Book of Extended Abstracts

Doctoral Consortium of the
12th Iberian Conference on Pattern Recognition and Image Analysis

Coimbra, Portugal • June 30, 2025

PRIA


INSTITUTO DE SISTEMAS E ROBÓTICA
UNIVERSIDADE DE COIMBRA


CISUC

TITLE

Book of Extended Abstracts of the 12th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2025)

EDITORS

Bernadete Ribeiro, CISUC, University of Coimbra (Portugal)
Catarina Silva, CISUC, University of Coimbra (Portugal)
Nuno Gonçalves, Institute of Systems and Robotics - University of Coimbra (Portugal)
Gustavo Bongiovi, Institute of Systems and Robotics - University of Coimbra (Portugal)

Version 1 - Available from July 14, 2025

This book has been published independently with non-commercial purposes only at:

https://visteam.isr.uc.pt/publications/ibpria-2025-book-of-extended-abstracts/

IbPRIA is an international conference co-organized by the Portuguese APRP and Spanish AERFAI chapters of the IAPR International Association for Pattern Recognition. The conference consist in one day of Tutorials and Doctoral Consortium, and three days of full-paper presentations.

The main conference consists of high-quality, previously unpublished papers, presented either orally or as a poster, intended to act as a forum for research groups, engineers and practitioners, to present recent results, algorithmic improvements and promising future directions in pattern recognition and image analysis. The accepted papers appears in the conference proceedings and get published in Springer Lecture Notes in Computer Science Series.

The Doctoral Consortium provides a unique opportunity for PhD students to present their ongoing work in a poster format and to interact with other students and experienced researchers. Submissions are welcome regardless of whether the work has been previously presented, is intended for future presentation, or is still in the early stages of development. Participants are required to submit a two-page paper outlining their research. The Extended Abstracts of the Doctoral Consortium are published independently in this *Book of Extended Abstracts*, ensuring visibility for the contributions.

https://ibpria.org/2025

Preface

The 12th Iberian Conference on Pattern Recognition and Image Analysis, whose acronym is IbPRIA 2025, took place in Coimbra, Portugal, from June 30th to July 3rd, 2025. This twelfth edition was organized by the Institute of Systems and Robotics - University of Coimbra (ISR-UC), with the collaboration of the Center for Informatics and Systems of the University of Coimbra (CISUC), in partnership with the Portuguese APRP and Spanish AERFAI chapters of the IAPR International Association for Pattern Recognition.

The IbPRIA 2025 Doctoral Consortium kicked off the conference on June 30, featuring 13 dynamic presentations from PhD students in a dedicated two-hour A1 poster session. Attendees had the opportunity to engage peers and experts and discuss ideas.

The first day also had four Tutorials. These Tutorials presented very interesting topics such as “Tutorial 1—Data-Efficient Strategies for Object Detection”, organized by a team from the Center for Informatics and Systems of the University of Coimbra, “Tutorial 2—On the Turning Away: Enhancing Stroke Survivors’ Rehabilitation with Virtual Reality”, organized by a team from University of Aveiro, “Tutorial 3—pyMDMA: An Open-Source Multimodal Framework for Enhanced Auditing of Real and Synthetic Data”, organized by a team from Fraunhofer-AICOS, and “Tutorial 4—Error Estimation in Pattern Recognition”, organized by a team from Polytechnic University of Valencia.

The venue was the Congress Center of the Hotel Quinta das Lágrimas close to Coimbra city center. This historical place was the stage of one of the most Romantic stories of the Portuguese Monarchy in the XIV century, when Dom Pedro, Infant of Portugal (later King of Portugal) and

his beloved Dona Inês de Castro, a noble Dame born in Galicia, Spain, were brutally separated. Today, the echoes of this story endure, while the space is still visited and considered one of the iconic places of the Center of Portugal. We invite you to take a moment to read their mysterious story.

Saudações Académicas / Academic greetings,
Nuno Gonçalves,
Bernadete Ribeiro,
Catarina Silva

DOCTORAL CONSORTIUM COMMITTEE

Doctoral Consortium and Tutorials Co-chairs

Bernadete Ribeiro, CISUC, University of Coimbra (Portugal)

Catarina Silva, CISUC, University of Coimbra (Portugal)

General Co-chairs

Hélder P. Oliveira, APRP Co-chair, INESC TEC, University of Porto (Portugal)

Joan Andreu Sánchez, AERFAI Co-chair, Polytechnic University of Valencia (Spain)

Local Chair

Nuno Gonçalves, Institute of Systems and Robotics - University of Coimbra (Portugal)

Local Committee

Paulo Menezes, Institute of Systems and Robotics - University of Coimbra (Portugal)

Paulo Peixoto, Institute of Systems and Robotics - University of Coimbra (Portugal)

Cristiano Premebida, Institute of Systems and Robotics - University of Coimbra (Portugal)

Joel Arrais, Institute of Systems and Robotics - University of Coimbra (Portugal)

João Marcos, Institute of Systems and Robotics - University of Coimbra (Portugal)

Secretariat

Gustavo Bongiovi, Institute of Systems and Robotics - University of Coimbra (Portugal)

Volunteers

Ajnas Muhammed

Allan Freitas

Carlos Roxo

Guilherme Schardong

Iurii Medvedev

Miguel Leão

ORGANIZATION

Institute of Systems and Robotics - University of Coimbra (ISR-UC)

Centre for Informatics and Systems of the University of Coimbra (CISUC)

Associação Portuguesa de Reconhecimento de Padrões (APRP)

Asociación Española de Reconocimiento de Formas y Análisis de Imágenes (AERFAI)

SPONSORS

Entrust, Corp

Faculty of Sciences and Technology of the University of Coimbra

Coimbra City Council

Table of Contents

Extended Abstracts

A Journey Through Steganography Security Marks: Tracing Innovations from StegaStamp to StampOne

Farhad Shadmand
farhad.shadmand@isr.uc.pt
Institute of Systems and Robotics,
University of Coimbra,
Portugal

Luiz Schirmer
luizschirmer@unisinos.br
University of the Sinos
River Valley Rio de Janeiro,
Brazil

Nuno Gonçalves
nunogon@deec.uc.pt
Institute of Systems and Robotics,
University of Coimbra, Portugal
INCM, Lisbon, Portugal

INTRODUCTION

Modern machine-readable travel documents (MRTDs) and industrial authentication systems increasingly integrate a combination of biometric identifiers and security markings to prevent unauthorized replication or tampering. Beyond MRTDs, security pattern technologies are widely employed in other domains, particularly in brand protection and tax validation. Their use is prevalent in industries like luxury goods, wine and spirits, pharmaceuticals, and medical packaging.

This abstract introduces a set of advanced techniques for visual information embedding, drawing on the principles of steganography and digital watermarking. These methods are built upon modern deep learning frameworks and are designed to meet stringent requirements for security, robustness, and imperceptibility. The proposed solutions enable secure data integration within visual media and are applicable to high-stakes scenarios such as identity verification, counterfeit prevention, and brand authentication.

Steganography is the practice of embedding information within another medium in a manner that conceals the very existence of the hidden data. In digital image-based steganography, this typically involves two independent image-to-image neural networks: an encoder and a decoder. The encoder learns to embed a secret payload—such as text, another image, or binary code—into a cover image, producing an encoded image that remains perceptually similar to the original. The decoder, operating independently, is trained to extract and reconstruct the hidden message from the encoded image, even under conditions of distortion or noise.

The first highly robust steganography model in this field was StegaStamp [8], which introduced a pioneering approach by simulating a wide range of digital and physical distortions—including printer and scanner noise—during the training process. Its architecture employed a specialized U-Net encoder with a bottleneck layer for embedding the payload. However, while StegaStamp demonstrated resilience to various distortions, it struggled to preserve the structural integrity of input images. This limitation became particularly evident on semantically sensitive datasets, such as frontal face images, where it exhibited poor perceptual performance.

To overcome these challenges, we propose CodeFace [5] as a next-generation steganography framework that significantly improves the perceptual quality of encoded face images. Designed specifically for frontal facial data, CodeFace integrates a face-aware pipeline combining face detection and deep feature extraction to guide the encoding process. This enables the model to minimize perceptual discrepancies between the original and encoded images. We further deploy CodeFace as a security-enhancing layer for face images in Machine-Readable Travel Documents (MRTDs), offering both robustness and high visual fidelity in identity-sensitive applications.

