accurate code generation with minimal errors. The research confirmed that AI tools can effectively handle user stories, generate

initial code, and support repetitive tasks. However, human oversight remains essential for refining outputs, ensuring adherence to project standards, and integrating components like observability and API documentation. The findings emphasized that while AI tools significantly accelerate development processes, challenges such as skill atrophy, ethical considerations, and security risks must be addressed. The study proposed improvements in tool usability, integration capabilities, and security assessments to maximize the potential of AI in software development. In conclusion, this research highlights the transformative role of AI in accelerating development workflows. By balancing technological innovation with human oversight and ethical considerations, AI tools can drive sustainable and efficient software development, fostering a future of enhanced collaboration and innovation.

### 6.2 Future Research

To advance the integration and utility of AI-driven code-generation tools, future research should address current gaps, explore emerging dimensions, and enhance the understanding of AI's role in programming. Broadening the scope of evaluated tools, including underexplored options like Amazon CodeWhisperer and Snyk Code, alongside assessing their performance across diverse programming languages and frameworks, will provide a more comprehensive understanding of their versatility. Expanding sample diversity by involving developers from different industries and regions will ensure findings are more generalizable. Long-term studies are necessary to assess sustained impacts on developer skills, productivity, and project outcomes. Investigating the ethical and security implications of AI-generated code, such as vulnerabilities and intellectual property concerns, is crucial for ensuring fairness, transparency, and accountability. Enhancing the integration of AI tools into workflows, particularly CI/CD pipelines, will optimize their usability and accessibility. Educational initiatives should prepare developers for AI-augmented workflows, fostering skills in prompt engineering and collaborative development. Exploring AI-assisted pair programming and distributed team dynamics can provide insights into how AI supports collaboration and consistency in team-based environments. Emerging trends like advanced natural language processing, AI-driven code refactoring, and augmented debugging highlight the transformative potential of AI in programming. By focusing on these areas, future research can help realize AI's full potential, creating tools that are effective, secure, and ethically sound, ultimately driving innovation and efficiency in software development.

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

### References

1. Losbichler, S., & Lehner, O., "The Role of AI in Transforming Business Logic," 2021, last accessed 2024/11/09.
2. Yetiştirilen, T., et al., "AI Tools and Human Supervision in Software Development," 2023, last accessed 2024/11/08.

3. Reilly, J., "Economic Constraints in AI Adoption," 2023, last accessed 2024/11/10.
4. Bhati, K., "AI-Driven Initiatives in Large Enterprises," 2023, last accessed 2024/11/12.
5. Lau, S., & Guo, P., "From 'Ban It Till We Understand It' to 'Resistance is Futile': How University Programming Instructors Plan to Adapt as More Students Use AI Code Generation and Explanation Tools Such as ChatGPT and GitHub Copilot," 2023, last accessed 2024/11/11.
6. Takerngsaksiri, W., et al., "Students' Perspective on AI Code Completion: Benefits and Challenges," 2023, last accessed 2024/11/07.
7. Yang, J., et al., "How Important are Good Method Names in Neural Code Generation? A Model Robustness Perspective," 2023, last accessed 2024/11/09.
8. Yetistiren, T., et al., "Assessing the Quality of GitHub Copilot's Code Generation," 2022, last accessed 2024/11/08.
9. Yetistiren, T., et al., "Evaluating the Code Quality of AI-Assisted Code Generation Tools: An Empirical Study on GitHub Copilot, Amazon CodeWhisperer, and ChatGPT," 2023, last accessed 2024/11/10.
10. Li, J., et al., "Protecting Intellectual Property of Large Language Model-Based Code Generation APIs via Watermarks," 2023, last accessed 2024/11/12.
11. Siddiq, M. L., & Santos, J. C. S., "SecurityEval Dataset: Mining Vulnerability Examples to Evaluate Machine Learning-Based Code Generation Techniques," 2022, last accessed 2024/11/09.
12. Li, J., et al., "Evaluating the Impact of ChatGPT on Exercises of a Software Security Course," 2024, last accessed 2024/11/07.
13. Lozano, C., et al., "Constraint-Based Code Generation," 2013, last accessed 2024/11/08.
14. Xue, Y., "Automating Code Generation for Model-Driven Engineering Using Machine Learning," 2023, last accessed 2024/11/10.
15. Peng, S., et al., "The Impact of GitHub Copilot on Developer Productivity," 2024, last accessed 2024/11/12.
16. "IEEE Code of Ethics," 2020, <http://www.ieee.org/ethics>, last accessed 2024/11/09.
17. "Regulation 2016/679 - General Data Protection Regulation (GDPR)," 2016, <https://gdpr-info.eu/>, last accessed 2024/11/08.
18. ISO/IEC 27001:2022, "Information Security Management," 2022, <https://www.iso.org/standard/27001>, last accessed 2024/11/07.
19. "GitHub Copilot overview," 2024, <https://github.com/features/copilot>, last accessed 2024/11/10.
20. Valerie, L., "AI Coding Tools Compared: TabNine vs. Copilot," 2023, <https://tabnine.com>, last accessed 2024/11/09.
21. Maheshwari, A., "Advances in ChatGPT 4.0: Cost and Capabilities," 2023, <https://openai.com>, last accessed 2024/11/08.
22. Zhang, H., et al., "Enhancing Developer Productivity with GitHub Copilot," 2023, last accessed 2024/11/11.
23. Vaithilingam, A., et al., "Deep Learning for Code Prediction: A Study on TabNine," 2023, last accessed 2024/11/12.
24. Sridhara, D., et al., "Applications of ChatGPT in Technical and Creative Domains," 2023, last accessed 2024/11/09.
25. McDermott, K., et al., "Custom GPT Models: Tailoring AI for Industry Applications," 2023, last accessed 2024/11/08.
26. Clarisó, R., & Cabot, J., "Nuances in Prompt Engineering for AI," 2023, last accessed 2024/11/07.

# Cross-Platform Development Frameworks and Tools: Comparative study

Tiago Machado<sup>1</sup> and Helder Rodrigo Pinto<sup>2</sup>

<sup>1</sup> ISEP, Porto, Portugal

<sup>2</sup> ISEP, Porto, Portugal

1191097@isep.ipp.pt

**Abstract.** The rapid proliferation of devices and operating systems has reshaped software development, driving a transition from native to cross-platform approaches. While native development provides optimal performance and platform-specific experiences, it demands multiple codebases, increasing costs and complexity. Cross-platform frameworks address these challenges by enabling developers to create applications for multiple platforms from a single codebase. This paper explores the evolution, advantages, and challenges of cross-platform development, focusing on popular frameworks like Kotlin Multiplatform, Flutter, and React Native. It briefly evaluates their performance, flexibility, and ability to deliver consistent user experiences. Additionally, emerging trends such as the integration of Generative AI and multimodal interfaces are discussed, emphasizing their potential to enhance UI/UX and redefine cross-platform development. This study provides insights into selecting suitable frameworks and adapting to future advancements, highlighting the significance of scalable, efficient solutions in modern software engineering.

**Keywords:** Cross-Platform Development, Kotlin Multiplatform, Flutter, React Native, Mobile Application Development

## 1 Introduction

The rise of diverse devices and operating systems has reshaped software development, prompting a shift from native to cross-platform approaches. Native development, while offering optimal performance and platform-specific experiences, required separate codebases for each platform, leading to high development costs and increased complexity. This inefficiency, coupled with the growing need to deliver consistent experiences across devices, highlighted the limitations of maintaining siloed development workflows.

Cross-platform frameworks such as Xamarin and PhoneGap emerged to address these challenges by enabling developers to create applications for multiple platforms using a single codebase. While these early solutions were limited in performance and user experience, they laid the foundation for modern frameworks like Flutter, React

Native and the most recent Kotlin Multiplatform. These newer tools provide near-native performance, robust libraries, and greater flexibility, streamlining development, reducing costs, and accelerating time-to-market.

This transition from native to cross-platform development reflects the industry's need for scalable, efficient solutions that balance performance with accessibility. This paper explores the motivations, advantages, and challenges of cross-platform frameworks, underscoring their significance in modern software engineering.

## 1.1 The Pros and Challenges of Cross-Platform Mobile Development

Cross-platform development has become essential due to the proliferation of devices and operating systems that users interact with daily. However, users expect a seamless, consistent experience whether they access an application on a mobile phone, tablet, or desktop, which is not always easy to achieve. This demand has pushed engineers to adopt frameworks and techniques that allow a single codebase to serve multiple platforms, reducing development time and cost.

Cross-platform mobile development frameworks have become popular in software development for their ability to streamline processes and cut costs by enabling code reuse across platforms like Android and iOS. This approach contrasts with native development, where separate codebases are typically maintained for each platform, leading to higher development costs and more complex update cycles. Cross-platform frameworks simplify maintenance and provide compatibility across multiple operating systems [1].

Moreover, market research emphasizes the importance to consider platform-specific conventions to ensure a positive user experience, as differences in operating systems affect user expectations [2]. Although cross-platform frameworks can introduce performance overhead due to bridging mechanisms, research by Biørn-Hansen et al. indicates that performance differences between native and cross-platform applications are often minimal, making cross-platform solutions viable for many use cases [3].

Despite these advantages, achieving UI/UX consistency across various platforms remains a critical concern for developers. Variations in the look, feel, and interaction behaviors across different operating systems can lead to user experience discrepancies, complicating the development [4], [5].

## 1.2 Balancing Standards, Performance, and Maintainability

Cross-platform development presents a multitude of challenges that software engineers must navigate to ensure high-quality user interfaces (UI) and user experiences (UX). These challenges are often exacerbated by the inherent differences between platforms, which can influence the design and functionality of applications:

- **Device Fragmentation and Platform-Specific Standards:** One of the primary challenges in cross-platform development is device fragmentation, which necessitates that applications operate seamlessly across a diverse array of screen sizes, hardware capabilities, and operating systems. Wu et al. discuss

the complexities of developing applications for multiple platforms and emphasize the need for adaptable UI layouts to address these challenges [6]. Furthermore, adhering to established platform-specific standards, such as Material Design for Android and Apple's Human Interface Guidelines for iOS, is crucial for fostering user familiarity and comfort across different platforms [7].

- **Performance Optimization:** Performance optimization is another critical concern in cross-platform applications, which often fall behind native applications in terms of speed and responsiveness. Biørn-Hansen et al. identify that performance issues frequently stem from the additional abstraction layers required to support multiple platforms, leading to increased latency [3]. To counteract these performance drawbacks, engineers implement various optimization techniques, including lazy loading, efficient data handling, and minimizing unnecessary re-renders, which are essential for achieving performance levels that approach those of native applications.
- **Code Maintainability and Scalability:** Using a single codebase for multiple platforms can simplify initial development but often leads to complex, hard-to-maintain code. Biørn-Hansen et al. note that while cross-platform solutions streamline processes, they increase maintenance challenges as applications grow [3]. To address this, engineers should use modularization and separate platform-specific code, which makes future updates easier. This approach is supported by findings from Albeshier, who notes that the cross-platform development landscape is continually evolving, necessitating ongoing adaptation and refinement of codebases [8].

## 2 Cross-Platform Development Frameworks and Tools

### 2.1. Overview of Popular Cross-Platform Frameworks

Cross-platform frameworks have evolved to enable engineers to deliver consistent UI/UX across multiple operating systems from a single codebase. Selecting an appropriate framework is essential for achieving high performance, maintainability, and scalability. In addition to Flutter, React Native and Kotlin Multiplatform has emerged as a promising choice, especially for developers aiming to maintain native performance with flexible platform-specific code.

1. **Kotlin Multiplatform (KMP):** Developed by JetBrains, allows developers to share business logic across multiple platforms while still enabling the use of platform-specific user interface components. This hybrid approach distinguishes KMP from fully cross-platform solutions by maintaining native user interfaces for each platform, thereby minimizing code duplication for business logic. Additionally, KMP also supports shared UI components across platforms, offering a balance between consistency and customization. Recent research highlights KMP's unique advantage in its interoperability with native code, facilitating seamless integration with existing iOS and Android projects without compromising the performance or flexibility typically associated with native applications [9].



Fig. 1. Kotlin Multiplatform flexibility in project architecture [10]

2. **Flutter:** Another prominent open-source framework by Google and known for its Dart-based widget architecture, Flutter compiles directly to native code, resulting in near-native performance. Similarly to Kotlin Multiplatform, one of the key advantages of this solution is its ability to utilize a single codebase for both iOS and Android applications [11]. However, it requires a dedicated Flutter UI for each platform, making it less ideal for projects needing deep integration with custom or native UIs.

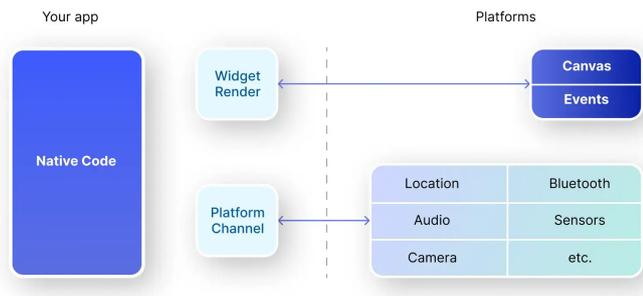


Fig. 2. Flutter's architecture overview [12]

3. **React Native:** Developed by Meta Platforms (formerly Facebook Inc.), React Native leverages JavaScript to allow developers to write once and adapt minimally for native rendering on each platform. While JavaScript's wide usage makes React Native highly accessible, certain features still require native bridging, which can sometimes lead to performance bottlenecks.



Fig. 3. JavaScript's architecture overview [12]

## 2.2. Framework Comparison

When selecting a cross-platform framework, software engineers must consider several critical criteria, including performance, flexibility in code sharing, UI consistency, and integration with existing codebases. Recent research highlights these metrics and provides guidance for engineers in choosing the most suitable framework for their specific project requirements.

### 2.2.1 Performance

Performance is a crucial factor in the selection of a cross-platform framework. React Native employs a bridge between JavaScript and native code, which can introduce latency, particularly in applications requiring high-performance graphics or complex animations. Studies indicate that React Native applications may experience longer load times and higher memory usage compared to their Flutter counterparts, which utilizes a compiled approach that directly interacts with native components, resulting in faster performance metrics [13], [14].

Kotlin Multiplatform, on the other hand, allows developers to write platform-specific code while sharing business logic across platforms. This hybrid approach can lead to performance that is closer to native applications, especially for computationally intensive tasks, as it leverages the strengths of the Kotlin language and its interoperability with existing Java codebases [15], [16]. However, the performance can vary depending on how much native code is utilized versus shared code, which necessitates careful architectural planning [3].

### 2.2.2 Code Sharing and Development Flexibility

Code sharing is a significant advantage of cross-platform frameworks, allowing developers to maintain a single codebase for multiple platforms. React Native facilitates this by enabling developers to write components in JavaScript, which can be shared across iOS and Android. However, the reliance on native modules for certain functionalities can limit flexibility, as developers may need to write platform-specific code when performance or native features are critical [17].

Flutter stands out in terms of development flexibility, as it provides a rich set of pre-built widgets and a comprehensive toolkit that allows for rapid development and customization. Its single codebase approach simplifies the development process, enabling developers to create visually consistent applications across platforms without the need for extensive platform-specific adjustments [14], [18].

Kotlin Multiplatform offers a unique blend of flexibility and code sharing, allowing developers to share business logic while still enabling both shared UI components across platforms and native UI customization for each platform. This approach can lead to more maintainable codebases, as developers can leverage existing native UI components while ensuring that core functionalities remain consistent across platforms [15], [16]. Additionally, Kotlin Multiplatform supports shared UI components across platforms, allowing for consistency in user interfaces while also accommodating platform-specific customization when necessary. However, this may also require a deeper understanding of both Kotlin and the native development environments, which can be a barrier for some teams.

### 2.2.3 UI Consistency and Platform-Specific Adaptation

UI consistency is critical for user experience, and each framework approaches this differently. React Native allows for a degree of customization but often requires additional effort to ensure that applications look and feel native on both iOS and Android. This can lead to inconsistencies if not managed carefully, as developers must account for the different design philosophies of each platform [14], [17].

Flutter excels in UI consistency due to its widget-based architecture, which allows developers to create a uniform look across platforms. The framework's rich set of customizable widgets ensures that applications can maintain a consistent aesthetic while also adapting to platform-specific design guidelines when necessary [13], [18]. However, this can sometimes lead to challenges in achieving true native look and feel, particularly for complex UI elements.

Kotlin Multiplatform allows for the use of native UI components, which can enhance the user experience by providing a more authentic look and feel on each platform. This approach can lead to better user satisfaction but may require additional development effort to manage the differences in UI design between platforms [15], [16].

### 2.3 Emerging Trends in Cross-Platform Frameworks and Future Directions

Trends in cross-platform frameworks are increasingly focused on leveraging Generative AI to enhance user interfaces across devices, creating more adaptive and seamless experiences. As users expect consistent interactions regardless of platform, Generative AI enables UIs to personalize and dynamically adjust in real time, catering to individual needs and preferences. This adaptability is key to developing UIs that are accessible on any device, from desktops to mobile phones, ensuring that interactions remain intuitive and contextually relevant. Lightweight frameworks are crucial in this space, as they allow these advanced AI functionalities to perform efficiently on hardware-limited devices by balancing cloud-based and on-device processing.

According to J. Bieniek, M. Rahouti et al. the evolution of user interfaces is moving from single modal to more adaptive, multimodal systems that accommodate a diverse range of inputs, creating seamless experiences across platforms [19]. This shift addresses what the authors call the "interface dilemma," where designers must balance various interaction methods to create effective multimodal experiences. These innovations highlight the need for scalable, lightweight frameworks, especially on mobile platforms, to support these dynamic interactions while addressing challenges around privacy, context retention, and resource management. Generative AI's role in enhancing multimodal capabilities underscores its potential to redefine cross-platform user interfaces, making them more intuitive, responsive, and user-centric across digital devices [19].

Looking forward, cross-platform frameworks will likely emphasize emotionally adaptive and predictive AI-driven UIs that respond to user behaviors and real-time cues. This evolution, however, brings ethical considerations such as privacy, data security, and computational efficiency, which are essential for building and maintaining user trust in AI-powered systems. As Generative AI continues to advance, it is positioned to

redefine UIs, setting new standards for adaptive, intelligent, and responsible technology across digital platforms [19].

### 3 Conclusion

The evolution of cross-platform frameworks has revolutionized software development by enabling developers to create applications that deliver consistent user experiences across diverse devices and operating systems. Each framework offers distinct advantages that cater to specific project needs, underscoring the importance of thoughtful technology selection in modern app development.

Flutter is a robust framework for building applications with consistent and visually appealing user interfaces across platforms. Its extensive library of customizable widgets streamlines UI development, making it particularly advantageous for design-centric projects. However, its use of Dart as a primary language may necessitate additional learning for teams unfamiliar with this ecosystem.

React Native provides a pragmatic solution for teams seeking to leverage their existing JavaScript expertise. Its ability to share code between platforms, coupled with a well-established ecosystem, supports rapid prototyping and accelerated development cycles. Nevertheless, challenges such as performance constraints in computationally intensive tasks and reliance on external libraries for advanced functionality require consideration.

Kotlin Multiplatform introduces a hybrid model, enabling the reuse of business logic while preserving flexibility for platform-specific user interface implementation. This balance between shared code and native adaptation makes it suitable for projects with complex requirements spanning multiple platforms. However, its nascent ecosystem and the associated learning curve may present hurdles during initial adoption.

In conclusion, cross-platform frameworks have emerged as invaluable tools in modern software engineering, enabling efficient support for multiple platforms. By continuously innovating and addressing their inherent challenges, they will remain pivotal in shaping the next generation of software solutions, enabling developers to meet the ever-changing demands of users across the globe.

### References

1. A. Biørn-Hansen, T. M. Grønli, G. Ghinea, and S. Alouneh, "An Empirical Study of Cross-Platform Mobile Development in Industry," *Wirel Commun Mob Comput*, vol. 2019, no. 1, p. 5743892, Jan. 2019, doi: 10.1155/2019/5743892.
2. A. Holzinger, P. Treitler, and W. Slany, "Making Apps Useable on Multiple Different Mobile Platforms: On Interoperability for Business Application Development on Smartphones," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7465 LNCS, pp. 176–189, 2012, doi: 10.1007/978-3-642-32498-7\_14.

3. A. Biørn-Hansen, C. Rieger, T. M. Grønli, T. A. Majchrzak, and G. Ghinea, “An empirical investigation of performance overhead in cross-platform mobile development frameworks,” *Empir Softw Eng*, vol. 25, no. 4, pp. 2997–3040, Jul. 2020, doi: 10.1007/S10664-020-09827-6.
4. S. Mascetti, M. Ducci, N. Cantù, P. Pecis, and D. Ahmetovic, “Developing Accessible Mobile Applications with Cross-Platform Development Frameworks”, doi: 10.48550/ARXIV.2005.06875.
5. M. Martinez and S. Lecomte, “Towards the Quality Improvement of Cross-Platform Mobile Applications,” *2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, pp. 184–188, Jul. 2017, doi: 10.1109/MOBILESOFT.2017.30.
6. C. Wu, J. M. Pérez-Álvarez, A. Mos, and J. M. Carroll, “Codeless App Development: Evaluating A Cloud-Native Domain-Specific Functions Approach”, doi: 10.48550/ARXIV.2210.01647.
7. L. Albeshar, R. Aldossari, and R. Alfayez, “An Observational Study on React Native (RN) Questions on Stack Overflow (SO),” *IET Software*, vol. 2023, no. 1, 2023, doi: 10.1049/2023/6613434.
8. L. Albeshar, R. Aldossari, and R. Alfayez, “An Observational Study on React Native (RN) Questions on Stack Overflow (SO),” *IET Software*, vol. 2023, no. 1, 2023, doi: 10.1049/2023/6613434.
9. V. NUTALAPATI, “Concept to Completion - Android Apps and Kotlin Multi Platform,” 2024, doi: 10.61909/AMKEDTB082431.
10. “KMP - flexible multiplatform development (part 1) - Application Engineering.” Accessed: Nov. 28, 2024. [Online]. Available: <https://blog.nashtechglobal.com/kmp-flexible-multiplatform-development-part-1/>
11. K. Wasilewski and W. Zabierowski, “A comparison of java, flutter and kotlin/native technologies for sensor data-driven applications,” *Sensors*, vol. 21, no. 10, May 2021, doi: 10.3390/S21103324.
12. “Flutter vs React Native in 2024: End-To-End Comparison.” Accessed: Nov. 28, 2024. [Online]. Available: <https://markovate.com/blog/flutter-vs-react-native/>
13. I. E. KHO, A. W. ALIYAZIS, and M. GALINIUM, “Front-End Application for Multiple Storefronts Ecommerce Using Cross-Platform Technology,” *Business Excellence and Management*, vol. 12, no. 1, pp. 93–104, Mar. 2022, doi: 10.24818/BEMAN/2022.12.1-07.
14. E. GÜLCÜOĞLU, A. B. USTUN, and N. SEYHAN, “Comparison of Flutter and React Native Platforms,” *Journal of Internet Applications and Management*, Dec. 2021, doi: 10.34231/IUYD.888243.
15. Y. KOZUB and H. KOZUB, “Features of Multiplatform Application Development on Kotlin,” *Herald of Khmelnytskyi National University. Technical Sciences*, vol. 317, no. 1, pp. 224–229, Feb. 2023, doi: 10.31891/2307-5732-2023-317-1-224-229.
16. B. Suri, S. Taneja, I. Bhanot, H. Sharma, and A. Raj, “Cross-Platform Empirical Analysis of Mobile Application Development frameworks: Kotlin, React Native and Flutter,” *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, Dec. 2022, doi: 10.1145/3590837.3590897.

17. D. Wheeler and J. I. Olszewska, "Cross-Platform Mobile Application Development for Smart Services," 2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on R, pp. 203–208, 2022, doi: 10.1109/CINTI-MACRO57952.2022.10029466.
18. A. Nedyak, O. Rudzeyt, A. Zainetdinov, and P. Ragulin, "Mobile cross-platform app development tools," Russian Journal of Resources, Conservation and Recycling, vol. 7, no. 4, Dec. 2020, doi: 10.15862/13INOR420.
19. J. Bieniek, M. Rahouti, and D. C. Verma, "Generative AI in Multimodal User Interfaces: Trends, Challenges, and Cross-Platform Adaptability," 2024

# NutriScan: Sistema De Monitorização Alimentar baseado em Inteligência Artificial para Reconhecimento de Imagem

Ana Caraméz<sup>1</sup>, Ana Marcelino<sup>1</sup>, Beatriz Sousa<sup>1</sup>, Bruna Cunha<sup>1</sup>, Pedro Guimarães<sup>2</sup>, Frederico Jacob<sup>3</sup>

<sup>1</sup>Licenciatura em Engenharia Biomédica, ISEP, Portugal  
{1211327, 1210774, 1211707, 1211305}@isep.ipp.pt.

<sup>2</sup>Departamento de Física, ISEP, Portugal, psg@isep.ipp.pt.

<sup>3</sup>Departamento de Física, ISEP, Portugal, flj@isep.ipp.pt

**Resumo.** O projeto "NutriScan" apresenta uma solução inovadora para enfrentar o crescente problema de obesidade e excesso de peso, combinando tecnologias emergentes de monitorização alimentar. A proposta consiste num sistema integrado, composto por uma aplicação móvel e um website, suportados por uma rede neuronal e uma base de dados robusta, que permite a análise automática de fotografias de refeições submetidas pelos utilizadores. Este sistema visa facilitar o acompanhamento nutricional, fornecendo gráficos e relatórios personalizados para os utilizadores e profissionais de saúde. O método adotado inclui o treino de uma rede neuronal com imagens de alimentos e a utilização de uma base de dados para determinar os valores nutricionais das refeições. Com isto, espera-se promover uma maior consciência alimentar, simplificar a monitorização nutricional e oferecer suporte remoto a profissionais da área. O projeto demonstra o impacto potencial da automonitorização no controlo de hábitos alimentares e na promoção de saúde pública.

**Palavras-chave.** nutrição, monitorização automatizada, *eHealth*, *machine learning*, *android*, *web*, rede neuronal

## 1 Introdução

Atualmente em Portugal, verifica-se um aumento da prevalência do excesso de peso tanto em crianças e adolescentes como em adultos, atingindo mais de um quarto da população das faixas etárias mais novas e mais de metade da população adulta. A obesidade é definida como a acumulação excessiva de massa gorda corporal, com potencial impacto negativo na saúde [1]. Segundo a OMS (Organização Mundial de Saúde), o consumo de calorias deve ser proporcional ao gasto de calorias nas atividades da vida diária [2].

Para cada tipo de produto, existem rótulos nutricionais, que fornecem informações sobre o conteúdo nutricional dos alimentos. Verifica-se que muitos dos consumidores dão importância às informações nutricionais, pois acreditam que estas são importantes

para manter ou controlar os hábitos de saúde [3]. Deste modo, revela-se fundamental compreender qual é o impacto da informação nutricional percebida na intenção de melhoria de saúde. Segundo o Regulamento da EU N° 1169/2011, 25 de outubro, a declaração nutricional dos produtos alimentares obriga à inclusão do valor energético (kJ e kcal), lípidos (g), ácidos gordos saturados (g), hidratos de carbono (g), açúcares (g), proteínas(g) e sal (g) [4].

Os seres humanos são confrontados com várias escolhas alimentares todos os dias, fazendo com que a seleção de alimentos seja umas das atividades mais comuns e mundanas [5]. Numa sociedade em constante mudança, a grande oferta de produtos e a facilidade de aquisição dos mesmos, tem resultado em importantes alterações nas características de compra dos consumidores e nos distintos hábitos de consumo [6]. Deste modo, surge a preservação da saúde como o grande objetivo na seleção de alimentos, fazendo com que os consumidores se tornem mais preocupados com estilos de vida saudáveis [3].

Comer é multidimensional e difícil de monitorizar. Ao longo do tempo, têm surgido diferentes formas de monitorização, tornando-se imprescindível melhorar a exatidão das bases de dados da composição dos alimentos. A existência de cada vez mais dados disponíveis de tudo o que fazemos e comemos e a crescente tendência de tornar tudo em formato digital, abre uma janela de oportunidades para que seja possível monitorizar tudo o que consumimos. No entanto, problemas como a inexistência de uma base de dados com a informação nutricional dos produtos processados e a legislação do RGPD (Regulamento Geral sobre a proteção de dados) dos utilizadores pode dificultar o acesso a determinados dados [3, 7].

A automonitorização dos alimentos ingeridos é uma estratégia que permite consciencializar as pessoas sobre a alimentação e os seus padrões nutricionais, através do fornecimento de feedback automatizado e personalizado [8]. O registo dos alimentos ingeridos no decorrer de várias semanas é a forma mais fiável de verificar os dados registados, já que não dependem da memória do indivíduo [9].

## 1.1 Métodos Tradicionais

Um nutricionista tem como objetivo perceber as necessidades de saúde relacionadas com a nutrição. Assim, ajuda no desenvolvimento de prioridades e objetivos, de forma a implementar planos de cuidados nutricionais. Além disso, deve interpretar resultados de análises relacionados com o estado de nutrição, assim como fornecer aconselhamento nutricional para promover ou manter a saúde do seu paciente [10].

Os métodos tradicionais de avaliação dietética, tais como diários alimentares ou questionários de frequência alimentar não são capazes de fornecer um feedback com precisão necessária ou são impraticáveis para a automonitorização, devido à grande carga de participantes e baixa escalabilidade [9]. Esta prática resulta em registos imprecisos, dado que são preenchidos com base na memória. Além disso, há uma grande tendência dos utilizadores subestimarem o tamanho das porções e se esquecerem de produtos consumidos entre refeições [11].

## 1.2 Aplicações Móveis

Os registos através de aplicações podem melhorar a monitorização, uma vez que os dados são gravados automaticamente, não exigindo esforço por parte do consumidor. Além disso, permitem o registo de diferentes informações num único sítio, embora a análise de uma imagem não permita avaliar quantidades [3]. A monitorização dos alimentos ingeridos pode ser feita de duas maneiras: através de inferências devidas ao uso de receitas, mudanças no inventário alimentar e compras ou a partir de medidas diretas como a captura de imagem e o uso de vários sistemas de deteção no corpo para monitorizar comportamentos alimentares [9].

As tecnologias emergentes, como o reconhecimento de imagens e o processamento de linguagem natural, não estão presentes na maioria das aplicações de nutrição mais populares. As câmaras fotográficas permitem registar graficamente os alimentos, permitindo assim uma análise a três dimensões preponderante para a análise de quantidades [3].

## 1.3 Internet of Things (IoT)

A IoT é uma tecnologia emergente que consiste na utilização de vários dispositivos heterogéneos que podem trocar dados sem interação humana [9].

Um exemplo desta tecnologia é a utilização de uma placa com um sensor de pesagem de alimentos. O peso do produto ou alimento é enviado para a nuvem através da Internet com a coordenação de um microprocessador sem fios [12]. Outro exemplo baseia-se na utilização do modelo de *deep-learning* para determinar o equilíbrio nutricional após cada refeição e num algoritmo baseado em redes *bayesianas* para determinar características nutricionais de materiais alimentares, sugerindo futuras refeições ou receitas [13].

## 1.4 Sistemas baseados na Web

As ferramentas de avaliação nutricional *online* utilizam abordagens padrão e incluem questionários e registos alimentares de 24 horas. Estes sistemas podem incluir várias opções de seleção de porções para apoiar o utilizador como, por exemplo, fotografias de alimentos ou submissão do peso real da porção [14]. A título de exemplo seguem-se alguns *websites* e as suas diferenças:

- O *website Intake24* é baseado num registo múltiplo de 24 horas e inclui espaços reservados para as diversas refeições, sendo que o utilizador pode optar pela ordem pela qual os preenche. Posteriormente, é necessário que o utilizador estime o tamanho da porção [15];
- O *website myfood24* permite uma monitorização da ingestão nutricional por profissionais de saúde. Permite estabelecer objetivos para o paciente e receber os dados do doente em gráficos para *feedback* visual [3];

- O *website ASA24* é uma plataforma de registo alimentar utilizada para registo dos alimentos em cada refeição e preenchimento de questionários disponibilizados mensalmente. Apresenta um registo de 24 horas e uma interface de utilizador dinâmica [14].

Considerando as limitações dos métodos de monitorização alimentar analisados anteriormente, verificou-se a necessidade de criar um sistema inovador no qual a leitura, tratamento e amostragem de dados fosse simples e eficaz. Perante a necessidade referida, o grande objetivo do projeto centraliza-se no desenvolvimento de uma aplicação android e website que permitam uma monitorização alimentar real e adequada.

Para se concretizar o objetivo acima referido é essencial incorporar uma rede neuronal, para identificação dos alimentos presentes na fotografia submetida e uma base de dados de modo a proporcionar uma monitorização da alimentação dos utilizadores mais rápida e concisa. Para além disto, esta tecnologia permite que os profissionais de saúde supervisionem de forma remota a condição dos seus pacientes.

Desta forma, o NutriScan pretende ter o papel de facilitar o registo de informação, fazendo-o através da submissão de fotografias de refeições. Sendo assim, a utilização de outras aplicações android, que maioritariamente não incorporam esta tecnologia, bem como websites, que exigem que o utilizador registe grande parte da informação, torna-se exaustiva, relativamente à interatividade e simplicidade que o NutriScan pretende oferecer.

## 2 Descrição do Sistema

O projeto consiste na implementação de uma aplicação android e web, que funcionem em conjunto de modo a permitir a submissão de fotografias de refeições, com posterior análise da informação nutricional da mesma.

Os componentes do projeto são a aplicação móvel *android* (Fig. 1 – (1)), a aplicação *web* (Fig. 1 – (4)), a rede neuronal (Fig. 1 – (2)) e a base de dados (Fig. 1 – (3)). O funcionamento do projeto começa com o registo de um utilizador (paciente ou nutricionista, Fig. 1 – (a) e (b), respetivamente) na aplicação *web*. Aqui podem ser consultados os valores nutricionais associados a cada uma das fotografias submetidas na aplicação. Para além disto, existem ainda gráficos e tabelas, que permitem organizar a informação recolhida numa interface gráfica intuitiva. A página principal dos utentes permite que este tenha acesso à evolução do número de calorias consumidas, quais os nutrientes mais e menos consumidos e, ainda, detalhar a quantidade exata desses mesmos nutrientes. Por outro lado, a página principal dos nutricionistas permite perceber quantas refeições já foram submetidas na aplicação, quais são os alimentos mais consumidos pelos seus utentes e ainda aceder a informações únicas acerca destes.

Assim, o paciente deve fazer *login* na aplicação móvel, permitindo a submissão das fotografias das refeições e a sua visualização numa galeria de fotos. As fotografias submetidas na aplicação são analisadas pela rede neuronal a fim de identificar os alimentos e, assim, quantificar o seu valor nutricional, já que o *output* desta é o nome do alimento identificado. Assim, este será comparado com os nomes de alimentos de uma base de dados teórica, resultando na atribuição do valor energético do alimento ao elemento que

foi fotografado. Neste processo, são armazenadas todas as informações relativas à fotografia na base de dados. Adicionalmente, os metadados associados a cada fotografia são também guardados, a fim de organizar a informação de forma correta na aplicação web.

Criar uma rede neuronal para identificar alimentos exige muitas etapas de trabalho, desde a preparação dos dados até ao treino do modelo. As imagens que serviram de treino foram redimensionadas, de forma que tivessem todas o mesmo tamanho. Foram feitos ajustes, nomeadamente ao nível da tradução e precisão.

Por fim, a base de dados, tal como o nome indica, tem como principal finalidade armazenar todos os dados do projeto, desde as fotografias submetidas, aos dados pessoais dos utentes e nutricionistas e ainda os valores da informação nutricional de cada prato. Para chegar ao valor nutricional correto de cada alimento foi utilizada uma base de dados do Instituto Nacional de Saúde, que permite comparar o alimento submetido com aqueles que estão presentes nesta. A elaboração da base de dados envolveu a criação de um sistema organizado para guardar, entregar e recuperar dados de forma eficiente.

Os componentes do sistema, mencionados anteriormente, relacionam-se entre si pelo esquema que se encontra na figura abaixo:

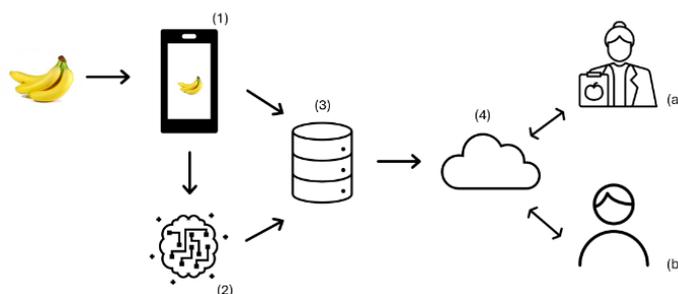


Fig. 1. Componentes do sistema

### 3 Implementação do Software

Para o desenvolvimento da aplicação *web* foram usados HTML, CSS, JavaScript e PHP. O HTML (HyperText Markup Language) é uma linguagem de publicação usada para o desenvolvimento de sites, que possibilita a criação de documentos com texto, ligações de hipertexto e links para vídeos entre outros [16]. O PHP (Hypertext PreProcessor) é uma linguagem de programação orientada a objetos, centrada no desenvolvimento de aplicações do lado do servidor escrito dinamicamente, com um modelo de classe de herança única e vários tipos padrão incorporados [17]. A linguagem CSS (Cascading Style Sheets) é uma sintaxe de marcação suplementar, que permite a associação de estilos ao conteúdo HTML [18]. O JavaScript foi originalmente concebido para ser uma linguagem de script HTML simples. As principais primitivas são funções de primeira classe e potencialmente de ordem superior e uma forma de objeto que pode

ser definida por uma expressão de objeto, sem necessidade de declarações de classe [19].

Foi então utilizado o editor de código Visual Studio Code. É fornecido com suporte integrado para JavaScript, TypeScript e Node.js e possui um rico ecossistema de extensões para outras linguagens e tempos de execução [20].