Subsequently, RiemStega [3] was introduced to further enhance the performance of both the encoder and decoder components. This model incorporates a novel covariance-based loss function that operates in a Riemannian geometry space, encouraging the preservation of statistical consistency between original and encoded image features. In addition, RiemStega replaces the bottleneck structure used in StegaStamp's U-Net [4] with a self-attention mechanism, enabling more effective global feature interactions and improving the model's ability to embed and recover information with higher fidelity.

RoSteALS [1] is a lightweight and highly robust steganography framework based on generative adversarial networks (GANs), comprising only $300k$ parameters. Despite its compact architecture, the model exhibits strong resilience against various digital noise simulations, making it well-suited for purely digital communication scenarios. However, a key limitation of RoSteALS lies in its decoder's inability to accurately recover hidden messages from printed and re-scanned images, thereby restricting its applicability in print-based or physical media steganography.

Finally, we introduced StampOne [7], a steganography framework that bridges the gap between robust and non-robust models by placing greater emphasis on enhancing print-ability and resilience to real-world distortions. StampOne proposes a novel Reinforcement High-Frequency Strategy, designed to improve the robustness of embedded messages against transformations introduced by printing and scanning processes. The model incorporates a dedicated analysis-and-conversion module that preprocessed input data before encoding and decoding. This module aims to optimize the spectral distribution of features—specifically by enhancing high-frequency components and ensuring balanced frequency representations, thereby improving both visual fidelity and message recoverability in the final encoded images.

KEY CHALLENGES AND DESIGN TRADE-OFFS IN STEGANOGRAPHY STAMPS

The methods presented in this dissertation are grounded in the intersection of steganography, digital watermarking, and deep learning. By harnessing the advanced feature extraction and representational power of neural networks, we propose techniques that strive to optimize the trade-offs among several key performance criteria:

- **Perceptual Quality:** Maintaining a high degree of visual similarity between the encoded and original images, thereby concealing the presence of embedded information and preserving the natural appearance of the cover image.
- **Robustness:** Ensuring reliable extraction of the hidden payload under a wide range of digital and physical perturbations, including compression artifacts, additive noise, geometric distortions, and surface damage such as scratches or folds.
- **Capacity:** Maximizing the volume of information that can be embedded without degrading perceptual quality, while maintaining decoder reliability.
- **Security:** Providing strong protection against unauthorized decoding or tampering by designing models that resist reverse engineering, brute-force extraction, and adversarial attacks.

These objectives guide the design of the proposed frameworks, enabling the development of practical, scalable, and secure steganography systems suitable for real-world deployment.

PERFORMANCE COMPARING BETWEEN MODELS

Table 1 (A) presents the perceptual quality evaluation of encoded images. Among the compared models—CodeFace, StegaStamp, and Stam-

Table 1: (A) Quantitative evaluation of encoded image quality using perceptual similarity metrics. (B) Decoding performance on 40 printed encoded images, captured using a Samsung S22 Ultra smartphone. Models M1 and M2 correspond to the StampOne architecture employing Attention-VNet and UNetPlus backbones, respectively. Model M3 represents a non-robust baseline constructed with two independent Attention-VNet networks. The first four rows present results from high-robustness models, while the final two rows provide non-robust references, serving as a benchmark for evaluating decoder reliability under real-world print-capture conditions.

	(A) Encoded images quality			(B) **Bit acc (%)** - VGGFace2 [2]				
Methods	SSIM (⇑)	LPIPS (⇓)	ColorHisto (⇓)	6×6 cm	5×5 cm	4×4 cm	3×3 cm	2×2cm
StegaStamp [8]	0.93 ± 0.001	4.92 ± 1.6	6.11 ± 10.5	78	72	70	65	48
CodeFace [6]	0.95 ± 0.0002	3.06 ± 0.9	7.32 ± 6.1	55	55	50	38	15
StampOne (M1)	$\mathbf{0.98 \pm 0.00002}$	$\mathbf{1.25 \pm 0.4}$	$\mathbf{5.38 \pm 4.9}$	**100**	**100**	**100**	**95**	**62**
StampOne (M2)	0.96 ± 0.00007	2.74 ± 2.38	6.30 ± 4.07	88	85	72	63	43
Non-robust (M3)	0.92 ± 0.001623	1.04 ± 1.69	2.80 ± 60.8	0	0	0	0	0
RoSteALS [1]	0.95 ± 0.0006	$\mathbf{0.04 \pm 0.0003}$	$\mathbf{0.09 \pm 0.003}$	0	0	0	0	0

Table 2: Impact of three types of image under different noise types. 1000 images from COCO test dataset are used for the decoder performance evaluation. Bit accuracy (%) during decoding from encoded images is evaluated under various types and levels of noise. M1 and M2 represent StampOne models utilizing the Attention−VNet and UNetPlus architectures, respectively. On the other hand, M3 refers to a non-robust model constructed through the utilization of two instances of Attention−VNet.

	JPEG (%)			Gaussian (Std 0 to 1)			Resolution (Pixel)		
Methods	70	60	50	0.08	0.06	0.04	(60 × 60)	(80 × 80)	(100 × 100)
StegaStamp [8]	**100**	**100**	**100**	**100**	**100**	**100**	55	80	91
CodeFace [6]	80	88	88	55	75	86	2	11	36
RoSteALS [1]	87	90	94	23	35	53	**96**	97	98
StampOne (M1)	**100**	**100**	**100**	98	**100**	**100**	74	**98**	**100**
StampOne (M2)	97	99	**100**	88	96	99	72	94	99
Non-robust (M3)	0	0	0	13	46	84	0	0	22

pOne—StampOne demonstrates superior overall performance, particularly when using the Attention-VNet and UNetPlus backbones, as indicated in the table. Further evaluations of StampOne with alternative architectures are provided in the supplementary material.

In terms of SSIM, StampOne consistently achieves the highest scores among all robust models, indicating strong structural preservation. Although RoSteALS achieves slightly better results in the Color Histogram and LPIPS metrics, it fails to recover any messages from printed encoded images, limiting its practical applicability. In contrast, StampOne maintains both high perceptual quality and robustness to real-world printing conditions.

For print-based evaluation, a set of forty frontal face images from the VGGFace2 dataset was encoded and printed at various physical sizes, ranging from 2×2 cm to 6×6 cm (width $\times$ height), using a standard consumer-grade Brother L3270CDW color printer. To simulate real-world deployment conditions, decoding was conducted under uncontrolled lighting environments, with video recordings captured using a Samsung S22 Ultra smartphone.

The decoding performance of our proposed models—employing AttentionVNet and UNetPlus architectures—was benchmarked against established methods, including StegaStamp and CodeFace. As presented in Table 1(B), the Attention-VNet–based model consistently achieved the highest recovery accuracy from printed images, demonstrating superior robustness and confirming its effectiveness for printer-resilient steganographic applications. Additional cross-device results obtained using different smartphones are included in the supplementary material.

To assess decoder performance under real-world distortions, we conducted a series of experiments involving various noise conditions, including JPEG compression, Gaussian noise, resolution reduction, and contrast and brightness variations. Decoder effectiveness was quantified by the percentage of successfully recovered messages from the encoded images.

The results, summarized in Table 2, indicate that StampOne consistently outperforms competing models across most distortion scenarios. Notably, StegaStamp exhibits comparable robustness to StampOne under specific conditions, particularly JPEG compression and Gaussian noise, highlighting its resilience in digitally degraded environments.

REFERENCES

[1] Tu Bui, Shruti Agarwal, Ning Yu, and John Collomosse. Rosteals: Robust steganography using autoencoder latent space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 933–942, 2023.

[2] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face and Gesture Recognition*, 2018.

[3] Aniana Cruz, Guilherme Schardong, Luiz Schirmer, João Marcos, Farhad Shadmand, and Nuno Gonçalves. Riemstega: Covariance-based loss for print-proof transmission of data in images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Tucson, USA, 2025.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 234–241. Springer, 2015.

[5] Farhad Sadmand, Iurii Medvedev, and Nuno Gonçalves. Codeface: a deep learning printer-proof steganography for face portraits. *IEEE Access*, pages 1–1, 2021. doi: 10.1109/ACCESS.2021.3132581.

[6] Farhad Shadmand, Iurii Medvedev, and Nuno Gonçalves. Code face: A deep learning printer-proof steganography for face portraits. *IEEE Access*, 9:167282–167291, 2021.

[7] Farhad Shadmand, Ivan Medvedev, Lucas Schirmer, Joao Marcos, and Nuno Gonçalves. Stampone: Addressing frequency balance in printer-proof steganography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4367–4376, 2024.

[8] Matthew Tancik, Ben Mildenhall, and Ren Ng. Stegastamp: Invisible hyperlinks in physical photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2117–2126, 2020.

Assessing Prototype Generation strategies for Data Reduction in multilabel classification: A comparison between direct and adapted methods

Antonio Requena
antonio.requena@ua.es
Antonio Javier Gallego
jgallego@dlsi.ua.es
Jose J. Valero-Mas
jjvalero@dlsi.ua.es

Pattern Recognition and Artificial Intelligence Group,
University of Alicante,
San Vicente del Raspeig, Alicante, Spain

Abstract

Prototype Generation (PG) is one of the main approaches for data reduction. This family of techniques aim at decreasing the size of the training set while maintaining strong classification performance. Although PG strategies have been largely studied in multiclass scenarios, their direct application to multilabel classification (MLC) has been scarcely explored. This work analyzes the feasibility of applying PG mechanisms in multilabel contexts, comparing direct approaches with adaptations from the multiclass paradigm. Results show that methods specifically designed for MLC offer a better balance between efficiency and accuracy, especially in cases with strong label dependencies.

1 Introduction

Prototype Generation (PG) has become one of the main strategies for data reduction in supervised learning. By generating synthetic instances that aim to preserve the decision boundaries of the original dataset, PG allows for reducing both temporal and spatial costs, without remarkably compromising predictive performance [2]. Although these techniques have been widely studied in the context of multiclass classification, their application to multilabel learning has received considerably less attention.

Multilabel classification (MLC) introduces additional challenges compared to the multiclass setting, as each instance can be associated with multiple simultaneous labels, which may exhibit complex dependencies [1]. This increases the difficulty of preserving label structure during the reduction process. To address this, recent work has proposed adapting PG methods to operate directly within the multilabel space, leading to what we refer to as Multilabel Prototype Generation (MPG). These methods aim to reduce training set size while maintaining classification accuracy and preserving label correlation.

An alternative research direction explores the possibility of reusing PG methods originally designed for multiclass classification by first transforming the multilabel problem into one (or more) multiclass scenarios. Specifically, two representative strategies are identified: one based on decomposing the problem into multiple binary subproblems using the Binary Relevance paradigm (BR-PG), and another that transforms the multilabel dataset into a multiclass one through the Label Powerset technique (LP-PG). These transformation-based approaches allow for the application of established PG techniques in multiclass contexts, potentially simplifying implementation and tuning.

This study conducts a comparative analysis between the direct multilabel approach (MPG) and the transformation-based strategies (BR-PG and LP-PG), evaluating their performance in terms of data reduction, classification accuracy, and computational efficiency. For this purpose, the k-Nearest Neighbour (kNN) classifier is used as a reference, given its widespread adoption as a non-parametric method and its suitability for PG techniques, making it a representative testbed for evaluating the effectiveness of the proposed strategies [3].

2 Methodology

This work evaluates two transformation-based strategies for applying prototype generation (PG) techniques in multilabel classification: **BR-PG** and **LP-PG**. These strategies enable the reuse of well-established methods in multiclass contexts by converting the multilabel problem into a format compatible with such algorithms. The objective of this study is to analyze the effectiveness of both strategies in terms of data reduction, computational efficiency, and classification performance, and to compare them with methods specifically designed to operate in the multilabel space, such as MPG, whose implementation is detailed in the experimentation section.

2.1 BR-PG: Binary Relevance with Multiclass Reduction

The **BR-PG** strategy is based on the *Binary Relevance* paradigm, which decomposes the multilabel problem into a collection of binary classification subproblems, one for each label in the dataset. Each subproblem is then independently processed using a multiclass PG algorithm, aiming to reduce the training set for the k-NN classifier. After the reduction and independent classification of each subproblem, the binary predictions are combined to reconstruct the predicted label set for each instance.

Figure 1 illustrates the proposed approach. The multilabel (ML) training set $\mathcal{T}^{ml}$ is split into L—number of labels—multiclass (MC) subsets (ML $\rightarrow$ MC), each corresponding to a binary classification task for an individual label λ_i. These subsets are then reduced using the multiclass PG method, resulting in a compact set $\mathcal{T}_R^{\lambda_i}$ for each label. Subsequently, classification is performed independently using the k-NN algorithm. The predicted outputs $\hat{y}^{\lambda_1}, \hat{y}^{\lambda_2}, \ldots, \hat{y}^{\lambda_L}$ are finally combined through an aggregation procedure (MC $\rightarrow$ ML) to reconstruct the complete multilabel prediction for each query sample.

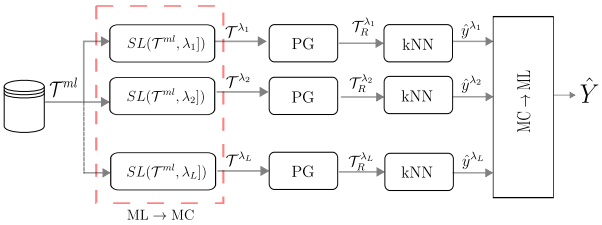


Figure 1: Scheme of the BR-PG strategy: binary decomposition and reduction with multiclass PG.

This approach offers important advantages, such as the direct reuse of existing multiclass methods without the need for redesign, and high scalability when the number of labels is moderate. However, a notable limitation is that it does not model label correlations, which may lead to inconsistent or suboptimal predictions in domains with strong label dependencies.

2.2 LP-PG: Label Powerset Transformation with Multiclass Reduction

The **LP-PG** strategy transforms the multilabel classification problem into a multiclass classification problem using the well-known *Label Powerset* (LP) technique. This transformation assigns a unique class to each combination of labels present in the training set, thereby reducing the multilabel task to a traditional multiclass scenario.

While this technique allows the direct application of multiclass algorithms, such as PG methods and conventional k-NN classifiers, its main drawback is the exponential growth in the number of classes as the number of original labels increases. More precisely, LP transforms the initial L-size multilabel space into a multiclass space with 2^L classes.

After performing the Label Powerset (LP) transformation, a multiclass prototype generation (PG) method is applied to the transformed data, yielding a reduced prototype set $\mathcal{T}_R^{LP}$.

Then, a multiclass k-NN classifier is directly applied to the transformed instances, and the predicted classes are mapped back to their corresponding original multi-label sets.

Figure 2 illustrates the overall flow of this strategy. The multilabel training set $\mathcal{T}^{ml}$ is transformed using the LP method, reduced through multiclass PG, and finally classified to produce multiclass predictions that are later reinterpreted as multilabel outputs.

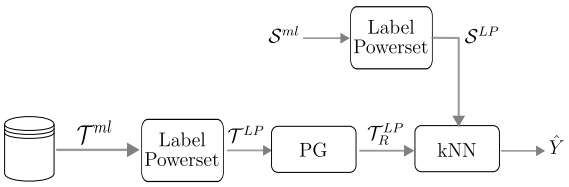


Figure 2: Scheme of the LP-PG strategy: LP transformation, multiclass PG reduction, and classification.

The main advantage of this approach lies in its ability to leverage the extensive knowledge developed for multiclass classification methods, as well as the optimized PG mechanisms established in that context. However, its performance may degrade due to class space fragmentation, particularly as the number of label combinations increases substantially.

3 Experimentation

This section compares the data reduction strategies for multilabel classification, focusing on the transformation-based methods **BR-PG** and **LP-PG**, and the direct multilabel approach **MPG**.