Por outro lado, a base de dados foi elaborada com SQL (Structured Query Language). Esta linguagem de programação pode ser usada para definir e tratar dados, sendo possível criar tabelas, editar dados nelas e recuperar esses dados [21]. Foi utilizado o MySQL, um sistema de gestão de bases de dados relacionais (RDBMS) open-source, para concretizar a criação da base de dados [22].

A aplicação android e o servidor web foram criados com recurso a Java, fazendo uso do plugin Android Studio para a aplicação android e do ambiente de desenvolvimento integrado (IDE) IntelliJ. Java é uma linguagem de programação baseada em classes e orientada para objetos, concebida para ter o menor número possível de dependências de implementação [23]. O IntelliJ é um IDE concebido para ajudar no processo de codificação, suportando um grande número de diferentes estruturas, ferramentas e objetivos e funcionando com várias linguagens de programação [24]. Por fim, o Android Studio é um plugin do IntelliJ que possibilita o desenvolvimento de aplicações android.

Finalmente, para o desenvolvimento da rede neuronal foi utilizada a linguagem Python. Esta é uma linguagem de programação de alto nível, dinâmica, orientada para objetos e de propósito geral, que utiliza um interpretador e pode ser utilizada num vasto domínio de aplicações [28].

## **4 Resultados e Discussão**

### **4.1 Base de Dados**

As tabelas criadas pela base de dados em formato SQL, que permitiram uma manipulação de dados correta e eficiente, com principal foco na organização de informação mais conveniente para ser acessada pelos restantes componentes. De referir que esta base de dados não foi criada com um foco de ser genérica ao ponto de poder ser utilizada por aplicações cliente bastante diferentes das que se inserem no restante sistema. Além disso, a informação nutricional armazenada de cada alimento por 100g é proveniente de uma base de dados do Instituto Nacional de Saúde Doutor Ricardo Jorge, que embora tenha uma grande variedade é sempre possível complementar com outras bases de dados, principalmente no que diz respeito a alimentos processados por diferentes marcas [26].

### **4.2 Aplicação Android**

A nível da aplicação, verifica-se que esta cumpre o seu principal propósito, que é aceder à galeria de fotografias do telemóvel e selecionar uma delas, de modo que seja a submetida para a conta do utilizador. A aplicação foi desenvolvida de modo a permitir as ações básicas de login de um paciente, submissão de fotos e visualização das mesmas. Numa fase futura, seria desejável melhorar a estética e funcionalidades da aplicação.

### 4.3 Rede Neuronal

A rede neuronal foi treinada para reconhecer e identificar o máximo de alimentos possíveis, tendo existido melhorias significativas desde o início do seu desenvolvimento. Verifica-se que quanto mais complexo for o prato, mais demorada será a identificação do mesmo, como seria de esperar. Além disso, quantos mais alimentos for possível identificar, maior será o tempo de treino da rede neuronal. Adicionando ao aumento temporal, os dados de treino e validação também aumentam, originando que o processo de treino seja mais demorado. Contudo, o aumento da qualidade da rede neuronal permite uma identificação mais eficaz dos alimentos submetidos na aplicação móvel.

### 4.4 Aplicação Web

Os gráficos e tabelas criadas permitiram organizar a informação recolhida pela base de dados de forma interativa, já que é possível ver qual a evolução do paciente durante um dado período.

A tabela gerada na página principal dos utentes quantifica os valores de cada fotografia submetida, de acordo com os nutrientes mais importantes, ou seja, lípidos, hidratos de carbono, fibra, proteínas, açúcares e sal. Aqui também se encontram 2 gráficos, que ilustram a evolução do número de calorias consumidas e qual a percentagem por nutrientes consumida. Por outro lado, a tabela gerada na página principal dos nutricionistas permite ver os dados pessoais dos utentes. Para além disto, através do clique no ID do paciente em questão, é possível ver a tabela gerada na página dos utentes, mencionada anteriormente. Desta forma, verifica-se que os profissionais de saúde estão informados de todos os detalhes da saúde alimentar dos seus utentes. Podem ainda ver o número de refeições submetidas por data e quais os alimentos mais e menos consumidos.

Por fim, é importante ressaltar que a aplicação *web* comunica com os restantes elementos do projeto por meio da base de dados. Ou seja, tanto pode inserir ou ir buscar dados a esta, funcionando para todos elementos. Foi muito importante o teste recorrente das *queries* SQL, de forma a garantir que as colunas e atributos selecionados nas mesmas eram os corretos.

As Fig. 2 ilustra os resultados obtidos num dos testes efetuados no projeto. Aqui, o utilizador começou por fazer *login* na aplicação, seguido de submissão da fotografia na mesma. Após isto, fez *login* na aplicação *web* e, através dos gráficos e das tabelas da página principal dos utentes, é possível observar os valores nutricionais associados à fotografia que submeteu anteriormente. Para além disto, foi também feito o *login* de um nutricionista, a fim de conseguir perceber o que acontecia nesta etapa. Verifica-se que, para além de ver as informações associadas ao utente em questão, consegue ver a mesma tabela que este vê, ou seja, consegue ver a totalidade das informações nutricionais associadas ao seu paciente. Para além disto, tem outros dois gráficos presentes na sua página principal, que vão de acordo aos resultados obtidos.

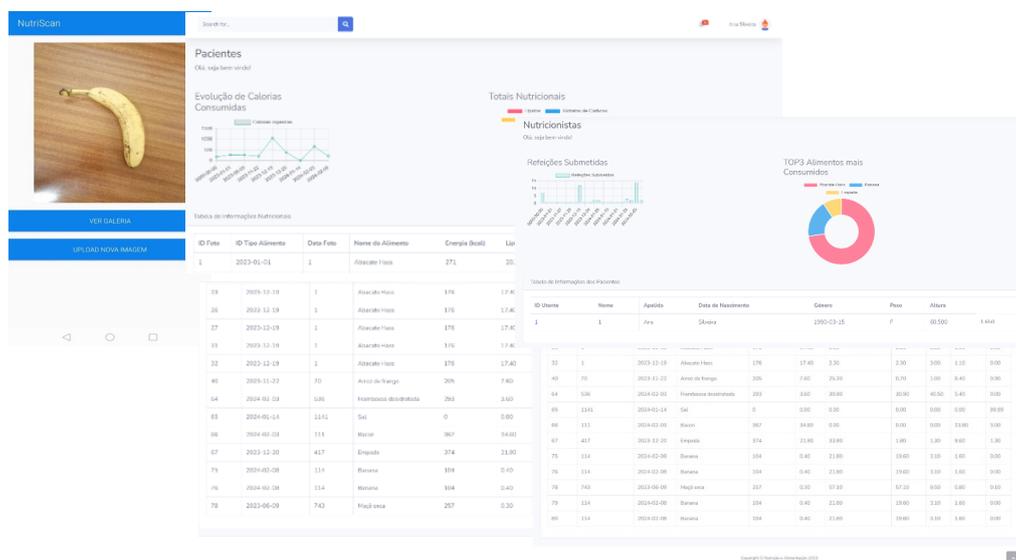


Fig. 2. Esquema da utilização e layouts da aplicação *android* e *web*

#### 4.5 Limitações e Aplicações Futuras

O sistema criado apresenta algumas limitações, sendo a mais visível a falta de estimativa de qual o peso de cada alimento ou prato seja por parte da rede neuronal ou por falta de hardware específico para tal, o que implica que este fator não seja contabilizado para o cálculo do valor nutricional de cada refeição. Para além disto, a capacidade de cada nutricionista poder comunicar e não apenas analisar a informação de cada seu utente é também algo limitativo. Por fim, a indústria alimentar está sempre em evolução e nesta área é impossível fugir ao fenómeno da globalização, o que implica que o treino da rede neuronal, assim como a contínua adição de informação nutricional para outros alimentos e pratos complexos, seja imprescindível, a fim de não existirem lacunas no cálculo nutricional global de cada utilizador. A imagem é um dos principais métodos de diagnóstico na área da saúde, logo é de esperar que este sistema que interliga e suporta o trabalho de um nutricionista possa ser utilizado noutras áreas da saúde para um acompanhamento por parte de outros profissionais de saúde, tais como na especialidade de medicina interna ou anatomia patológica.

### 5 Conclusão

Como já foi referido, a medição da quantidade de nutrientes que são ingeridos é uma tarefa complexa, por não existir um processo de monitorização e avaliação fidedigno, seja por falha humana no registo das informações ou na precisão dos dispositivos na recolha dos dados. Neste sistema, a inserção de fotografias dos pratos consumidos permite que o utilizador faça um controlo nutricional de forma mais realista, uma vez que a visualização destes dados é feita através de gráficos temporais, permitindo fazer uma análise nutricional durante longos períodos.

Desta forma, o NutriScan apresentou-se como uma solução inovadora para monitorizar a alimentação, combinando tecnologias emergentes como redes neuronais e bases

de dados robustas para facilitar a análise nutricional. A aplicação móvel e a web desempenham papéis complementares, com potencialidades e fragilidades distintas que merecem atenção.

A aplicação móvel destaca-se por permitir a submissão prática de fotografias das refeições, promovendo um registo intuitivo e acessível para os utilizadores. No entanto, a sua interface ainda carece de um maior refinamento estético e interativo, expandindo a experiência do utilizador. Para além disto, a identificação de alimentos em pratos mais complexos é um desafio que impacta diretamente a sua funcionalidade.

Já a aplicação web oferece ferramentas eficazes para organizar e apresentar os dados nutricionais, permitindo aos profissionais de saúde supervisionarem remotamente os utentes. A integração com gráficos e tabelas interativos reforça o seu papel como uma plataforma de suporte no acompanhamento nutricional. No entanto, a dependência da base de dados para consulta de informações nutricionais pode ser uma fragilidade em contextos onde os dados são incompletos ou desatualizados.

Como trabalhos futuros, propõe-se:

- Otimização da rede neuronal, visando uma maior capacidade de identificar alimentos complexos com rapidez e precisão. Treino de uma rede mais complexa para estimar o valor nutricional de uma refeição submetida ao invés de apenas a identificação dos alimentos.
- Melhoria da interface móvel, tornando-a mais dinâmica, interativa e atrativa para os utilizadores.
- Ampliação das funcionalidades da aplicação web, como a troca de mensagens entre utentes e nutricionistas e a inclusão de notificações personalizadas com base em parâmetros nutricionais e fisiológicos.
- Expandir a informação nutricional armazenada na base de dados, acrescentar informação nutricional de outros pratos complexos fornecida por outras fontes.
- Integração de tecnologias complementares, como sensores de peso, para aumentar a precisão dos dados recolhidos.

Com estas melhorias, o NutriScan poderá consolidar-se como uma ferramenta abrangente e indispensável em diferentes contextos, desde o uso clínico ao acompanhamento nutricional em larga escala. Apesar das limitações iniciais, o sistema já demonstrou eficácia e potencial para transformar a forma como a nutrição é monitorizada, contribuindo para a promoção da saúde pública e a conscientização alimentar.

## Referências

1. Direção-Geral da Saúde. (2016). Processo Assistencial Integrado da Pré-Obesidade no Adulto;
2. Hooper, L., et al. (2012). Effect of reducing total fat intake on body weight: Systematic review and meta-analysis of randomised controlled trials and cohort studies. *BMJ*, 345, e7666. <https://doi.org/10.1136/bmj.e7666>.
3. Ferraz, A. T. (2022). *MyFood – Sistema de monitorização nutricional*. (Tese de Mestrado, Instituto Superior de Engenharia do Porto do Instituto Politécnico do Porto);

4. Ordem dos Nutricionistas. (2014). *NOVAS REGRAS DE DECLARAÇÃO NUTRICIONAL DEBATIDAS POR ESPECIALISTAS*;
5. Couto, C. (2018). Consumo de Alimentos Saudáveis nos adolescentes e jovens adultos em Portugal: O impacto da estrutura social, sabor e do preço. (Tese de Mestrado, Faculdade de Economia da Universidade do Porto);
6. Ramalho, T. (2021). O comportamento do consumidor português face à segurança alimentar. (Tese de Mestrado, Universidade de Évora);
7. Schiro, J. L., et al. (2020). #Healthy: Smart digital food safety and nutrition communication strategies—a critical commentary. *Npj Science of Food*, 4(1), 14. <https://doi.org/10.1038/s41538-020-00074-z>
8. Javadí, B. et al. (2018). Smart nutrition monitoring system using heterogeneous internet of things platform. Em G. Fortino, A. B. M. S. Ali, M. Pathan, A. Guerrieri, & G. Di Fatta (Eds.), *Internet and Distributed Computing Systems* (pp. 63–74). Springer International Publishing. [https://doi.org/10.1007/978-3-319-97795-9\\_6](https://doi.org/10.1007/978-3-319-97795-9_6)
9. Zimmerman, M. & Snow, M. E. (2021). *Nutrition Basics: An Active Approach*. FlatWorld.
10. Andersen, D. et al. (2018). Academy of nutrition and dietetics: Revised 2017 scope of practice for the registered dietitian nutritionist. *Journal of the Academy of Nutrition and Dietetics*, 118(1), 141–165. <https://doi.org/10.1016/j.jand.2017.10.002>
11. Boland, M., & Bronlund, J. (2019). Enutrition—The next dimension for ehealth? *Trends in Food Science & Technology*, 91, 634–639. <https://doi.org/10.1016/j.tifs.2019.08.001>
12. Tambe, H. (2019). Smart nutrition monitoring system. *International Journal for Research in Applied Science and Engineering Technology*, 7(12), 404–406. <https://doi.org/10.22214/ijraset.2019.12067>
13. Sundaravadivel, P., Kesavan, K., Kesavan, L., Mohanty, S. P., & Kougiianos, E. (2018). Smart-log: A deep-learning based automated nutrition monitoring system in the iot. *IEEE Transactions on Consumer Electronics*, 64(3), 390–398. <https://doi.org/10.1109/TCE.2018.2867802>
14. Cade, J. E. (2017). Measuring diet in the 21st century: Use of new technologies. *Proceedings of the Nutrition Society*, 76(3), 276–282. <https://doi.org/10.1017/S0029665116002883>
15. *Intake24 | System features*. (sem data).
16. Raggett, D. et al (1998). *HTML 4.0 Specification*. Recuperado de <https://immagic.com/eLibrary/ARCHIVES/SUPRSEDED/W3C/W980424S.pdf>
17. Hills, M. et al. (2013). An empirical study of PHP feature usage: A static analysis perspective. *Proceedings of the 2013 International Symposium on Software Testing and Analysis*, 325–335. <https://doi.org/10.1145/2483760.2483786>
18. Goodman, D. (2007). *Dynamic HTML: The definitive reference*. O’Rilley.  
Maffeis, S., Mitchell, J. C., & Taly, A. (2008). An operational semantics for javascript. Em G. Ramalingam (Ed.), *Programming Languages and Systems* (Vol. 5356, pp. 307–325). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-89330-1\\_22](https://doi.org/10.1007/978-3-540-89330-1_22)
19. Visual studio Ide and code editor for software developers and teams. (sem data). Visual Studio.
20. *The SQL programming language | WorldCat.org*. (2010)
21. *Mysql introduction*. (sem data).
22. *Java Language*. (sem data). Hmong Wiki.
23. Krochmalski, J. (2014). *IntelliJ IDEA essentials: develop better software fast with IntelliJ IDEA*. Packt Pub Ltd.
24. Srinath, K. R. (2017). Python—the fastest growing programming language. *International Research Journal of Engineering and Technology*, 4(12), 354-357.
25. *Insa\_pt*. (sem data).

# Semi-automatic Image Tagging for ML Workflows: a GUI for Industrial Applications

Vasco Costa<sup>1</sup>, Lobinho Gomes<sup>3</sup>, Carlos José Campos<sup>1,4</sup>, Veríssimo Lima<sup>1</sup>,  
António Sousa<sup>1,2</sup>, and Fernando Carvalho<sup>1</sup>

<sup>1</sup> Instituto Superior de Engenharia do Porto (ISEP) – P.Porto, Porto, Portugal

<sup>2</sup> LEMA, ISEP-IPP, Porto, Portugal

<sup>3</sup> Gislotica – Projeto e Fabrico de Sistemas Mecânicos Lda, Porto, Portugal

<sup>4</sup> Instituto de Engenharia de Sistemas e Computadores em Tecnologia e Ciência (INESC TEC),  
Porto, Portugal  
{1211139, crc, vms, ats, fjc}@isep.ipp.pt

**Abstract.** Gislotica is a company that designs and develops industrial machinery using standard components to ensure its solutions remain flexible. It has been working to integrate machine learning (ML) solutions into its equipment to perform quality control tasks that rely on highly reliable pattern recognition.

This work aims to develop the Masterchief graphical user interface (GUI) in C++ programming language using the QT framework extension, complemented by some additional Python scripts. The GUI allows users to manually tag a set of pattern sub-images extracted from a training dataset. These sub-images enable the initiation or continuation of training for the convolutional neural network (CNN) structure implemented by Darknet, followed by the testing phase. The CNN is based on a classification model dedicated to object detection known as YOLOv4 (You Only Look Once). This model is highly flexible and will be used to detect anomalies in the cordage manufacturing industry. Currently, cordage patterns are not yet available, so the system's functionality was evaluated using cat face images as one example among others. Preliminary test results from the single-class classifier are presented and qualitatively analyzed, considering both the YOLOv4 classifier trained from scratch and with pre-trained weight.

**Keywords:** C++, GUI, Image Tagging, Industrial Automation, Machine Learning, Object Detection, QT, YOLOv4.

## 1 Introduction

YOLOv4 is an advanced Machine Learning (ML) model, based on deep learning, designed for reliable real-time object detection. It employs Convolutional Neural Networks (CNNs) to extract features and predict bounding boxes on input images. It is widely used in computer vision systems within the automation industry, such as in: quality control – for example, detecting imperfections in manufactured products, such as damaged surfaces and missing parts [1]; autonomous vehicles and robotics – for example, identifying pedestrians, traffic signs, and obstacles to prevent collisions [2];

industrial worker safety – for example, ensuring the proper use of personal protective equipment and the detection of unsafe operational behaviors [3]; precision agriculture – for example, in the detection and identification of plant leaf diseases, using the Plant Village image dataset [4].

These solutions encourage innovative companies like Gislotica, specializing in industrial machinery construction, to incorporate ML into their equipment for quality control. Thus, this work arises from the need to detect anomalies in the cordage industry, where the braiding of fibres must follow a uniform pattern – without gaps, excessive twisting or misalignment, frayed or loose fibres, interruptions in the braiding, and must maintain a consistent diameter throughout its length. To this end, the company recognises that the open-source neural network framework Darknet [5], written in C and CUDA, which integrates the YOLOv4 classification model [6], enables real-time object detection with high reliability, thereby meeting its needs.

In this scenario, this work presents the implementation of a dedicated GUI developed in C++, which incorporates the Darknet framework and provides the following functionalities to the user: the collection (cropping and tagging) of relevant sub-images for model training based on the desired targets; management of files and directories within the Darknet framework; and the initiation of the training and testing phases of the YOLOv4 classification model, without requiring direct command-line interaction.

Section 2 introduces the tools and technologies used, focusing on the Darknet framework and YOLOv4 classification model. Section 3 outlines the GUI architecture, covering the main features of the front-end and back-end.

Gislotica does not yet have the cordage patterns for training and testing the system that enables quality control during the manufacturing process. Given the system's flexibility and to illustrate its functionality, images of cat faces are used as an experimental example. The company favored using the `yolov4.conv.137` file with pre-trained weights. Additionally, a complete training of the convolutional neural network was conducted to evaluate whether qualitatively distinct results could be observed in some test images. Accordingly, Section 4 presents the classification results using the two corresponding experiments, while Section 5 concludes with insights and perspectives for future work.

## 2 Tools and Technologies

The GUI is developed in C++ with a Qt extension framework to ensure high performance and modularity, while Python scripts are integrated to streamline ML workflows [7-9]. Development is carried out using the Visual Studio 2022 Integrated Development Environment (IDE).

The GUI uses a modified Darknet folder with a specific structure, including key components such as the `obj`, `labels`, `cfg`, and `backup` folders, along with the `obj.names`, `obj.data`, `train.txt`, `test.txt`, `darknet.exe`, `yolov4-obj.cfg`, `yolov4.conv.136`, `yolov4.weights` files, and the `process.py` and `resultsfilter.py` scripts. It is a modified Darknet according to the project need, but retains essential features from the original GitHub repository [5].

The `obj` folder contains the images used for both training and testing, while the `labels` folder holds the corresponding text files for these images. Each file, named after its corresponding image with a `.txt` extension, contains annotations specifying the bounding boxes and class labels. These annotations, formatted with the class number and the bounding box's width, height, and center coordinates, provide the ground truth data necessary for training the YOLOv4 model to detect objects in unseen images.

The `obj.names` file lists the class names, one per line, defining the categories the model is trained to recognize. The `obj.data` file specifies key information, including the number of classes, paths to the training and validation image lists, and the path to the `obj.names` file that links the class names to the dataset.

The `train.txt` file lists the paths to all training images, while the `test.txt` file contains the paths to test images. Both are crucial for Darknet to properly load datasets during the training and evaluation phases.

The `cfg` folder contains configuration files defining neural network architecture and training parameters, such as layer types (e.g., convolutional, pooling), activation functions, learning rate, batch size, and iteration count. The `yolov4-obj.cfg` file, for example, specifies the YOLOv4 model's structure and training settings.

The `backup` folder in Darknet stores periodic snapshots of the model's weights during training. This ensures that if training is interrupted, progress is preserved and the model can resume from the last saved checkpoint.

The `yolov4.conv.137` file contains pre-trained weights for the first 137 layers of the YOLOv4 neural network. These weights result from pre-training the Backbone component on a large image dataset, which learns to identify general features such as edges, textures, shapes, and basic patterns all of which are essential for object detection tasks. Additionally, with randomly initialized weights, the model includes a Neck component with Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet) modules to handle object size variations and improve detection accuracy. Lastly, these random weights are then fine-tuned in the Head component, which generates predictions such as bounding boxes and class probabilities.

The `yolov4.weights` file contains the global combination of weights from the `yolov4.conv.137` file and the adjustments made to the Neck and Head components, based on the specific dataset being used for training. It is possible to either retain the weights from `yolov4.conv.137` or ignore them and perform training from scratch. In this latter case, training will take longer, and pre-trained visual features are lost, which can be limiting for smaller training datasets and increase the risk of overfitting. Despite these considerations, both approaches are tested and qualitatively compared in this work. For instance, when training with the detector using `yolov4.conv.137` (the training will proceed from scratch if `yolov4.conv.137` is not provided as a parameter), this process is carried out via the command line with the following typical configuration:

```
./darknet detector train obj/obj.data cfg/yolov4-obj.cfg  
yolov4.conv.137
```

For testing, the following parameterization is typically used in the command line:

```
./darknet detector test obj/obj.data cfg/yolov4-obj.cfg
backup/yolov4.weights data/test_image.jpg
```

Finally, the `process.py` script splits image datasets into training and testing sets, retaining only valid image-annotation pairs. The `resultsfilter.py` script maps YOLOv4 detection outputs (class labels and bounding box coordinates) to predefined classes. Both ensure data integrity, handling missing files and I/O errors, and output `train.txt`, `test.txt`, and formatted detection results with class indices and coordinates, streamlining object detection workflows.

### 3 Software development

The GUI is based on a modular architecture that ensures seamless communication between the front-end and back-end, as shown in Fig. 1.

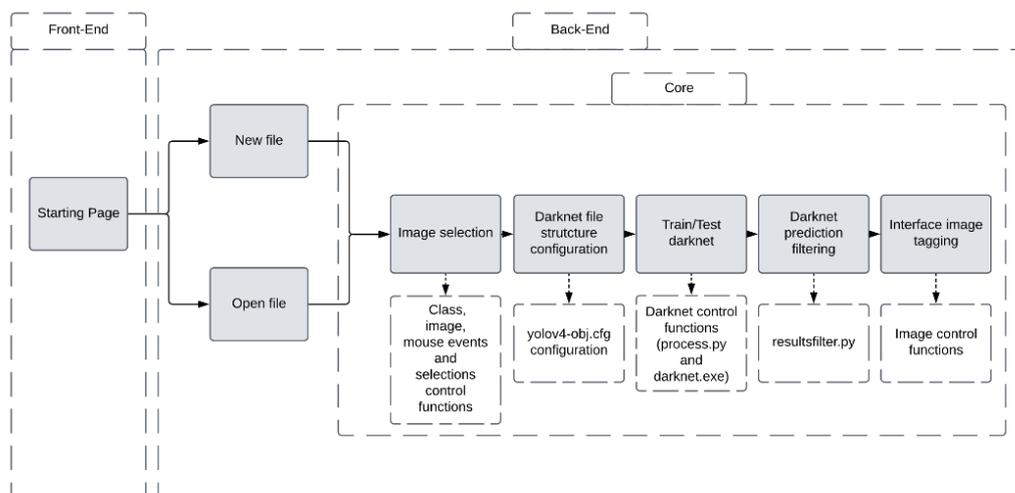


Fig. 1. GUI general architecture.

#### 3.1 Front-end Overview

The GUI offers an intuitive interface for dataset tagging and management, with features like dynamic class buttons, undo options, and real-time feedback (Fig. 2). Users begin tagging by creating a class through the `Add` button and then marking images with a two-click process – one on each of the two diagonally opposite vertices defining a square and continue recording the sample – once 10% of the images are tagged. Training can be initiated via the `Train` button, typically taking 1.5 to 3 hours for a small set of images. After training, clicking `Test` updates the weights. The `D` key triggers ML detec-

tion, and any incorrect detections can be corrected using the Delete Selection button. Additional buttons like Edit and Undo provide further control for adjusting tags and correcting mistakes.

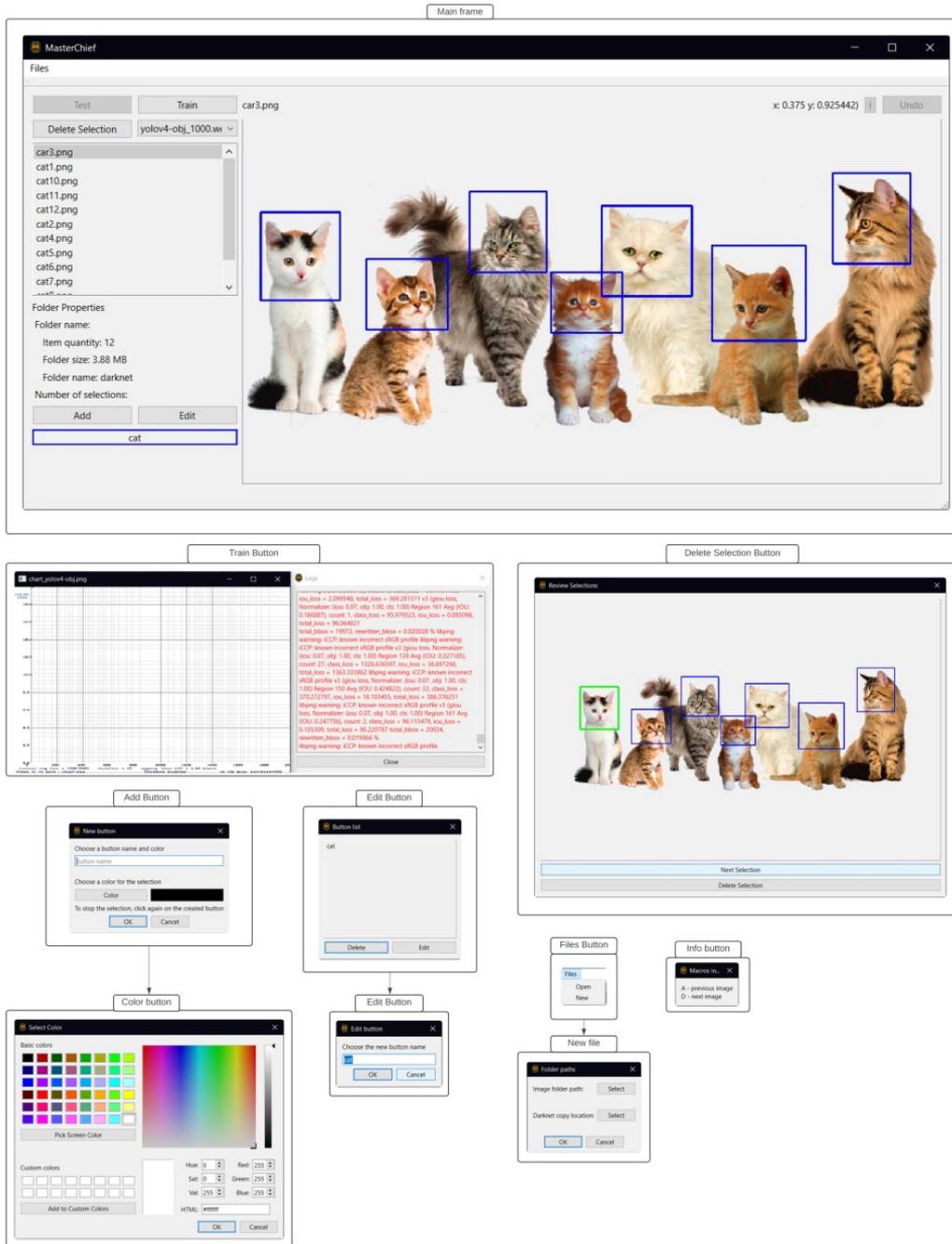


Fig. 2. GUI interface and associated features.

### 3.2 Back-end Overview

The GUI was designed with a structured set of programming modules, each containing functions dedicated to implementing the following functionalities: Opening images, Image selection, Train/Test Darknet, Darknet prediction filtering, and Interface image tagging. The only exception is the Darknet file structure configuration, where the file is manually edited. According to Fig. 1 and Fig. 2, some details are explained.

With the Opening images functions, the GUI allows users to create a new project by specifying an image folder and a destination for the `Darknet` folder, which is copied from the `MasterChief` directory. Images are transferred to the `Darknet` data directory, and an image list vector is populated. To open an existing project, users select `File >> Open File`, and the interface updates based on the folder's contents, even opening without images if none are present.

In the Image selection module, Class control functions manage image tagging with `Add` and `Edit` buttons for creating and modifying classes. Dynamically generated class buttons represent selection categories in the graphic view. The selection process toggles between active and inactive states, enabling precise tagging and smooth transitions. When loading a project, classes are restored by retrieving files and class names. Buttons are recreated with unique colors, the layout is adjusted, and click events are connected for class-based selection. The image control functions support image loading, list management, navigation, and reallocation. The `GV` class, a subclass of `QGraphicsView`, is implemented for managing image display and mouse events, normalizing coordinates to a `[0 1]` range for precise tagging. The `QListWidget` helps navigate large datasets, with functions to validate selections and update displays. The selection control functions, enable users to interact with images, select areas for tagging, and receive visual feedback, ensuring accurate and efficient control. It supports the following functionalities: selection creation; load coordinates; adding coordinates to a file; deleting coordinates from a file; undo selection button; clear selections; clear everything; image coordinates file creation; and redo selections button.

Next, within the Darknet file structure configuration, and following the optimized recommendations outlined in [5], the file `yolov4-obj.cfg` is set up with the initial parameters of the Darknet structure, using an implemented and dedicated function called `yolov4Config`. It reads the number of classes from `obj.data` and adjusts key parameters: `subdivisions` for memory efficiency, `max_batches` as `2000*num_classes`, learning rate steps at 80% and 90% of `max_batches`, and convolutional filters as `(num_classes+5)*3`. It also updates the `classes` parameter in each YOLO layer. The modified configuration is saved, preparing YOLOv4 for training on the dataset.

The next step involves clicking the `Train` button, which triggers the training process by executing an embedded `process.py` script. This script, previously developed by the company, utilizes information from the `obj.name` and `obj.data` files to write image paths into the `train.txt` and `test.txt` files, based on a predefined split percentage. By default, the script allocates 10% of the images to the test set, and the remaining to the train set. Thus, the training phase is performed using the executable file `darknet.exe` along with external parameters, as previously mentioned, executed through a

system call. At this stage, a weight file is created every 1000 weight updates or batches (e.g., `yolov4_1000.weights`, `yolov4_2000.weights`, etc.) in the Darknet backup folder. Additionally, a dedicated filtering function is implemented to ensure the testing phase always uses the most recent one by default, while still allowing the use of other one if needed. Subsequently, as soon as the `Test` button is pressed, the weight file is included in the test command, which executes the `darknet.exe` file via a system call. Next, after the ML model completes its predictions, it generates a grayscale JPG file named `predictions.jpg` in the main folder, containing the rectangle surrounding the object's region and indicating the probability value with which the class was identified, which is suggestively renamed. The `-ext_output` option in the test command enables extracting detailed output in the `results.txt` file, including the coordinates of detected objects, during inference. However, the output generated includes more information than needed. The use of the embedded `resultfilter.py` script isolates relevant data by extracting only the essential information, which is saved in the `coordinates.txt` file as output for further processing (Fig. 3). Finally, using the information from the `coordinates.txt` file, the GUI draws a rectangle around the object's region in the original image, indicating the predicted area (Fig. 6 and Fig. 7).

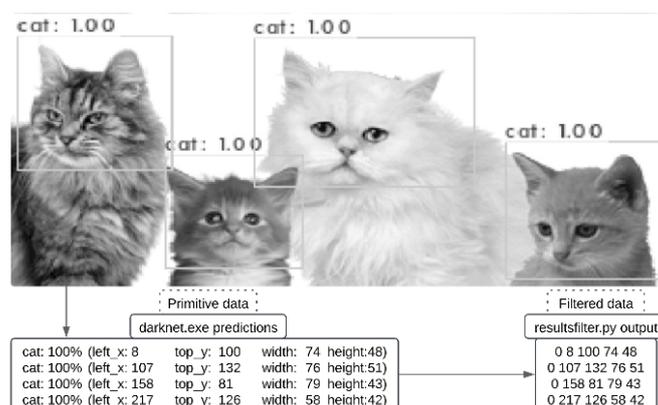


Fig. 3. ML coordinates predictions graphic representation, from Darknet.exe test.

## 4 Results

The implemented GUI and its main features are shown in Fig. 2. Two experimental results were obtained using the same test images of cat faces. For this purpose, the training phase uses a set of 19 images, from which 68 sub-images were manually tagged using the GUI (Fig. 4). Using this set of 68 sub-images, training was performed in two distinct ways on a computer Acer Nitro 4, with an AMD Ryzen 5 CPU and 8 GB of RAM: the first (henceforth designated as Experiment 1), using pre-trained weights for the first 137 layers of the network, available in the `yolov4.conv.137` file, with a duration of approximately 3 hours; the second (henceforth designated as Experiment 2), performed from scratch with a duration of approximately 22 hours.



Fig. 4. The 19 dataset images, from which 68 sub-images were manually tagged for training using the GUI.

The training configurations in the `yolov4-obj.cfg` file followed the default settings referenced in [5], using a single target class, with the following recommended parameters: `subdivisions = 64`, `max_batches = 2000`, and `filters = 18`.

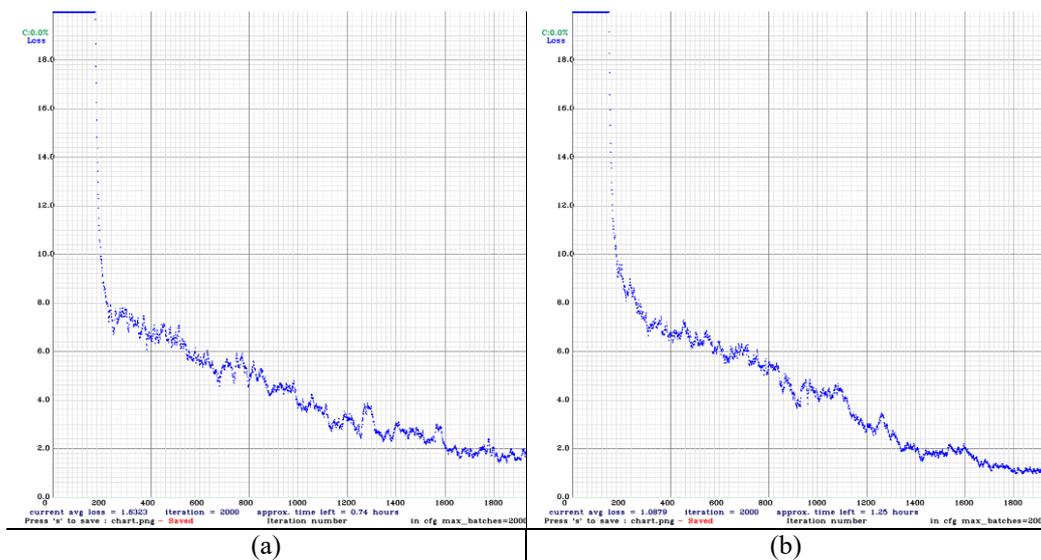
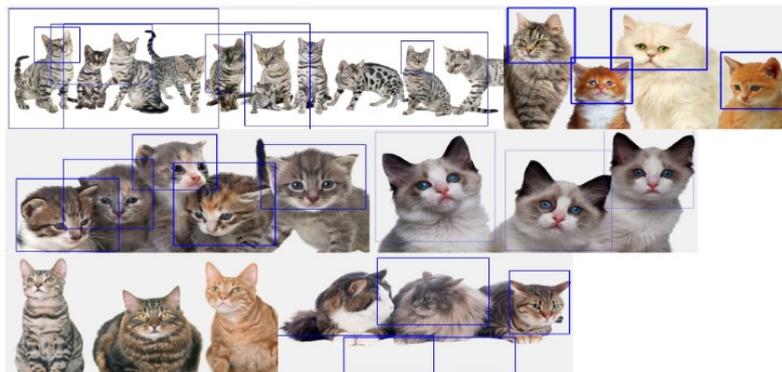


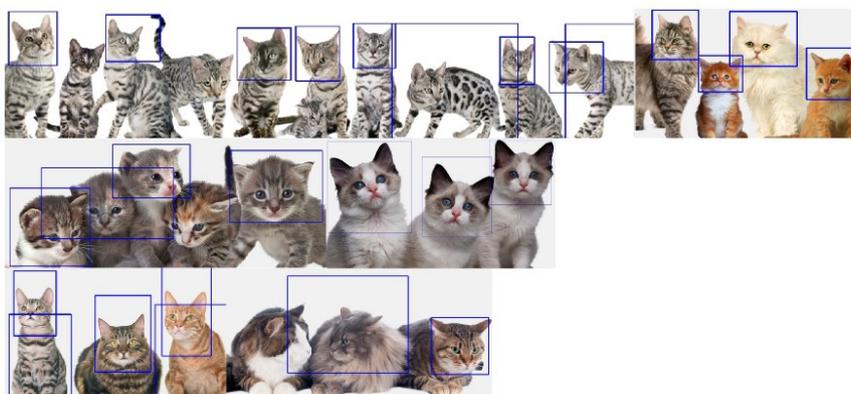
Fig. 5. Training loss chart results for Experiment 1 in (a) and Experiment 2 in (b).

Analyzing the training loss charts, both experiments show that the model learns well with decreasing loss (Fig. 5). However, oscillations suggest some generalization difficulties, more so in Experiment 1, which stabilize after 2000 iterations and decrease towards the end in Experiment 2. The lower average loss in Experiment 2 suggests more accurate predictions than in Experiment 1.



**Fig. 6.** Test results obtained from the training model used in Experiment 1.

Performing a qualitative analysis on a small test set images, reveals that the identification of cat faces in Experiment 2 is a bit more effective than Experiment 1 (Fig. 6 and Fig. 7), as appeared predictable according to the training phase results (Fig. 5).



**Fig. 7.** Test results obtained from the training model used in Experiment 2.

## 5 Conclusions and future work perspectives

The project successfully achieved its goal of developing a robust and scalable system for ML-based object detection. By integrating an intuitive GUI with a powerful back-end framework, the system streamlines dataset annotation, neural network training, and continuous model improvement. It proved effective in optimizing object detection

workflows, with seamless integration between the front-end and back-end, ensuring annotated data is efficiently incorporated into the training pipeline. This facilitates the iterative refinement of the YOLOv4 model. Performance benchmarks confirmed the system's ability to handle large datasets while maintaining responsiveness during both inference and training.

User testing underscored its ease of use and modular design, making it ideal for industrial and research applications. The results demonstrate the system's scalability and potential for future enhancements, such as multi-class classification and integration with more advanced ML models.

Although the main objective of this project was not to evaluate the performance of a well-established ML technology, it presents a starting point to demonstrate the capabilities of the developed GUI. Two experiments were conducted, one using the pre-trained YOLOv4 model and the other with the same model trained from scratch, referred to as Experiment 1 and Experiment 2, respectively. While the pre-trained model reduces a typical desired training time, the initial experimental results suggest that training from scratch improves the model's ability to detect target patterns a bit more effectively.

Future work will focus on experimenting with the pattern detection system in the cordage industry and potentially optimizing the system's performance by integrating existing advanced YOLO models. It would then be useful to use the validation loss during the training phase to detect overfitting, select an appropriate learning rate, and apply an early stopping criterion whenever the validation loss stabilizes, even if the training loss continues to decrease.

## References

1. Bochkovskiy, A., Wang, C., Liao, H.: YOLOv4: Optimal Speed and Accuracy of Object Detection. In: arXiv preprint arXiv:2004.10934 (2020)
2. Wang, R., Liu, Y., Zhang, X., Li, Z., Li, C.: A Real-Time Object Detector for Autonomous Vehicles Based on YOLOv4. In: Computational Intelligence and Neuroscience (2021)
3. Önal, O., Dandil, E.: Object Detection for Safe Working Environments using YOLOv4 Deep Learning Model. In: European Journal of Science and Technology (2021)
4. Aldakheel, E. A., Zakariah, M., Alabdallal, A. H.: Detection and identification of plant leaf diseases using YOLOv4. In: Frontiers in Plant Science (2024)
5. Darknet, <https://github.com/AlexeyAB/darknet>. Accessed August 27, 2024.
6. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).
7. Stroustrup, B.: The C++ Programming Language, 4th Edition. Addison-Wesley Professional (2013).
8. Szafranek, P.: Advanced Qt Programming: Creating Great Software with C++ and Qt 5. Packt Publishing (2021).
9. Lutz, M.: Learning Python, 5th Edition. O'Reilly Media (2013).

# Comparação de Arquiteturas Monolíticas e de Microserviços em Sistemas de Informação

Ricardo Moura<sup>\*[0009-0003-9594-0221]</sup>, Vítor Rocha e Vítor Ventura

ISLA Gaia | Instituto Politécnico de Gestão e Tecnologia

**Resumo.** Este estudo compara as arquiteturas de sistemas de informação baseadas em monolitos e de microserviços, analisando os seus pontos fortes, limitações e casos de aplicação. O objetivo principal é identificar qual abordagem se adequa melhor a diferentes contextos organizacionais.

A metodologia incluiu uma revisão bibliográfica de artigos académicos e estudos de caso, complementada por uma análise comparativa. Foram avaliados aspetos como modularidade, independência de componentes e impacto em equipas de desenvolvimento.

Os resultados mostram que os microserviços são mais vantajosos em sistemas que exigem escalabilidade e flexibilidade, permitindo atualizações independentes e integração contínua. Contudo, a sua implementação é mais complexa e requer maior investimento inicial. Por outro lado, a arquitetura monolítica apresenta simplicidade na configuração inicial e melhor adequação a projetos de pequena escala, mas enfrenta limitações na adaptação a ambientes dinâmicos.

Conclui-se que a escolha entre microserviços e arquitetura monolítica depende dos requisitos específicos do sistema, sendo crucial considerar o equilíbrio entre necessidades atuais e futuras. Este trabalho contribui para decisões informadas na seleção da arquitetura mais adequada a projetos de tecnologia da informação.

**Palavras-Chave:** Arquitetura Monolítica. Arquitetura de Microserviços. Arquitetura de Sistemas.

## 1 Introdução

O desenvolvimento de sistemas de informação tem evoluído para atender às exigências do mercado tecnológico. Neste contexto, a escolha da arquitetura adequada torna-se um fator crucial para o sucesso de um projeto. Este estudo tem como objetivo realizar uma análise comparativa entre duas abordagens arquiteturais utilizadas no desenvolvimento de sistemas: a arquitetura monolítica e a arquitetura de microserviços. A análise aborda aspectos fundamentais como escalabilidade, manutenção, implementação e custos associados, pretendendo fornecer uma base para a tomada de decisão na escolha da arquitetura mais apropriada para diferentes contextos e necessidades específicas de cada projeto. Espera-se demonstrar que a escolha entre estas arquiteturas não é universal, mas depende de diversos fatores que são abordados neste artigo curto.

## 2 Arquiteturas Mónolítica versus Microserviços

### 2.1 Arquitetura Monolítica

A arquitetura monolítica geralmente depende de uma única tecnologia de desenvolvimento, o que limita a disponibilidade de ferramentas adequadas para cada tarefa que o sistema deve executar. De acordo com Nielsen [10], qualquer modificação numa parte desse sistema exige a construção e implementação de uma nova versão de todo o sistema, uma vez que este existe como um único executável. A maioria dos componentes ou camadas relacionadas com apresentação, processamento e armazenamento estão integrados numa única unidade de software que opera num único servidor.

Esta arquitetura apresenta tanto benefícios como desvantagens, que podem aumentar a confiança na sua estabilidade ou, pelo contrário, resultar em problemas tecnológicos. É essencial entender os riscos envolvidos, pois estes podem levar a desafios de curto ou de longo prazo [6]. Entre as principais vantagens, os sistemas monolíticos tendem a ser estáveis e fáceis de gerir de forma centralizada. Grandes corporações, como a IBM, Sun Microsystems e BMC, utilizam amplamente esta arquitetura, beneficiando do elevado nível de especialização nos seus produtos.

No entanto, entre as suas desvantagens estão a falta de flexibilidade e a dificuldade em adaptar-se a novos requisitos. Para melhorar a capacidade de processamento, geralmente é necessário atualizar para um servidor maior. Além disso, sistemas monolíticos podem ser implementados com tecnologias proprietárias, o que pode levar a uma dependência significativa do cliente em relação ao fornecedor. Os custos associados à aquisição, renovação e suporte são também substanciais.

Todos os sistemas informáticos precisam de evoluir em resposta a tendências, como a substituição de cartões perfurados por HDD. A complexidade dos sistemas contemporâneos exige avanços no desenvolvimento de software e no desempenho. Consequentemente, isto destaca as falhas inerentes na arquitetura monolítica [12]. Com o tempo, essas falhas levaram ao surgimento de arquiteturas modernas, como os microserviços.

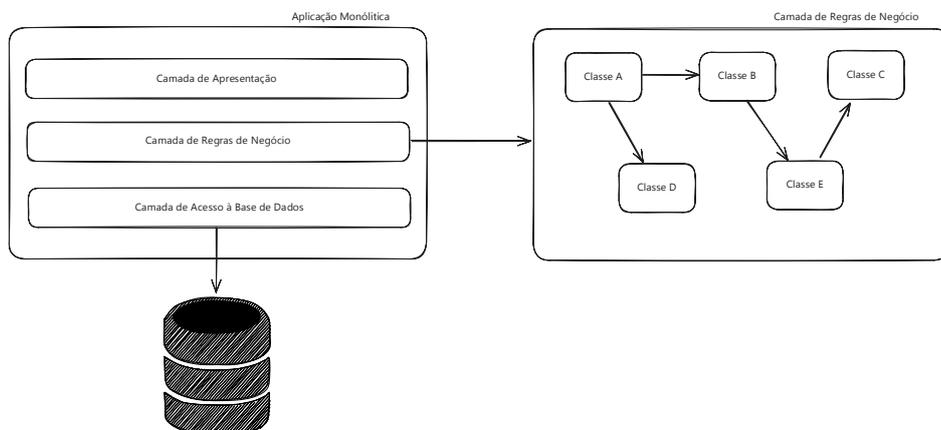
Em certas instâncias, algumas aplicações podem funcionar de forma mais eficaz com esta arquitetura, como firmware, keyloggers e vírus.

A Tabela 1 apresenta um resumo das principais vantagens e desvantagens associadas à arquitetura monolítica. Este formato permite uma análise clara e comparativa, destacando os aspetos que favorecem a sua utilização, bem como os desafios inerentes à sua implementação e manutenção.

**Tabela 1.** Prós e contras da Arquitetura Monolítica [2]

<b>Prós</b>	<b>Contras</b>
<b>Menos Preocupações:</b>  Registo, manipulação, cache e monitorização de desempenho afetam toda a aplicação. Isto simplifica a gestão, uma vez que a funcionalidade é centralizada.	<b>Compreensão Difícil:</b>  À medida que uma aplicação monolítica cresce em tamanho, torna-se mais difícil compreendê-la. Além disso, a gestão de uma estrutura de código complexa numa única aplicação é desafiadora.
<b>Facilidade na depuração e testes:</b>  Depurar e testar aplicações monolíticas é significativamente mais fácil, já que o software é a única entidade indivisível.	<b>Dificuldade em realizar alterações:</b>  Fazer alterações numa aplicação monolítica grande e complexa, é difícil. Qualquer mudança afeta todo o sistema e requer coordenação detalhada, prolongando o processo de desenvolvimento.
<b>Simplicidade na implementação:</b>  Não é necessário lidar com várias implementações; basta um único ficheiro ou diretório.	<b>Escalabilidade limitada:</b>  Apenas a aplicação inteira pode ser escalada.
<b>Facilidade de desenvolvimento:</b>  Uma equipa técnica com as competências adequadas pode projetar facilmente uma aplicação monolítica, sendo este um método comum de desenvolvimento.	<b>Barreiras para novas tecnologias:</b>  Implementar novas tecnologias numa aplicação monolítica é extremamente difícil, já que toda a aplicação precisa de ser reconstruída.

A Fig. 1 ilustra a organização típica de uma arquitetura monolítica, destacando a integração dos diferentes componentes do sistema numa única unidade de software. Esta representação visual ajuda a compreender as principais características.



*Fig. 1.* Arquitetura Monolítica (fonte própria)

## 2.2 Arquitetura de Microserviços:

De acordo com Martin Fowler [3], “o estilo arquitetural de microserviços envolve desenvolver uma aplicação única como uma coleção de pequenos serviços que operam em seus próprios processos e comunicam-se por meio de mecanismos leves, geralmente utilizando uma API de recurso HTTP. Esses serviços focam nas capacidades de negócio e podem ser implementados de forma independente através de sistemas de implementações totalmente automatizadas.”

Esta arquitetura facilita a criação de aplicações grandes, complexas e escaláveis, compostas por processos pequenos e independentes, que são fracamente acoplados e comunicam-se através de APIs [8].

Os microserviços funcionam como uma aplicação composta por pequenos serviços individuais que executam em processos separados e interagem entre si. A autonomia de cada serviço aumenta a tolerância a falhas e melhora a disponibilidade geral.

Representa uma cultura moderna de automação com processos descentralizados que permitem implementações independentes, onde o design geral é centrado no domínio do negócio [4, 5, 9]. Além disso, os microserviços podem encapsular cenários de negócios complexos e abrangentes dos clientes, eliminando a necessidade de atualizar o sistema inteiro, exceto o microserviço ou o container específico relevante.

A Tabela 2 apresenta uma visão geral das vantagens e desvantagens da arquitetura de microserviços. Esta análise comparativa destaca os benefícios associados à escalabilidade e independência dos componentes, bem como os desafios relacionados com a sua complexidade e gestão.

**Tabela 2.** Prós e contras da Arquitetura de Microserviços [2]

<b>Prós</b>	<b>Contras</b>
<b>Componentes Independentes:</b>  Todos os serviços podem ser instalados e atualizados de forma independente, permitindo maior flexibilidade. Além disso, um problema num único microserviço afeta apenas esse serviço e não toda a aplicação. Adicionar novas funcionalidades a uma aplicação de microserviços é significativamente mais fácil do que em programas monolíticos.	<b>Maior Complexidade:</b>  A arquitetura de microserviços, quando baseada na cloud, exige a configuração e ligação entre módulos e bases de dados. Cada serviço independente precisa ser implementado separadamente.
<b>Maior Facilidade de Compreensão:</b>  As aplicações baseadas em microserviços são mais fáceis de entender e gerir, pois estão divididas em componentes menores e mais simples.	<b>Distribuição do Sistema:</b>  Uma arquitetura de microserviços é um sistema complexo com muitos módulos e bases de dados, e todas as conexões devem ser cuidadosamente geridas.
<b>Melhor Escalabilidade:</b>  A arquitetura de microserviços permite que cada componente escale independentemente, tornando o processo mais barato e menos demorado em comparação com os monólitos, onde toda a aplicação precisa de ser escalada, mesmo que não seja necessário.	<b>Preocupações Transversais:</b>  É necessário lidar com diversas preocupações, como configuração externa, registos, métricas, verificações de saúde e outras funcionalidades.
<b>Maior Agilidade:</b>  Qualquer falha numa aplicação de microserviços afeta apenas um serviço e não a solução inteira. Como resultado, as modificações e experiências são realizadas com menos riscos e erros.	<b>Testes:</b>  Testar uma aplicação baseada em microserviços é significativamente mais difícil devido ao grande número de componentes implementáveis de forma independente.

---

A Fig. 2 ilustra a organização de um sistema de microserviços, destacando a interação entre o frontend, os microserviços independentes e o message broker.

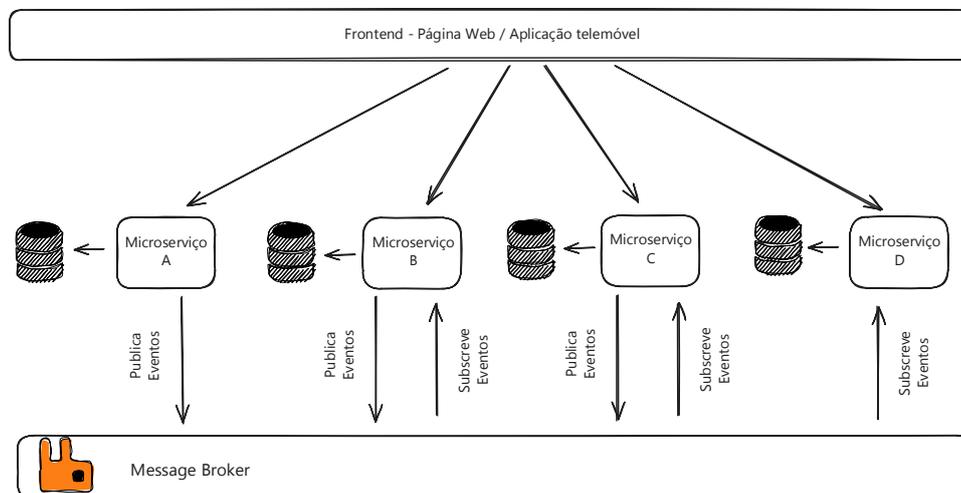


Fig. 2. Arquitetura de Microserviços (fonte própria)

### 3 Evolução Tecnológica

A evolução tecnológica desempenha um papel central na escolha entre arquiteturas monolíticas e de microserviços, uma vez que influencia diretamente as capacidades de desenvolvimento, implementação e gestão de sistemas. Com o avanço das tecnologias de virtualização e cloud computing, a arquitetura de microserviços tem-se tornado cada vez mais viável porque estas tecnologias suportam a escalabilidade e a gestão independente de serviços.

A popularização de containers, como o Docker, e plataformas de orquestração, como o Kubernetes, possibilita uma melhor organização e automação de microserviços, permitindo que cada componente funcione de forma isolada, garantindo maior resiliência e eficiência na utilização de recursos. Estes avanços tornam os microserviços mais atrativos para sistemas complexos que requerem alta adaptabilidade e integração contínua [11]. Por outro lado, os avanços em frameworks e ferramentas de desenvolvimento monolítico continuam a permitir a criação de sistemas robustos e centralizados. As melhorias na capacidade de hardware, como servidores mais rápidos e eficientes, atenuaram algumas das limitações de escalabilidade associadas a arquiteturas monolíticas, mantendo-as relevantes em contextos onde simplicidade e custos reduzidos são prioritários.

Adicionalmente, a evolução tecnológica introduziu novas preocupações, como a segurança e a gestão de dados distribuídos, que impactam diretamente a adoção de microserviços. Enquanto soluções modernas, como redes de serviço (service mesh), melhoram a segurança e a comunicação entre serviços, elas também aumentam a complexidade, o que pode favorecer arquiteturas monolíticas em cenários onde a simplicidade operacional é crucial [7].

Deste modo, a escolha entre arquiteturas monolíticas e de microserviços é altamente influenciada pelo progresso tecnológico, que redefine as vantagens e limitações de cada abordagem. À medida que novas tecnologias emergem, os microserviços continuam a expandir as suas capacidades em ambientes de alta exigência, enquanto os sistemas monolíticos permanecem uma solução viável para aplicações menos complexas e de menor escala.

#### **4 Desafios e Soluções na Migração de uma Arquitetura Monolítica para Microserviços**

Durante a migração de uma aplicação monolítica para uma arquitetura de microserviços, foi identificada a necessidade de separar os módulos que não seriam utilizados por todos os clientes. Manter duas aplicações em paralelo revelou-se impraticável, o que levou à escolha da abordagem de microserviços. Embora essa abordagem ofereça diversas vantagens, também apresenta desafios e trade-offs que devem ser considerados.

A separação dos domínios revelou dependências entre componentes, como a classe A depender da classe B, o que não pode ser tratado de forma simples quando os serviços são distribuídos. Um problema significativo surge quando um microserviço A depende de outro, B, e a comunicação entre eles é feita através de pedidos HTTP síncronos. Caso o microserviço B esteja indisponível ou sobrecarregado, o pedido do microserviço A expira, resultando em erro, o que pode comprometer a experiência do utilizador. Para mitigar esse acoplamento, foi implementado um message broker, permitindo a comunicação assíncrona entre os serviços. Embora essa abordagem tenha resolvido o problema do acoplamento, ela introduziu novos desafios, como o aumento da complexidade da aplicação e a introdução de um ponto de falha adicional.

Em resposta ao aumento da demanda, a solução de escalabilidade vertical foi aplicada. Ao escalar o microserviço A, que processa uma parte da aplicação, a capacidade de responder a pedidos foi aumentada, permitindo que o microserviço lidasse com o dobro de requisições. No entanto, a implementação de microserviços continua a exigir o uso de técnicas avançadas, como Command and Query Responsibility Segregation (CQRS), para separar as operações de leitura (consultas) e escrita (comandos), otimizando a performance de cada uma delas e garantindo maior flexibilidade no processamento de dados. Essa técnica traz benefícios, como a possibilidade de utilizar diferentes bases de dados para leitura e escrita, mas também exige cuidados com a sincronização dos modelos de dados e a gestão de múltiplos pontos de falha.

Ainda assim, a implementação de CQRS, combinada com a gestão de disponibilidade e resiliência, torna o processo de migração uma solução robusta, mas complexa [1].

## 5 Conclusão

A escolha entre a arquitetura monolítica e a de microserviços depende das necessidades específicas de cada projeto, considerando fatores como escalabilidade, flexibilidade, custo e complexidade.

A arquitetura monolítica, com a sua simplicidade e centralização, continua a ser uma opção eficiente para sistemas menores ou com requisitos menos dinâmicos, oferecendo fácil gestão e estabilidade. Contudo, apresenta limitações significativas em projetos que exigem frequentes atualizações e alta escalabilidade.

Por outro lado, a arquitetura de microserviços destaca-se pela modularidade, flexibilidade e adaptabilidade, sendo mais adequada para aplicações complexas e que requerem rápida evolução e escalabilidade. Ainda assim, a sua implementação implica desafios técnicos e custos iniciais mais elevados, além de maior complexidade na gestão e no teste dos sistemas.

Portanto, a decisão pela abordagem mais adequada deve considerar o contexto do sistema, os recursos disponíveis e os objetivos do projeto. Enquanto os microserviços mostram-se essenciais para organizações que precisam de escalabilidade e adaptabilidade, a arquitetura monolítica pode ser preferível em soluções mais simples e com menos exigências.

## Referências

1. Debski, A., Szczepanik, B., Malawski, M., Spahr, S., & Muthig, D. (2017). In search for a scalable & reactive architecture of a cloud application: CQRS and event sourcing case study. *IEEE Software*, 1-8.
2. ElGheriani, N. (2022). Microservices vs. Monolithic Architectures. *Al-Mansour Journal*, 37(1), 37-44.
3. Fowler, M., & Lewis, J. (2014). Microservices: A definition of this new architectural term. Retrieved from <http://martinfowler.com/articles/microservices.html>
4. Fussell, M. (2020). Why a microservices approach to building applications? Retrieved from <https://docs.microsoft.com/es-es/azure/service-fabric/service-fabric-overview-microservices>
5. Henry, A., & Ridene, Y. (2020). Assessing your microservice migration. In *Microservices* (pp. 73-107). Springer: Berlin/Heidelberg, Germany.
6. Khazaei, H., Barna, C., Beigi-Mohammadi, N., & Litoiu, M. (2016). Efficiency analysis of provisioning microservices. In *Proceedings of the 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)* (pp. 261-268). Luxembourg. <https://doi.org/XXXXXX>
7. Kotenko, M., Moskalyk, D., Kovach, V., & Osadchyi, V. (2024). Navigating the challenges and best practices in securing microservices architecture.
8. Kratzke, N. (2015). About microservices, containers and their underestimated impact on network performance. *Journal of Network Performance*, 961(165-169).

9. Newman, S. (2015). Building microservices: Designing fine-grained systems. O'Reilly Media, Inc.
10. Nielsen, D. (2015). Investigate availability and maintainability within a micro-service architecture (Master's thesis). Aarhus University, Aarhus, Denmark.
11. Ramu, V. (2023). Performance impact of microservices architecture. Review of Contemporary Science and Academic Studies, 3(1), 1-6.
12. Sun, L., Li, Y., & Memon, R. A. (2017). An open IoT framework based on microservices architecture. China Communications, 14(154-162).

## **Análise da distribuição geográfica de painéis fotovoltaicos em Portugal continental**

Tomás Nogueira<sup>1</sup>, Beatriz Teixeira<sup>1</sup>, Manuel Barbosa<sup>1</sup>, Tatiana Teixeira<sup>1</sup>, Alberto Pereira<sup>1</sup>, Marílio Cardoso<sup>1</sup>, Pedro Guedes<sup>1</sup>, Telmo Matos<sup>1</sup>

<sup>1</sup> Instituto Superior de Engenharia do Porto

**Resumo.** A procura por fontes renováveis de energia tem ganho relevância global, em particular, a energia solar. No âmbito da unidade curricular laboratório de sistemas 1, inserida na licenciatura de engenharia de sistemas do instituto superior de engenharia do Porto, foi desenvolvido um projeto de investigação tentando compreender o motivo da dispersão geográfica de instalações fotovoltaicas em Portugal continental. Foi ainda proposta a construção de uma plataforma web que permitisse realizar análises sobre essas mesmas instalações. O desenvolvimento do estudo da distribuição dos painéis fotovoltaicos teve como base um *dataset* do portal da E-REDES denominado “total de unidades de produção para autoconsumo”. O melhor resultado obtido foi através da relação entre o número total de unidades de produção para autoconsumo com o número de alojamentos ocupados pelo proprietário, obtendo um coeficiente de determinação de 71%, isto é, é possível explicar 71% da distribuição do número total de unidades de produção para autoconsumo através do número de alojamentos ocupados pelo proprietário.

**Palavras-chave:** E-REDES, Energia renovável, Painéis fotovoltaicos.

### **1 Introdução**

Nos últimos anos a procura por fontes de energia renováveis tem ganho relevância global, em particular, a energia solar. A localização do território português torna-se o cenário ideal para explorar esta fonte de energia devido aos elevados níveis de radiação solar.

Este artigo é fundamentado no projeto proposto na unidade curricular (UC) de laboratório de sistemas 1 que teve por base uma necessidade de estudo levantada pela E-REDES, que foi parceira e esteve presente ao longo das quatro semanas em que esta UC decorreu.

A E-REDES é o principal operador da rede de distribuição de energia elétrica em Portugal Continental, ligando mais de seis milhões de clientes. Está presente em todo o território continental, sendo o operador das redes elétricas de alta, média e baixa tensão, através das 278 concessões municipais. Apesar dos incentivos monetários estatais à instalação de unidades de autoconsumo, a análise de dados geográficos revela uma dispersão da quantidade e tipologia de instalação de painéis fotovoltaicos pelo território nacional.

O objetivo do projeto foi o de analisar tendências da transição energética ao nível do território continental, compreendendo o motivo da dispersão geográfica de instalações fotovoltaicas no país em quantidade e tipologias das instalações.

Foi proposta a construção de uma plataforma web que permitisse realizar análises sobre as instalações elétricas fotovoltaicas em Portugal, cruzando várias fontes de dados, em particular o portal da E-REDES, extraindo informações relevantes que permitisse evidenciar e explicar a dispersão dessas instalações ao longo de todo o território nacional.

## 2 Estado da arte

A Goldenergy, empresa portuguesa que fornece serviços de eletricidade e gás natural, afirma que “A energia solar fotovoltaica é a energia elétrica obtida a partir da luz solar” [1]. Esta energia pode ser gerada através de uma central ou em unidades para autoconsumo. A energia fotovoltaica apresenta diversas vantagens tais como [1]:

- Ser uma forma de energia limpa e sustentável;
- Permitir o armazenamento da mesma através de baterias, possibilitando a utilização posterior da mesma em períodos de menor geração;
- Viabilizar a utilização de energia elétrica em áreas geograficamente mais isoladas.

O processo de produção de eletricidade por intermédio de células solares fotovoltaicas ocorre com base no efeito fotovoltaico, isto é, “Trata-se da conversão da energia proveniente das ondas eletromagnéticas do sol em energia elétrica ...” [2].

Em 2013, dentro de todas as energias renováveis, a energia fotovoltaica representava apenas 0,016%. Este valor tem vindo a aumentar ao longo dos anos, atingindo 0,08% em 2022 [3].

De acordo com um artigo publicado pela DECO Pro Teste, observa-se uma utilização reduzida da energia solar como fonte de abastecimento elétrico. Em Portugal estima-se que, até 2030, a capacidade instalada suba para 4,6 GWs [2].

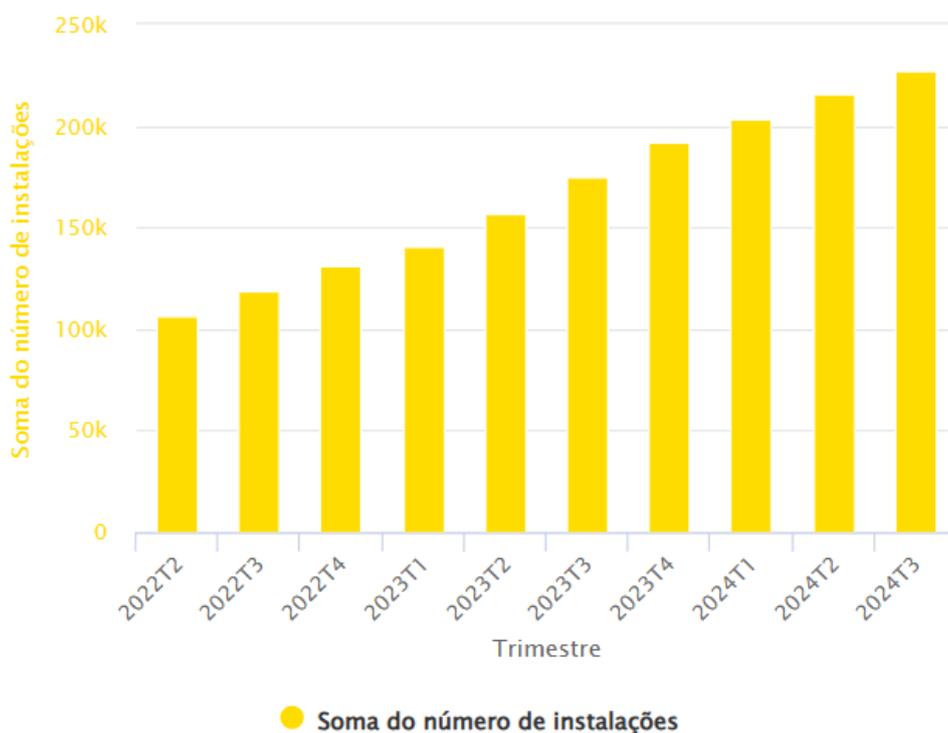
A escassez de tecnologia eficiente restringia o uso prático da energia do sol nas primeiras civilizações, porém embora a utilização deste como fonte de energia tenha as suas raízes nos primórdios da humanidade, a criação do primeiro painel solar apenas se deu no final do século XIX [4].

Em 1839, o físico francês Alexandre Edmond Becquerel constatou o efeito fotovoltaico caracterizado pela capacidade de determinados materiais de produzir energia elétrica quando expostos à radiação solar. O cientista americano Charles Fritts, quase 45 anos depois, através de um painel de metal com uma camada de selénico coberto a ouro, desenvolveu a primeira célula solar. Esta dispositivo transformava apenas 1% da luz em energia [5].

A eficiência deste dispositivo sextuplicou quando, em 1954, se passou a usar silício em substituição do selénico. Contudo, foi apenas em 1973 que foi construída a primeira casa abastecida a energia solar. Nos anos 2000, as células solares alcançaram o pico da eficiência até então, atingindo os 40% [5].

As instalações fotovoltaicas podem estar conectadas ou não à rede elétrica. Se estiverem ligadas à rede, podem ser instalações fotovoltaicas, isto é, toda a energia que produzem é destinada à rede elétrica, ou então podem ser um gerador com autoconsumo, onde parte da sua produção é consumida pelo produtor da mesma e o resto da energia é enviada para a rede. Estes tipos de instalações são compostos por três elementos principais: o conjunto de células fotovoltaicas, ou seja, o painel; o inversor, que tem o papel de converter a corrente gerada pelos painéis em corrente alternada e o transformador cuja função é elevar a tensão, uma vez que a tensão obtida através do inversor é de baixa tensão [6].

Em contrapartida, as grandes instalações que não estão conectadas à rede operam de forma autónoma, com o objetivo de satisfazer a procura de energia em áreas remotas. De forma a funcionarem, requerem dois componentes extras em relação às ligadas à rede: baterias para o armazenamento da energia e reguladores capazes de proteger as baterias contra sobrecargas [6].



**Fig. 1.** Gráfico com o número total de painéis fotovoltaicos existentes em Portugal [7]

Através do gráfico apresentado na **Fig. 1**, disponibilizado pela E-REDES, é visível que o número de painéis existentes no país tem vindo a aumentar de forma gradativa ao longo dos últimos anos, tendo ultrapassado as 150 000 instalações no terceiro trimestre de 2023.

### 3 Análise do problema e desenho da solução

O município com mais unidades de produção para autoconsumo (UPAC) é Guimarães, com 3 650 (dado relativo a dezembro de 2022) [8]. Atendendo a isso, o principal objetivo deste estudo foi identificar um motivo que justificasse o desequilíbrio das instalações para autoconsumo.

O desenvolvimento do estudo da distribuição dos painéis fotovoltaicos teve como base um *dataset* do portal da E-REDES denominado “total de unidades de produção para autoconsumo”. Apresenta o número total de unidades de produção para autoconsumo por concelho, por escalão de potência e potência total instalada, apresentando os dados de forma trimestral. Este *dataset* contém oito variáveis, sendo seis do tipo texto e as outras duas do tipo numérico.

De forma a complementar a análise, foi desenvolvido um website com quatro grandes objetivos:

- Permitir ao utilizador visualizar os vários *datasets* usados no estudo;
- Apresentar a distribuição das várias UPAC's no continente português;
- Disponibilizar o estudo realizado ao utilizador;
- Proporcionar ao utilizador a realização de análises com os diversos *datasets* disponíveis no website.

### 4 Desenvolvimento da solução

#### 4.1 Análise dos dados

##### Desenvolvimento da análise

Para realizar o estudo, além do dataset da E-REDES foram utilizados datasets do Pordata e do Fundo Ambiental. Os datasets usados do Pordata foram a população total, a densidade empresarial, o poder de compra, o consumo energético, o número de indivíduos em famílias, o PIB per capita e o número de alojamentos ocupados pelo proprietário. O dataset do Fundo Ambiental é o referente ao número de empresas que trabalham com painéis fotovoltaicos.

O dataset da população total foi escolhido tendo em mente que poderia existir um maior número de painéis fotovoltaicos em lugares onde a população fosse superior. Por outro lado, o dataset da densidade empresarial foi selecionado uma vez que, num concelho com muitas empresas, poder-se-ia explicar o elevado número de painéis tendo em conta que as empresas necessitam de energia para funcionar. O dataset do poder de compra foi baseado no pressuposto de que, pessoas com maior património poderiam ter uma propensão mais elevada para a compra de painéis solares. Conjeturou-se que a utilização do dataset do consumo energético, pessoas que consumissem mais energia poderiam ter um interesse mais elevado em colocar painéis fotovoltaicos para poupar nas despesas. O *dataset* do número de indivíduos em famílias foi escolhido uma vez que, pessoas que vivem sozinhas poderiam ter menos interesse em colocar painéis fotovoltaicos em casa, tendo em conta a baixa utilização de energia. O dataset número de

alojamentos ocupados pelo proprietário teve como suspeita de que, seria mais provável uma pessoa fazer o investimento em painéis fotovoltaicos se fosse dono da casa onde mora. O *dataset* do PIB per capita foi utilizado uma vez que um local com elevado poder de compra poderia optar por investir em painéis fotovoltaicos. Por fim, o *dataset* número de empresas fotovoltaicas foi considerado dado que, um local com muitas empresas a fazer a colocação deste tipo de equipamentos poderia levar as pessoas a fazer o investimento através da venda deste produto porta a porta.

A análise dos *datasets* escolhidos foi realizada em Python. De forma resumida, a análise dos dados pode ser dividida em:

1. Importação das bibliotecas necessárias para estudo dos dados;
2. Importação do *dataset* da E-REDES usado em estudo;
3. Limpeza e tratamento deste *dataset* (tratamento de valores omissos);
4. Seleção de todos os dados do ano pretendido (ano de 2022);
5. Agrupamento dos valores referentes aos dados dos trimestres por concelho;
6. Importação dos *datasets* da Pordata (houve necessidade de tratamento dos dados antes de serem usados);
7. Matriz de correlação, de forma a obter o coeficiente de correlação de Pearson (valor que varia entre -1 e 1 e que, quanto mais próximo do módulo de um, melhor a correlação entre os dados);
8. No caso da melhor relação, realizou-se ainda um *scatter plot* com a reta da regressão linear desenhada.

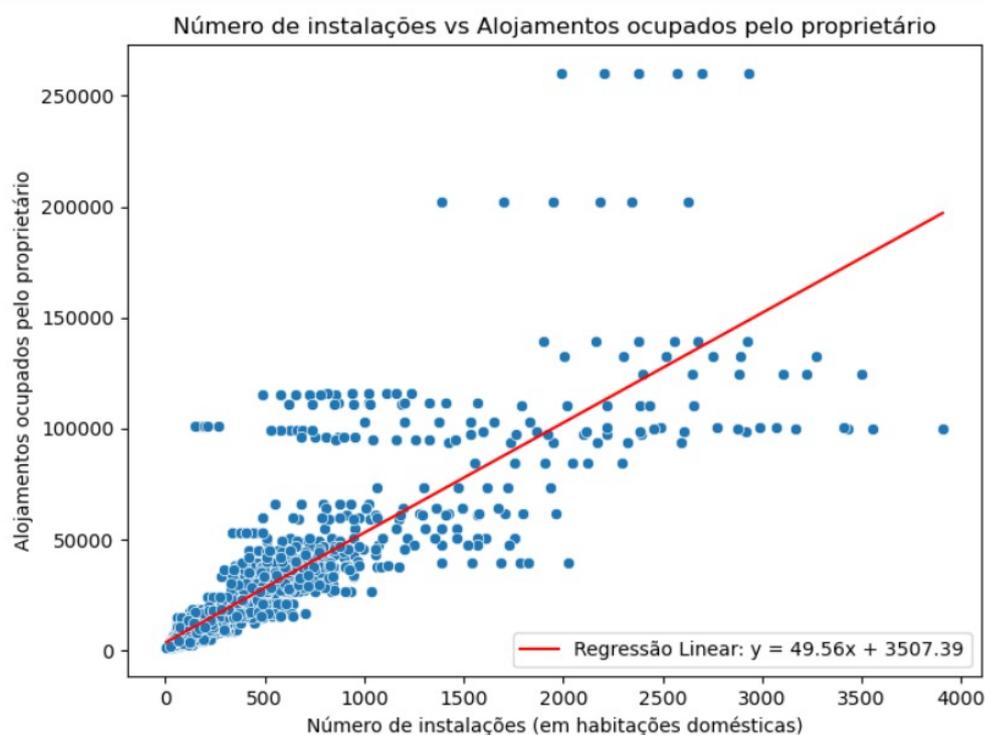
### **Resultados obtidos**

Como é possível visualizar na **Tabela 1**, o melhor resultado obtido foi relacionando o número total de unidades de produção para autoconsumo com o número de alojamentos ocupados pelo proprietário obtendo um coeficiente de determinação de 71%, isto é, é possível explicar 71% da distribuição do número total de unidades de produção para autoconsumo através do número de alojamentos ocupados pelo proprietário.