To this end, a set of twelve datasets from the *Mulan* library has been used, including *birds*, *Corel5k*, *emotions*, *genbase*, *medical*, *rcv1subset1*, *rcv1subset2*, *rcv1subset3*, *rcv1subset4*, *scene*, and *yeast*. These datasets collectively encompass several thousand instances and show a high degree of variability in terms of attributes, labels, and inter-label dependencies, allowing for a robust evaluation.

The **MPG** strategy operates directly in the multilabel space, avoiding transformations into binary or multiclass schemes. It employs specific algorithms to group the original data and generate representative prototypes with aggregated labels, aiming to preserve inter-label dependencies—an aspect crucial for maintaining predictive quality, at least in principle. Subsequently, the ML-*k*NN multilabel classifier is applied to the reduced set $\mathcal{T}_R^{ml}$. Figure 3 illustrates the complete MPG process.

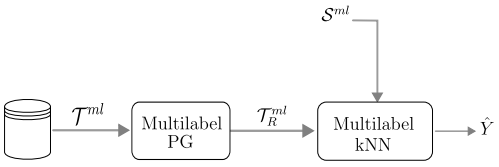


Figure 3: General scheme of the MPG strategy: direct reduction in the multilabel space and classification with ML-*k*NN.

To ensure comparability between approaches, all configurations were evaluated under identical experimental conditions. The same procedures for partitioning, reduction, and classification were applied across all datasets, using a fixed random seed. Performance was assessed using a broad set of metrics to ensure reproducibility and support robust conclusions about each method's suitability. However, due to space constraints, we report results only for the Macro F1 (F_1^M) score.

4 Results

4.1 Effectiveness of the algorithms

The results obtained (see Fig. 4) allow us to identify some general trends in the behavior of the evaluated strategies.

The **MPG** approach shows acceptable data reduction rates while maintaining adequate levels of classification performance.

Both **BR-PG** and **LP-PG** exhibit consistent behavior across the evaluated datasets, suggesting that transformation-based adaptations represent a viable alternative for applying traditional prototype generation techniques in multi-label settings.

The direct application of prototype generation to the multi-label domain, as implemented in **MPG**, appears to better preserve the original data structure, which could be beneficial in scenarios where label relationships constitute a valuable source of information.

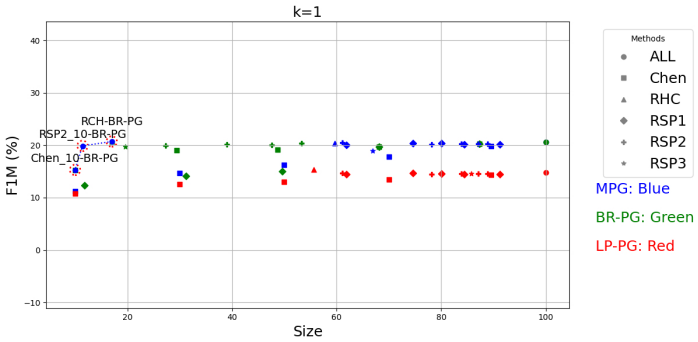


Figure 4: Results in terms of classification performance (F_1) and resulting set size of the MPG, BR-PG, and LP-PG frameworks.

4.2 Computational cost analysis

Beyond predictive performance and data reduction, assessing the computational cost of each strategy is also essential. In this regard, Table 1 provides a qualitative comparison of these computational aspects.

	Preprocess			
Method	**Transformations**	**Reduction**	**Classification**	**Labels**
MPG	–	1	1	L
BR-PG	L	L	L	2
LP-PG	1	1	1	2^L

Table 1: Comparative summary of computational cost across transformation, reduction, and classification steps for each approach.

As it may be observed, **BR-PG** requires repeating the transformation, reduction, and classification processes L times—once for each label—while **LP-PG** and **MPG** perform each step only once. However, **LP-PG** operates in a label space that can grow exponentially with L, whereas **MPG** avoids transformation by working directly in the multilabel space.

5 Conclusions and Future Work

Preliminary results show that **MPG (Multi-label Prototype Generation)** outperforms transformation-based strategies in both efficiency and the preservation of multilabel structure. By operating directly in the multilabel space, MPG avoids the loss of information inherent in conversions to binary or multiclass formats, leading to more accurate and consistent predictions.

Its strong performance under class imbalance also suggests greater robustness, which will be further explored by introducing artificial noise in the labels. This future work aims to validate the resilience of MPG methods to label noise and investigate adaptive extensions that consider local label distributions during prototype generation.

Acknowledgments. This work was partially funded by the Generalitat Valenciana through project CIGE/2023/216 and the Spanish Ministerio de Ciencia, Innovación y Universidades through project PID2023-148259NB-I00 (LEMUR).

References

[1] Á. Arnaiz-González, J. F. Díez-Pastor, J. J. Rodríguez, and C. García-Osorio. Study of data transformation techniques for adapting single-label prototype selection algorithms to multi-label learning. *Expert Systems with Applications*, 109:114–130, 2018.

[2] J.J. Valero-Mas, A.J. Gallego, P. Alonso-Jiménez, and X. Serra. Multilabel prototype generation for data reduction in k-nearest neighbour classification. *Pattern Recognition*, 135:109190, 2023.

[3] M.L. Zhang and Z.H. Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048, 2007.

Automatic defect detection in ornamental rocks

Marco Tereso
d41655@alunos.uevora.pt
Teresa Gonçalves
tcg@uevora.pt
Luís Rato
lmr@uevora.pt

Universidade de Évora
Évora, PT

INTRODUCTION

Innovation, improvement, and simplification of processes are, nowadays, keywords for the industry. In an era where digital transformation and artificial intelligence are the order of the day, it is understood that companies are beginning to take seriously the advantages brought by these two areas [2].

This research work addresses a real problem in the stone industry and seeks to automate processes that are currently manual. The main objective is to automatically mark defects in the stone. According to data from COMPETE 2030, Portugal is the seventh largest producer of natural stone in the world [1]. The same source states that the Portuguese stone industry generated 1.2 billion euros in 2023 and employs 14,000 people, numbers that represent 0.6% of the total exports, and demonstrate the importance of this sector of activity for our country. Companies are more competitive if they can deliver quality products, optimize processes, and shorten delivery times. It is on this principle that this research is based, to develop automatic detection models that support a very important economic activity for our country.

RESEARCH PROBLEM

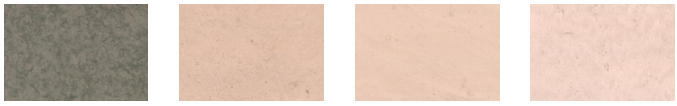
Stone blocks are extracted from quarries and transported to processing units. In the processing units, they are sawn using cutting devices made up of diamond cables, or with cutting discs, slicing the block into plates. After this abrasive cutting process, the plates are taken to a polishing machine, responsible for polishing one of the faces. After this process, the plates are transported to the CNC cutting machines. Transport within the manufacturing unit is always carried out using overhead cranes installed at the top of the facilities. To support this and other projects, a scanning scanner was installed, which reproduces an image of stone plate as it slides out of the polisher. These images are saved on a data server and accessible, and are the basis of all this research. In CNC, it is the operator's responsibility to select the defects in the plate, using contrast and image scanning techniques, carried out. The nesting process can later be carried out, which consists of positioning the stones to be cut, on the CNC computer over the image of the plate, in which it is the operator's responsibility to try to waste as little stone as possible, avoiding defective areas. Defect marking, being a process that is done manually by the operator of each cutting CNC, can differ from operator to operator. This differentiation in defect marking is essentially related to the experience of operator, as well as his eye health.

The main objective of this research is to make this classification homogeneous, guaranteeing a final product of higher quality, optimizing time and resources, with all processing being done while the stones move to the cutting CNC, allowing the nesting process to also be automatic, and in this way reduce the waste of stone from each sheet. Once the problem was identified and the objective defined, we began to obtain a set of images to build a dataset to support an automatic defect marking model.

The major initial goal of this research is related to the detection of all defective areas in the image. In the long term, it is intended to classify the defects according to their type. Implementing the possibility of defining which types of defects may be included in the marking, allowing classification of first or second category works, as well as, deselecting defects that may not be defects in certain classes of stone.

METHODOLOGY

Initially, a survey was carried out to select the types of rocks more commercialized by the company supporting this project. From a list of 18 classes, the 4 most representative were selected and a set of plate images of those types of rocks was assembled. Figure 1 illustrates the types of rocks selected and Table 1 presents the number of plate images that compose the dataset. The images have variable sizes, with average values being 2600x1400 px, with the largest images measuring 3200x2000 px.



(a) (b) (c) (d)

Figure 1: This figure illustrates a Selection of illustrative images of the rock types selected for this work: (a) CADOICO; (b) SBM; (c) SBR; (d) VMF.

Rock Name	CADOICO	SBM	SBR	VMF
Total images	12	37	6	21

Table 1: Total number of images analyzed, distributed by classes.

After collecting the images, it was necessary to mark the defects in the images manually, with the help of an operator with 18 years of experience, using a graphic editing program. The defect marking process, included carrying out an inventory of defects to understand their categories and speciations. The types of defects considered were: Glass Lines; Crack; Fossils; Holes; Finishing Defects; Color spots.

The adopted approach began with marking defects in the original images, in this case, the color blue was used (Figure 3b). Next, the image was divided into patches. Then, the patches were classified as defective or non-defective. Then, the image was divided into patches (similar to what is represented by Figure 2). A size of 28x28 was defined for the patches as it is considered the minimum size for a concentration of pixels not to be deemed a defect, approximately the size of a 0.05€ coin, a measure defined by professionals. The patches named with the name of the original image, row number, and column number of the crop (for example: CADOICO-01_1_1.jpg), facilitating the subsequent process of reconstructing the original image. An algorithm was then implemented that analyzed the percentage of pixels marked as defective (blue pixels), considering all patches that contained more than 1% of pixels marked in blue as defective. This process allowed for the creation of an Excel file with the names of all defective images, so that the original images could be cropped, and the patches separated by the GOOD and DEFECTIVE classes. In terms of comparison, images were recreated that overlaid this last analysis with the previous one (marked in blue), resulting in Figure 3c (red overlay of patches with more than 1% defective pixels over the blue markings).After all this initial work, which was essential in building the dataset, the classification model was applied to the patches. Finally, a classification model is applied to the dataset. After the model is executed, the original image is reconstructed by replacing, in this case, the patches classified as defective with completely green patches (Figure 3d).

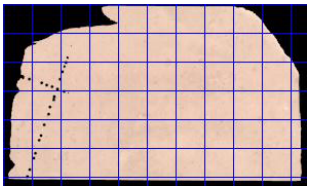


Figure 2: Cropping of the original image into 28x28 px patches

Rock Name	CADOICO	SBM	SBR	VMF	Total	%
Total Images	58785	189820	28605	89228	366438	—
GOOD Images	56719	183462	25998	83355	349534	95.4%
DEFECTIVE Images	2066	6358	2607	5873	16904	4.6%
%	16%	51.8%	7.8%	24.4%	100%	—

Table 2: Total number of images analyzed, distributed by classes.

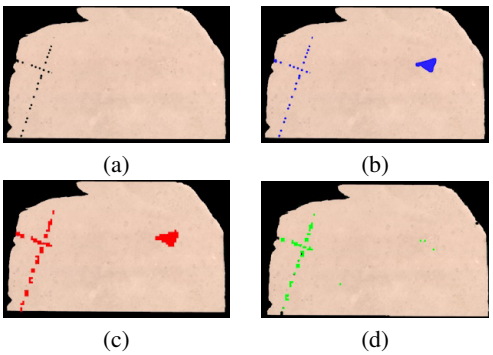


Figure 3: Illustration of representative images of process: (a) original image; (b) image marked by experienced operator; (c) image with the markings of all images with at least 1% of defective pixels; (d) image classified by model.

After the clipping process, a dataset of 366,438 patches was obtained. Table 2 presents its statistics detailed by type of rock. As can be observed, the dataset is highly imbalanced (only 4.6% of the total patches are defective).

The second step aimed to go through the folders of each of the classes, of the dataset of marked images, and move to a DEFECTIVE folder all Then, patches with at least 1% of pixels being defective were labelled as defective, and the rest to a folder called GOOD.

Once the images were moved, with the help of an algorithm that used the data from the Excel file, all the pieces of the original image dataset were distributed, using the same type of separation, that is, all the pieces with defects, but original, were moved without any marking to the DEFECTIVE folder and all the others to the GOOD folder.

Finally, a simple classification model was trained. In which the main objective was to evaluate its performance without major parameter adjustments, to subsequently reconstruct the image piece by piece and signal the defects marked by the algorithm. In this process, an algorithm was created that, through the nomenclature of each piece, reconstructed the image, taking the pieces classified as good and replacing the missing images with equal-sized red squares. An overlay was made on the marked images, in order to have a better view of the final result. Figure 3 illustrates an example, using the same image, showing the original format, the manual marking and the overlap of the manual marking with the marking performed by the algorithm.

Classification Model Used

To build the classification model, a split of 70-15-15 for training, validation and test purposes was made, keeping all the patches of the same image in the same subset to prevent data leakage.

The RGB values of each pixel were used as input to the Random Forest algorithm with SMOTE resampling.

The classification model used in this first experiment was a Random Forest, based model with SMOTE (Synthetic Minority Over-sampling Technique) re-sampling. The first stage of the pipeline is the use of SMOTE, an oversampling technique aimed at correcting the class imbalance, commonly used in binary classification problems [5], [3]. The second stage consists of the inclusion of the Random Forest classifier, which is a method based on multiple decision trees [4]. The default hyperparameters of the algorithm were used, and the make_pipeline function was used so that the defined steps were applied in a sequential and coherent manner.

Analyzing the results of Table 3, we conclude that the model has excellent performance in classifying patches of the GOOD class, but poor performance in DEFECTIVE class. The weak performance of precision (DEFECTIVE class) represents a high number of false positives; however, visually it is possible to confirm that these occur alongside other true positives. Thus, in performance evaluation, it will be necessary in the future to consider metrics (using some form of post-processing) that take into account whether false positive errors (and false negatives) occur in isolation (a more serious situation) or alongside, possibly on the edges, of areas correctly identified as DEFECTIVE. The precision indicates that we have many false positives and the recall shows that the model only classifies half of the actual defective examples. As for the accuracy values, they cannot be taken into account due to the fact that we are dealing with unbalanced data, which can be, and is in this case, a misleading metric.

Class	Precision	Recall	F1-score
GOOD	0.97	0.84	0.90
DEFECTIVE	0.13	0.50	0.21
Accuracy			0.82

Table 3: Results of the evaluation of the implemented model.

PARTIAL RESULTS

After running the model and reconstructing the images to their original dimensions, it was possible to visually analyze the results. Figure 3 provides an overview of the steps followed in the strategy. The algorithm is this preliminary version has difficulty on correctly classifying patches with a smaller number of defective pixels (cracks) as well as defective areas with colors very similar to the predominant color of the rock.

In Figure 3d we can observe that some of the holes in the stone are not completely green, which is due to the fact that in our dataset we have many patches that are completely black, from the area of no interest in the image, which may be causing the model not to classify completely black patches as defects.

CONCLUSIONS AND FUTURE WORK

These are the data from a preliminary work, whose objective was essentially to test the methodology. The next steps involve creating a more capable ML model. The fact that this is a problem with unbalanced data poses some challenges in the approach and implementation of improvements to the model presented here. As future work, we will seek to create a more stable ML model in order to achieve better results and move closer to the optimal.

REFERENCES

[1] Compete 2030 and C. S. Pinto. Portugal É o 7o maior produtor mundial de pedra natural, 2024. Accessed in: 19-05-2025.

[2] Liliana Isabel Esteves Gomes. Transformação digital e inteligência artificial nos serviços de informação: inovação e perspectivas para a ciência da informação no mundo pós-pandemia. *Revista Ibero-Americana de Ciência da Informação*, 15(1), 2022.

[3] Mehdi Imani, Ali Beikmohammadi, and Hamid Reza Arabnia. Comprehensive analysis of random forest and xgboost performance with smote, adasyn, and gnus under varying imbalance levels. *Technologies*, 13(3):88, 2025.