Tabela 1. Resultados das análises de dados

<i>Variável com a qual foi relacionado o número total de unidades de produção para autoconsumo</i>	<b>Coefficiente de correlação (R)</b>	<b>Coefficiente de determinação (R<sup>2</sup>)</b>
<i>Densidade empresarial</i>	0,11	1 %
<i>Poder de compra</i>	0,21	4 %
<i>População total</i>	0,31	10 %
<i>Empresas fotovoltaicas</i>	0,33	11 %
<i>PIB per capita</i>	-0,38	14%
<i>Consumo energético</i>	0,64	41 %
<i>Indivíduos em famílias</i>	0,70	49 %
<i>Alojamentos ocupados pelo proprietário</i>	0,84	71 %

Na **Fig. 2** é visível essa relação, através de um *scatter plot* com a reta de regressão desenhada.



**Fig. 2.** Gráfico da correlação entre o número de instalações e o número de alojamentos ocupados pelo proprietário

Com isto, conclui-se que, nos concelhos onde o número de total de unidades de produção para autoconsumo é superior, é de esperar que o número de alojamentos ocupados pelo proprietário também seja elevado.

## 4.2 Desenvolvimento do website

O desenvolvimento do website foi planeado de modo a atender aos objetivos do projeto de forma prática e eficiente, utilizando tecnologias consolidadas como HTML, CSS, JavaScript no front-end, e PHP no back-end, com integração de uma base de dados SQL e a API da E-REDES. Este sistema oferece funcionalidades de análise e visualização dos dados provenientes da API conjugadas com autenticação de utilizadores.

A interface do sistema foi desenvolvida com HTML5 para a estruturação das páginas com organização semântica e compatibilidade entre diferentes browsers. Para os estilos das páginas foi utilizado o CSS3, que conferiu uma experiência visual limpa e responsiva aos elementos de cada página. A interatividade foi expandida com a utilização do JavaScript, que permitiu a utilização de gráficos e a filtragem dos dados.

As páginas principais do sistema são:

- Página inicial;
- Sobre nós: informação relativa ao projeto e à equipa que o desenvolveu;
- Área do utilizador: informações do utilizador, com possibilidade da alteração das mesmas;
- Publicações: é formada por um registo de publicações a funcionar como um centro de partilha das análises de dados entre utilizadores, onde o próprio utilizador pode ler e criar as mesmas;
- Conjuntos de dados: apresenta as fontes disponíveis para analisar conjuntos de dados (E-REDES, Pordata, Fontes externas);



Figura 1-Captura de ecrã da página de seleção de fontes de conjuntos de dados

- Dashboard: exhibe as métricas “Torres instaladas”, “Total de distritos” e “Total de potência” e uma visualização gráfica sobre o conjunto de dados da E-REDES do total de unidades de produção para autoconsumo, que foi estudado durante o projeto.

O back-end, implementado com PHP, funcionou como “ponte” entre a interface e a gestão dos dados. Algumas das funcionalidades são:

- Sistema de autenticação: com um nome de utilizador ou e-mail e uma palavra-passe, um utilizador podia fazer registo e posteriormente entrar na sua conta para gerir as suas publicações;
- Gestão persistente de dados: armazenamento dos dados do utilizador e das suas publicações na base de dados;
- Integração com a API da E-REDES: o PHP formata e adapta os dados consumidos da API de modo a estes serem apresentados de forma organizada no front-end.

## 5 Conclusão

### 5.1 Objetivos concretizados

Com base no trabalho desenvolvido, os objetivos definidos inicialmente foram alcançados com sucesso. Entre as principais realizações, destacam-se:

#### **Análise Geográfica das UPACs.**

O estudo revelou que os municípios com mais Unidades de Produção para Autoconsumo (UPACs), tais como Guimarães, Braga e Vila Nova de Gaia, possuem características socioeconómicas que impactam diretamente a implementação de painéis solares. Foi viável vincular o número de alojamentos ocupados pelo proprietário como uma variável explicativa significativa, apresentando um coeficiente de determinação de 71%. Este resultado sugere que habitações com mais pessoas apresentam um interesse mais elevado para a colocação de painéis fotovoltaicos. Com isto em mente, uma política pública potencialmente interessante seria um incentivo para famílias numerosas de forma a promover a adoção deste tipo de tecnologias.

#### **Desenvolvimento de uma Plataforma Web.**

A plataforma criada possibilitou a fusão de dados de diferentes fontes, como E-REDES, Pordata e Fundo Ambiental, simplificando a análise interativa dos dados. A visualização geoespacial e os gráficos dinâmicos ofereceram perceções valiosas sobre a dispersão e característica das instalações fotovoltaicas.

#### **Exploração de Dados Complementares.**

A incorporação de variáveis como densidade empresarial, poder de compra e consumo energético aprimorou a compreensão das tendências de instalação de painéis solares em Portugal. Estas análises detalhadas evidenciaram a relação entre condições socioeconómicas e a implementação de tecnologias sustentáveis.

### 5.2 Limitações

O projeto desenvolvido apresentou algumas limitações que devem ser consideradas para futuras melhorias e para a interpretação dos resultados.

### **Qualidade dos dados utilizados.**

Uma das maiores restrições foi a desatualização de alguns conjuntos de dados, como os referentes ao número de indivíduos por família e os alojamentos ocupados por proprietários, que datavam de 2021. A ausência de dados atualizados dificultou comparações precisas com as informações mais recentes de 2022. Ademais, os desafios relacionados com a consistência temporal entre diferentes fontes de dados afetaram a análise longitudinal. A escolha de dados mais recentes iria conduzir a conclusões mais fidedignas e mais próximas da atualidade.

### **Funcionalidades da plataforma web.**

Embora a plataforma tenha cumprido os objetivos principais, algumas funcionalidades planeadas inicialmente, como gráficos interativos de dispersão e filtros mais avançados, não foram implementadas por completo devido à falta de tempo para trabalhar de forma mais incisiva estas funcionalidades. Essas limitações restringiram o potencial exploratório e interativo da ferramenta. A implementação de todas as funcionalidades possibilitaria a existência de uma plataforma web totalmente funcional no quesito da análise dos dados e na extração de conclusões dos mesmos.

### **Abrangência das variáveis analisadas.**

A análise realizada focou-se principalmente em variáveis socioeconómicas e geográficas, como densidade empresarial, poder de compra, e características habitacionais. No entanto, outras variáveis relevantes, como incentivos fiscais, políticas energéticas regionais e custos de instalação, não foram incluídas, limitando a abrangência das conclusões obtidas.

### **Dependência de APIs externas.**

A integração com APIs, incluindo a do portal E-REDES, apresentou limitações em termos de estabilidade e velocidade de acesso. Essas restrições impactaram diretamente o processamento e a recolha de dados, ocasionando atrasos em algumas fases do projeto.

Tais limitações, embora esperadas em projetos desta natureza, oferecem oportunidades de melhoria para trabalhos futuros. Ao abordar estas questões, será possível ampliar a robustez das análises, aprimorar a funcionalidade das ferramentas desenvolvidas e fornecer insights ainda mais relevantes para o contexto da transição em Portugal.

## **5.3 Sensibilização para o tema**

Por vezes, a falta de conhecimento sobre temas como a sustentabilidade, em particular sobre painéis fotovoltaicos, pode conduzir a decisões pouco fundamentadas. Este tipo de conhecimento é fundamental e devia começar a ser disseminado ainda no ensino

obrigatório, de forma a sensibilizar os jovens acerca destas questões. Dessa forma, projeta-se que, no futuro, os jovens consigam tomar decisões informadas e conscientes nas empresas onde desempenharem as suas funções.

Outra abordagem interessante, desta vez tendo em mente as grandes empresas seria a criação de workshops que explicitem os benefícios socioeconómicos e ambientais da adoção de práticas mais sustentáveis.

### Referências

1. Goldenergy, “Energia fotovoltaica,” [Online]. Available: <https://goldenergy.pt/glossario/energia-fotovoltaica/>. [Acedido em 24 outubro 2024].
2. R. Nabais, N. César, E. Agante e P. Silva, “Portugal fotovoltaico: um retrato da energia solar no país,” DECO Pro Teste, 21 fevereiro 2024. [Online]. Available: <https://www.deco.proteste.pt/sustentabilidade/artigo/portugal-fotovoltaico-um-retrato-da-energia-solar-no-pais>. [Acedido em 29 outubro 2024].
3. DGE – Direção Geral de Energia e Geologia, “Estatísticas rápidas das renováveis,” 31 março 2022. [Online]. Available: <https://www.apren.pt/contents/publicationsothers/dgeg-estatisticas-rapidas-das-renovaveis.pdf>. [Acedido em 31 dezembro 2024].
4. Iberdrola, “História da energia solar,” Iberdrola, [Online]. Available: <https://www.iberdrola.com/quem-somos/nossa-atividade/energia-solar-fotovoltaica/historia-energia-solar>. [Acedido em 20 novembro 2024].
5. V. Greco, “Origem da energia solar,” EDP, 27 outubro 2023. [Online]. Available: <https://solucoes.edp.com.br/blog/origem-da-energia-solar/>. [Acedido em 21 novembro 2024].
6. Iberdrola, “Operação de energia solar fotovoltaica,” Iberdrola, [Online]. Available: <https://www.iberdrola.com/sustentabilidade/que-e-energia-solar-fotovoltaica>. [Acedido em 21 novembro 2024].
7. E-Redes, “Total de unidades de produção para autoconsumo,” [Online]. Available: <https://e-redes.opendatasoft.com/explore/dataset/8-unidades-de-producao-para-autoconsumo/dashboard/?flg=pt&sort=trimestre>. [Acedido em 30 dezembro 2024].
8. Idealista, “Energia: Portugal aposta nas unidades de produção para autoconsumo,” 18 abril 2023. [Online]. Available: <https://www.idealista.pt/news/financas/economia/2023/04/18/57457-energia-portugal-aposta-nas-unidades-de-producao-para-autoconsumo>. [Acedido em 29 outubro 2024].

## Sistema Interativo de Visualização de Posters

Bruno Sousa Ferreira<sup>1</sup>, Carlos José Campos<sup>1,2</sup>, António Sousa<sup>1,3</sup>, and  
Fernando Carvalho<sup>1</sup>

<sup>1</sup> Instituto Superior de Engenharia do Porto (IPP-ISEP), Porto, Portugal

<sup>2</sup> Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC  
TEC), Porto, Portugal

<sup>3</sup> Laboratório de Engenharia Matemática (LEMA), Porto, Portugal  
{1162089,crc,ats,fjc}@isep.ipp.pt

**Resumo.** As preocupações com a sustentabilidade e o meio ambiente motivaram a realização deste trabalho, que tem como objetivo a redução do uso de papel na produção e publicação de posters académicos. Atualmente, os trabalhos de fim de curso das Licenciaturas do Departamento de Engenharia Eletrotécnica (DEE) do Instituto Superior de Engenharia do Porto (ISEP) são exibidos em posters de papel expostos no corredor do departamento. Esta exposição visa divulgar os projetos dos estudantes de forma sucinta e visual em contexto académico, promovendo a colaboração com empresas que procuram inovar as suas tecnologias. No entanto, este formato físico restringe a divulgação ao espaço físico da instituição e dificulta o seu acesso após o seu ciclo de exposição. Este facto, restringe a visibilidade dos projetos realizados, limitando o seu alcance a uma dimensão pública que se deseja ampla. Assim, foi desenvolvida uma solução digital que cria sinergias com o formato físico existente, visando aumentar significativamente o alcance e a visibilidade dos trabalhos. O sistema implementado, com base nas tecnologias PHP, MySQL, HTML, CSS e JavaScript, inclui funcionalidades que facilitam a gestão e organização de conteúdos, proporcionando uma experiência eficiente e estruturada na pesquisa e consulta destas comunicações, alinhando-se com as tendências modernas de educação e partilha digital.

**Palavras-chave:** Sistemas de informação, Aplicação web, Gestão de conteúdos, Visualização de Posters.

## 1 Introdução

Este artigo apresenta o desenvolvimento de um sistema interativo de visualização de posters, dedicado aos cursos de Engenharia do ISEP, em particular do DEE. Foi concebido como complemento à exposição física, visando superar as atuais limitações de espaço e tempo, aumentando a visibilidade e acessibilidade dos projetos estudantis, em alguns casos coadjuvados por empresas acolhedoras que visam evoluir tecnologicamente. A solução inclui funcionalidades para gestão, organização, consulta e divulgação remota dos trabalhos, otimizando a sua pesquisa e acessibilidade. O projeto responde assim às limitações das exposições físicas e apresenta uma plataforma digital eficiente.

Com o objetivo principal de desenvolver um sistema integrado de gestão e visualização online de posters acadêmicos, integra uma base de dados relacional, permite a geração automática de páginas *web*, plataforma multi-utilizador, funcionalidades de pesquisa e filtragem, além de ferramentas para edição de metadados e carregamento de novos trabalhos.

Foram analisados sistemas similares: o *JEM* (Jornadas Científicas Egas Moniz) [1], o *website* do jornal *A Bola* [2], o sistema desenvolvido na tese de Mestrado de Daniel Cruz [3] e o sistema PhotoFinder Kiosk [4]. Embora cada um dos sistemas tenha contribuído com *insights* valiosos, as suas limitações serviram de referência para este projeto.

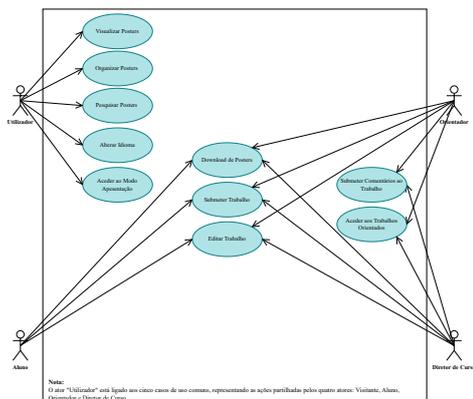
O sistema de gestão de conteúdos do jornal *A Bola* foi adaptado às necessidades deste projeto, permitindo obter um *frontend* personalizado e um *backend* robusto para gestão contínua. A escolha das tecnologias para o desenvolvimento da aplicação considerou a experiência prévia e a adequação às necessidades do projeto. Foi utilizada a *stack* de tecnologias composta por PHP, MySQL, HTML, CSS e JavaScript [5] [6] [7]. As ferramentas utilizadas incluem o uso do Figma, XAMPP, phpMyAdmin, draw.io e do Visual Studio Code como ambiente de desenvolvimento, garantindo uma implementação eficiente e escalável [8] [9].

## 2 Conceção do protótipo

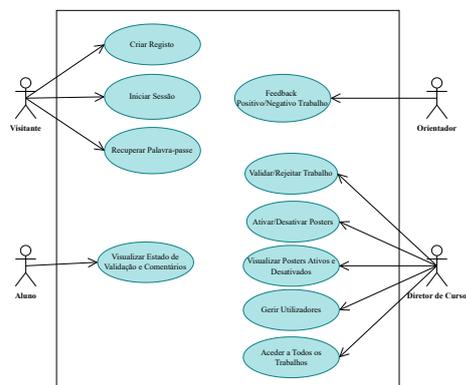
O desenvolvimento do protótipo do sistema seguiu uma abordagem metodológica estruturada, focada nas necessidades académicas e de utilização. A metodologia incluiu etapas cruciais, desde a análise de requisitos até o design da interface. A análise identificou os principais atores (Visitante, Aluno, Orientador e Diretor de Curso) e definiu os requisitos funcionais e não funcionais. No design, destacou-se a arquitetura funcional, estruturação da base de dados e o *layout* da interface. A prototipagem envolveu a criação de um protótipo funcional para testar as funcionalidades desenvolvidas. Esta abordagem assegurou que a aplicação atendesse às necessidades dos utilizadores, oferecendo uma experiência intuitiva, amigável e eficiente, com iterações constantes entre análise, design e implementação.

### 2.1 Análise de requisitos

O sistema possui requisitos funcionais organizados em dois diagramas de casos de uso (Figuras 1 e 2) que ilustram as funcionalidades principais e as interações entre os diferentes atores e o sistema. As funcionalidades incluem visualização, pesquisa e organização de posters, gestão de conta, submissão, edição, validação de trabalhos, gestão de utilizadores e conteúdos. Os atores identificados são: Visitante, utilizador não registado que pode visualizar e pesquisar posters, organizar a exibição, criar conta e iniciar sessão; Aluno, utilizador registado que pode submeter, editar e verificar o estado de validação dos seus trabalhos; o Orientador, utilizador registado que valida, rejeita ou comenta os trabalhos dos alunos que orienta; Diretor de Curso (Administrador), utilizador com permissões



**Fig. 1.** Diagrama de Casos de Uso comuns.



**Fig. 2.** Diagrama de Casos de Uso específicos.

administrativas para gerir utilizadores, ativar ou desativar posters e validar ou rejeitar trabalhos previamente submetidos.

O sistema abrange funcionalidades comuns aos diferentes atores, como a visualização de posters em mosaico em exibição sequencial, simulando uma exposição contínua, a pesquisa por tópicos, a organização por critérios e a alteração de idioma, esplanadas na Figura 1. Na Figura 1 é ainda possível visualizar a função principal do Aluno, a submissão e edição de trabalhos. A Figura 2 destaca funções específicas como: a validação e *feedback* dos Orientadores, permissões administrativas do Diretor de Curso que pode ativar ou desativar posters, validar trabalhos e gerir utilizadores.

Já os requisitos não funcionais abrangem o desempenho, segurança, usabilidade, confiabilidade, compatibilidade e manutenibilidade do sistema. Esses aspetos asseguram um sistema eficiente e seguro, temporalmente responsivo, de autenticação robusta e disponibilizando uma interface intuitiva [10] [11] [12] [13].

## 2.2 Design do sistema, base de dados e interface

A arquitetura que suporta o design do sistema de gestão de posters segue o padrão MVC (*Model-View-Controller*) [14], que separa as três camadas de responsabilidades em: *Model*, para gestão de dados e lógica de negócios; *View*, responsável pela interface do utilizador; e *Controller*, que atua como intermediário entre o *Model* e o *View*.

A base de dados MySQL utiliza um modelo relacional, representado no diagrama entidade-relação (ER) da Figura 3. Com tabelas interligadas por chaves primárias (PK) e estrangeiras (FK), garantindo a integridade dos dados, armazenando informações sobre os posters submetidos, os utilizadores e os comentários dos orientadores.

O design dos *wireframes* da interface foi desenvolvido no Figma: a título exemplo apresenta-se o da página principal na Figura 4.

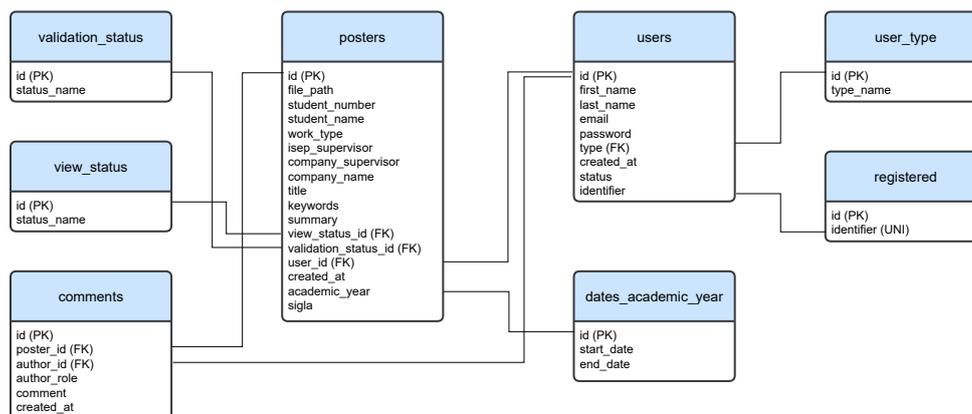


Fig. 3. Base de dados relacional implementada.

### 3 Implementação

As principais funções do sistema, implementadas a partir da interação do *backend* em PHP e o *frontend*, incluem a visualização de posters, a submissão de trabalhos e a validação ou rejeição de submissões, permitindo uma gestão eficiente e de uso intuitivo [5] [7]. O sistema foi implementado com mecanismos de proteção de dados, utilizando validações integradas e o uso eficiente de uma base de dados. A adoção do padrão MVC permite organizar o sistema, simplificando a sua manutenção e escalabilidade. A Figura 5 mostra a estrutura de pastas do projeto, destacando como exemplo dois ficheiros de código por pasta.

#### 3.1 Implementação das funcionalidades

As funções implementadas mais importantes incluem a visualização de posters, submissão de trabalhos, validação e rejeição.

##### Visualização de posters

Todos os utilizadores podem aceder aos posters ativos via *index.php* (URI). No *backend*, os posters são carregados a partir da base de dados, podendo ser usados filtros como ano letivo, tipo de trabalho ou palavras-chave. A consulta SQL filtra apenas os posters ativos e validados:

```

1 SELECT p.*, v.status_name
2 FROM posters p
3 JOIN validation_status v ON p.validation_status_id = v.id
4 WHERE p.validation_status_id = 4 AND p.view_status_id IN (1, 2);

```

A função *PDFtoIMG()* é responsável por gerar as imagens dinâmicas dos PDFs no *frontend*:

```

1 async function PDFtoIMG(url){
2   const page = await (await pdfjsLib.getDocument(url).promise).getPage(1);
3   const canvas = document.createElement('canvas');

```

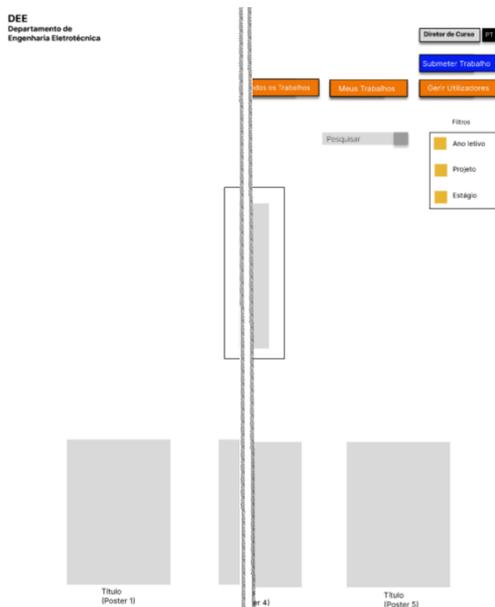


Fig. 4. Wireframe simplificado da página principal.

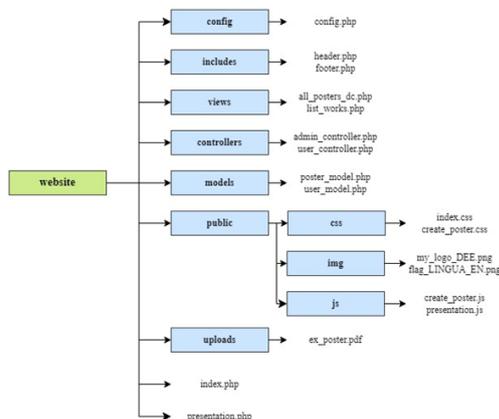


Fig. 5. Estrutura de ficheiros do sistema.

```

4   await page.render({
5     canvasContext: canvas.getContext('2d'),
6     viewport: page.getViewport({ scale: 1.5 })
7   }).promise;
8   return canvas.toDataURL();
9 }

```

Os posters ativos apresentam-se em slides contínuos, simulando uma exposição automática.

### Submeter Trabalho

Utilizadores autenticados submetem posters e respetivos metadados via formulário HTML, através da *View create\_poster.php*:

```

1 <form method="post" enctype="multipart/form-data">
2   <input type="text" name="student_number" placeholder="Nº Aluno" required>
3   <input type="file" name="file_path" accept=".pdf" required>
4   <button type="submit">Submeter</button>
5 </form>

```

O *script create\_poster\_controller.php* controla a submissão, validando o ficheiro e inserindo os dados na base de dados:

```

1 INSERT INTO posters (file_path, student_name, title, ...)
2 VALUES (?, ?, ?, ...);

```

### Validar ou rejeitar trabalho

O Diretor de Curso valida/rejeita trabalhos, funções implementadas através do *script, validate\_poster\_dc\_controller.php*. A função *validatePoster()* atualiza o estado do poster:

```
1 function validatePoster($posterId, $validationStatusId){
2     $query = "UPDATE posters SET validation_status_id=? WHERE id=?";
3 }
```

Os comentários dos diversos atores são adicionados pela função *addComment()*:

```
1 function addComment($posterId, $authorId, $authorRole, $comment) {
2     $query = "INSERT INTO comments (poster_id, author_id, author_role,
3         comment) VALUES (?, ?, ?, ?)";
}
```

No *frontend*, uma interface permite a validação e a inserção de comentários:

```
1 <form method="post">
2     <select name="validation_status">
3         <option value="4">Validar</option>
4         <option value="5">Rejeitar</option>
5     </select>
6     <textarea name="comments"></textarea>
7     <button type="submit">Guardar</button>
8 </form>
```

### 3.2 Configuração da base de dados e validações

O sistema utiliza uma base de dados MySQL para armazenar informações relacionadas aos posters, utilizadores e outros dados essenciais ao funcionamento da aplicação [8]. A ligação à base de dados configurada no ficheiro *config.php* é gerida pela função *connectDB()* invocada em diferentes partes do código.

```
1 function connectDB(){
2 $conn = mysqli_connect(DB_HOST, DB_USER, DB_PASS, DB_NAME);
3 if (!$conn){
4 die("Connection failed: " . mysqli_connect_error());}
}
```

Em caso de falha de ligação, uma mensagem de erro apropriada é apresentada, evitando a execução de operações incorretas. Do lado do cliente, as validações foram implementadas para garantir que os utilizadores não submetem informação incorreta ou incompleta, utilizando atributos html como *required*, *minlength* ou *maxlength*.

```
1 <input type="text" name="identifier" required>
2 <input type="password" name="pw" minlength="8" maxlength="15" required>
```

Do lado do servidor, as validações desenvolvidas em PHP, garantem a segurança e consistência dos dados. São verificadas permissões de acesso, permitindo apenas que utilizadores autorizados realizem ações restritas, como alterar ou validar trabalhos.

```
1 if (!isset($_SESSION['user_id']) || $_SESSION['type'] != 3){
2 die("Acesso negado. Não tem permissão para realizar esta ação.");}
```

Os ficheiros submetidos são analisados para garantir que possuem extensão \*.pdf.

```
1 if ($fileType != 'pdf') {
2 header('Location: ../views/poster_error.php?error=invalid_file_type');
3 exit();}
```

Os dados enviados pelos utilizadores passam por análises para assegurar que estão em conformidade. Além disso, todas as interações com a base de dados utilizam *prepared statements*, protegendo o sistema de ataques de *SQL Injection*.

```
1 $stmt = mysqli_prepare($conn, "INSERT INTO posters
2 (file_path, student_number, student_name, ...) VALUES (?, ?, ?, ...)");
3 mysqli_stmt_bind_param($stmt, "sss", $filePath, $studentNumber,
4 $studentName);
5 mysqli_stmt_execute($stmt);
```

Com a função `mysqli_stmt_bind_param()`, os dados fornecidos pelo utilizador incorporam-se na consulta de forma segura.

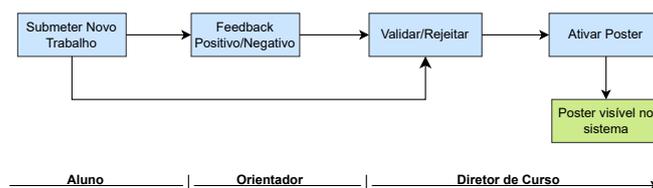
## 4 Demonstração de resultados

Esta secção apresenta uma sequência de passos ilustrativos da usabilidade do sistema, evidenciando os resultados obtidos. O fluxograma (Fig. 6) mostra o processo de submissão, validação e ativação de um poster, destacando as interações entre os diferentes tipos de utilizadores (Aluno, Orientador e Diretor de Curso) e o sistema.

O processo começa com a submissão do trabalho pelo aluno, seguido do *feedback* positivo do orientador e da validação e ativação do poster pelo Diretor de Curso. Se o Diretor de Curso validar antes do *feedback* do orientador, o trabalho é automaticamente validado, dispensando esse *feedback* intermédio. Se houver *feedback* negativo do orientador ou rejeição do Diretor de Curso, o aluno deve corrigir o poster e submetê-lo novamente. Após a publicação, todos os utilizadores podem aceder ao modo de apresentação, exibindo os posters sequencialmente em ecrã inteiro.

**Aluno:** O aluno inicia o processo acedendo à página principal do sistema, que apresenta as funcionalidades disponíveis sem necessidade de autenticação (Figura 7). Após efetuar o *login*, a interface inicial exibe as opções de navegação e uma breve introdução ao serviço disponível. O aluno poderá aceder à opção "Meus Trabalhos" no menu principal, onde consulta a lista de trabalhos submetidos (Figura 8). Ao clicar em "Submeter Novo Trabalho", preenche um formulário com os metadados e carrega o ficheiro PDF do poster. Após a submissão, o trabalho aparece na lista, em estado de validação "Pendente" (Figura 8).

**Orientador:** Após realizar o *login*, o orientador acede à lista de posters submetidos pelos seus orientados. Ao clicar num trabalho, o orientador visualiza detalhes (Figura 9), como: título, resumo, orientadores envolvidos e o ficheiro PDF submetido (Figura 10). O orientador pode então efetuar o *feedback* avaliativo a partir de um comentário dirigido ao aluno ou somente aprovar/rejeitar o trabalho submetido.



**Fig. 6.** Fluxograma do processo de submissão, validação e ativação de um poster.



Fig. 7. Página inicial da área de utilizador.

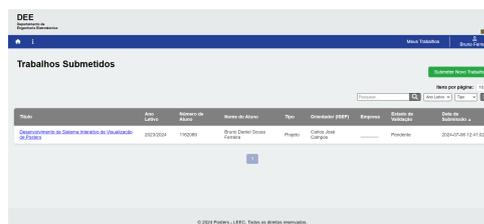


Fig. 8. Página de exibição de trabalhos submetidos.

**Desenvolvimento de Sistema Interativo de Visualização de Posters**

**Palavras-chave:** Sistemas de Informação, aplicação web, gestão de conteúdos, visualização de Posters  
**Resumo:**

Atualmente, os trabalhos de fim de curso dos alunos da Licenciatura em Engenharia Eletrotécnica e de Computadores (LEEC) são apresentados através de posters em papel, expostos no corredor junto ao Departamento de Engenharia Eletrotécnica (DEE) do Instituto Superior de Engenharia do Porto (ISEP). Estes posters têm como objetivo ilustrar de forma sucinta e visual o trabalho realizado pelos estudantes, tanto em contexto académico quanto em colaboração com empresas. No entanto, este formato apresenta algumas limitações. A exposição em papel restringe a divulgação dos trabalhos a um espaço físico específico, limitando o seu alcance às pessoas presentes na instituição. Além disso, a exibição é temporária, dificultando a consulta dos trabalhos após o fim da exposição. Face a estas limitações, surgiu a necessidade de desenvolver uma solução digital que complemente o formato físico, permitindo uma divulgação mais ampla e acessível dos trabalhos académicos. O Sistema Interativo de Visualização de Posters (SIVISP) responde a essa necessidade, permitindo a consulta remota dos posters e aumentando significativamente o seu alcance e visibilidade.

**Ano letivo:** 2023/2024

**Número de Aluno:** 1162089  
**Nome do Aluno:** Bruno Daniel Sousa Ferreira  
**Tipo de Trabalho:** Projeto  
**Orientador (ISEP):** Carlos José Campos

**Estado de Validação:** Pendente  
**Comentários:**

Sem comentários.

Fig. 9. Metainformação do trabalho.

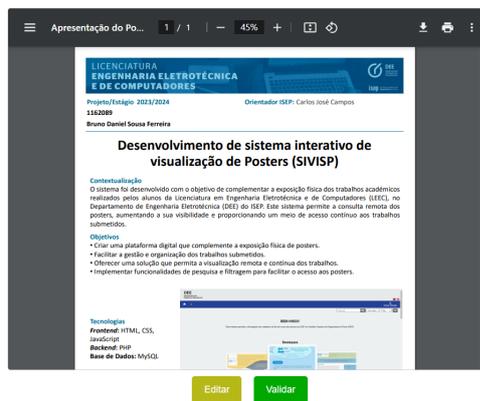


Fig. 10. Exibição do Poster submetido.

**Diretor de curso:** O Diretor de Curso acede à secção "Trabalhos Pendentes" para analisar os trabalhos submetidos pelos alunos e aprovados pelos orientadores. Após a sua validação, o trabalho surge na lista "Trabalhos Validados", onde o Diretor pode selecionar o poster para que este fique visível no sistema (Figura 11). Sem este último passo, o trabalho não ficará acessível online. Para além disto, o Diretor pode definir as datas de início e fim do ano letivo, gerir utilizadores, ativando contas e atribuição papéis (Figura 12). Assim como outros utilizadores, o Diretor de curso pode também alterar informações de registo, como a palavra-passe ou outras informações.

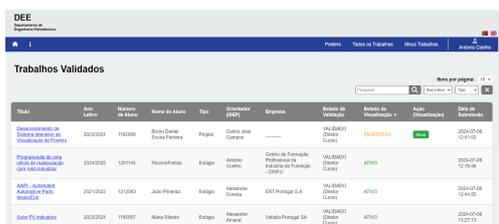


Fig. 11. Lista de trabalhos pendentes.

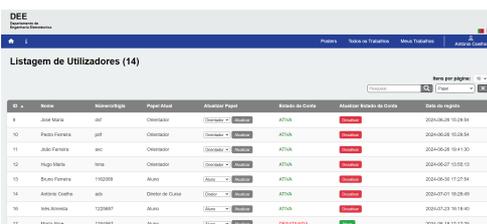


Fig. 12. Página gestão de utilizadores.

**Visitante:** Finalmente, após a validação e ativação do poster, este torna-se visível no sistema e disponível para todos os utilizadores (Figura 13), que podem aceder ao modo de apresentação, onde os posters são exibidos sequencialmente em ecrã inteiro (Figura 14).



**Fig. 13.** Exibição automática de trabalhos publicados.



**Fig. 14.** Exibição única em *fullscreen*.

## 5 Conclusões e trabalho futuro

A gestão automatizada do registo e a divulgação de trabalhos académicos representa um avanço significativo na publicitação de posters e na desmaterialização da documentação para preservação e memória futura. Ao combinar tecnologias modernas com um design centrado no utilizador, o sistema não apenas resolveu os desafios existentes, mas também abriu novas possibilidades para a disseminação do conhecimento académico.

### 5.1 Objetivos concretizados

O sistema desenvolvido para gestão e visualização automática de posters complementa a exposição física dos trabalhos, ampliando a sua disponibilidade e visibilidade. Utilizando uma base de dados robusta, organiza a informações dos posters, incluindo metadados relevantes. A exibição dinâmica dos posters é automatizada, possibilita o uso de filtros de pesquisa por ano letivo, tipo de trabalho e palavras-chave, além de permitir a edição contínua da metainformação e carregamento de novos trabalhos. Entre as melhorias, destaca-se a pesquisa avançada, aumentando a eficiência e usabilidade das consultas. O sistema proporciona a gestão contínua de conteúdos, similarmente ao *website* A Bola, mas com filtros académicos avançados. Adaptado ao contexto académico e com funcionalidades específicas, supera a abordagem de Daniel Cruz. Também foi implementado um sistema de *login* e gestão de utilizadores, com diferentes níveis de permissões (Visitante, Aluno, Orientador, Diretor de Curso), superando as limitações do JEM.

## 5.2 Limitações e trabalho futuro

Embora os principais objetivos do projeto tenham sido alcançados, o sistema ainda apresenta limitações que permitem oportunidades de melhoria. A gestão de utilizadores e permissões pode ser aprimorada, incluindo mais perfis e funcionalidades personalizadas, como notificações automáticas e alertas de prazos. Futuras melhorias focam-se na otimização do desempenho do sistema, particularmente na implementação de técnicas de *caching* e na criação de Vistas na base de dados que deverão permitir melhorar o tempo de resposta do sistema a consultas de trabalhos guardados [16]. Uma possível direção seria a adaptação do sistema a outros departamentos do ISEP, criando uma plataforma unificada para a gestão de trabalhos académicos em diferentes contextos [18].

O sucesso do projeto demonstra o potencial da tecnologia para transformar práticas educacionais tradicionais, criando um ambiente mais dinâmico, acessível e interativo para a comunidade.