[4] Nour El Islem Karabadji, Abdelaziz Amara Korba, Ali Assi, Hassina Seridi, Sabeur Aridhi, and Wajdi Dhifli. Accuracy and diversity-aware multi-objective approach for random forest construction. *Expert Systems with Applications*, 225:120138, 2023.

[5] Abdoulaye Sakho, Emmanuel Malherbe, and Erwan Scornet. Do we need rebalancing strategies? a theoretical and empirical study around smote and its variants. *arXiv preprint arXiv:2402.03819*, 2024.

Deep Learning in Mild Cognitive Impairment Diagnosis using Eye Movements and Image Content in Visual Memory Tasks

Tomás Silva Santos Rocha[1]
tomasrocha01@tecnico.ulisboa.pt
José Santos-Victor[1]
jose.santos-victor@tecnico.ulisboa.pt
Anastasiia Mikhailova[1, 2]
amikhailova@uchicago.edu
Moreno I. Coco[3]
moreno.coco@uniroma1.it

[1] Institute for Systems and Robotics, LARSys
Instituto Superior Técnico, Universidade de Lisboa
Lisbon, Portugal

[2] Department of Psychology & Institute for Mind and Biology
University of Chicago
Chicago, USA

[3] Department of Psychology
Sapienza Università di Roma
Rome, Italy

INTRODUCTION

The WHO's "Global Status Report on the Public Health Response to dementia" [10] highlights dementia as a growing global health issue, affecting over 55 million people, projected to rise to 139 million by 2050, with a disproportionate impact expected in middle-income countries. This emphasizes the urgent need for diagnostic tools and interventions, particularly as populations age. Dementia is characterized by cognitive decline that interferes with daily life, including memory loss and impaired thinking. Alzheimer's disease is the most common form, though other types include vascular, frontotemporal, and Lewy body dementia [2]. Before developing dementia, patients go through a stage called Mild Cognitive Impairment (MCI), experiencing mild cognitive changes that do not significantly interfere with daily life activities, *e.g.* impacts on memory function. Early detection of MCI provides a critical window for interventions that may delay progression to dementia [11].

RESEARCH PROBLEM

MCI is influenced by age, genetics (*e.g.* APOE $\varepsilon 4$ allele), lifestyle, and cardiovascular health and is associated with biomarkers such as hippocampal atrophy and amyloid-beta accumulation [6, 8]. However, diagnosing it in a timely manner remains challenging, as the limited availability of specialists restricts the widespread use of diagnosing tools, such as standard cognitive tests like MMSE and MoCA or neuroimaging tools. When combined with delayed symptom recognition, this often results in MCI being detected only when cognitive decline is noticeable, limiting early intervention opportunities [1, 3].

THEORETICAL FRAMEWORK

To tackle these challenges, digital assessments on computers, tablets, and smartphones are emerging as practical and cost-effective alternatives [13]. When integrated with wearable sensors, cameras, or eye-tracking devices, these tools capture many relevant data without adding time costs, enhancing their diagnostic potential [7].

Eye-tracking integration in MCI diagnosis has gained attention, with studies suggesting that eye movement patterns could serve as early indicators of cognitive decline [19].

Eye-movements: A window to the memories

Eye movements, essential for visual perception, offer insight into cognitive processing and memory. Key movements include saccades, rapid gaze shifts between points, and fixations, where the eyes pause to process information. Saccades allow quick repositioning but temporarily suppress visual intake, while fixations enable detailed analysis, with durations varying by task [12]. Fixations are particularly useful, as their longer periods allow for easier detection. Eye movements significantly impact memory mechanisms, aiding encoding and retrieval. Increased fixations and shifts in gaze while encoding a scene improve memory recall, as eye movements facilitate scene exploration [5]. During encoding, eye movement patterns differ notably between MCIs and healthy individuals. MCI patients often have delayed saccadic responses and less accurate saccade targeting, leading to inefficient visual exploration. Their fixations are longer but fewer, possibly indicating attention and visual processing difficulties. This reduced fixation frequency and longer saccadic latency contribute to limited scene exploration, impacting their ability to recognize and remember visual details [19]. In the literature, we also see that image content influences eye movements by engaging in a process where our eyes selectively focus on elements perceived as necessary, consequently influencing the memory processes [15].

Deep Learning and MCI

To enhance the digital tools deep learning algorithms are being employed, improving diagnostic accuracy and scalability, thus making the timely detection of MCI a possibility [19].

Algorithms based on various methods have been employed, such as Recursive Neural Networks (RNN), Convolutional Neural Networks (CNN), autoencoders, transformers, and others [14]. The transformer based architectures usually perform better than the others, however they require large amounts of data to be properly trained [9, 14]. One problem with recent studies is the fact that, typically there is no differentiation between MCI patients and patients with more advanced dementias, such as Alzheimer's, which can bring biases to the results obtained [14].

RESEARCH OBJECTIVES

Building upon recent advancements, this study explores deep learning models that use eye-tracking data collected from MCI patients and Healthy Controls (HCs). The participants perform a visual long-term active memory task to predict MCI, which have been shown to have better diagnostic accuracy [17, 19]. Additionally, this research investigates incorporating image content, a topic that has not been extensively researched, and memory performance to try to enhance the diagnostic process further.

PROPOSED METHODOLOGY

The models will use the data from MCI patients and Healthy Controls (HCs) collected by us and also the data available in Coco *et al.* [4] performing a similar task (44 participants: 24 MCI patients — 71.92 ± 9.06 years old, 9.83 ± 4.50 schooling years; 20 HCs — 68.50 ± 8.79 years old, 11.05 ± 5.10 schooling years).

The tasks is a visual long-term active memory task divided in two sections: an encoding phase, where, in both studies, the participants are displayed a series of images that they are asked to try to remember; and a recognition phase, where in our task the participants are shown again a series of images, however, this time, they are asked to say whether they remember seeing the image. In Coco *et al.* [4] they are presented with two images at the same time, and they have to choose which of the images they remember seeing before.

While this task is being performed, the eye movements of the participants are being recorded. The data from the encoding phase will then be used to train variations of the VTNet, a model that as already shown to have the capability of predicting if a person as Alzheimer's[16]. This model receives two inputs: a visual representation of the eye-movements, which is then processed via a CNN; and a time-series representation of the eye-movements, which is processed via a RNN.

Method	Sensitivity	Specificity
Sriram *et al*. [16]	70 ± 0.02	73 ± 0.02
Ours	68.42 ± 27.26	76.47 ± 27.64

Table 1: Models sensitivity and specificity (mean $\pm$ standard deviation) for Sririam *et al*. [16], using Alzheimer's disease patients and Scanpaths, and our results using MCI patients and gaze heatmaps.

WORK PLAN

In the first part of the project we started by training various models with on the data collected from Coco *et al*. [4]. For each model we experimented with various types of visual inputs: scanpaths, gaze heatmaps, image content (*i.e.* image seen) and a combination of the gaze heatmaps and the image content. We observed that gaze heatmaps lead to better performing models, having been able to reach comparable results to Sriram *et al*. [16], while performing under more challenging conditions, as Alzheimer's patients are easier to differentiate than MCI patients (Tab. 1).

After our first experiment we proved that this architecture can be used to predict if a person as MCI. We are now continuing to collect more data, having already collected data from 21 participants (9 HCs and 12 MCI patients). The participants need to be between 50 and 85 years old and have at least 4 years of schooling in order to account for cognitive biases related to both age and education. Also, the patients need to be clinically diagnosed with neurodegenerative MCI and not any more severe condition. We will then combine the two datasets to have a more significant sample size and a more robust model. Also, to address the significant standard deviation we are looking into methods such as bootstrap aggregation[18].

EXPECTED CONTRIBUTIONS

With this study, we pretend to create a diagnosis test capable of doing an initial triage of people with MCI. We also believe that it has the capability of being adapted to work on laptops using their respective built-in cameras increasing the scalability and availability of the diagnosis.