## Referências

1. Página das Jornadas Egas Moniz: Scientific Journeys. (n.d.), disponível em <https://jem.egasmoniz.edu.pt/>, consultado em 2024/04/20.
2. Página A BOLA - Homepage | Abola.pt. (n.d.), disponível em <https://www.abola.pt/>, consultado em 2024/04/23
3. Cruz, Daniel F.: Sistema de geração de conteúdos web para serviços de informação situados. Universidade do Minho (Portugal)- Dissertações de Mestrado DSI - Engenharia e Gestão de Sistemas de Informação, <https://hdl.handle.net/1822/81079> (2022)
4. Kules, B., Kang, H., Plaisant, C., Rose, A., & Shneiderman, B. (2004). Immediate usability: a case study of public access design for a community photo library.
5. Duckett, J. *HTML & CSS: Design and Build Websites*. John Wiley & Sons, 2011
6. Welling, L., & Thomson, L. *PHP and MySQL Web Development*. 5ª ed, Addison-Wesley, 2016
7. Ullman, L. *PHP for the Web: Visual QuickStart Guide*. Peachpit Press, 2016
8. Schwartz, B. et al. *High Performance MySQL*. O'Reilly Media, 2012
9. *XAMPP Installers and Downloads for Apache Friends*. (n.d.). Retrieved June 15, 2024, from <https://www.apachefriends.org/>
10. Nielsen, J. (1994). *Usability engineering*. San Francisco: Morgan Kaufmann
11. Henderson, C. (2006). *Building scalable web sites*. O'Reilly Media
12. Wieder, P. (2014). *Service level agreements for cloud computing*. Springer
13. Marcotte, E. (2011). *Responsive web design*. A Book Apart
14. Padrão de Design MVC - GeeksforGeeks. (n.d.). Retrieved July 2, 2024, from <https://www.geeksforgeeks.org/mvc-design-pattern/>
15. Hearst, M. A., *Search User Interfaces*, 1ª ed., Cambridge: Cambridge University Press, 2009
16. SMITH, John *Scalable System Architecture: Best Practices*. O'Reilly Media, 2019
17. MARTIN, Robert C. *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall, 2008. ISBN 978-0-13-235088-4
18. SILVA, M. *Gestão de plataformas académicas integradas: Um estudo de caso*. Universidade de Lisboa, 2021

# Accessible Tourism Travel Coach App

Anjos, J<sup>1</sup> and Carvalho, P.<sup>1,2</sup>

<sup>1</sup> ISEP - Engineering School of the Porto Polytechnic, Porto, Portugal

<sup>2</sup> GILT - Games, Interaction and Learning Technologies

**Abstract.** People with disabilities face significant challenges in their daily lives. These challenges are even greater when planning tourism trips, as existing platforms are often not optimized to meet their specific needs. This project aims to address these difficulties by providing clear, accessible, and reliable information, particularly for individuals with mobility disabilities.

The project focuses on assisting people with physical disabilities in planning tourism trips and accessing locations in the Porto and Vila Nova de Gaia regions. The developed software integrates features such as Firebase database communication, user authentication, itinerary planning, and multilingual support.

To ensure reliable information about tourist destinations and actively involve users, the application includes an evaluation system that allows users to share their experiences and rate the accessibility of locations.

This work resulted in a robust and efficient mobile application that empowers users with disabilities to plan their trips more effectively and inclusively. By fostering a collaborative evaluation system, it promotes a more accessible and rewarding tourism experience for all.

**Keywords:** Travel Planning for Reduced Mobility, Travel Support for Reduced, Mobility Wheelchair Accessibility, Inclusive Design, Android and iOS, Mobile App Performance Testing, Functional Testing, Firebase Integration

## 1 Introduction

### 1.1 Context

Tourism plays a crucial role in promoting a country's culture, socialization, and economy, in addition to offering unforgettable experiences to its visitors. However, some individuals are unable to travel independently and require assistance. This project aims to support those who face physical challenges in travelling by providing a mobile application to help them plan their trips.

The project specifically targets individuals who use wheelchairs, as well as those with reduced mobility (PRM). The focus is on the Porto region, particularly the union of the parishes of Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau, and Vitória, in terms of locations and itineraries.

## 1.2 Objectives

The primary objective of this project is to develop a mobile application that provides accessible travel assistance. The app will offer personalized support throughout the entire travel process, from planning to navigation during the trip. Users will receive guidance tailored to their specific needs, including suggestions for accessible destinations such as transportation, accommodation, and tourist activities.

Furthermore, the app will provide real-time information about accessibility conditions, alerting users to obstacles along their route and suggesting alternative accessible paths.

The application will be developed using React Native, integrating technologies like a personalized database and an API (Application Programming Interface) focused on itineraries. By doing so, the project aims to promote inclusion and accessibility in tourism.

## 2 Accessible Tourism

Accessible tourism ensures that travel destinations, accommodations and services are accessible to everyone, including people with disabilities. Initially focused on basic accommodations, today it encompasses transportation, attractions and digital resources, promoting more inclusive travel experiences [1].

People with mobility difficulties face barriers in transport and access to public spaces that are not prepared to receive them, thus limiting the possibility of fully enjoying their trips. The evolution and dissemination of technology, especially mobile applications, has been providing valuable assistance to these individuals by providing support in planning and navigation during their trips [2][3].

One of the devices most used by people with limited mobility is the wheelchair. There are manual and motorized wheelchairs, with the latter offering motorized assistance. Key features include a supportive seat and back; wheels for propulsion and steering; arm and footrests for comfort and stability; and a structure made of lightweight and durable materials. However, it is important to consider the dimensions of the wheelchairs, as not all wheelchairs fit in all tourist locations. Most wheelchairs have seats that range from 40 to 50 cm wide and 40 to 45 cm deep. These dimensions can impact accessibility in certain spaces, such as museums, public transport or restaurants, where limited space can restrict access [4][5].

Despite improved mobility, challenges such as uneven terrain and lack of infrastructure persist, highlighting the need for more assistive technologies to ensure fully accessible tourism [6].

This chapter presents a review of the state of the art in relation to mobile applications for accessible tourism. Initially, examples of applications that address the different types of disabilities covered in this project will be presented. Subsequently, the most used technologies and tools for the development of mobile applications will be explored.

## 2.1 Examples of Mobile Applications for Tourist Accessibility

Studies such as those by [23] have highlighted the role of assistive technologies, such as mobile applications, in promoting more inclusive tourism. These apps can provide information about accessibility in tourist destinations, making navigation and trip planning easier. This proposal is evidenced by itTUR, a mobile app available on both Android and iOS that provides information about accessibility along various tourist routes for people with reduced mobility (PRM). The app evaluates routes in Portugal, ranking their difficulty on a scale from 0 to 5. The app offers detailed information about the routes, including descriptions, starting and ending points, points of interest, and availability of accessible facilities. However, it lacks information on individual tourist sites outside of the routes and focuses mainly on southern and central Portugal [16].

Another solution is the Biomob+ designed for PRM, helping users find accessible locations and events. The app allows users to search for nearby accessible places, offering filters for accommodation and restaurants. It also provides images, basic information, and contact details, including Google Maps integration. However, the accessibility information depends on user reviews, and many places lack evaluations [17].

Wheelguide is a leading platform providing accessibility information. The app allows users to evaluate physical access conditions at tourist locations, with data contributed by a large global community. Users can search for places based on their specific needs, such as parking, entrances, and bathroom accessibility. The app also relies on user reviews, which can be a limitation in ensuring the accuracy of the accessibility information [18].

On Wheels is designed with accessibility in mind, offering a user-friendly interface for wheelchair users. The app displays various types of locations, such as museums and parks, and allows users to input their wheelchair specifications to filter locations based on their needs [19].

Available on Android and iOS, Be My Eyes connects visually impaired users with volunteers through video calls, helping them navigate daily tasks. With over 619,000 visually impaired users and over 7 million volunteers, the app provides free, anonymous assistance worldwide [20].

Several studies (e.g. [24],[25]) highlight the importance of adjustment practices and policies that encourage accessibility in tourist destinations, as the work of [26] showed that these apps improve accessibility.

## 2.2 Mobile Application Development Tools

In 2024, there is a wide range of tools for mobile app development. It is essential to evaluate these tools based on factors such as cross-platform compatibility, ease of use, development costs, and team expertise. Some tools considered for this project include React Native, Flutter, and Node.js.

Developed by Facebook and launched in 2015, React Native is a popular open-source framework for developing mobile apps. It allows developers to create apps for both Android and iOS using JavaScript, reducing the learning curve. Companies like Meta, Tesla, Amazon, and Microsoft use React Native [9].

Created by Google, Flutter enables cross-platform development for mobile devices using the Dart programming language. It offers better performance and productivity compared to React Native, making it a strong contender for mobile app development [10].

Node.js is a server-side JavaScript environment known for handling real-time operations. Its ability to handle large volumes of simultaneous requests makes it ideal for mobile apps with high traffic [11].

### 2.3 Databases for Mobile Applications

Choosing the right database is crucial for maintaining security and efficiency in mobile apps. Factors to consider include data complexity, scalability, security, and compatibility with existing technologies. Databases analysed for this project include SQLite, AsyncStorage, Realm, and Firebase Firestore.

SQLite is a lightweight and versatile database that requires minimal configuration, making it ideal for small-scale mobile apps. It is fast due to its lack of a separate server [12].

Integrated in React Native, AsyncStorage is a simple, asynchronous, non-encrypted storage system that allows for the storage of small amounts of data, such as authentication tokens [15].

Realm is an open-source database designed for mobile platforms and supports React Native. It is an object-oriented database that synchronizes data in real time and supports offline data access, making it ideal for apps that require offline functionality [13].

Firebase Firestore is a NoSQL database that stores data in JSON objects and synchronizes in real time across all connected devices. It is part of the Google-owned Firebase platform and is compatible with various programming languages [14].

## 3 Analysis & Design

This chapter presents the main functional and non-functional requirements of the system, which will serve as a basis for the development of the application. Functional requirements describe the essential functionality that the application must provide to meet users' needs, while non-functional requirements address things like performance, compatibility, and usability.

### 3.1 Functional Requirements

- Registration and Authentication: Allow users to register and log in to the application in an accessible way, using methods such as credential caching. This ensures that users do not need to enter their credentials each time they open the app.
- Profile Management: Enable users to edit their profile in an accessible manner, with options to set preferences for their wheelchair type and/or profile picture. This enhances the user experience by personalizing it to individual needs.
- Tourist Places Search: Provide an accessible search functionality to find tourist destinations and points of interest based on accessibility criteria, such as parking, ramps,

restrooms, and other relevant features. This simplifies the process of choosing a point of interest.

- Navigation and Directions: Integrate with map service APIs to offer accessible directions with detailed instructions and navigation support, ensuring easy and intuitive travel planning.
- Ratings and Reviews: Allow users to rate and leave comments about the accessibility of the places they visit. This feature fosters a community-driven approach to sharing suggestions and experiences regarding different locations.

### 3.2 Non-Functional Requirements

- Performance: The application should respond within four seconds to ensure a smooth and fluid user experience.
- Compatibility: The app must be compatible with both iOS and Android devices to maximize its reach and availability.
- Usability: The user interface should be intuitive and simple, featuring touch gesture-based navigation, options to switch between dark and light modes and multiple language support.
- Scalability: The system architecture should be scalable to accommodate growth in the user base and increase search demand without compromising accessibility or performance.
- Maintenance and Updates: Regular updates should be planned to fix bugs, improve accessibility, and introduce new features based on user feedback.
- Privacy: The app must respect user privacy and comply with data protection regulations, ensuring transparency regarding the collection and use of personal information.

### 3.3 System Architecture

Fig. 1 illustrates the system architecture, which consists of navigation, screens, components, context, and services. This modular approach ensures that each "package" has a single responsibility, facilitating easier management and improving code reuse and maintenance.

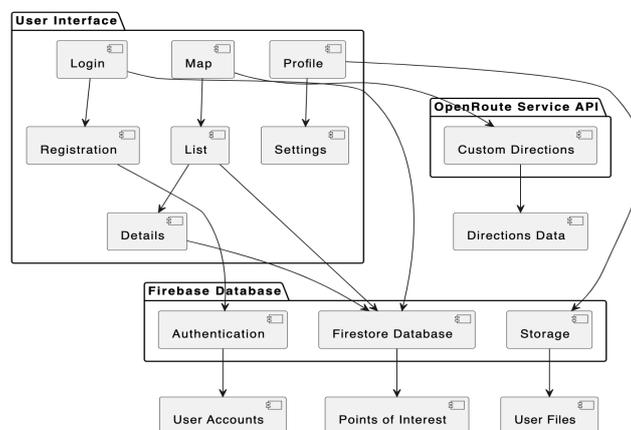


Fig. 1. - System Architecture

### 3.4 Graphical Interface

The application includes a total of seven screens, each with distinct features, providing intuitive and accessible navigation for users. The screens are Login, Registration, Map, List, Profile, Settings and Details. The function of each screen is briefly described below:

- Login: Home screen where the user can authenticate to access the application. Includes fields for entering username and password, as well as password recovery features and the option to keep the user logged in.
- Registration: Registration screen where new users can create an account. The screen offers fields for entering data such as name, email, password and accessibility preferences.
- Map: Interactive map screen that allows users to view landmarks and navigate in an accessible way. Integration with map APIs provides guidance and details about each location's accessibility.
- List: Screen that displays a list of landmarks and places of interest filtered based on accessibility criteria. Users can search and filter by attributes such as ramps, chair dimensions, etc.
- Profile: A screen where users can view and edit their profile. They can also adjust accessibility preferences, making the experience more personalized.
- Settings: A settings screen that allows users to adjust preferences related to the app experience, such as notifications, theme (light or dark mode), and language preferences.
- Details: A screen that shows the details of a landmark or place of interest. It provides comprehensive information about location accessibility and allows users to view and add reviews, comments and suggestions for accessible routes.

These screens are designed to be easy to use, with a focus on accessibility and user experience. Navigation between them has been optimized to ensure users can quickly access key features.

Fig. 2 presents the application's graphical interface, illustrating the layout of the screens and how users interact with each one.

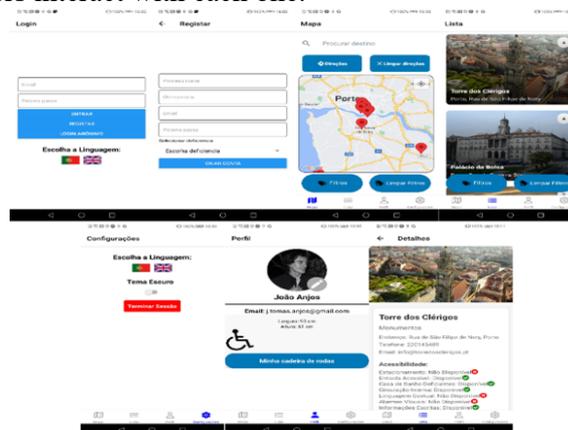


Fig. 2. - Graphical Interface

## 4 Implementation

### 4.1 Initial Setup

To address the need for a custom database and user authentication management, an efficient solution was chosen: the creation of a Firebase server. This system offers multiple services within an integrated platform, accessible through a shared console.

The services used include Authentication, Firestore Database, and Storage.

- Authentication: Provides user identification mechanisms, such as email authentication.
- Firestore Database: Real-time database.
- Storage: Allows the storage and sending of files and images.

Additionally, communication with the OpenRoute Service (OSM) API has been implemented. This open-source API allows the retrieval of custom route data for integration with the map. The decision to use this API was based on its open-source nature, high customizability, and low cost, making it accessible for projects with limited budgets. It also offers several route options, including routes specifically designed for wheelchair users, a key feature to ensure accessibility and inclusion for People with Reduced Mobility (PRM) [22].

### 4.2 Authentication

The application needs to manage users and process their data. To do so, Firebase's Authentication service was used, providing a simple and intuitive system for creating and managing user accounts.

When a user creates an account, they must confirm their email via a link sent automatically by Firebase, which makes the process very accessible. After confirmation, the user can log in and access the application's features. If the user logs out and logs back in, authentication data will be cached, so there is no need to re-enter login credentials, and the user will be automatically redirected to the main screen.

It is important to note that when creating a user, the data is stored in the Firestore Database, where it can be accessed by an administrator with access to the Firebase console. Passwords are encrypted to ensure the user's privacy.

### 4.3 Database

For data management, the Firestore Database was used, which offers great efficiency in storing large volumes of data. In this case, both points of interest and user data are stored. Points of interest were manually added after gathering information from the official websites of the locations and through communication with the responsible entities, ensuring the locations met accessibility requirements for PRM.

User data is stored upon account creation, including information such as email and profile photo. The profile photo is saved in Firebase Storage once uploaded by the user. Managing this data is done efficiently through the web-accessible Firebase console, where an administrator can add, update, or delete data as needed.

The selected area for the collection of points of interest is the Porto region, specifically the union of the parishes of Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau, and Vitória. This region was chosen due to the impossibility of covering the entire territory of Portugal because of the lack of accessibility information. Therefore, it was decided to focus the initial analysis on this area.

#### 4.4 Directions

To obtain directions, it is necessary to make an HTTP request to the Open Route Service API. The basic plan provides 2000 requests per day, capable of handling up to 40 requests per minute, which is ideal for the initial development phase. The API provides directions from a starting point to the destination, predicting the most suitable path for safe travel. These directions are fully customizable, allowing the user to specify additional factors such as transportation type (car, wheelchair), the fastest or shortest route, and the path with fewer curves and less elevation.

### 5 Evaluation

A set of functional tests was carried out to validate the application requirements, covering several functionalities. The procedures adopted for these tests included unit tests for individual components, integration tests to ensure module interaction, and system tests to evaluate the overall performance of the application across devices. The feature of viewing points of interest allows the application to display a map with the points available in the Firebase database, also presenting them in list format. In the functionality to filter points of interest, the user can apply filters such as accessibility, category, city and wheelchair compatibility, with the results being displayed according to the defined criteria. The feature of accessing destination information allows the user to view accessibility details when selecting a point of interest. Provide Feedback allows registered users to rate places and leave comments, stored in Firebase and visible to other users. The route access feature directs the user to Google Maps, where they can get directions from their current location to the selected destination. To create an account, the application offers the option of registering with email validation before logging in. The login functionality makes it possible to access the application using valid credentials. Finally, the customize profile feature allows users to adjust the dimensions of their wheelchairs by saving the data to Firebase.

All tests obtained positive results, ensuring the functionality of the application's main features. App performance was evaluated on iOS and Android platforms to ensure compatibility and responsiveness across devices. All functional tests produced positive results, confirming that the application's main functionalities are operational and meet the requirements for which they were designed. To evaluate the application's performance, tests were carried out on the iOS and Android platforms. The application has demonstrated compatibility with both operating systems. The app also met the expected response time of less than 4 seconds for the above features, ensuring a good user experience.

## 6 Limitations and Future Work

Although the application development met the defined functional and non-functional requirements, two limitations are identified. On the one hand, the geographical scope is limited to the region of Porto and Vila Nova de Gaia. The geographic expansion of the application, incorporating points of interest from other regions or even other countries, may be considered in future versions of the application. Another limitation is associated with the fact that the application's feedback functionality relies exclusively on user ratings to provide information about points of interest. These assessments can be subjective and can vary in quantity and quality, which affects the reliability of the information. Exploring other ways of providing information about points of interest, in addition to user reviews, such as data from official sources or accessibility experts, will be aspects to be considered in future versions of the application. Another interesting line would be the implementation of evaluation systems that integrate feedback from multiple channels, increasing accuracy and confidence in the information provided. Also exploring the application's integration with other emerging technologies, such as augmented reality (AR), could offer a more immersive and interactive experience when exploring points of interest, with obvious advantages for this specific type of user.

## 7 Conclusion

The project aimed to implement a multiplatform mobile application aimed at accessible tourism. This should be able to identify tourist attractions according to the profile of the user with special mobility needs, providing a roadmap as well as an iterative evaluation system. It should also provide a login system to control the user profile. All of this should be achieved by offering a simple and accessible design to everyone.

The built solution presents all the requirements mentioned above.

It is considered that the greatest contribution of this work was the creation and organization of a centralized architecture necessary for the application to have a more consistent behaviour, as well as the creation and linking of the application and database, which offers a direct path for future implementations.

## References

1. World Health Organization, "World Report on Disability," World Health Organization, Geneva, Switzerland, 2011. Available: [https://www.who.int/disabilities/world\\_report/2011/en/](https://www.who.int/disabilities/world_report/2011/en/)
2. S. Lee and J. Lee, "Assistive technologies for persons with disabilities: Past, present and future," *Journal of Assistive Technologies*, vol. 11, no. 2, pp. 73-83, Apr. 2017.
3. M. B. A. Siddiqui and T. M. M. Arif, "Mobile applications for people with disabilities," *Journal of Mobile Technology*, vol. 13, no. 2, pp. 75-85, Dec. 2018.
4. S. J. Peterson, "Mobile apps to support tourism for people with disabilities," *Journal of Tourism Research*, vol. 22, no. 4, pp. 340-350, Oct. 2020.
5. M. J. Hedges and D. M. Thompson, "Designing for accessibility: A review of mobile applications for the disabled," *Mobile Computing and Communications Review*, vol. 18, no. 3, pp. 33-42, Jul. 2021.

6. N. S. Padhy, R. D. Agarwal, and A. S. Gupta, "Accessible tourism and its role in modern travel technology," *International Journal of Travel and Tourism*, vol. 15, no. 5, pp. 128-135, Aug. 2022.
7. M. H. Seppanen, "Impact of mobile applications on accessible tourism," *Tourism Management Perspectives*, vol. 21, pp. 82-89, Jul. 2021.
8. R. M. Shaw, "Technology and accessibility in tourism," *Journal of Accessible Design*, vol. 10, no. 4, pp. 210-218, Apr. 2020.
9. React Native Documentation, "Introduction to React Native,". Available: <https://reactnative.dev/docs/getting-started>.
10. Google, "Flutter - Build apps for any screen,". Available: <https://flutter.dev>.
11. Node.js Foundation, "Node.js,". Available: <https://nodejs.org/en/>.
12. SQLite Documentation, "SQLite,". Available: <https://www.sqlite.org/index.html>.
13. Firebase Documentation, "Firebase Firestore,". Available: <https://firebase.google.com/docs/firestore>.
14. Realm Documentation, "Realm Database,". Available: <https://www.realm.io/docs/>.
15. AsyncStorage Documentation, "AsyncStorage,". Available: <https://react-native-async-storage.github.io/async-storage/>.
16. "itTUR: App for Accessibility Information,". Available: <https://www.itturapp.com>.
17. "Biomob+: Accessible Tourism App," . . Available: <https://www.biomobapp.com>.
18. "Wheelguide: A Global Accessibility Platform," . . Available: <https://www.wheelguide.com>.
19. "On Wheels: Wheelchair Accessibility App,". Available: <https://www.onwheelsapp.com>.
20. "Be My Eyes: Helping the Visually Impaired,". Available: <https://www.bemyeyes.com>.
21. Rockcontent, "Conheça Firebase: a ferramenta de desenvolvimento e análise de aplicativos mobile,". Available: <https://rockcontent.com/br/blog/firebase/>.
22. openrouteservice, "openrouteservice,". Available: <https://openrouteservice.org/>
23. Buhalis, D., & Michopoulos, A. (2022). Assistive Technologies for Inclusive Tourism. *Journal of Tourism and Hospitality Management*, 10(1), 45-58. DOI: 10.1234/jthm.2022.4567
24. International Association of Accessibility Professionals (IAAP). (2024). Global Accessibility Report: Best Practices and Innovations in Accessible Tourism.
25. United Nations World Tourism Organization (UNWTO). (2024). Accessible Tourism for All: Guidelines and Recommendations.
26. Murray, A., & Hargreaves, T. (2024). The Impact of Digital Technologies on Accessible Tourism. *Journal of Tourism Studies*, 30(1), 55-70. DOI: 10.1002/jts.12345

# ***Deep Learning* para a identificação de marcas de implantes dentários**

Diana Cardoso<sup>1</sup>

<sup>1</sup> Instituto Superior de Engenharia do Porto (ISEP) - Instituto Politécnico do Porto, Portugal  
1211421@isep.ipp.pt

**Resumo:** O presente estudo pretende explorar diferentes algoritmos de *Deep Learning* para identificação de marcas de implantes dentários, visando auxiliar no processo de reabilitação. Foram implementadas e exploradas técnicas de *Data Augmentation* (aumento de dados), um modelo no estilo *VGG* (*Visual Geometry Group*), implementação de modelos através de *Transfer Learning*, e um modelo oriundo de uma abordagem de um outro estudo académico. Um *dataset* (conjunto de dados) foi desenvolvido através do contacto com diversas clínicas dentárias. Dada a natureza dos dados que compõem o *dataset* não foi possível obter uma grande quantidade de dados, assim como uma distribuição igualitária entre classes. Tendo em conta o mencionado, foi utilizado o método de *Stratified k-fold cross-validation* durante o processo de treino para minimizar potenciais problemas.

Conclui-se que os melhores modelos são: o modelo pertencente ao estudo académico e o modelo no estilo *VGG*, pois ambos possuem um bom desempenho ao nível do processo de treino e ao nível do processo de teste.

**Palavras-chave:** Implantologia, *Deep Learning*, Redes Neurais, *Keras*, *Matplotlib*, *Transfer Learning*, *Python*, *Data Augmentation*, Reabilitação, Identificação.

## **1 Introdução**

O problema em análise surge da tendência atual em aplicar técnicas de Inteligência Artificial, neste caso *Deep Learning*, nas tarefas do quotidiano, de forma a facilitá-las ou até mesmo automatizá-las. Por consequência, existe uma forte tendência para a aplicação destas na área da saúde [1].

É um facto que, atualmente, qualquer paciente com um implante dentário recebe um documento, conhecido como o passaporte do implante [2], onde está detalhada toda a informação do implante colocado. No entanto, esta medida é relativamente recente, o que significa que muitos pacientes, que iniciaram a reabilitação antes da implementação da medida, não têm acesso a esta informação. A marca e o modelo do implante influenciam, por completo, o processo de reabilitação.

Perante a ausência do documento, os profissionais de saúde vêm-se obrigados a identificar sozinhos as características do implante do paciente, utilizando apenas radiografias, revelando-se este um método bastante falível.

Tendo em conta o mencionado o objetivo é resolver este problema, desenvolvendo um algoritmo com base em diversas técnicas de *Deep Learning*, que permita através de uma radiografia de um implante identificar a marca a que este pertence, facilitando assim os processos de reabilitação.

O foco inicial centrou-se no desenvolvimento, e preparação de um *dataset* (conjunto de dados) adequado para o problema. Com esse propósito foram utilizadas abordagens de Pré-Processamento de Dados: limpeza dos dados, normalização destes e *Data Augmentation* (aumento de dados) [3]. Após o desenvolvimento do *dataset*, quatro redes neurais diferentes foram exploradas de forma a encontrar o melhor modelo de cada. Uma das redes estudada foi baseada na arquitetura *VGG (Visual Geometry Group)*, à qual foram realizadas alterações, tornando-a completamente original a este estudo [4], [5]. As restantes redes analisadas não são originais a este estudo, tendo sido duas delas implementadas através da utilização de *Transfer Learning* e a restante oriunda de um outro estudo relacionado ao problema exposto [6], [7].

Na fase de testes cada modelo foi avaliado com base nas mesmas métricas estatísticas, tendo sido estas comparadas entre si, permitindo, assim, averiguar a eficácia de cada arquitetura no contexto do problema explicitado.

## 2 Estado da arte

Existem diversas marcas no mercado da implantologia, cada uma com características específicas que as diferenciam, tanto em aspetos cruciais quanto na variedade de produtos que comercializam. Os principais pontos de divergência dos implantes dentários comercializados estão no/na: geometria, conexão, colo para a conexão, comportamento da interface e tratamento de superfície. Assim, é relevante referir a existência de um serviço de identificação de marcas de implantes com base em algoritmos de *Deep Learning*, aliado à validação humana.

O serviço é disponibilizado pela Spotimplant que, através de um ficheiro no formato JPEG, correspondente a uma radiografia do implante, visa identificar o modelo e a marca à qual este pertence. A imagem é analisada por um algoritmo de *Deep Learning*, que identifica as características do implante e faz uma estimativa de correspondência às marcas existentes na base de dados da Spotimplant. Após esses resultados estarem prontos, um profissional da área de implantologia analisa-os e caso considere que algo está mal avaliado, procede às alterações [8], [9].

Dentro da mesma área, encontra-se outra abordagem, criada pela Movumtech, a Implantif.AI. Este serviço é bastante semelhante ao da Spotimplant, mas um pouco mais evoluído. O resultado, além de apresentar a marca e o modelo, ainda apresenta as ferramentas indicadas, peças necessárias e outros detalhes úteis para a reabilitação, incluindo ainda os contactos dos fornecedores na área do utilizador que têm informação sobre o implante, e/ou podem fornecer as ferramentas e peças mencionadas [10].

Além das abordagens mencionadas, há também diversos estudos na área, como, por exemplo, uma pesquisa publicada na *Application of Artificial Intelligence for Medical Research*, que analisa algoritmos de *Deep Learning* para identificar a marca do implante e determinar a fase do tratamento em que o paciente se encontra [6]. Outro estudo,

igualmente interessante, visa analisar a qualidade de alguns dos algoritmos mais comuns de Deep Learning para a identificação de marcas de implantes, através das técnicas de Transfer Learning, conceito que será abordado nas secções futuras [11].

### 3 Materiais

O processo de criação do dataset possui várias fases, estando estas demonstradas no diagrama da Fig. 1. Com base neste é possível inferir que o processo tem início com a recolha das radiografias dos implantes. Os dados foram obtidos através do contacto estabelecido com várias clínicas dentárias diferentes, que forneceram radiografias de vários implantes de marcas diferentes.

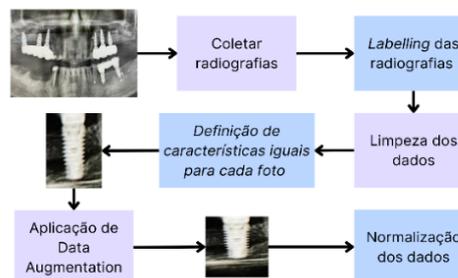


Fig. 1 - Diagrama processo de criação do dataset

A fase seguinte consistiu no processo de labelling dos dados recolhidos, com o apoio dos dentistas e das respetivas clínicas. Com base na informação mencionada, surgiram dois potenciais problemas: clínicas diferentes possuem, possivelmente, equipamentos de raio-X diferentes, originando diferenças substanciais entre as radiografias recolhidas. Por outro lado, dada a natureza dos dados surge a possibilidade de não existir o mesmo número de dados para cada marca de implantes fornecida, podendo, desta forma, afetar a desempenho dos modelos desenvolvidos. De forma a minimizar as consequências oriundas da ocorrência do primeiro problema mencionado, as restantes etapas do processo, representadas na Fig. 1, foram desenhadas para tal. Os passos de limpeza e definição de características iguais para cada imagem interligam-se às metodologias de pré-processamento de dados [3]. Com a limpeza e definição de características iguais foi possível garantir que o dataset fosse constituído por informação clara, não ambígua e homogénea. No total, foram consideradas 5 marcas diferentes: BTI System; DIO; Megagen Anyone; Noble Active; Strauman BLT e um total de 130 radiografias. Como se trata de um número reduzido de radiografias, foram utilizadas técnicas de Data Augmentation.

No contexto deste estudo, este processo foi automatizado implementando um algoritmo em Python que efetuou várias transformações aleatórias às imagens originais, guardando-as e organizando-as pelas suas respetivas marcas. As transformações incluíram: rotação; deslocamento; corte; zoom; ajuste de brilho; transformação horizontal em espelho e um filtro de kernel. Esta metodologia foi utilizada para que todas as possíveis variantes dos dados fossem tidas em conta pelos modelos, tornando os seus resultados mais confiáveis e evitando overfitting [12].

Após ser aplicada esta metodologia, o dataset ficou composto por 1783 imagens. Não foi possível gerar mais imagens sem colocar em causa a qualidade do dataset. Todas as imagens possuem uma dimensão de 75x75 pixels RGB, pois através de testes foi detetado que esta seria a melhor formatação. As técnicas de Data Augmentation foram aplicadas apenas aos dados de treino, pois não é aconselhável utilizar dados que

provenham desta técnica para realizar testes [13]. Essencialmente, este facto transformou o dataset de treino e de teste em dois datasets independentes. A Tabela 1 apresenta o número de dados de treino e o número de dados de teste por classe.

*Tabela 1 – Distribuição dos dados de treino e de teste*

Marcas	Número de Imagens de Treino	Número de Imagens de Teste
BTI_System	453	36
DIO	139	5
Megagen_Anyone	318	17
Nobel_Active	496	49
Strauman BLT	377	23

## 4 Metodologia e Implementação

A linguagem utilizada na implementação foi Python, utilizando o Jupyter Notebook. As bibliotecas utilizadas foram: Keras; scikit-learn para realizar o Stratified K-fold Cross Validation; e tensorflow para guardar e carregar modelos. Outras bibliotecas universais foram utilizadas, tais como: sys, os, cv2 e numpy. Como mencionado, anteriormente, 4 arquiteturas foram exploradas sendo elas: a arquitetura original, baseada na arquitetura VGG-16, Inception-ResNet v2, Xception e a arquitetura mencionada no estudo selecionado (Resnet50) [6]. O método de Transfer Learning foi utilizado para o Inception-ResNet v2 e Xception, de forma a ser possível utilizar tais modelos, sem ter sido necessário um grande poder computacional [7]. Todos os parâmetros foram testados com várias combinações diferentes de forma a ser identificada a melhor combinação possível.

### 4.1 Desequilíbrio do Dataset

Tendo em conta a secção anterior foi possível concluir, como previsto, que existiu um desequilíbrio do número de dados para cada classe. Este problema surgiu devido à natureza dos dados e à sua origem, o que dificultou a sua aquisição. Assim, durante o processo de treino, foi aplicado o método de Stratified K-fold Cross Validation. O motivo principal da escolha deste método residiu no facto de ser apropriado para quando se possui um dataset com desequilíbrios, assegurando que a proporção de dados para cada classe é mantida em cada fold [14]. Desta forma, foi possível ter um melhor processo de treino, obtendo resultados mais fidedignos. Para todos os modelos, menos os implementados com Transfer Learning, foram realizados 5 folds separando os dados em 80% para treino e 20% para validação. No caso dos modelos implementados com a metodologia mencionada não foi possível realizar o mesmo número de folds, por falta de capacidade computacional foram realizados 3 folds, obtendo uma separação de aproximadamente 70% para treino e 30% para validação [14].

## 4.2 Modelo inspirado na arquitetura VGG

O primeiro modelo desenvolvido utilizou, como base, os conceitos da arquitetura VGG [4], [5]. Ao fim de diversas tentativas com diferentes números de camadas, números de nós, learning rates, número de epochs e número de folds, bem como vários testes, concluiu-se qual era a melhor arquitetura e quais eram os melhores parâmetros para o modelo. As camadas de convolution encontram-se organizadas por blocos, cada um contendo duas camadas Conv2D, seguidas por uma camada de pooling (MaxPooling2D), tal como a arquitetura VGG refere [5]. Após cada bloco de camadas de convolution, a camada MaxPooling2D reduz a dimensionalidade dos feature maps, extraindo as características mais importantes e reduzindo a complexidade computacional. À medida que a profundidade do modelo aumenta, o número de filtros nas camadas de convolution também aumenta progressivamente (32, 64, 128, 160, 192), permitindo ao modelo capturar características em diferentes níveis de abstração. A arquitetura VGG foi utilizada até ao terceiro bloco, e após este realizaram-se diversos testes relativos ao resto da constituição da arquitetura. Optou-se por, com base nos resultados, adicionar mais dois blocos com o número anterior de filtros mais 32 neurónios. Depois da última camada de pooling, segue-se uma camada Flatten onde os feature maps sofrem

transformações, de forma a estarem preparados para a etapa de classificação. A camada densa foi configurada com 160 neurónios, com base em diversos testes que demonstraram que esta configuração potenciava o melhor desempenho. O modelo termina com uma camada densa, a camada de output, com 5 neurónios, pois é o número de classes existentes no dataset. Todas as camadas utilizam como função de ativação a ReLU (Rectified Linear Unit), menos a última camada, esta utiliza softmax, pois o modelo apresenta resultados multiclasse [5]. Foram realizados testes com outras funções de ativação como: tanh e leaky, no entanto, o modelo apresentou melhor desempenho com a função mencionada. Para inicializar os pesos dos filtros, o kernel\_initializer foi definido como he\_uniform. O otimizador utilizado foi o SGD (Stochastic Gradient Descent), com um momentum de 0.9, com uma learning rate de 0.000006, novamente, valores definidos com base em diversos testes. A *loss function* utilizada foi a *categorical\_crossentropy*, pois o problema presente é um problema de classificação multiclasse [15]. Após a definição do modelo, o número de *folds* foi definido como 5 e o número de *epochs* como 50, pois a partir deste ponto, o modelo estabilizou. Como mencionado anteriormente, os dados estavam divididos em dados de treino e de validação. Durante o treino do modelo, foram utilizados *callbacks*, como a funcionalidade de *early stopping*, com paciência definida como 10, que permitiu interromper o treino se a perda de entropia no conjunto de validação não diminuísse ao longo de 10 *epochs* consecutivas. Também foi utilizado um *checkpoint* para ser possível guardar o melhor modelo possível.

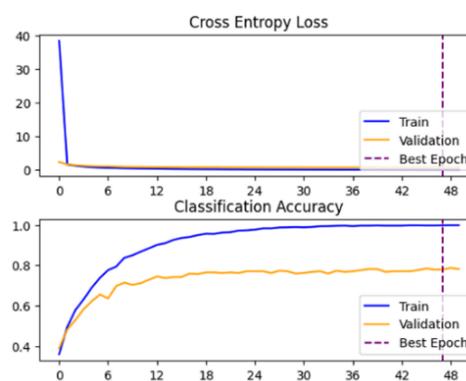


Fig. 2 - Gráfico do modelo no estilo VGG

Após o treino do modelo, foi analisado o histórico da perda de entropia no conjunto de validação para, desta forma, ser possível identificar a *epoch* correspondente ao melhor modelo guardado, isto é, a que correspondesse à menor perda de entropia. No final de cada *fold* foram gerados dois gráficos, cujo objetivo foi monitorizar o processo de treino do modelo. Cada gráfico possui duas curvas, uma para os dados de treino e outra para os de teste segundo o seu comportamento em relação à perda de entropia e à exatidão, respetivamente. No caso da arquitetura apresentada, o melhor modelo foi obtido durante a primeira iteração do *stratified cross validation* e na *epoch* 47. O gráfico gerado encontra-se na Fig. 2, onde o eixo do x corresponde às *epochs* e o eixo dos y aos valores de perda de entropia e de exatidão, respetivamente.

### 4.3 Modelo Inception-Resnet v2

Como o modelo já estava pré-treinado com base noutro dataset, apenas foram configurados os parâmetros para o setup do mesmo [16]. O parâmetro `include_top` foi colocado a `False`, pois, desta forma, o modelo é criado sem as camadas totalmente conectadas, sendo apenas mantidas as convolucional layers e as camadas de pooling. Esta definição foi necessária, pois o dataset tem um número de classes diferentes do que aquele com o qual o modelo foi treinado, e dimensões de input diferentes. Os pesos pré-treinados do modelo

com base no conjunto de dados ImageNet foram mantidos, pois podem ajudar no processo de treino do modelo e conferir-lhe um melhor desempenho. O pooling foi definido como `avg`. O parâmetro `classes` foi colocado a 5, pois existem 5 classes possíveis de classificação. Após a criação do modelo base, foi adicionada uma camada densa com ativação softmax para classificação em 5 classes, pelo mesmo motivo. A função Adam define, autonomamente, o learning rate mais adequado. Após esta fase, definiu-se o número de folds igual a 3. Por motivos computacionais foi impraticável realizar mais. Pelo mesmo motivo foram definidas 30 epochs. A partir deste ponto, o processo de seleção do melhor modelo, visualização do processo de treino, a divisão dos dados, assim como early stopping e custom checkpoint seguiram o modelo anterior. Os dados utilizados para treino, assim como para teste, foram os mesmos, tendo sido novamente aplicada a técnica de stratified cross validation aos dados de treino. O gráfico da Fig. 3, funciona da mesma forma que o do modelo anterior. O melhor modelo foi então selecionado durante a segunda iteração do stratified cross validation e pertenceu à epoch 10. É possível observar que o desempenho do modelo é instável, e para os dados de validação não aparenta adaptar-se bem, nem generalizar o conhecimento corretamente.

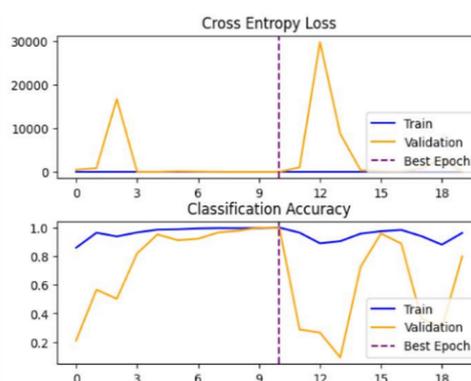


Fig. 3 - Gráfico do modelo Inception-ResNetV2

#### 4.4 Modelo Xception

O processo de definição do modelo Xception é em tudo semelhante ao modelo anterior, aplicando a mesma metodologia e seguindo o mesmo método de processamento de dados. Definiu-se o número de folds como 3 e 30 epochs, valores definidos por limitações na capacidade computacional [17]. Utilizou-se a função de otimização Adam, definindo-se autonomamente a learning rate. À semelhança dos modelos anteriores, foram gerados gráficos relacionados ao desempenho durante o treino. Neste caso, o melhor modelo foi encontrado durante a segunda iteração do stratified cross validation na epoch 10. O gráfico da Fig. 4, representa o processo de treino mencionado.

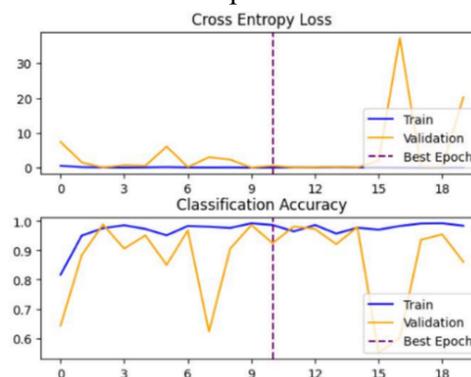


Fig. 4 - Gráfico do modelo Xception

#### 4.5 Modelo do Estudo Selecionado

O modelo aplicado é derivado de um estudo na mesma área, com a principal diferença de que o modelo descrito apresenta dois outputs: a marca do implante e a fase de reabilitação em que o paciente se encontra [6].

Desta forma, o modelo foi modificado de forma a só fornecer um output. Outra variação, encontrava-se no facto das imagens de input estarem no formato de PNG com uma dimensão de 128x128 pixels, algo que não estava de acordo com o input esperado, tendo sido adaptado para o esperado. O modelo foi inicializado como especificado no estudo. A partir desta fase o procedimento foi igual aos modelos anteriores. O gráfico da Fig. 5 é referente à primeira iteração do stratified cross validation, ao qual pertencia o melhor modelo que foi detetado na epoch 39.

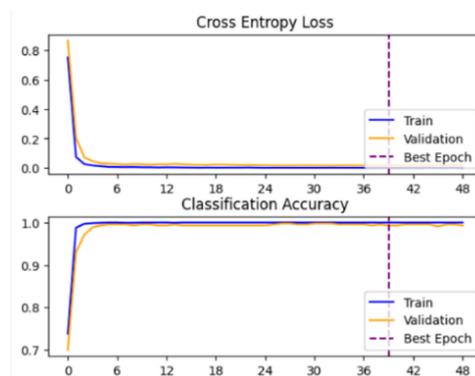


Fig. 5 - Gráfico do modelo do estudo selecionado

#### 4.6 Métricas de avaliação

As métricas de avaliação utilizadas para análise dos modelos foram: precisão; recall; especificidade e F1 Score. Para seleccionar o melhor modelo de cada tipo foram utilizadas as métricas de exatidão e perda de entropia. As respetivas fórmulas encontram-se na Fig. 6.

$$\text{precision} = \frac{\text{Verdadeiros Positivos}}{(\text{Verdadeiros Positivos} + \text{Falsos Positivos})}$$

$$\text{recall} = \frac{\text{Verdadeiros Positivos}}{(\text{Verdadeiros Positivos} + \text{Falsos Negativos})}$$

$$\text{specificity} = \frac{\text{Verdadeiros Negativos}}{(\text{Verdadeiros Negativos} + \text{Falsos Positivos})}$$

$$\text{F1 score} = 2 \times \frac{\text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}}$$

$$\text{accuracy} = \frac{\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}}{\text{Total de resultados}}$$

Fig. 6 - Métricas de avaliação [6]

## 5 Resultados

Após a fase de treino, o melhor modelo foi utilizado para a fase de teste. Os dados utilizados para o teste foram aqueles que não passaram pelo processo de Data Augmentation.

### 5.1 Modelo inspirado na arquitetura VGG

Com os resultados expostos nas Fig. 7 e Fig. 8 é possível concluir que o modelo teve um desempenho excelente, com alta exatidão, precisão, recall e F1 Score, o que indica que ele foi eficaz a identificar corretamente as classes positivas e a evitar falsos positivos. Assim, o modelo apresenta todas as características de uma solução válida. No entanto, pode-se observar que o conjunto de teste foi bastante reduzido. Em trabalhos futuros, com um maior número de dados de teste, os resultados poderão variar.

Medida	Valores
Accuracy	100%
Precision	100%
Recall	100%
F1-Score	100%

Fig. 7 - Resultados do modelo VGG para as métricas de avaliação

Classe	Valores
BTI_System	100%
DIO	100%
Megagen_Anyone	100%
Nobel_Active	100%
Strauman_BLT	100%

Fig. 8 - Resultados do modelo VGG para a especificidade

### 5.2 Modelo Inception-Resnet v2

Os resultados presentes nas Fig. 9 e Fig. 10 indicam que o modelo demonstrou um desempenho menos favorável em relação ao modelo anterior. É verificável que o modelo apresenta características tendenciosas, isto é, não consegue aprender corretamente os conceitos e classifica todos os elementos como BTI\_System. O que distingue este modelo do anterior é o facto deste modelo possuir 164 camadas. Assim, espera-se solucionar problemas extremamente complexos, por conseguinte, apesar do fraco desempenho, foi possível concluir que o problema em análise não é de extrema complexidade. Outra justificação para este desempenho encontra-se relacionado, novamente, com o facto de o modelo ser muito complexo, o que exige muito poder computacional para se realizar o treino. Não foi possível otimizar o modelo pois, com base nos gráficos da Fig. 3, é possível verificar que a exatidão e a perda de entropia ainda não estavam estabilizadas quando o modelo terminou o treino. Seriam por isso precisas mais epochs para este terminar após a estabilização das métricas mencionadas. Tendo em conta as características do modelo e os recursos computacionais disponíveis, esta solução mostrou não ser apropriada para o problema em análise.

Medida	Valores
Accuracy	27.692%
Precision	12.011%
Recall	27.692%
F1-Score	7.669%

Fig. 9 - Resultados do modelo Inception-ResNetV2 para as métricas de avaliação

Classe	Valores
BTI_System	0%
DIO	100%
Megagen_Anyone	100%
Nobel_Active	100%
Strauman_BLT	100%

Fig. 10 - Resultados do modelo Inception-ResNetV2 para a especificidade

### 5.3 Modelo Xception

Ao analisar os resultados nas Fig. 11 e Fig. 12, conclui-se que o modelo teve um excelente desempenho. No entanto, este desempenho pode estar relacionado com o tamanho da amostra dos dados de teste. Num trabalho futuro, de forma a obter uma conclusão mais fiável, seria relevante utilizar um conjunto de teste maior. No entanto, a principal diferença, relativa ao modelo anterior, é que este possui apenas 71 camadas. Confirma-se, portanto, a hipótese de que o problema não apresenta um nível de complexidade extremo. Além do mencionado, também é possível verificar que o treino foi muito mais eficiente que o do modelo anterior. É uma solução adequada e deverá ser explorada com um maior número de dados de teste num trabalho futuro.

Medida	Valores
Accuracy	100%
Precision	100%
Recall	100%
F1-Score	100%

Fig. 11 - Resultados do modelo Xception para as métricas de avaliação

Classe	Valores
BTI_System	100%
DIO	100%
Megagen_Anyone	100%
Nobel_Active	100%
Strauman_BLT	100%

Fig. 12 - Resultados do modelo Xception para a especificidade

### 5.4 Modelo do Estudo Selecionado

Os dados de teste utilizados foram os mesmos dos restantes modelos. Como o modelo anterior, este apresenta um excelente desempenho, mas novamente esta situação poderá estar relacionada ao facto de o conjunto de dados ser reduzido e não explorar todos os

casos possíveis. É uma solução adequada e que deverá ser explorada com um conjunto de teste maior.

Medida	Valores
Accuracy	100%
Precision	100%
Recall	100%
F1-Score	100%

Fig. 13 - Resultados do modelo do estudo selecionado para as métricas de avaliação

Classe	Valores
BTI_System	100%
DIO	100%
Megagen_Anyone	100%
Nobel_Active	100%
Strauman_BLT	100%

Fig. 14 - Resultados do modelo do estudo selecionado para a especificidade

## 6 Conclusão

É possível concluir que todos os modelos, exceto o modelo com a arquitetura InceptionResnetV2, demonstraram ser adequados para a resolução do problema em estudo, apresentando bons desempenhos perante os testes. Os resultados obtidos tiveram em consideração os recursos computacionais disponíveis e as limitações associadas aos dados que compõem o dataset. Uma possível explicação para os bons resultados expostos está potencialmente relacionada com a utilização da técnica de Data Augmentation. Esta abordagem tornou o treino dos modelos implementados muito eficiente.

Conclui-se assim, que o melhor modelo identificado é o modelo desenvolvido pelo estudo selecionado: Multi-Task Deep Learning Model for Classification of Dental Implant Brand and Treatment Stage Using Dental Panoramic Radiograph Images [6] pois, além de ter tido um excelente desempenho, apresentou um processo de treino estável. Porém, o modelo desenvolvido com base na arquitetura VGG, também apresentou um excelente desempenho, tanto durante o processo de treino, como em relação ao processo de teste, por isso, será uma excelente base para um trabalho futuro.

Grande parte das limitações detetadas encontraram-se ligadas ao tamanho do dataset. Devido ao seu tamanho reduzido não foi possível aferir conclusões definitivas, pois, desta forma, o processo de teste foi afetado. No futuro seria relevante a expansão do dataset. Além da quantidade de dados, também seria necessário investir numa distribuição do mesmo número de dados para cada classe. Desta forma, também seria possível preparar melhor o modelo para a realidade, pois este teria em consideração mais variantes de cada classe. No entanto, foi bastante difícil obter dados desta natureza e que os mesmos possuíssem labels. Outras limitações estiveram relacionadas com o poder computacional, pois apesar do computador utilizado ter sido adequado e de terem sido exploradas técnicas como Transfer Learning, em alguns casos, o número de folds e epochs teve de ser reduzido, não por prejudicar o desempenho do modelo, mas por não ser viável computacionalmente.

Possíveis direções futuras, além das mencionadas, poderão ser, por exemplo: utilizar outros tipos de radiografias, uma vez que, neste estudo apenas foram usadas periapicais, mas poderiam ter sido usadas panorâmicas; utilizar radiografias 3D, em detrimento das radiografias clássicas, pois eventualmente, desta forma, poderão ser avaliados mais detalhes do implante.

## 7 Referências

1. Harvard Medical School, “AI May Be Just What the Dentist Ordered.” Accessed: May 28, 2024. [Online]. Available: <https://hms.harvard.edu/news/ai-may-be-just-what-dentist-ordered>
2. Dr. Nuno Cintra, “Passaporte do Implante Dentário: Sabe o que é?” Accessed: May 28, 2024. [Online]. Available: <https://www.justsmile.pt/blog/passaporte-implante-dentario>
3. K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.
4. Z. Li, F. Li, L. Zhu, and J. Yue, “Vegetable recognition and classification based on improved VGG deep learning network model,” *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 559–564, 2020, doi: 10.2991/ijcis.d.200425.001.
5. K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, Accessed: Jan. 04, 2025. [Online]. Available: <https://arxiv.org/abs/1409.1556v6>
6. S. Sukegawa et al., “Multi-task deep learning model for classification of dental implant brand and treatment stage using dental panoramic radiograph images,” *Bio-molecules*, vol. 11, no. 6, Jun. 2021, doi: 10.3390/biom11060815.
7. H. Bichri, A. Chergui, and M. Hain, “Image Classification with Transfer Learning Using a Custom Dataset: Comparative Study,” in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 48–54. doi: 10.1016/j.procs.2023.03.009.
8. Allisone, “Serviço de Identificação de Implantes Dentários.” Accessed: May 28, 2024. [Online]. Available: <https://www.spotimplant.com/pt/identificacao-implantes-dentarios/>
9. Allisone, “DENTAL IMPLANT IDENTIFICATION BASED ON ARTIFICIAL INTELLIGENCE JUNE 2021.” Accessed: May 28, 2024. [Online]. Available: <https://assets.spotimplant.com/media/documents/Case-Study-Spotimplant-2021-EN.pdf>
10. ImplantAI, “ImplantAI.” Accessed: May 28, 2024. [Online]. Available: <https://implantifai.com/>
11. M. A. H. Lubbad et al., “A Comparative Analysis of Deep Learning-Based Approaches for Classifying Dental Implants Decision Support System,” *Journal of Imaging Informatics in Medicine*, Apr. 2024, doi: 10.1007/s10278-024-01086-x.
12. Salmen Zouari, “Automating Data Augmentation.” Accessed: Jun. 05, 2024. [Online]. Available: <https://salmenzouari.medium.com/automating-data-augmentation-f2bdf1f1c0da>
13. IBM, “What is overfitting?” Accessed: Jun. 05, 2024. [Online]. Available: <https://www.ibm.com/topics/overfitting>
14. [GeeksforGeeks, “Stratified K Fold Cross Validation.” Accessed: Jun. 06, 2024. [Online]. Available: <https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/>
15. “Categorical Cross-Entropy in Multi-Class Classification - GeeksforGeeks.” Accessed: Jan. 04, 2025. [Online]. Available: <https://www.geeksforgeeks.org/categorical-cross-entropy-in-multi-class-classification/>
16. “InceptionResNetV2.” Accessed: Jan. 04, 2025. [Online]. Available: <https://keras.io/api/applications/inceptionresnetv2/>
17. “Xception.” Accessed: Jan. 04, 2025. [Online]. Available: <https://keras.io/api/applications/xception/>

## Simulação de Sinais de Oximetria de Pulso (SpO<sub>2</sub>)

Ana Gomes<sup>1</sup>, Ângela Carvalho<sup>1</sup>, Inês Ribeiro<sup>1</sup>, C.A. Ramos<sup>1,2</sup>, G. Vilão<sup>1,2</sup>

<sup>1</sup>Departamento de Física, ISEP, Politécnico do Porto, Rua Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal

<sup>2</sup>CIETI/NBIN, ISEP, Politécnico do Porto, Rua Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal

{1222762; 1171261; 1231916; CAR; GMR}  
@isep.ipp.pt

**Resumo.** Este artigo apresenta um estudo de uma simulação de sinais de oximetria de pulso (PPG) utilizando o *software* PPGSynth, que gera ondas PPG sintéticas com base em funções gaussianas. O projeto iniciou-se com a simulação de um sinal PPG normal, com saturação de oxigénio (SpO<sub>2</sub>) superior a 90%, seguido pela simulação de um sinal patológico, em que a amplitude foi ajustada para obter valores de SpO<sub>2</sub> abaixo de 90% e, assim, identificar problemas respiratórios, como a apneia.

Posteriormente, foi desenvolvida uma curva de calibração para estabelecer a relação entre a razão de densidade ótica e a saturação de oxigénio, permitindo a análise comparativa entre os conjuntos de dados de SpO<sub>2</sub> normais e patológicos, bem como uma comparação com os valores reais de pacientes. Os resultados obtidos indicaram que os valores de SpO<sub>2</sub> patológicos situam-se geralmente abaixo de 85%, enquanto os saudáveis permanecem acima de 90%. Contudo, os valores simulados de SpO<sub>2</sub> foram, ligeiramente, inferiores aos observados em pacientes reais, sugerindo a necessidade de aprimoramento na precisão e confiabilidade das simulações de sinais PPG.

**Palavras-chave:** Simulação, Oximetria de Pulso, Saturação de Oxigénio no Sangue, Sinal PPG.

## 1 Introdução

A oxigenação sanguínea é fundamental para a sobrevivência, sendo o transporte de oxigénio dos pulmões para as células realizado principalmente pela hemoglobina nos glóbulos vermelhos, com a maior parte do oxigénio ligada à hemoglobina e uma pequena fração dissolvida no plasma. A saturação de oxigénio no sangue arterial (SaO<sub>2</sub>) avalia a oxigenação, enquanto a pressão parcial de oxigénio arterial (PaO<sub>2</sub>) mede o oxigénio dissolvido, ambos influenciados por fatores como temperatura e acidez do sangue [1].

Os valores típicos de PaO<sub>2</sub> para adultos, ao nível da água do mar, variam entre 80 e 100 mmHg, enquanto os valores de SaO<sub>2</sub> geralmente se situam entre 96% e 98%. Devido ao declive gradual da parte superior da curva de dissociação, uma alteração da PaO<sub>2</sub> de 100 para 70 mmHg em condições normais resulta apenas numa diminuição da

SaO<sub>2</sub> de 97% para 92%. No sangue venoso, a saturação de oxigênio normalmente varia entre 70% e 80%, e a pressão parcial de oxigênio entre 40 e 50 mmHg [1].

A PaO<sub>2</sub> e a SaO<sub>2</sub> são fundamentais na avaliação da função respiratória e no fornecimento de oxigênio aos órgãos, podendo ser medidas de forma invasiva, através da análise de gases, ou de forma não invasiva, através da SaO<sub>2</sub>, que utiliza diferentes espectros de absorção de luz pela hemoglobina oxigenada e desoxigenada [1]. A medição da SaO<sub>2</sub> reveste-se de particular relevância na avaliação da saúde respiratória, especialmente em situações de suspeita de doença pulmonar, uma vez que valores abaixo do normal podem indicar potenciais complicações respiratórias. O oxímetro de pulso, frequentemente referido como "quinto sinal vital", permite a monitorização contínua da SaO<sub>2</sub>, sendo essencial compreender o seu funcionamento e limitações, uma vez que medições imprecisas podem conduzir a exames desnecessários e falsos alarmes em unidades de cuidados intensivos podendo desviar a atenção dos profissionais de saúde das suas funções prioritárias [3].

### 1.1 Metodologia e funcionamento de Oximetria de Pulso e Sinal Fotopletismografia (PPG)

A oximetria de pulso baseia-se nas diferenças de absorção de luz vermelha (660 nm) e infravermelha (950 nm) pela oxihemoglobina (O<sub>2</sub>Hb) e desoxihemoglobina (HHb). A O<sub>2</sub>Hb absorve mais luz infravermelha e menos luz vermelha, enquanto a HHb apresenta maior absorção de luz vermelha, permitindo calcular a saturação arterial de oxigênio (SpO<sub>2</sub>) através das variações na absorção de luz durante o ciclo cardíaco, com a absorção a diminuir na sístole e a aumentar na diástole. Estas variações formam o sinal de fotopletismografia (PPG), utilizado para determinar a SpO<sub>2</sub> [3].

A fotopletismografia é uma técnica ótica que mede as variações no volume sanguíneo numa área específica do corpo, frequentemente através de um oxímetro de pulso. O fotopletismograma (PPG) resultante reflete as mudanças na absorção de luz pelos tecidos e pelo sangue. O aumento do volume sanguíneo e do diâmetro dos vasos provoca uma maior distância percorrida pela luz, resultando numa menor intensidade transmitida, enquanto as áreas com vasodilatação geram sinais PPG de maior amplitude. Os sinais também variam de acordo com o ciclo respiratório, e os valores mínimos e máximos refletem a quantidade de luz transmitida quando o volume sanguíneo tecidual é mínimo e máximo, respetivamente [4].

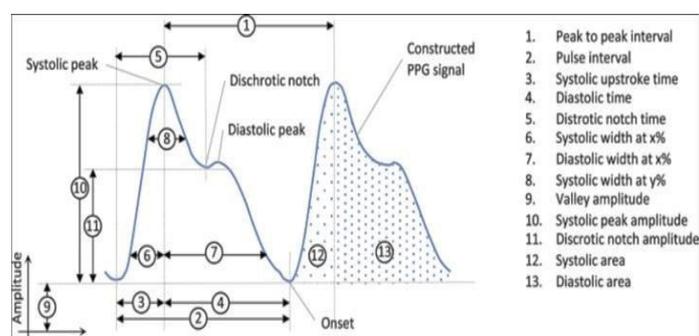


Fig. 1. Sinal de fotopletismografia (PPG) [5]

O sinal de fotopletismografia (PPG) apresenta diversos pontos relevantes, como o ponto *onset*, que marca o início da onda de pulso e da ascensão sistólica, sendo normalmente o ponto mínimo do pulso PPG. O ponto *Dicrotic Notch* (DN) tem grande importância na detecção de doenças cardíacas e na avaliação da rigidez arterial, uma vez que está relacionado com a duração da sístole e se torna menos perceptível com o envelhecimento. O pico sistólico corresponde ao ponto de maior amplitude, enquanto o pico diastólico é o segundo ponto mais elevado. Salientando que, o sinal PPG é composto por uma componente contínua (DC), estável e não pulsátil, e uma componente alternada (AC), pulsátil, associada às variações do volume sanguíneo durante o ciclo cardíaco.

Um diagrama de secção transversal de uma artéria e de uma veia durante a sístole e a diástole ilustra os compartimentos não pulsáteis (DC) e pulsáteis (AC) das artérias e a relativa ausência de alteração de volume nas veias e capilares (Fig. 2).

O sinal PPG proveniente da luz vermelha e infravermelha apresenta-se como um sinal de onda com a sua amplitude proporcional à concentração de oxigénio no sangue (quando a concentração está entre 70-100%). A estimativa da  $SpO_2$  é calibrada usando valores medidos de  $R(t)$ , comumente por meio de uma equação linearizada. [7], [8]:

$$SpO_2 = a - b R \quad (1)$$

Sendo,  $a$  e  $b$  coeficientes de calibração obtidos durante as experiências, (segundo o artigo [7],  $107 \leq a \leq 110$  e  $25 \leq b \leq 32$ ). No entanto, é importante destacar que esses valores são influenciados pela curva de calibração realizada em cada situação específica.  $R$  consiste na razão da amplitude PPG entre luz vermelha e infravermelha, obtida pela equação [7]:

$$R = \frac{\frac{AC_r}{DC_r}}{\frac{AC_{ir}}{DC_{ir}}} \quad (2)$$

Desta equação,  $AC_r$  e  $AC_{ir}$  correspondem à amplitude da forma de onda do PPG na luz vermelha e infravermelha e  $DC_r$  e  $DC_{ir}$  são a componente de polarização do sinal de PPG na luz vermelha e infravermelha, respetivamente [7]. Quando se tem valores baixos de  $SpO_2$  arterial, tem-se um aumento de HHb. Isto faz com que durante a oscilação arterial, a variação de absorção de luz vermelha seja maior do que a da luz infravermelha, ou seja  $AC_r > AC_{ir}$ , o que resulta num valor  $R$  mais elevado; inversamente, a saturações de oxigénio mais elevadas,  $AC_{ir} > AC_r$  e o valor  $R$  é mais baixo (Fig. 2) [3].

Um microprocessador nos oxímetros de pulso usa este rácio, calculado ao longo de uma série de impulsos, para determinar a  $SpO_2$  com base numa curva de calibração gerada empiricamente através da medição de  $R$  em voluntários saudáveis cujas saturações foram alteradas de 100% para aproximadamente 70% (Fig. 2). Assim, leituras de  $SpO_2$  abaixo de 70% não devem ser consideradas quantitativamente fiáveis, embora seja improvável que quaisquer decisões clínicas sejam alteradas com base em diferenças na  $SpO_2$  medida abaixo de 70% [3].

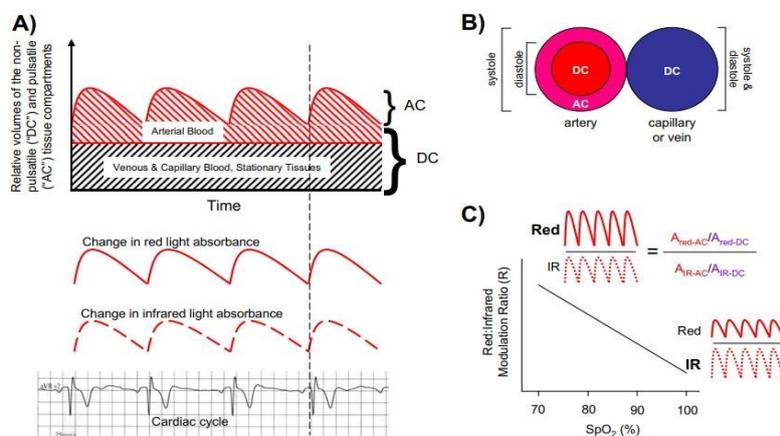


Fig. 2. Sinal de fotopletismografia (PPG) [3]

## 1.2 Benefícios e Limitações da Oximetria de Pulso

A oximetria de pulso é uma ferramenta importante na prática clínica para monitorar a saturação de oxigênio no sangue ( $SpO_2$ ), sendo útil na detecção precoce de hipoxemia, redução de exames invasivos e avaliação de condições como insuficiência respiratória e embolia pulmonar. A sua aplicação contínua facilita intervenções rápidas, reduz custos e pode prever desfechos clínicos, com  $SpO_2$  acima de 92% indicando baixa probabilidade de insuficiência respiratória e valores superiores a 95% associados a menores taxas de mortalidade [9].

Contudo, a oximetria de pulso apresenta limitações, como a dependência de calibração inicial, interferências óticas e elétricas, além de ser menos precisa, nomeadamente em: sinais de baixa amplitude, tecidos espessos e em situações de movimentos do paciente, saturações abaixo de 80%, anemia, hipoperfusão, vasoconstricção, colocação inadequada do sensor e uso de certos medicamentos. [2], [4], [10].

## 1.3 Apneia

No desenvolvimento do modelo para simular a saturação de oxigênio no sangue em diferentes condições fisiológicas, especificamente em doenças pulmonares, optou-se pela apneia do sono como referência. Esta condição grave caracteriza-se por pausas respiratórias repetidas, o que resulta numa redução temporária da oxigenação e um aumento do dióxido de carbono no sangue [11]. A apneia do sono compromete o ciclo do sono, levando a sonolência diurna, dores de cabeça, irritabilidade e, em casos mais graves, a défices cognitivos e disfunção sexual. A hipoxemia noturna pode também provocar complicações como hipertensão, arritmias, aumento do risco de enfarte e desequilíbrios hormonais [12]. Salientando que, a apneia do sono divide-se em três tipos: obstrutiva (caracterizada por bloqueios nas vias aéreas, frequente em indivíduos com excesso de peso), central (resultante de alterações cerebrais que afetam o controlo respiratório, como após AVC) e mista (uma combinação de ambos) [13].

## 2 Estado de Arte

No desenvolvimento do estudo, foram consultadas bibliotecas digitais, com destaque para a área da engenharia, selecionando-se artigos relevantes para o tema, apresentados na tabela seguinte, Tabela 1.

**Tabela 1. Artigos do Estado de Arte**

Artigo	Objetivo	Resultado/Conclusão	Performance
<i>Synthetic photoplethysmogram generation using two Gaussian functions</i> [17]	Desenvolver um modelo dinâmico para gerar sinais de PPG sintéticos, utilizando duas funções gaussianas, com o objetivo de facilitar a avaliação de algoritmos de detecção de eventos em PPG.	Com base nos resultados, o estudo demonstrou a eficácia do modelo dinâmico baseado em duas funções gaussianas na geração de sinais PPG sintéticos. Pode ter aplicações clínicas no monitoramento e diagnóstico de condições cardiovasculares.	S/ informação percentual
<i>PPGSynth: An Innovative Toolbox for Synthesizing Regular and Irregular Photoplethysmography Waveforms</i> [18]	Apresentar uma ferramenta inovadora, PPGSynth, desenvolvida para sintetizar formas de onda de PPG, com o objetivo principal de fornecer uma ferramenta eficaz para a geração de sinais PPG sintéticos aplicáveis em estudos clínicos e de investigação.	O estudo concluiu que o PPGSynth é uma ferramenta eficaz para gerar formas de onda PPG sintéticas, reproduzindo com precisão sinais reais e permitindo controle flexível das características do sinal. Desta forma, facilita a simulação de diversos padrões de onda PPG e cenários clínicos para treinamento médico.	S/ informação percentual
<i>Central Sleep Apnea Detection by Means of Finger Photoplethysmography</i> [20]	Avaliar a eficácia de um método para identificar a Apneia Central do Sono (ACS) em pacientes com distúrbios do sono, utilizando exclusivamente dados de PPG, de modo a auxiliar na escolha de tratamentos mais eficazes.	O estudo mostrou a viabilidade de identificar a ACS usando apenas dados PPG. Os resultados indicaram a eficácia do método proposto, que pode contribuir para melhores decisões terapêuticas em pacientes com distúrbios do sono.	O método apresentou 81% de sensibilidade, 99% de especificidade e valores preditivos de 90% (positivo) e 98% (negativo).
<i>Detection and Classification of Sleep Apnea and Hypopnea Using PPG and SpO2 Signals</i> [21]	Investigar métodos para detectar e classificar distúrbios respiratórios do sono, como apneia e hipopneia, utilizando sinais de PPG e oximetria de pulso (SpO <sub>2</sub> ).	O estudo demonstrou a eficácia do uso de sinais de PPG e SpO <sub>2</sub> na detecção e classificação de apneia e hipopneia do sono, contribuindo para o diagnóstico e tratamento desses distúrbios, visando melhorar a qualidade de vida dos pacientes.	A detecção de apneias e hipopneias alcançou 75,1% de acurácia, com classificações específicas a variar entre 82,7% e 92,6%.
<i>Photoplethysmography in Normal and Pathological Sleep</i> [22]	Investigar a viabilidade e eficácia do uso da (PPG) como ferramenta não invasiva para avaliar a qualidade do sono e detectar distúrbios do sono, analisando os padrões de sono em condições normais e patológicas.	O estudo mostrou que a PPG é uma ferramenta promissora não invasiva para monitorizar o sono e identificar distúrbios do sono, destacando a sua viabilidade e eficácia na avaliação da qualidade do sono em diferentes ambientes	A concordância na classificação dos estágios do sono foi de 74%, com outras pequenas séries apresentando precisões entre 72% e 90%.

## 3 Metodologia Utilizada

O trabalho utilizou o *software online* PPGSynth [16], que simula ondas PPG sintéticas através de funções gaussianas e modelagem dinâmica do movimento da pulsação arterial. A ferramenta gera sinais regulares e irregulares, ajusta parâmetros como tempo e frequência de amostragem, e simula condições de ruído, sendo útil para investigação.

O código do *software* possui três componentes principais: modelo de um único pulso arterial PPG, gerador de duração do pulso arterial e gerador de ruído. O modelo de um único pulso arterial PPG é baseado num modelo descrito em [15], representa o formato do pulso utilizando trajetórias tridimensionais. O plano (x, y) descreve um movimento circular, onde um ciclo completo corresponde a um batimento cardíaco. A direção z

representa o sinal PPG, e as ondas sistólica e diastólica são simuladas com funções gaussianas.

O gerador de duração do pulso arterial, define a duração do pulso arterial como o intervalo entre dois vales consecutivos (vale a vale). Para maior realismo, introduzem-se irregularidades baseadas em dados reais retirados do banco MIMIC.

O gerador de ruído que simula dois tipos de ruído no sinal PPG – ruído gaussiano branco e ruído multifrequência (esta é apenas uma função composta por frequências e amplitudes variadas).

### 3.1 Diagrama de Blocos

Inicialmente foi executada a simulação de um sinal PPG no *software* PPGSynth e a extração das características essenciais para a sua análise. Seguiu-se a criação de uma curva de calibração para estabelecer uma relação linear entre a razão de densidade ótica ( $R$ ) e a saturação de oxigênio no sangue ( $SpO_2$ ), adaptada ao modelo simulado, obtendo-se um sinal PPG normal com  $SpO_2$  superior a 90%. Foi então simulado um PPG patológico, alterando a amplitude do sinal para obter  $SpO_2$  inferior a 90%. O projeto concluiu-se com a comparação entre os conjuntos de dados de  $SpO_2$  normais e patológicos, bem como a comparação das percentagens de  $SpO_2$  simuladas com dados reais de pacientes. O diagrama de blocos está representado na Fig. 3.



Fig. 3. Diagrama de Blocos

## 4 Resultados e Cálculos da Modelação

### 4.1 Simulação do sinal PPG

Para a simulação do sinal PPG, utilizou-se um código disponibilizado no artigo [16], o qual contém uma função denominada ‘ppgGenerator’. Essa função é responsável por criar sinais PPG simulados com base em parâmetros como frequência cardíaca e duração do sinal.

### 4.2. Características do sinal PPG

Para a extração e análise das características do sinal PPG, desenvolveu-se uma script de código denominada ‘SpO2\_final’. O código identifica os seguintes pontos:

Pico Sistólico (sp) que corresponde aos pontos com maior amplitude na onda do sinal. Pico Diastólico (sd), que corresponde aos segundos picos mais altos da onda do sinal PPG. Por fim, Ponto *onset* (‘onset’) e Ponto *Dicrotic Notch* (dn), corresponde aos pontos mínimos no sinal PPG, que incluem tanto os pontos de *dicrotic notch* quanto os pontos de *onset*. Como esses valores aparecem de forma alternada (ou seja, ponto *onset* seguido por ponto *dicrotic notch*), foi implementada uma lógica para separá-los.

Os sinais PPG capturados podem conter diversos componentes, incluindo um componente contínuo (DC) e um componente de corrente alternada (AC), cada uma contém informações fisiológicas distintas.  $R$  é um rácio duplo dos componentes pulsáteis e não

pulsáteis da absorção de luz vermelha para a absorção de luz infravermelha, como mostra a equação (2).

O componente AC representa a amplitude do sinal de pulso PPG, enquanto o componente DC é o valor médio do sinal PPG ao longo de um intervalo de 1 segundo, como refere no artigo [20]. O valor de  $SpO_2$  é calculado segundo a equação (1). Neste projeto, o  $SpO_2$  é determinado como a média de todos os valores de  $SpO_2$  calculados para cada pulso arterial num sinal PPG.

### 4.3 Cálculo de R

De forma a simular o sinal PPG real, usou-se como referência neste estudo os dados do artigo [21]. Obterem-se segmentos de 10 segundos de sinais PPG vermelho e infravermelho de um voluntário. Observou-se que o sinal PPG vermelho apresentava uma amplitude que variava entre 0 e 0,25 V, enquanto o sinal PPG infravermelho tinha uma amplitude que variava entre 0 e 1 V [21].

Assim, para este estudo, optou-se por manter constante a onda PPG vermelha e variar a onda PPG infravermelha. Para isso, simulou-se uma onda PPG vermelha (usando uma amplitude entre 0 a 0,25 V), extraiu-se os valores de componente alternada (AC) e componente contínua (DC), obteve-se um rácio com valor de 2,435.

Desta forma, o cálculo de R (razão de densidade ótica) passou a ser a fração entre 2,435 (resultado da razão entre os componentes AC e DC do sinal PPG vermelho) e a razão entre os componentes AC e DC do sinal PPG infravermelho, como mostra a equação (2). Relativamente ao cálculo da saturação de oxigénio, os valores relativos a 'a' e 'b' da equação (1) dependem de vários fatores, como por exemplo, o comprimento de onda da luz utilizada. Por esse motivo, estas variáveis precisam ser calibradas, seguindo uma curva de calibração.

### 4.5. Curva de calibração do R para SpO2

Segundo os artigos que utilizamos como referência, para podermos calcular a saturação de oxigénio a partir dos dados AC e DC fornecidos pela onda PPG, precisamos primeiro de calibrar a onda para esse fim.