ACKNOWLEDGEMENT: This study is funded by Lisbon ELLIS Unit, the Center for Responsible AI (PRR), LARSyS FCT funding (DOI: 10.54499/LA/P/0083/2020, 10.54499/UIDP/50009/2020, and 10.54499/UIDB/50009/2020)

REFERENCES

[1] Alissa Bernstein Sideman, Tala Al-Rousan, Elena Tsoy, Stefanie D Piña Escudero, Maritza Pintado-Caipa, Suchanan Kanjanapong, Lingani Mbakile-Mahlanza, Maira Okada de Oliveira, Myriam De la Cruz-Puebla, Stelios Zygouris, et al. Facilitators and barriers to dementia assessment and diagnosis: Perspectives from dementia experts within a global health context. *Frontiers in neurology*, 13: 769360, 2022. doi: 10.3389/fneur.2022.769360.

[2] John C. S. Breitner. Dementia—epidemiological considerations, nomenclature, and a tacit consensus definition. *Journal of Geriatric Psychiatry and Neurology*, 19(3):129–136, 2006. doi: 10.1177/0891988706291081. PMID: 16880354.

[3] Yi Chen, Melinda C Power, Francine Grodstein, Ana W Capuano, Brittney S Lange-Maia, Ali Moghtaderi, Emma K Stapp, Joya Bhattacharyya, Raj C Shah, Lisa L Barnes, et al. Correlates of missed or late versus timely diagnosis of dementia in healthcare settings. *Alzheimer's & Dementia*, 20(8):5551–5560, 2024. doi: 10.1002/alz.14067.

[4] Moreno I Coco, Gabriella Merendino, Giuseppe Zappalà, and Sergio Della Sala. Semantic interference mechanisms on long-term visual memory and their eye-movement signatures in mild cognitive impairment. *Neuropsychology*, 35(5):498, 2021. doi: 10.1037/neu0000734.

[5] Claudia Damiano and Dirk B. Walther. Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, 184:119–129, 2019. ISSN 0010-0277. doi: 10.1016/j.cognition.2018.12.014.

[6] Clifford R. Jack Jr., David A. Bennett, Kaj Blennow, Maria C. Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M. Holtzman, William Jagust, Frank Jessen, Jason Karlawish, Enchi Liu, Jose Luis Molinuevo, Thomas Montine, Creighton Phelps, Katherine P. Rankin, Christopher C. Rowe, Philip Scheltens, Eric Siemers, Heather M. Snyder, Reisa Sperling, Contributors, Cerise Elliott, Eliezer Masliah, Laurie Ryan, and Nina Silverberg. Nia-aa research framework: Toward a biological definition of alzheimer's disease. *Alzheimer's & Dementia*, 14(4):535–562, 2018. doi: 10.1016/j.jalz.2018.02.018.

[7] Aoyu Li, Jingwen Li, Dongxu Zhang, Wei Wu, Juanjuan Zhao, and Yan Qiang. Synergy through integration of digital cognitive tests and wearable devices for mild cognitive impairment screening. *Frontiers in Human Neuroscience*, 17:1183457, 2023. doi: 10.3389/fnhum.2023.1183457.

[8] Gill Livingston, Jonathan Huntley, Kathy Y Liu, Sergi G Costafreda, Geir Selbæk, Suvarna Alladi, David Ames, Sube Banerjee, Alistair Burns, Carol Brayne, et al. Dementia prevention, intervention, and care: 2024 report of the lancet standing commission. *The Lancet*, 404(10452):572–628, 2024. doi: 10.1016/S0140-6736(24)01296-0.

[9] Sumit Madan, Manuel Lentzen, Johannes Brandt, Daniel Rueckert, Martin Hofmann-Apitius, and Holger Fröhlich. Transformer models in biomedicine. *BMC Medical Informatics and Decision Making*, 24 (1):214, 2024. doi: 10.1186/s12911-024-02600-5.

[10] World Health Organization et al. Global status report on the public health response to dementia, 2021. URL `https://digitalcommons.fiu.edu/srhreports/health/health/65/`.

[11] Ronald C Petersen. Mild cognitive impairment. *CONTINUUM: lifelong Learning in Neurology*, 22(2):404–418, 2016. doi: 10.1212/CON.0000000000000313.

[12] Daniel Reisberg. *The Oxford handbook of cognitive psychology*, chapter 5. Eye Movements, pages 69–82. OUP USA, 2013. doi: 10.1093/oxfordhb/9780195376746.001.0001.

[13] Marwan N Sabbagh, M Boada, S Borson, M Chilukuri, PM Doraiswamy, B Dubois, J Ingram, A Iwata, AP Porsteinsson, KL Possin, et al. Rationale for early diagnosis of mild cognitive impairment (mci) supported by emerging digital technologies. *The journal of prevention of Alzheimer's disease*, 7:158–164, 2020.

[14] Hasnain Ali Shah, Salman Khalil, Sami Andberg, Anne M. Koivisto, and Roman Bednarik. Eye tracking based detection of mild cognitive impairment: A review. *Information Fusion*, 122: 103202, 2025. ISSN 1566-2535. doi: j.inffus.2025.103202.

[15] David Souto and Dirk Kerzel. Visual selective attention and the control of tracking eye movements: a critical review. *Journal of Neurophysiology*, 125:1552–1576, 2021. doi: 10.1152/jn.00145.2019.

[16] Harshinee Sriram, Cristina Conati, and Thalia Field. Classification of alzheimer's disease with deep learning on eye-tracking data. In *Proceedings of the 25th International Conference on Multimodal Interaction*, pages 104–113, 2023. doi: 10.1145/3577190.3614149.

[17] Koh Tadokoro, Toru Yamashita, Yusuke Fukui, Emi Nomura, Yasuyuki Ohta, Setsuko Ueno, Saya Nishina, Keiichiro Tsunoda, Yosuke Wakutani, Yoshiki Takao, Takahiro Miyoshi, Yasuto Higashi, Yosuke Osakada, Ryo Sasaki, Namiko Matsumoto, Yuko Kawahara, Yoshio Omote, Mami Takemoto, Nozomi Hishikawa, Ryuta Morihara, and Koji Abe. Early detection of cognitive decline in mild cognitive impairment and alzheimer's disease with a novel eye tracking test. *Journal of the Neurological Sciences*, 427: 117529, 2021. ISSN 0022-510X. doi: 10.1016/j.jns.2021.117529.

[18] Ioanna Vourlaki, Costas Balas, George Livanos, Manos Vardoulakis, George Giakos, and Michalis Zervakis. Bootstrap clustering approaches for organization of data: Application in improving grade separability in cervical neoplasia. *Biomedical Signal Processing and Control*, 49:263–273, 2019. doi: 10.1016/j.bspc.2018.12.014.

[19] Alexandra Wolf, Kornkanok Tripanpitak, Satoshi Umeda, and Mihoko Otake-Matsuura. Eye-tracking paradigms for the assessment of mild cognitive impairment: a systematic review. *Frontiers in Psychology*, 14, 2023. ISSN 1664-1078. doi: 10.3389/fpsyg.2023.1197567.

Effort Reduction through Interactive Machine Translation and Quality Estimation: Innovations and Applications

Ángel Navarro[1]
annamar8@prhlt.upv.es
Francisco Casacuberta[12]
fcn@prhlt.upv.es
Roberto Paredes[1]
rparedes@prhlt.upv.es

[1] PRHLT
Universitat Politècnica de València
Valencia, Spain

[2] ValgrAI
Camí de Vera s/n, 46022
Valencia, Spain

INTRODUCTION

The automatic processing of natural language has become a central challenge in machine learning, combining the complexity of human communication with the need for scalable, data-driven solutions. As a subfield of artificial intelligence, Natural Language Processing (NLP) aims to develop computational models capable of understanding, generating, and translating human language. This area has witnessed remarkable progress thanks to advances in deep learning, the availability of large datasets, and the design of increasingly sophisticated architectures.

Among the most impactful applications of NLP is Machine Translation (MT), which seeks to enable cross-linguistic access to information. From its early rule-based systems to current neural architectures, MT has evolved into a mature field with systems that achieve fluent and accurate translations in many language pairs [10]. However, these models still face key limitations, such as their dependence on large annotated corpora, difficulties adapting to specific domains or user needs, and the generation of plausible but incorrect outputs [2].