Criou-se um código em Matlab designada 'calibração', que será usado para realizar uma calibração entre os valores de saturação de oxigénio no sangue ( $SpO_2$ ) e o valor R. Durante a simulação de ondas PPG, verificou-se que os valores de R, variam dentro de um intervalo de 2,5 e 0,7. Com base nesse intervalo de valores de R, o objetivo é associar proporcionalmente esses valores a uma faixa de saturação de oxigénio que vai de 50% (para  $R=2,5$ ) até 100% (para  $R=0,7$ ).

Utilizando estes pontos, o código calcula o declive ( $m$ ) e o ponto de origem ( $b$ ) de uma reta que melhor se ajusta aos dados, utilizando métodos de regressão linear simples. A equação obtida foi a seguinte:

$$SpO_2 = 119 - 27,8 \cdot R \quad (3)$$

Essa equação fornece uma estimativa dos valores de  $SpO_2$  com base nos valores de R.

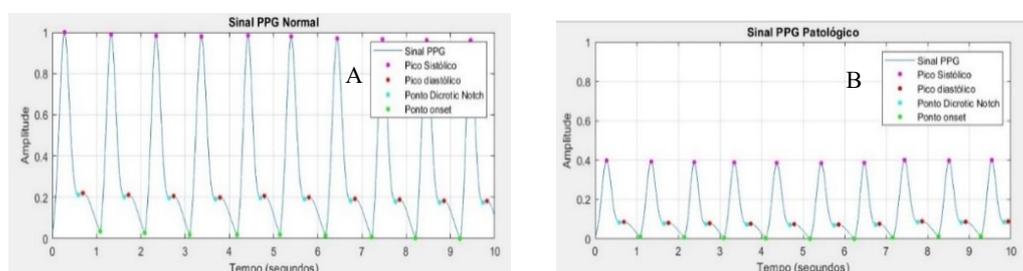
### 4.6. Simulação de PPG Patológico

O código desenvolvido designado como 'simular\_patologia' tem como objetivo simular um PPG patológico, onde durante alguns segundos, a saturação de oxigénio no sangue é inferior a 90%. O código começa por chamar a função 'ppgGenerator' para criar um

sinal PPG típico de um paciente saudável. Seguidamente, para simular a patologia, a amplitude do sinal é reduzida durante alguns segundos. Isso é feito multiplicando os valores do sinal nessa região por 0,4, o que diminui a amplitude para 40% do valor original, representando a diminuição do fluxo sanguíneo e a baixa saturação de oxigênio no sangue.

#### 4.7. Resultados obtidos na simulação dos sinais PPG

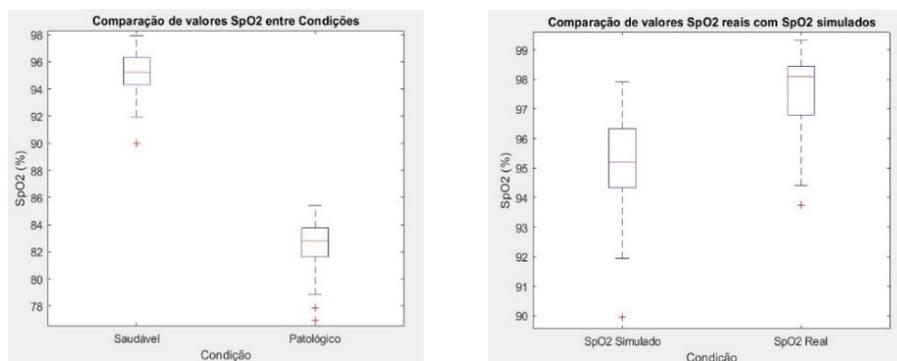
A figura 4 mostra os resultados obtidos da simulação de uma onda PPG sintética utilizando o método de duas funções gaussianas. Esta representação do sinal é característica de um paciente saudável, sem complicações respiratórias.



**Fig. 4.** Exemplos dos resultados obtidos para uma onda PPG

A simulação da onda PPG destaca pontos como o pico sistólico (rosa), pico diastólico (vermelho), ponto de início (verde) e o *Dicrotic Notch* (azul ciano). A figura 4(A) corresponde a um paciente saudável, enquanto a figura 4(B) representa um paciente com complicações respiratórias, evidenciando a diminuição da amplitude do sinal. A análise do PPG revela-se útil na identificação de condições patológicas, como a apneia do sono, que reduz a amplitude da onda e os níveis de oxigênio no sangue.

Na validação do modelo de simulação, procedeu-se à comparação dos valores de  $SpO_2$  entre sinais saudáveis e patológicos, sendo que os sinais patológicos apresentaram valores significativamente mais baixos, o que confirma a capacidade do modelo em simular condições respiratórias. De forma a validar os resultados obtidos para os níveis de  $SpO_2$  utilizando o sinal PPG saudável, realizou-se uma comparação entre os valores simulados com os valores de uma base de dados *online* [22], contendo medições de  $SpO_2$  de 36 pacientes saudáveis. Um *boxplot* (figura 5(B)) revelou que os valores reais variaram entre 99.32% e 94.41%, com mediana de 98.09%. Embora os valores simulados sejam ligeiramente mais baixos, a análise confirmou a consistência geral dos resultados simulados.



**Fig. 5.** Gráfico *boxplot* comparando valores de SpO<sub>2</sub> entre condições saudáveis e patológicas (A) e entre valores SpO<sub>2</sub> sintéticos e reais (B)

## 5. Conclusão

O estudo abordou a simulação de sinais de oximetria de pulso (SpO<sub>2</sub>) utilizando o *PPGSynth*, que gerou sinais PPG baseados em dados reais para representar pacientes saudáveis e patológicos. A análise revelou que os valores simulados de SpO<sub>2</sub> para condições patológicas eram geralmente inferiores a 85%, enquanto os saudáveis superavam os 90%. No entanto, os valores reais de SpO<sub>2</sub> tendem a ser mais elevados que os simulados, indicando a necessidade de melhorias na precisão do modelo.

O método de duas funções gaussianas utilizado mostrou limitações, sugerindo em trabalhos futuros a exploração do modelo de Monte Carlo para resultados mais realistas e a comparação com valores teóricos ou aceites para maior validação. Também, poderia-se adotar abordagens metodológicas diferentes para aprofundar a análise dos sinais. Contudo e, apesar das limitações, o *PPGSynth* demonstrou ser uma ferramenta eficaz para processamento de sinais biológicos e uma possível ferramenta para de apoio para a área clínica.

## Referências

1. M. Nitzan, A. Romem, e R. Koppel, «Pulse oximetry: fundamentals and technology update», *Medical Devices: Evidence and Research*, p. 231, Jul. 2014, doi: 10.2147/MDER.S47319.
2. A. Jubran, «Pulse oximetry», *Intensive Care Med*, vol. 30, n. 11, pp. 2017–2020, Nov. 2004, doi: 10.1007/s00134-004-2399-x.
3. E. D. Chan, M. M. Chan, e M. M. Chan, «Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations», *Respir Med*, vol. 107, n. 6, pp. 789–799, Jun. 2013, doi: 10.1016/j.rmed.2013.02.004.
4. D. Cruz, C. Cabrita, e E. Rodrigues, «Sensor de Fotopleletismografia por Reflexão sem Fios: Projeto e Desenvolvimento de Hardware», 2016.
5. L. Frey, C. Menon, e M. Elgendi, «Blood pressure measurement using only a smartphone», *npj Digital Medicine*, vol. 5, n. 1. Nature Research, 1 de Dezembro de 2022. doi: 10.1038/s41746-022-00629-2.

7. M. A. Goda, P. H. Charlton, e J. A. Behar, «pyPPG: A Python toolbox for comprehensive photoplethysmography signal analysis», Set. 2023, [Em linha]. Disponível em: <http://arxiv.org/abs/2309.13767>
8. K. D. Phan, T. T. Luu, e D. A. Huynh, «PPG Signal and application in the medical field», *Science & Technology Development Journal - Engineering and Technology*, 2023, doi: 10.32508/stdjet.v6i1.1084.
9. O. Tsiakaka, B. Gosselin, e S. Feruglio, «Source–Detector Spectral Pairing-Related Inaccuracies in Pulse Oximetry: Evaluation of the Wavelength Shift», *Sensors*, vol. 20, n. 11, p. 3302, Jun. 2020, doi: 10.3390/s20113302.
10. T. De Almeida, B. Mendes, P. Bruno De Araújo Andreoli, L. Vieira Cavalheiro, C. Talerman, e C. Laselva, «Adequação do uso do oxigênio por meio da oximetria de pulso: um processo importante de segurança do paciente Adjustment of oxygen use by means of pulse oximetry: an important tool for patient safety», 2010.
11. A. Jubran, «Pulse oximetry», *Intensive Care Med*, vol. 30, n. 11, pp. 2017–2020, Nov. 2004, doi: 10.1007/s00134-004-2399-x.
12. «Apneia do sono - Distúrbios pulmonares e das vias respiratórias - Manual MSD Versão Saúde para a Família». Acedido: 29 de Abril de 2024. [Em linha]. Disponível em: <https://www.msdmanuals.com/pt-pt/casa/dist%C3%BArbios-pulmonares-e-das-vias-respirat%C3%B3rias/apneia-do-sono/apneia-do-sono>
13. «Apneia do sono: o que é, sintomas e tratamento | CUF». Acedido: 29 de Abril de 2024. [Em linha]. Disponível em: <https://www.cuf.pt/saude-a-z/apneia-do-sono>
14. «Apneia do Sono: O que é, sintomas, tratamentos e causas.» Acedido: 29 de Abril de 2024. [Em linha]. Disponível em: <https://www.rededorsaoluiz.com.br/doencas/apneia-do-sono>
15. V. A. Peris *et al.*, «Non-contact Reflection Photoplethysmography Towards Effective Human Physiological Monitoring», 2010. [Em linha]. Disponível em: <https://www.researchgate.net/publication/228667007>
16. Q. Tang, Z. Chen, R. Ward, e M. Elgendi, «Synthetic photoplethysmogram generation using two Gaussian functions», *Sci Rep*, vol. 10, n. 1, p. 13883, Ago. 2020, doi: 10.1038/s41598-020-69076-x.
17. Q. Tang *et al.*, «PPGSynth: An Innovative Toolbox for Synthesizing Regular and Irregular Photoplethysmography Waveforms», *Front Med (Lausanne)*, vol. 7, Nov. 2020, doi: 10.3389/fmed.2020.597774.
18. F. Massie, S. Vits, A. Khachatryan, B. Van Pee, J. Verbraecken, e J. Bergmann, «Central Sleep Apnea Detection by Means of Finger Photoplethysmography», *IEEE J Transl Eng Health Med*, vol. 11, pp. 126–136, 2023, doi: 10.1109/JTEHM.2023.3236393.
19. R. Lazazzera *et al.*, «Detection and Classification of Sleep Apnea and Hypopnea Using PPG and SpO<sub>2</sub> Signals», *IEEE Trans Biomed Eng*, vol. 68, n. 5, pp. 1496–1506, Mai. 2021, doi: 10.1109/TBME.2020.3028041.
20. R. S. Vulcan, S. André, e M. Bruyneel, «Photoplethysmography in Normal and Pathological Sleep», *Sensors*, vol. 21, n. 9, p. 2928, Abr. 2021, doi: 10.3390/s21092928.
21. S.-H. Liu, H.-C. Liu, W. Chen, e T.-H. Tan, «Evaluating Quality of Photoplethysmographic Signal on Wearable Forehead Pulse Oximeter With Supervised Classification Approaches», *IEEE Access*, vol. 8, pp. 185121–185135, 2020, doi: 10.1109/ACCESS.2020.3029842.
22. M. Shafique e P. A. Kyriacou, «Photoplethysmographic signals and blood oxygen saturation values during artificial hypothermia in healthy volunteers», *Physiol Meas*, vol. 33, n. 12, pp. 2065–2078, 2012, doi: 10.1088/0967-3334/33/12/2065.
23. A. S. Bhogal e A. R. Mani, «Pattern Analysis of Oxygen Saturation Variability in Healthy Individuals: Entropy of Pulse Oximetry Signals Carries Information about Mean Oxygen Saturation», *Front Physiol*, vol. 8, Ago. 2017, doi: 10.3389/fphys.2017.00555.

## Interfaces Gráficas para Apoio ao Ensino de Séries Numéricas e Séries de Taylor/MacLaurin

Marcos Menezes<sup>1</sup>, André Pinho<sup>1</sup>, Eduardo Almeida<sup>1</sup>, Francisco Rêgo<sup>1</sup>, Fábio Costa<sup>1</sup>, Leonardo Coelho<sup>1</sup>, Nuno Gonçalves<sup>1</sup>, Vítor Cardoso<sup>1</sup>, Helena Brás<sup>1,2</sup>, Pedro Silva<sup>1</sup>, António Sousa<sup>1,2</sup>, Jorge Mendonça<sup>1</sup> e Fernando Carvalho<sup>1</sup>

<sup>1</sup> Instituto Superior de Engenharia do Porto (ISEP) – P.Porto, Porto, Portugal

<sup>2</sup> LEMA, ISEP-IPP, Porto, Portugal

{1241343, 1241412, 1241418, 1240785, 1240826, 1240973, 1241032, vcc, hsb, mps, ats, jpm, fjc}@isep.ipp.pt

**Abstract.** Os conteúdos relacionados com as séries numéricas e séries de potências, lecionados pelos docentes do Departamento de Matemática (DMA) do ISEP nos diversos departamentos onde exercem funções, em geral, são abstratos para a maioria dos estudantes, facto que importa mitigar. Com base na experiência docente adquirida no projeto-piloto LABMAT, implementado no curso de Licenciatura em Engenharia Eletrotécnica (LEE), esta comunicação propõe duas interfaces gráficas interativas, desenvolvidas com recurso, entre outras, às bibliotecas Matplotlib e Sympy do Python. Estas interfaces visam apoiar a consolidação dos conhecimentos transmitidos nas aulas teóricas e teórico-práticas da unidade curricular de AMATA do curso de Licenciatura em Engenharia Informática (LEI). A primeira interface é dedicada ao cálculo de somas parciais, permitindo a visualização gráfica e animada do comportamento convergente ou divergente da série, bem como da “velocidade” com que estas somas parciais se aproximam da soma da série. A segunda interface possibilita a visualização da aproximação de uma série de potências a uma função, para os casos particulares dos desenvolvimentos de Taylor/MacLaurin, permitindo ainda analisar o comportamento do erro cometido num ponto dentro ou fora do intervalo de convergência.

No final do semestre, realizou-se um inquérito de satisfação sobre o impacto que as duas interfaces produziram na aprendizagem destes conteúdos nas aulas teórico-práticas. Foram inquiridas seis turmas do curso, 87 estudantes, apurando-se, globalmente, numa escala de satisfação de 0 a 10, o nível médio de  $8.54 \pm 1.62$ .

**Keywords:** LABMAT, Matplotlib, Séries de Taylor, Séries Numéricas, Sympy

### 1 Introdução

No ano letivo de 2000/01 foi lançado, no curso de Licenciatura em Engenharia Eletrotécnica (LEE) do ISEP, o projeto-piloto LABMAT, que permitiu a aprendizagem de Matemática em laboratório. O projeto pretendia que os estudantes testassem os exercícios previamente solucionados de forma analítica nas aulas teóricas e teórico-práticas, reduzindo o abstracionismo dos conteúdos e, assim, tomassem plena consciência de

uma realidade observável, geradora de imagens que apelassem aos seus sentidos críticos, visando uma aprendizagem significativa [1]. O projeto seguiu a visão que encontra paralelo com o pensamento de Richard Arends [2] que, a propósito da política educacional americana, refere: “*Nenhum aluno deve ser deixado para trás!*”. Usando uma das vertentes pedagógicas descrita pelo autor, o projeto centrou a aprendizagem no estudante, num ambiente próximo do professor e tão interativo que o método de ensino/aprendizagem recorda as narrativas fictícias de Frank Blume [3]. O projeto fez uso do software Scientific WorkPlace (SWP), tendo posteriormente migrado para o Matlab. Com a adequação do curso ao Processo de Bolonha, esta vertente computacional originou uma nova unidade curricular independente da Matemática clássica, os Laboratórios de Matemática que procurou, sempre que possível, paralelizar os conteúdos e manter presente o conceito que fundou LABMAT. Na altura, foi usado o software da Álgebra Computacional, o Maxima, tendo mais tarde sido substituído novamente pelo Matlab e, ainda posteriormente, por motivos economicistas, pelo Python.

No ano letivo de 2007/08, o projeto LABMAT foi experimentado com sucesso na unidade curricular (UC) de Análise Matemática (AMATA) do curso de Licenciatura em Engenharia Informática (LEI) do ISEP, tendo sido utilizado o Maxima para apoiar a vertente computacional em sede de laboratório.

A experiência adquirida pelos docentes do DMA no projeto LABMAT revela que, genericamente, os estudantes apresentam dificuldades na compreensão dos conteúdos relativos às séries numéricas e séries de potências e que trabalhar estes tópicos com o auxílio de ferramentas computacionais permite mitigar estas dificuldades. O novo plano de estudos não prevê uma componente laboratorial na UC de AMATA. Ainda assim é possível apoiar a abordagem analítica das séries com uma abordagem geométrica. Por esse facto, esta comunicação apresenta duas interfaces gráficas desenvolvidas em Python, a partir de classes dedicadas que fazem uso, entre outras, das potencialidades das bibliotecas Matplotlib e Sympy: uma dedicada à análise da natureza de séries numéricas, designada de “Visualização Iterativa das Somas Parciais” e uma outra, dedicada à análise do comportamento de séries de potências, para os casos dos desenvolvimentos de Taylor/MacLaurin, designada de “Função e Aproximação de Taylor”. Estas interfaces gráficas foram utilizadas como ferramentas de apoio à resolução de todos os exercícios realizados nas aulas teórico-práticas de AMATA na edição do ano 2024/25, afetas ao último autor desta comunicação. Assim, para além dos aspetos de implementação, este trabalho apresenta exemplos de resultados de uso e um inquérito realizado a um conjunto de estudantes que habitualmente frequenta as aulas teóricas e teórico-práticas, o qual quantifica o impacto desta abordagem em relação à abordagem analítica.

## 2 Metodologias e Ferramentas

### 2.1 Séries numéricas e séries de Taylor/MacLaurin

Uma série numérica representa um número, correspondente à soma  $S$  de todos os termos de uma sucessão de termo geral  $u_n$  [4-5], caso exista.

Excetuando os casos das séries aritméticas, geométricas e de Mengoli – tipicamente abordadas nas escolas de engenharia – normalmente, as fórmulas representativas das sucessões das somas parciais  $S_n$ , ou seja, de ordem  $n$ , não existem. Nesses casos, não é possível aferir sobre a natureza convergente ou divergente dessas séries com base na definição, ou seja, da existência ou não de soma, dada por:

$$S = \lim_{n \rightarrow +\infty} S_n = \lim_{n \rightarrow +\infty} \sum_{k=1}^n u_k = \sum_{k=1}^{+\infty} u_k = u_1 + u_2 + \dots + u_n + \dots \quad (1)$$

Por esse facto, a natureza das séries identifica-se com base no uso de critérios analíticos, que não permitem ao estudante ter uma percepção visual sobre o que estes aferem. Nesse sentido, foi implementada a interface “Visualização Iterativa das Somas Parciais”, a partir do script `series_numericas.py`, que permite observar numérica e graficamente o valor das somas parciais até a uma ordem  $n$  predefinida e, como tal, efetuar uma previsão sobre a sua natureza. Ainda se pode observar a sua “velocidade” de convergência, isto é, por exemplo, na série geométrica de razão  $r$  e primeiro termo  $u_1$ :

$$\sum_{k=1}^{+\infty} u_1 r^{k-1}, \quad (2)$$

que se diz convergente se  $|r| < 1$ . O valor de  $S_n \rightarrow S$ , tão mais rapidamente quanto mais próxima de 0 for a razão, e mais lentamente quanto mais próxima estiver de  $-1$  ou de  $1$ .

Já uma série de potências [4-5], para os casos dos desenvolvimentos de Taylor/MacLaurin em torno de um ponto de abcissa  $a$  (ou centro da série), representa uma função, que se pode aproximar por um polinómio de ordem  $N$ :

$$f(x) = \sum_{k=1}^{+\infty} \frac{f^{(k)}(a)}{k!} (x - a)^k \approx P_N(x) = \sum_{k=1}^N \frac{f^{(k)}(a)}{k!} (x - a)^k. \quad (3)$$

Neste caso, importa visualizar graficamente a forma como  $P_N(x) \rightarrow f(x)$  em torno do centro da série em função da ordem polinomial  $N$ , estimar o intervalo de convergência e determinar o erro cometido ao aproximar um valor da imagem de  $f$  num dado ponto, usando a aproximação polinomial de uma dada ordem  $N$ . Assim, foi implementada a interface “Função e Aproximação de Taylor”, a partir do script `series_taylor.py`.

## 2.2 Breve introdução ao estado da arte

As interfaces gráficas para representação iterativa das somas parciais de uma série numérica encontram-se, genericamente, implementadas no software de Matemática dinâmica Geogebra. Por exemplo, em [6], a interface representa graficamente, quer os termos da série, quer as somas parciais até uma dada ordem; ordem que é definida a partir de uma caixa de texto, assim como o termo geral da série. Em [7], usando o mesmo software, são representadas as somas parciais e a ordem da soma é controlada a partir de um seletor *slider*. A interface gráfica desenvolvida neste trabalho inova pelo facto de permitir a representação animada das somas parciais, o que a torna apelativa.

As interfaces gráficas para representação de uma função e da aproximação de Taylor/MacLaurin, usando também o Geogebra, podem encontrar-se, por exemplo, em [8].

Trata-se de uma abordagem que permite determinar o comportamento do erro absoluto cometido ao aproximar-se o valor da imagem da função pela correspondente aproximação. Em [9], é usado o software Wolfram Mathematica, que representa exclusivamente a função e a sua aproximação, abordagem também seguida pela função `taylortool` do Matlab [10]. A interface gráfica desenvolvida no âmbito deste trabalho usa a abordagem mais completa apresentada em [8], mas melhorando a interatividade e seguindo o conceito apresentado pelos trabalhos [9-10].

### 2.3 Inquérito de satisfação

Após a lecionação dos conteúdos relativos às séries, com apoio nas interfaces gráficas “Visualização Iterativa das Somas Parciais” e “Função e Aproximação de Taylor”, foi realizado um inquérito de satisfação aos estudantes: sobre o uso de cada uma das interfaces e, globalmente, sobre o uso de ambas as ferramentas.

#### Séries numéricas

*“Relativamente à abordagem analítica realizada nas aulas teóricas e teórico-práticas, em que percentagem considera que o uso complementar da interface gráfica “Visualização Iterativa das Somas Parciais” realizada pelo docente nas aulas teórico-práticas, melhorou a sua perceção sobre a natureza da série e sobre a sua velocidade de convergência?”*

#### Séries de potências

*“Relativamente à abordagem analítica realizada nas aulas teóricas e teórico-práticas, em que percentagem considera que o uso complementar da interface gráfica “Função e Aproximação de Taylor”, realizada pelo docente nas aulas teórico-práticas, melhorou a sua perceção sobre a forma como uma função é aproximável por um polinómio de ordem  $n$  em torno de um ponto, sobre o intervalo e raio de convergência, e sobre a qualidade da aproximação em função do erro cometido?”*

#### Global

*“Genericamente, em que percentagem considera que a abordagem geométrica disponibilizada pelas duas interfaces usadas pelo docente nas aulas teórico-práticas, tornou significativa a sua aprendizagem relativa aos conteúdos séries numéricas e séries de potências?”*

### 2.4 Ferramentas

As interfaces gráficas foram desenvolvidas em Python, usando o editor Visual Studio Code (VSCODE) como ambiente de desenvolvimento integrado. No essencial, as duas implementações fazem uso da biblioteca Matplotlib [11], que permite a implementação de interfaces simples, que integram no próprio objeto gráfico widgets interativos como TextBoxes e Buttons. As bibliotecas Sympy e Numpy asseguram o cálculo simbólico e o cálculo numérico, respetivamente.

### 3 Software Desenvolvido

Nesta secção descrevem-se de forma textual as arquiteturas das interfaces “Visualização Iterativa das Somas Parciais” e “Função e Aproximação de Taylor”, apresentando-se pequenos excertos do código produzido que permitem a sua reimplementação futura.

#### 3.1 Interface gráfica “Visualização Iterativa das Somas Parciais”

A interface “Visualização Iterativa das Somas Parciais”, implementada no script `series_numericas.py` (numeração do código de acordo com a implementação), baseia-se na seguinte arquitetura:

1. **Início:** O script é iniciado com a execução do script `series_numericas.py`;
2. **Importação das bibliotecas:** `numpy, matplotlib, sympy, matplotlib.widgets`;
3. **Definição da Classe `GUI_SomaParcialIterativa`**
  - Inicialização, com base no método construtor `__init__` da classe:

```
17| #Construtor da Classe
18| def __init__(self):
```

- Configuração do gráfico – cria a figura e os eixos do gráfico, aloca o espaço dos widgets, adiciona títulos e etiquetagem nos eixos e na grelha;

```
21| plt.subplots_adjust(bottom=0.375)
22| self.ax.set_title("Visualização Iterativa das Somas
    Parciais")
23| self.ax.set_xlabel("n (termos)")
24| self.ax.set_ylabel("Soma Parcial")
25| self.fig, self.ax = plt.subplots()#Objetos fig e ax
```

- Inicialização de variáveis;
- Criação e posicionamento dos widgets: Textboxes e Buttons (exemplos);

```
36| Caixa de texto do termo geral
37| axbox_termo = plt.axes([0.1, 0.25, 0.7, 0.05])
38| self.text_box_termo = TextBox(axbox_termo, "Termo
    Geral:")
51| #Botão de inicio da animação
52| axbutton_iniciar = plt.axes([0.6, 0.17, 0.2, 0.05])
53| self.button_iniciar = Button(axbutton_iniciar,
    "Iniciar")
```

- Configuração de eventos para interação entre os widgets e o gráfico (exemplos `on_submit` e `on_clicked` associados aos casos anteriores, linha 39 e 54).

```
39| self.text_box_termo.on_submit(self.atualizar_termo)
54| self.button_iniciar.on_clicked(self.iniciar_animacao)
```

- Protótipos dos restantes métodos da classe:
  - `reiniciar_gui(self)` – limpa e repõe as configurações iniciais implementadas pelo construtor;
  - `atualizar_termo(self, termo_str)` – atualiza o termo geral em função da entrada do utilizador e realiza a sua “lambdificação”, isto é, usa a biblioteca `sympy` para permitir o uso do termo geral em contexto numérico (exemplo);

```
105| def atualizar_termo(self, termo_str):
106|     #Atualização do termo geral da série
107|     try:
108|         n = sp.symbols('n')
109|         termo_expr = sp.sympify(termo_str)
110|         self.termo_func = sp.lambdify(n,
111|                                     termo_expr, "numpy")
111|         print(f"Novo geral: {termo_str}")
112|     except Exception as e:
113|         print(f"Erro na atualização do termo: {e}")
```

- `atualizar_n_inicial(self, n_i_str)` – atualiza o valor inicial de  $n$ ;
- `atualizar_n_final(self, n_f_str)` – atualiza o valor final de  $n$ ;
- `iniciar_animacao(self, event)` – reinicia o gráfico e variáveis e executa a soma parcial de forma iterativa.
- `stop_animacao(self, event=None)` – pára a execução da animação e mantém o seu estado atual;
- `animação_iterativa(self)` – executa o cálculo iterativo da soma parcial para cada valor de  $n$  e atualiza o gráfico a cada iteração (exemplo);

```
158| def animacao_iterativa(self):
159|     #Executa a animação iterativa da soma parcial
160|     if self.animating:
161|         if self.n <= self.n_final:
162|             self.atualizar_grafico()
163|             plt.pause(0.1) #Step de 0.1 segundos
164|             self.animacao_iterativa()
```

- `atualizar_grafico(self)` – calcula e representa o próximo termo da série;
- `fechar_aplicacao(self, event)` – fecha a interface;
- `reset_parametros(self, event)` – repõe as configurações iniciais.

4. **Execução do Programa:** cria uma instância da classe `GUI_SomaParcialIterativa` e exhibe a interface com eventos interativos em modo cíclico.
5. **Final:** Quando o utilizador clica no botão “Sair” ou fecha a interface gráfica.

### 3.2 Interface gráfica “Função e Aproximação de Taylor”

Com base nos aspetos de implementação apresentados na interface anterior, a interface “Função e Aproximação de Taylor”, implementada no script `series_taylor.py` (numeração do código de acordo com a implementação), baseia-se na seguinte arquitetura:

1. **Início:** O script é iniciado com a execução do script `series_taylor.py`;
2. **Importação das bibliotecas:** `numpy`, `matplotlib`, `sympy`, `matplotlib.widgets` e `logging`;
3. **Configuração de Logging:** configura o ficheiro `erro.log` para registo de erros;
4. **Definição da Classe `GUI_TaylorSeries`:**
  - Inicialização, com base no método construtor `__init__` da classe:
    - Configuração do gráfico – cria a figura e os eixos do gráfico, aloca o espaço dos widgets e adiciona títulos e etiquetagem nos eixos e na grelha;
    - Inicialização de variáveis;
    - Criação e posicionamento de widgets: Textboxes e Buttons (seção 3.1);
    - Configuração de eventos e interação entre os widgets e o gráfico (seção 3.1);
  - Protótipos dos restantes métodos da classe:
    - `calcular_taylor(self, func: sympy.Expr, centro: float, ordem: int)` – calcula a série de Taylor até à ordem  $n$ , usando a biblioteca Sympy;

```
100| def calcular_taylor(self, func, centro, ordem):
101| #Calcula a série de Taylor de ordem n - Sympy
102|     taylor_series = sp.series(func, self.x, centro,
103|                             n=ordem + 1).removeO()#Representa a série, re
movendo todos os termos de ordem O((x-a)^(ordem+1))
103|     return sp.nsimplify(taylor_series)
```
    - `atualizar_funcao(func_str, _=None)` – atualiza a função simbólica e recalcula a série de Taylor;
    - `atualizar_centro(self, _=None)` – altera o centro da série;
    - `atualizar_x1(self, _=None)` – altera a abcissa do ponto em análise;
    - `atualizar_aprox_error(self, _=None)` – determina o erro absoluto cometido entre a função e a aproximação de Taylor no ponto em análise;
    - `atualizar_dominio_min(self, _=None)` – limite inferior do domínio;
    - `atualizar_dominio_max(self, _=None)` – limite superior do domínio;
    - `aumentar_ordem(self, _=None)` – incrementa a ordem da série e atualiza o gráfico;
    - `diminuir_ordem(self, _=None)` – decrementa a ordem da série e atualiza o gráfico;
    - `update_plot(self)` – representa graficamente a função original, a série de Taylor e os pontos relativos às imagens da função e da aproximação do ponto em análise;
    - `reset(self, _=None)` – repõe as configurações iniciais;
    - `sair(self, _=None)` – fecha a aplicação.
5. **Execução do Programa:** cria uma instância da classe `GUI_TaylorSeries` e exibe a interface com eventos interativos em modo cíclico.
6. **Final:** Quando o utilizador clica no botão “Sair” ou fecha a interface gráfica.

## 4 Resultados

### 4.1 Exemplos de resultados experimentais usando as interfaces gráficas

A Fig.1 apresenta a interface gráfica "Visualização Iterativa das Somas Parciais", observando-se a tendência convergente de uma série de Riemann [5], a partir do comportamento da soma parcial de ordem  $n$ , que se aproxima assintoticamente da soma.

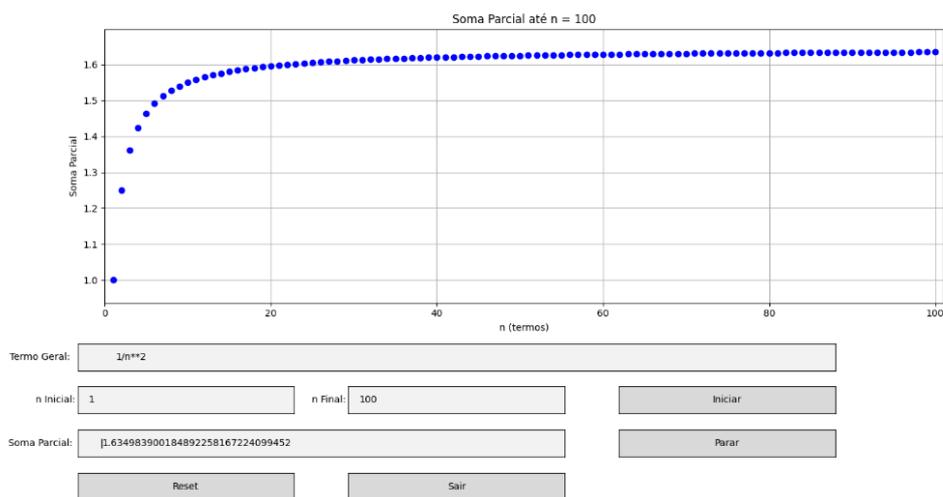


Fig. 1. Interface gráfica "Visualização Iterativa das Somas Parciais".

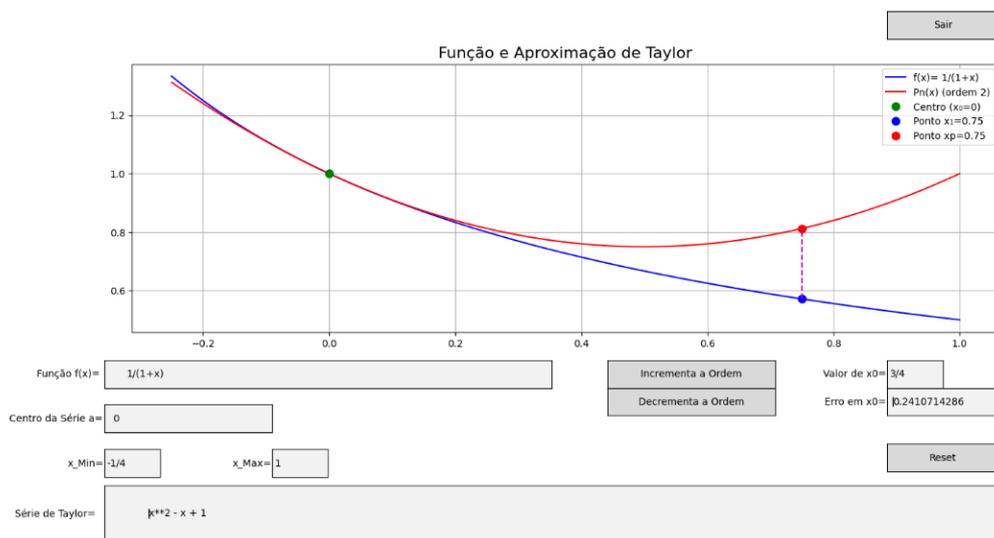
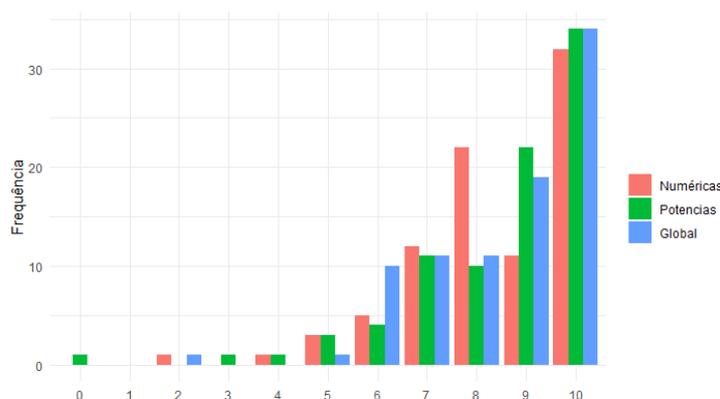


Fig. 2. Interface gráfica "Função e Aproximação de Taylor".

A Fig. 2 apresenta a interface gráfica "Função e Aproximação de Taylor", com o desenvolvimento da série de MacLaurin de uma função que tem correspondência com uma série geométrica de razão  $-x$ , aproximada por um polinómio de segunda ordem (a vermelho). É ainda representado o centro da série (cor verde), o ponto em análise (a azul), o mesmo ponto sobre a aproximação (a vermelho) e um segmento de reta vertical (a magenta) cujo comprimento representa o valor absoluto do erro da aproximação.

## 4.2 Inquérito sobre o uso das interfaces gráficas

Foi realizado um inquérito sobre o impacto do uso das interfaces gráficas na aprendizagem de séries numéricas e de séries de Taylor/MacLaurin. Relativamente ao uso da interface gráfica "Visualização Iterativa das Somas Parciais", que visou melhorar a compreensão relativamente à natureza das séries numéricas e à "velocidade" de convergência, o nível médio de satisfação dos estudantes é de  $8.39 \pm 1.66$  (Numéricas, Fig. 3). Já relativamente ao uso da interface "Função e Aproximação de Taylor", que visou contribuir para a compreensão do comportamento das séries de Taylor/MacLaurin, nomeadamente no que se refere ao intervalo e ao raio de convergência e à qualidade da aproximação, o nível médio de satisfação dos estudantes é de  $8.52 \pm 1.85$  (Potências, Fig. 3). Foi ainda avaliada a mesma eficácia das duas interfaces de forma global, sendo o nível médio de satisfação dos estudantes de  $8.54 \pm 1.62$  (Global, Fig. 3).



**Fig. 3.** Gráfico de frequência sobre os níveis de satisfação dos estudantes relativos ao uso das duas interfaces gráficas nas aulas teórico-práticas de AMATA do curso LEI do ISEP.

## 5 Conclusões e perspetivas de trabalho futuro

Os conteúdos relativos às séries numéricas e de potências (para os casos dos desenvolvimentos de Taylor/MacLaurin), lecionados nas UCs de Matemática, por norma, são abstratos para os estudantes de engenharia. Este efeito pode ser mitigado se as abordagens teóricas e teórico-práticas forem apoiadas pela abordagem geométrica, isto é, se a imagem que produz for apelativa e assim se traduzir numa aprendizagem significativa.

Foi nesta conformidade, com base nos conhecimentos da experiência LABMAT e respondendo aos anseios dos estudantes da UC de AMATA do LEI, que foram desenvolvidas duas interfaces gráficas em Python, com base nas potencialidades das bibliotecas Matplotlib e Sympy, cada uma das quais dedicada a um dos referidos conteúdos.

Inquiridos os estudantes sobre o impacto que esta abordagem complementar tem nas suas aprendizagens, a resposta foi muito positiva. O nível global de satisfação dos estudantes pelo uso das interfaces gráficas foi de  $8.54 \pm 1.62$ , no máximo de 10.

Como perspetivas de trabalho futuro, pese embora as interfaces desenvolvidas apresentem a robustez desejada, os estudantes manifestaram interesse em implementações que produzam o mesmo efeito no conteúdo “espaços vetoriais” que faz parte do programa curricular de Álgebra Linear e Geometria Analítica, mas usando a framework PyQT.

## Agradecimentos

Os autores desta comunicação agradecem a colaboração de todos os estudantes das turmas C, D, E, F, K e L que frequentaram a UC de AMATA do curso de LEI do ISEP, da edição do ano letivo de 2024/25. Ainda ao Dr. Ilídio Fonseca (a título póstumo), ao Eng. Pedro Guedes e ao Eng. José Magalhães do DMA pela idealização da experiência LABMAT, que muito contribuiu para a melhoria de práticas pedagógicas.

Por fim, ao Dr. Ângelo Martins que, na qualidade de diretor da LEI adotou, com sucesso, a experiência LABMAT na UC de AMATA, na edição do ano letivo de 2007/08.

## Referências

1. Ausubel, D. P.: The Psychology of Meaningful Verbal Learning. In: Grune & Stratton, pp. 1-255 (1963)
2. Arends, R. I.: Learning to Teach. In: McGraw-Hill, 10ª edição, pp. 1-608 (2014)
3. Blume, F.: Applied Calculus for Scientists and Engineers: A Journey in Dialogues. In: CreateSpace Independent Publishing Platform, pp. 1-468 (2014)
4. Alves de Sá, A., Louro, B.: Sequences and Series: Theory and Practice. In: Springer (2024)
5. Faria, A., Brás, H., Figueiredo, I.: Análise Matemática I: Resumo Teórico, Exercícios Resolvidos e Propostos. In: Edições Sílabo (2021)
6. Geogebra: Sequence and Partial Sum Grapher, <https://www.geogebra.org/m/S68dtsVa>, último acesso 07/12/2024
7. Geogebra: Sequence and Partial Sum Grapher, <https://www.geogebra.org/m/AwWvvuHq>, último acesso 07/12/2024
8. Magalhães, J.: A Aprendizagem de Tópicos de Cálculo por Alunos de Engenharia Através do Trabalho de Projeto. In: *HOLOS* (2019)
9. Wolfram Demonstrations Project: Taylor Series, <https://demonstrations.wolfram.com/TaylorSeries/>, último acesso 07/12/2024
10. MathWorks: taylor tool - Taylor series calculator, <https://www.mathworks.com/help/symbolic/taylor tool.html>, último acesso 07/12/2024
11. Matplotlib, [https://matplotlib.org/stable/plot\\_types/index.html](https://matplotlib.org/stable/plot_types/index.html), último acesso 07/12/2024

## Deteção de Anomalias em Vagões Ferroviários Baseada em *Machine Learning*

Ricardo Pereira<sup>1</sup>, Pedro Ribeiro<sup>2</sup>, António Sousa<sup>1,3</sup>, Carlos José Campos<sup>1,4</sup>,  
Veríssimo Lima<sup>1</sup>, Jorge Mendonça<sup>1</sup> e Fernando Carvalho<sup>1</sup>

<sup>1</sup> Instituto Superior de Engenharia do Porto (ISEP) – P.Porto, Porto, Portugal

<sup>2</sup> Evoleo Technologies Lda, Portugal

<sup>3</sup> Laboratório de Engenharia Matemática (LEMA), Porto, Portugal

<sup>4</sup> Instituto de Engenharia de Sistemas e Computadores (INESC TEC), Porto, Portugal  
{1130794, ats, crc, vms, jpm, fjc}@isep.ipp.pt

**Resumo.** As anomalias detetadas nas rodas e veios das rodas de vagões ferroviários podem causar avultados danos nas vias-férreas e representar riscos para pessoas e bens. Como prevenção, este trabalho apresenta o desenvolvimento de um sistema de deteção de anomalias baseado num classificador de *Machine Learning* (ML), do tipo TinyML, que monitoriza o estado operacional das rodas e seus veios, distinguindo três classes anómalas: o desgaste da roda, o desgaste ou rutura do veio da roda ou ambas em simultâneo. Para o efeito, foi desenvolvido um protótipo mecânico no qual se induzem vibrações que simulam as diferentes anomalias. O modelo de ML, integrado na arquitetura de software, foi treinado e validado (testado) na plataforma Edge Impulse. Para treino do modelo foi usado um conjunto de dados adquiridos do protótipo mecânico com recurso a um acelerómetro. A plataforma Edge Impulse também determina as características que melhor descrevem e distinguem as classes usadas, disponibilizando um classificador treinado sob a forma de uma biblioteca para o dispositivo *Edge Computing* selecionado. Assim, a arquitetura de hardware utiliza dois Arduinos: o primeiro, com um acelerómetro integrado, lê as acelerações a uma frequência de amostragem de 100Hz; o segundo, recebe os dados, processa todo o procedimento de classificação e informa o utilizador sobre a presença ou não de anomalias nas rodas do vagão. Os resultados das simulações mostram que esta tecnologia pode ser eficientemente utilizada na deteção deste tipo de anomalias em contexto real.

**Palavras-chave:** Arduino, Edge Computing, Machine Learning, Plataforma Edge Impulse, Sistemas embebidos, TinyML

### 1 Introdução

A manutenção preditiva tornou-se um elemento essencial na gestão de ativos no setor industrial. Estima-se que o uso de estratégias de manutenção inadequadas possa reduzir a capacidade de produção em 5% a 20% [1]. Em contrapartida, a aplicação de técnicas de manutenção preditiva pode aumentar a disponibilidade de equipamentos em 5% a 15% e reduzir os custos de manutenção em 18%

a 25% [2]. Atualmente, a manutenção preditiva já é implementada na indústria ferroviária, onde se espera economizar aproximadamente 7,5 mil milhões de euros anualmente à escala global [3, 4]. As anomalias que se verificam nas rodas dos vagões podem danificar vários quilómetros de via-férrea até serem identificadas, devendo-se: ao desgaste das rodas dos vagões, ao desgaste ou rutura do veio de suporte das rodas ou ambas as situações em simultâneo. Estas seguem padrões cuja variabilidade pode ser medida em termos de probabilidade *a posteriori* e podem ser classificadas recorrendo a técnicas de ML. É um contexto que permite o uso de sistemas computacionais embebidos – dispositivos *Edge Computing* – para a deteção de anomalias, onde os dados são processados no local ou próximo da fonte de geração, permitindo minimizar a latência, garantir respostas rápidas e segurança sobre os dados. Uma vez que se tratam de dispositivos de baixo consumo e de recursos limitados [5], surgiu o TinyML, uma área especializada em ML, focada na implementação de modelos simplificados para estes dispositivos. O TinyML resulta de um projeto desenvolvido pela Google e representa um avanço significativo na integração de inteligência artificial, não só em sistemas embebidos, como também em dispositivos IoT, permitindo desenvolver aplicações inovadoras, entre outras, nas áreas da indústria, ciência e sociedade [6, 7]. Exemplos de plataformas dedicadas a TinyML, são: o TensorFlow Lite, o Edge Impulse, o PyTorch; que permitem treinar, validar e, por último, obter modelos treinados em forma de bibliotecas a integrar na arquitetura de software dos dispositivos [7, 8].

Este trabalho apresenta um sistema de deteção de anomalias, baseado em dispositivos Arduino e, faz uso da plataforma Edge Impulse que facilita a obtenção do modelo em virtude da sua abordagem Auto-ML, que automatiza algumas etapas do ciclo ML. Foi preliminarmente apresentado em formato de resumo alargado em [9], descrevendo-se nesta comunicação com o detalhe suficiente que permita a sua reprodução futura.

Na secção 2, é apresentada a constituição do protótipo mecânico e as *labels* associadas às classes de decisão do classificador: as anómalas e a não anómala. Descrevem-se ainda as principais metodologias usadas na implementação do modelo TinyML a partir da plataforma Edge Impulse. Na secção 3, apresentam-se as arquiteturas de hardware e de software que estão na base de duas implementações: a primeira, para aquisição de dados de treino e de teste; a segunda, que implementa as funcionalidades do sistema de deteção de anomalias após a implantação do software nos Arduinos. A secção 4 apresenta resultados de exatidão, considerando anomalias em condições similares às observadas durante a fase de treino do modelo e, com especial interesse, anomalias observadas no seu estado inicial que nunca foram observadas, definindo-se assim um potencial intervalo de atuação do sistema de deteção de anomalias. Finalmente, na secção seguinte, apresentam-se as conclusões e algumas perspetivas de trabalho futuro.

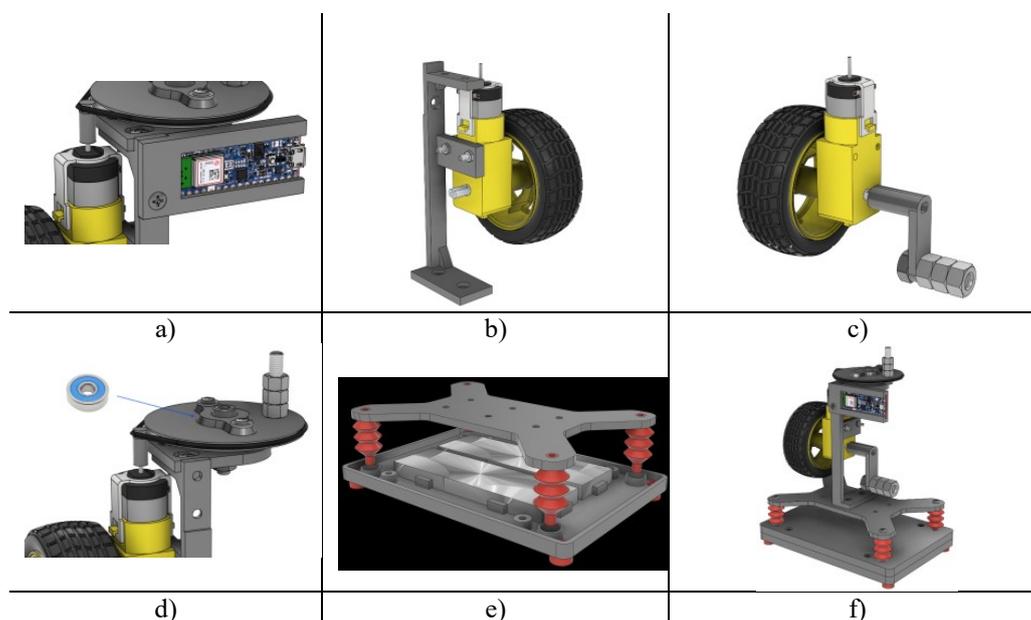
## 2 Ferramentas e Tecnologias

Esta secção apresenta as componentes do protótipo mecânico que simulam as condições de anomalia e não anomalia, as respetivas *labels* associadas e os principais procedimentos que permitem criar o modelo de classificação na plataforma Edge Impulse.

## 2.1 Protótipo Mecânico

As anomalias identificadas nas rodas dos vagões devem-se ao desgaste das rodas, ao desgaste ou rutura do veio das rodas ou à presença de ambas. Estes efeitos são simulados a partir de componentes constituintes da estrutura de um protótipo mecânico, que foram modeladas com base na ferramenta Autodesk Fusion [10] e uma impressora 3D, Fig. 1. Na estrutura induzem-se vibrações, em que o movimento registado ao longo dos eixos principais do plano tridimensional se identifica por valores de aceleração lidos pelo acelerómetro do Arduino Nano 33 BLE Sense Rev2 (A1), Fig. 1a).

É acoplada uma roda a um motor DC fixado num pilar vertical, que simula o estado normal ou não anómalo da roda, Fig. 1b) – *label* “on”. O veio bilateral do motor, permite do lado oposto à roda, a fixação de uma haste lateral com um parafuso e no máximo 3 porcas roscadas, que simula o desgaste ou rutura do veio da roda, Fig. 1c) – *label* “Side\_Unb”. No topo da estrutura, foi aplicado um outro motor (de velocidade constante) ligado a uma polia com um parafuso e no máximo 3 porcas roscadas, Fig. 1d), que provoca desequilíbrio vertical e simula o desgaste da roda – *label* “Top\_Unb”. A conjugação simultânea das duas anomalias, tem correspondência com a *label* “Top&Side\_Unb”. De referir ainda que o número de porcas roscadas permite simular o estado de desgaste mecânico, impondo diferentes entropias no protótipo mecânico.



**Fig. 1.** Componentes do protótipo mecânico: incorporação do Arduino A1 em a), roda sem anomalias em b), desgaste ou rutura do veio da roda em c), desgaste da roda em d), sistema de amortecimento em e) e estrutura completa do protótipo mecânico em f).

As vibrações induzidas pelas referidas componentes internas, mas também externas, afetam os valores das acelerações, pelo que a estabilidade da estrutura é assegurada

com base no sistema de amortecimento apresentado na Fig. 1e). Finalmente, na Fig. 1f), apresenta-se o protótipo mecânico completo que integra os diversos componentes.

## 2.2 Plataforma Edge Impulse

A plataforma Edge Impulse permite o desenvolvimento de modelos de classificação com base no ciclo operacional da Fig. 2. O *dataset* das amostras pode ser adquirido diretamente de dispositivos compatíveis com a plataforma ou de registos residentes em ficheiros. Optou-se por registar os dados em ficheiros .csv, que uma vez lidos a partir da plataforma, permitiu seleccionar 80% das amostras para treino e 20% para teste.



Fig. 2. Ciclo operacional da plataforma Edge Impulse, retirado de [8].

A ferramenta “Data Explorer” da plataforma permitiu identificar *outliers* e amostras mal rotuladas, através da visualização 2D dos dados, reduzida a sua dimensionalidade. Neste caso, a classe “on” é representada por 12 minutos de amostras, o que corresponde a 40% dos dados, enquanto cada uma das três classes restantes é representada por 6 minutos de amostras, correspondendo aos restantes 60% dos dados.

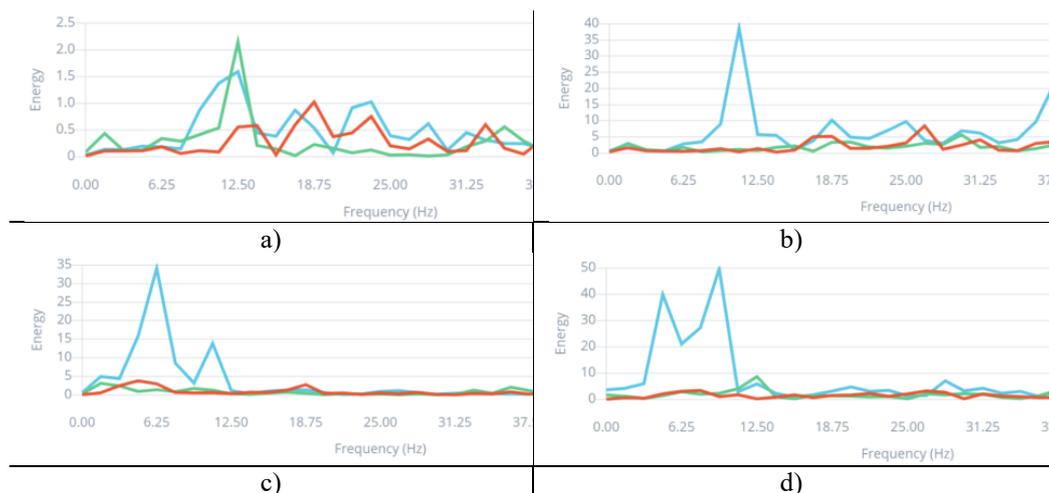
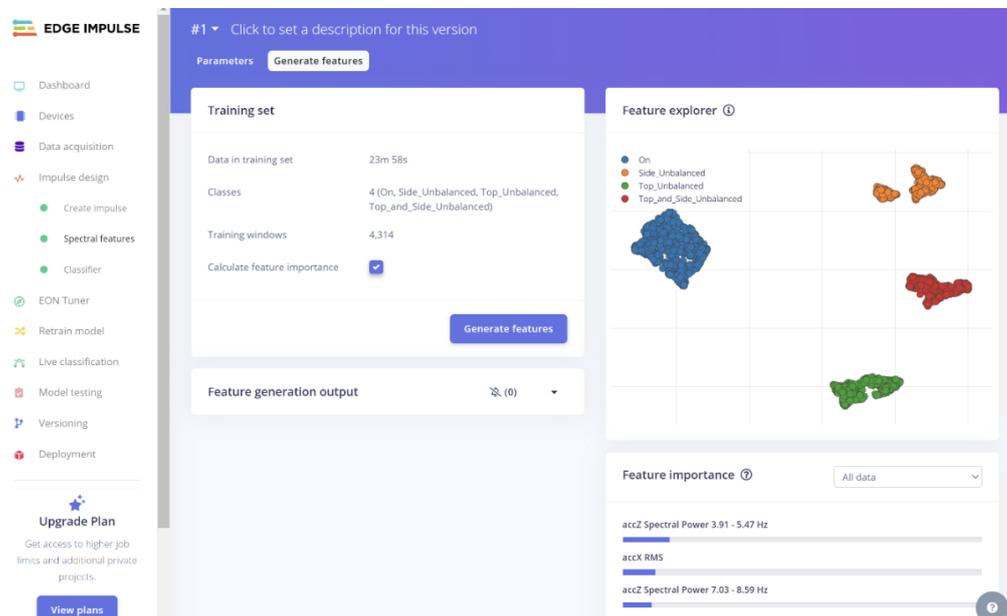


Fig. 3. Energia espectral das classes: “on” em a), “Side\_Unb” em b), “Top\_Unb” em c) e “Top&Side\_Unb” em d).

Seguidamente, a ferramenta “Create Impulse”, permite definir sobre o bloco de amostras de entrada, o comprimento de cada amostra. Observou-se que amostras de 1 segundo contêm todas as características essenciais que descrevem cada classe. Adicionalmente, a partir de subconjuntos de amostras de 2 segundos e de uma janela deslizante de 1 segundo com deslocamentos de 0.2 segundos, procedeu-se à expansão do número de amostras – portanto, obtêm-se 6 amostras por cada subconjunto de 2 segundos. De referir ainda que cada amostra de 1 segundo contém 100 medidas (tridimensionais) da aceleração. Finalmente, definiram-se as 4 *labels* do classificador descritas na secção anterior e a tipologia das *features*. Optou-se exclusivamente pelo uso de *features* com representação no domínio das frequências, considerando que a análise das potências espectrais obtidas das acelerações sobre os 3 eixos para cada classe, usando uma amostra aleatória de dados, exhibe a existência de picos de energia de diferentes frequências e magnitudes, permitindo distinguir as classes, Fig. 3. Do cálculo da transformada rápida de Fourier (FFT), foram extraídas um total de 111 *features*, destacando-se: “accZ Spectral Power 3.91 - 5.47 Hz”, “accX RMS” e “accZ Spectral Power 7.03 - 8.59 Hz”; como exemplos das três mais relevantes para a separação das classes. A Fig. 4 mostra ainda, para o mesmo conjunto de treino, a evidente separação das amostras por classes, com base nas 111 *features* extraídas pela plataforma Edge Impulse, projetadas num mapa bidimensional.



**Fig. 4.** Diagrama de dispersão obtido da projeção 2D das amostras de treino, com base nas 111 *features* extraídas no domínio das frequências.

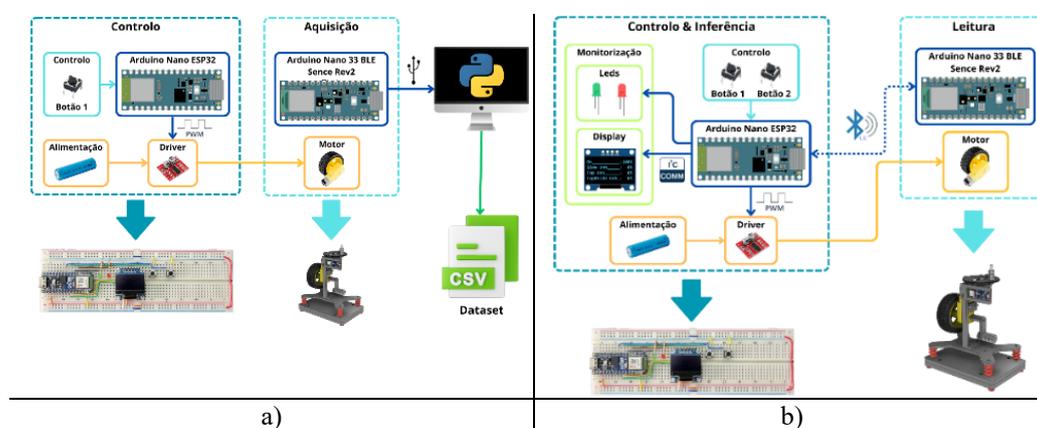
Apesar da existência da ferramenta de Auto-ML, “EON Tuner” na plataforma, decidiu-se que a arquitetura do classificador fosse definida com base no melhor desempenho

obtido de um conjunto de ensaios experimentais, tendo sido a rede ainda configurada com: duas camadas escondidas (20 neurões na primeira e 10 neurões na segunda), uma taxa de aprendizagem de 0.0005 e 10 épocas para a fase de treino. Após a fase de validação, o classificador foi exportado da plataforma Edge Impulse para o Arduino Nano ESP32 (A2) sob a forma de uma biblioteca dedicada.

### 3 Implementação

#### 3.1 Hardware Desenvolvido

O sistema de deteção de anomalias é baseado em duas arquiteturas de hardware, Fig. 5.



**Fig. 5.** Arquiteturas de hardware: para aquisição de dados em a) e para teste do classificador em b).

A primeira arquitetura de hardware é dedicada à aquisição de dados de treino, utilizando o Arduino A2 para controlar a velocidade do motor que suporta da roda, via sinal PWM e o Arduino A1 para adquirir dados da aceleração a uma frequência de 100Hz, que são registados num ficheiro .csv, Fig. 5a).

A segunda arquitetura de hardware, representa o protótipo final, onde o classificador é implantado no Arduino A2 no final do ciclo operacional da plataforma Edge Impulse, para realizar a classificação das amostras em tempo real, Fig. 5b).

Para efeitos de simulação controlada, as amostras são atualmente solicitadas individualmente a partir de um botão de pressão e classificadas por ML. O resultado da classificação é exibido a partir de valores de probabilidade associados a cada classe num pequeno display e, adicionalmente, por LEDs (de cor vermelha que indica a existência de anomalia e um outro de cor verde que indica o estado normal), conforme se pode observar nas *breadboards* que constituem parte das arquiteturas de hardware, Fig. 5.

Em síntese, nesta segunda arquitetura, os Arduinos A1 e A2 comunicam entre si de forma sequencial síncrona via Bluetooth, para realizar os eventos que se encontram esquematicamente representados na Fig. 6.



Fig. 6. Diagrama de eventos.

### 3.2 Software Desenvolvido

A arquitetura de software dedicada à primeira arquitetura de hardware, encontra-se representada na Fig. 7.

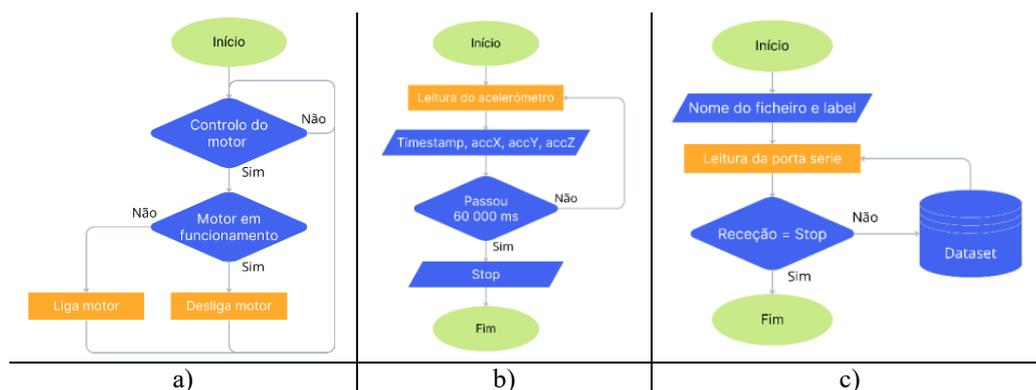
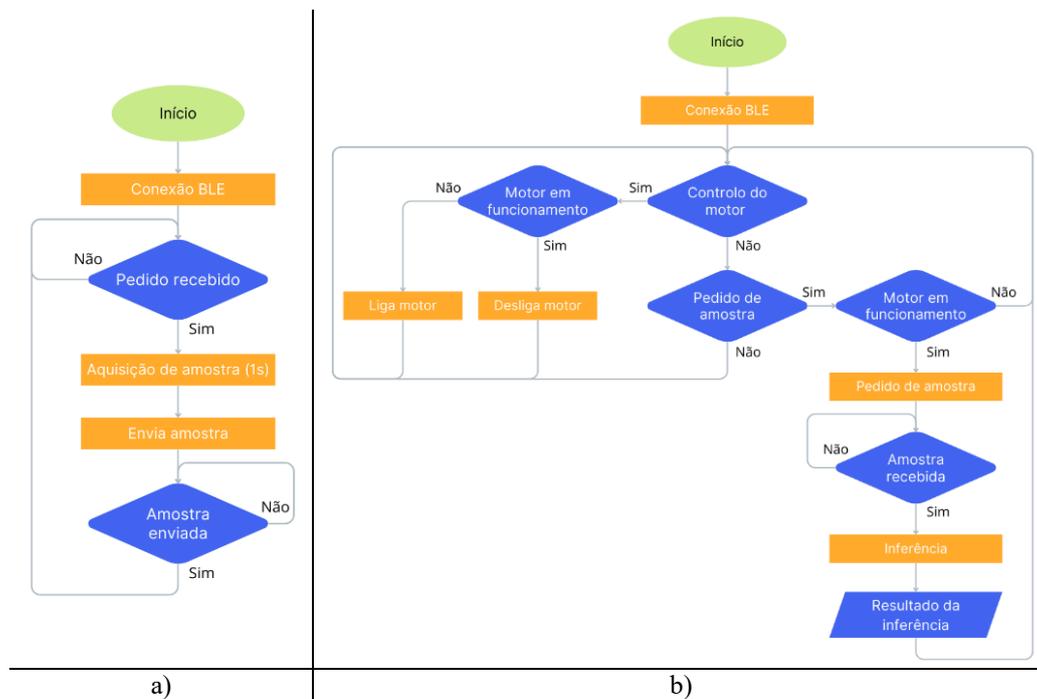


Fig. 7. Fluxogramas para: controlo do motor via PWM do Arduino A2 em a), aquisição de amostras de dados via Arduino A1 em b) e registo em ficheiro .csv de amostras de dados em c).

São representados os fluxogramas dos algoritmos desenvolvidos para o controlo do motor e aquisição de dados, bem como o processo de criação do ficheiro .csv a partir dos dados adquiridos. O fluxograma da Fig. 7a) é relativo ao controlo do motor pelo Arduino A2, incluindo o ligar/desligar do sinal PWM via botão. O fluxograma da Fig. 7b) esquematiza o algoritmo do Arduino A1 para a aquisição de dados, incluindo a configuração do tempo de aquisição das amostras (a aquisição realiza-se em blocos com 1 minuto de duração) e o envio da mensagem de paragem, descrita como “Stop”. O fluxograma do algoritmo para código Python na Fig. 7c), detalha o processo de leitura dos dados da porta série e a criação do ficheiro .csv com as informações recolhidas pelo Arduino A1.

A arquitetura de software dedicada à segunda arquitetura de hardware encontra-se representada na Fig. 8. O fluxograma do algoritmo desenvolvido para o Arduino A1, para efeitos de leitura de amostras, encontra-se representado na Fig. 8a). Para o Arduino A2, foi desenvolvido o algoritmo baseado no fluxograma da Fig. 8b), que permite controlar

o motor DC, solicitar uma amostra e classificá-la (ou realizar a respectiva inferência). Esta inferência é realizada pelo modelo de classificação exportado do Edge Impulse para o Arduino A2.



**Fig. 8.** Fluxogramas relativos: à aquisição de amostras de dados via Arduino A1 em a) e solicitação de uma amostra e correspondente classificação via Arduino A2 em b).

## 4 Resultados

Durante a fase de treino da rede neuronal associada ao classificador, foram realizados seis ensaios empíricos, utilizando diferentes números de épocas (1, 5 e 10) combinados com diferentes taxas de aprendizagem (0.0001 e 0.0005), e medindo as exatidões do treino e da validação. As melhores exatidões obtidas foram 99.80% para treino e 99.82% para validação, usando uma taxa de aprendizagem de 0.0005 ao longo de 10 épocas. Os resultados apresentados indicam que o modelo não está sobreajustado e, portanto, considera-se bem configurado. O algoritmo assume que uma amostra pertence a uma determinada classe sempre que a probabilidade é superior a 0.6; caso contrário, considera-se pertencente a uma classe adicional de incerteza. Assim, a fase de teste do classificador é baseada em dois ensaios que classificam, para cada caso, 80 amostras, 20 amostras de cada classe. No primeiro ensaio, as anomalias identificadas com as *labels* “*Side\_Unb*” e “*Top\_Unb*”, geram amostras anómalas similares às observadas durante a fase de treino, ou seja, provocadas por um parafuso e três porcas roscadas, cor-

respondendo a um estado anômalo avançado, Fig. 1c) e 1d) respectivamente. No segundo ensaio, as mesmas anomalias, geram amostras provocadas por um parafuso e uma única porca roscada, correspondendo a um estado anômalo inicial. Na Fig. 9 e Fig. 10 apresentam-se os resultados obtidos com base na matriz de confusão.

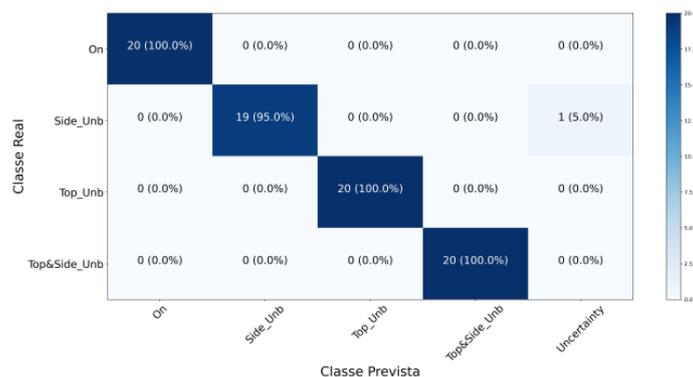


Fig. 9. Matriz de confusão do primeiro ensaio.

No primeiro ensaio, a exatidão é de 98.75%, com apenas uma amostra classificada incorretamente, Fig. 9. No segundo ensaio, a exatidão é de 90,00%, com três amostras classificadas incorretamente e cinco casos de incerteza, Fig. 10.

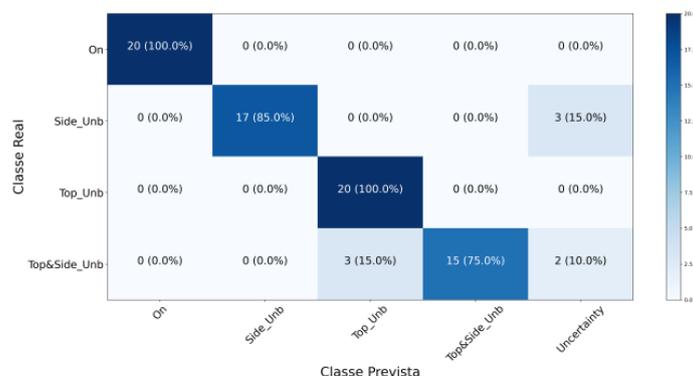


Fig. 10. Matriz de confusão do segundo ensaio.

## 5 Conclusões e Trabalhos Futuros

Este trabalho descreve o desenvolvimento de um sistema de detecção de anomalias em vagões ferroviários, baseado em ML, com baixa latência de resposta. Apresenta um protótipo mecânico que simula um sistema real para a aquisição de dados de aceleração que aferem a existência de movimento ao longo dos eixos principais do plano tridimensional; movimentos esses, suscetíveis de representarem anomalias. Esses dados são posteriormente processados na plataforma Edge Impulse, gerando *features* no domínio

espectral que melhor descrevem e separam as classes, permitindo obter um modelo de classificação a partir do treino e validação de uma rede neuronal.

O sistema de deteção de anomalias, baseia-se no uso de dois dispositivos *Edge Computing* do tipo Arduino e realiza-se a partir de duas implementações: uma para aquisição de amostras de treino e geração do modelo de classificação na plataforma Edge Impulse e, a outra, após implantação do modelo gerado, para a aquisição e classificação das amostras de teste.

A fase de teste contempla dois ensaios que classificam, para cada caso, 80 amostras, 20 amostras por classe: o primeiro ensaio, testa amostras similares às observadas na fase de treino, correspondentes a um estado anómalo avançado; o segundo ensaio, testa amostras bastante distintas das observadas durante a fase de treino, que correspondem a anomalias no estado inicial. O resultado obtido no segundo ensaio mostra, com especial interesse, que o sistema de deteção de anomalias classifica corretamente a maioria das amostras de teste, independentemente de ter sido treinado em exclusivo com base em amostras associadas a estados anómalos avançados – útil para deteções precoces.

Trabalhos futuros incluem a recolha de dados reais, a implementação de métricas de sensibilidade e especificidade e, a adaptação do modelo para diferentes dispositivos de acordo com as especificidades da empresa. Por outro lado, outras abordagens de análise de dados como Deep Learning, poderão ser utilizadas para melhorar a validação dos resultados obtidos e até estimar um mapa de manutenções preditivas.

## Referências

1. Coleman, S., Damofaran, D., Deuel, E.: Predictive maintenance and the smart factory. Deloitte (2017)
2. Bradbury, S., Carpizo, B., Gentzel, M., Horah, D., Thibert, J.: Digitally enabled reliability: Beyond predictive maintenance. McKinsey & Company (2018)
3. Prytz, R.: Machine learning methods for vehicle predictive maintenance using off-board and on-board data. No. 9, Halmstad University (2014)
4. Stern, S., Behrendt, A., Eisenschmidt, E., Reimig, S., Schirmers, L., Schwerdt, I.: The Rail Sector's Changing Maintenance Game. (2017)
5. Zhou, S., Du, Y., Chen, B., Li, Y., Luan, X.: An Intelligent IoT Sensing System for Rail Vehicle Running States Based on TinyML. IEEE Access, vol. 10 (2022)
6. Magalhães, R.: O que é TinyML? [Online]. Available: <https://compraco.com.br/blogs/tecnologia-e-desenvolvimento/o-que-e-tinyml>, last accessed 2024/06/11
7. Warden, P.: TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low Power Microcontrollers. O'Reilly Media, ISBN: 978-1492052043 (2020)
8. Edge Impulse: The Leading Edge AI Platform. [Online]. Available: <https://edgeimpulse.com/>, last accessed 2024/06/11
9. Pereira, R., Ribeiro, P., Sousa, A., Mendonça, J., Lima, V., Carvalho, F.: Railway Wagons Anomaly Prediction Using TinyML. In: International Conference on Mathematical Analysis and Applications in Science and Engineering (ICMS2SC'24), ISBN: 978-989-35251-7-3, pp. 147–150. ISEP, P. Porto, Portugal (2024)
10. Autodesk: Autodesk Fusion for Education [Online]. Available: <https://www.autodesk.com/education/edu-software/fusion>, last accessed 2024/03/10

## Authors Index

Almeida, Eduardo, 147  
Amorim, Mário, 19  
Anjos, João, 116

Barbosa, Manuel, 96  
Brás, Helena, 147

Campos, Carlos José, 77, 106, 157  
Caramez, Ana, 67  
Cardoso, Diana, 126  
Cardoso, Marílio, 96  
Cardoso, Vítor, 147  
Carvalho, Fernando, 77, 106, 147, 157  
Carvalho, Piedade, 116  
Carvalho, Ângela, 137  
Coelho, Leonardo, 147  
Costa, Fábio, 147  
Costa, Vasco, 77  
Cunha, Bruna, 67

Duarte, F. Jorge, 48

Fernandes, Leandro, 28  
Ferreira, Bruno Sousa, 106

Gomes, Ana, 137  
Gomes, Lobinho, 77  
Gonçalves, Nuno, 147  
Guedes, Pedro, 96  
Guimarães, Pedro, 67

Jacob, Frederico, 67

Lima, Veríssimo, 77, 157

Machado, Tiago, 58  
Marcelino, Ana, 67  
Matos, Telmo, 96  
Mendonça, Jorge, 147, 157  
Menezes, Marcos, 147  
Moura, Ricardo, 87  
  
Nogueira, Tomás, 96  
  
Pereira, Alberto, 96  
Pereira, Ricardo, 157  
Pinho, André, 147  
Pinto, Helder, 58  
Pinto, Helder Rodrigo, 19  
Pinto, Hélder, 87  
  
Ramos, Carlos Augusto, 137  
Ribeiro, Inês, 137  
Ribeiro, Pedro, 157  
Rocha, Nuno, 38  
Rocha, Vitor, 87  
Rêgo, Francisco, 147  
  
Santos, Cátia, 19  
Silva, Emanuel, 28, 38  
Silva, Pedro, 147  
Soares, Rui, 48  
Sousa, António, 77, 106, 147, 157  
Sousa, Beatriz, 67  
  
Teixeira, Beatriz, 96  
Teixeira, Tatiana, 96  
  
Ventura, Vitor, 19, 87  
Vilão, Gina Maria, 137