To address these challenges, hybrid approaches that integrate human expertise into the machine-learning loop have gained attention. One such approach is Computer Assisted Tools (CAT), where translators work alongside automatic systems using features like translation memories, terminology databases, and MT suggestions. Building on this, a more interactive paradigm has emerged: Interactive Machine Translation (IMT), in which the system dynamically adapts to user feedback during the translation process.

IMT reframes the translation task as a collaborative problem, where the human and the machine iteratively refine the output. Unlike post-editing, where the user corrects a static translation, IMT systems update their predictions in real-time based on partial user input. This setting poses new challenges for the design of machine learning systems capable of responding effectively to dynamic, sparse, and often implicit human feedback.

RESEARCH PROBLEM

Traditional machine translation systems —whether statistical or neural— are designed to generate complete translations autonomously, without user interaction during the generation process. In these systems, human feedback is typically limited to post-editing, where the user corrects the final output without influencing the underlying model's behavior in real-time. This creates a rigid workflow that does not exploit the full potential of human-machine collaboration.

IMT proposes a shift in this paradigm by incorporating the user into the inference loop. The system receives partial corrections or signals from the user and updates its translation hypothesis accordingly. From a machine learning perspective, this setting introduces a dynamic feedback loop, where the model must adapt to user inputs that are often sparse, noisy, and context-dependent.

The central challenge addressed in this thesis is how to model and optimize this interactive process to minimize the human effort required while maintaining or improving translation quality. This includes designing adaptive algorithms capable of integrating user feedback efficiently, determining what kind of feedback is most useful, and deciding when and how to solicit or respond to it.

Moreover, the problem involves investigating how auxiliary modules —such as Quality Estimation (QE)— can guide the interaction by identifying uncertain or potentially erroneous segments and how recent advances in generative language models can be adapted for this more collaborative and incremental task [5, 8, 12]. Unlike conventional generation tasks, interactive translation requires models to respond to user input, maintain coherence under partial constraints, and align closely with human intent —posing unique challenges for adaptive learning and real-time inference in NLP.

RESEARCH OBJECTIVES

This research aims to reduce human effort in IMT workflows by developing models and strategies that better integrate human feedback into the translation process. The thesis frames IMT as a human-in-the-loop learning problem and explores how to optimize interaction to improve overall system efficiency and usability.

To this end, the work is organized around three main research objectives:

1. **Exploration of Alternative Feedback.** Interactive systems have traditionally relied on explicit corrections (e.g., keystrokes) to guide the translation process. This objective explores alternative, less-intrusive forms of feedback that can be leveraged during user interaction. These include implicit behavioral cues such as mouse movements or hesitations and lightweight user actions (e.g., clicking on uncertain words) that do not require complete edits. The goal is to evaluate whether these novel signals can effectively guide model updates and reduce the overall correction burden.

2. **Effort Prioritization via Quality Estimation.** Not all words or segments in a machine-generated translation require user intervention. This objective investigates how QE techniques can identify high-risk elements at the word or sentence level and prioritize user attention accordingly. The hypothesis is that targeted interaction —focusing only on uncertain or likely erroneous words and sentences— can minimize cognitive load and editing time without compromising the final output.

3. **Integration of Large Language Models.** The emergence of Large Language Models (LLMs), such as mBART [7], has opened new possibilities for MT. However, these models are typically optimized for one-shot generation tasks and have not been extensively studied in interactive settings. This objective evaluates how LLMss behave in prefix-based and segment-based IMT setups and whether they can effectively incorporate incremental user feedback to improve translation hypotheses in real-time.

By addressing these objectives, the thesis aims to provide empirical evidence and design principles for building more adaptive, efficient, and user-centered IMT systems. Each line of inquiry is grounded in experimental evaluation, focusing on measurable reductions in user effort and improvements in interaction quality.

PROPOSED METHODOLOGY

The methodology adopted in this thesis is based on the empirical evaluation of various strategies aimed at reducing human effort within an IMT environment. The research follows an experimental approach structured into three stages, each corresponding to one of the main objectives.

An IMT system is implemented as a baseline using a Transformer-based architecture with attention mechanisms [11]. Due to time and resource constraints, instead of using real translators, we have implemented

two interaction paradigms to simulate user behavior: a prefix-based and a segmented model [3, 4]. The effort is measured using two standard metrics in the field: Mouse Action Ratio (MAR) and Word Stroke Ratio (WSR) [1, 9], alongside automated quality metrics for the initial system outputs (BLEU and TER).

A system that allows translation hypotheses to be updated based solely on implicit signals, such as cursor movement toward an erroneous word, is developed for the first objective. This form of non-intrusive interaction leverages user intent without requiring explicit corrections. The impact on translation effort is evaluated in comparison to the baseline system.

To address the second objective, a system is designed in which the user intervenes only on sentences or words flagged as potentially erroneous by a QE module. While this approach does not guarantee perfect translations, the final outputs are evaluated to ensure they meet a minimum acceptable quality threshold. The comparison with the baseline focused on verifying that effort is reduced without degrading the overall translation quality in less than 70 points of BLEU.

Finally, a system was developed for the third objective that integrates LLMs — such as mBART [7] and a fine-tuned version — into the IMT workflow. These models are tested using the prefix and segmented paradigms to evaluate their behavior in interactive settings. The results are analyzed regarding user effort to determine how these models can effectively adapt to real-time user feedback.

CONTRIBUTIONS

This thesis contributes to IMT by investigating and validating strategies to minimize human effort in translation tasks. The contributions are structured around three experimental axes corresponding to the research objectives and are supported by quantitative evaluations using standard metrics such as WSR and MAR. Among these, WSR is considered particularly significant, as translators may work exclusively via keyboard or other non-mouse input devices. Table 1 shows the best WSR obtained with each approach used in this thesis for the pair of languages Spanish-English of the Europarl corpus [6].

The first contribution establishes a baseline using a conventional IMT system built on a Transformer architecture. This reference system, which lacked any form of interactive optimization, achieved a WSR of 42.0% in a prefix approach, serving as a benchmark for subsequent enhancements.

The second contribution explores implicit user interaction — specifically, mouse movement toward errors and optional implicit actions (e.g., mouse clicks) to request translation updates without manual correction. These techniques proved effective, reducing WSR to 31.1% with non-explicit interactions and as low as 22.6% with implicit feedback (limited to five updates per error), underscoring the contextual value of minimal user signals.

The third contribution investigates the application of QE to restrict user involvement to segments identified as potentially erroneous. This strategy significantly reduced the number of required corrections, achieving a WSR of 18.3% while maintaining a final translation quality above 70 BLEU points. Combining automated error prediction and focused human intervention shows strong potential for improving translation efficiency.

Lastly, the thesis evaluates the integration of LLMs — including mBART and a fine-tuned variant — into IMT systems. Although their WSR scores (38.5% and 36.3%, respectively) did not match the best-performing specialized techniques, the results are competitive and highlight the promising role of LLMs in collaborative translation workflows.

In sum, this research demonstrates that translator effort in IMT environments can be significantly reduced through interaction-focused strategies, content prioritization, and integration of novel model architectures. The findings validate the effectiveness of these approaches and lay the groundwork for developing more adaptive, ergonomic, and user-centered translation systems.

Table 1: WSR Across IMT System Approaches. Spanish-English pair of languages from the Europarl corpus [6].

Approach	**WSR**
Baseline	42.0
Non-explicit mouse action	31.1
Implicit mouse actions	22.6
Quality Estimation	18.3
mBART	38.5
mBART fine-tune	36.3

REFERENCES

[1] Sergio Barrachina, Oliver Bender, Francisco Casacuberta, Jorge Civera, elsa Cubel, Shahram Khadivi, Antonio Lagarda, Hermann Ney, Jesús Tomás, enrique Vidal, and Juan-Miguel Vilar. Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1):3–28, March 2009. URL `https://aclanthology.org/J09-1002`.

[2] Raj Dabre, Chenhui Chu, and Anoop Kunchukuttan. A survey of multilingual neural machine translation. *ACM Computing Surveys (CSUR)*, 53(5):1–38, 2020.

[3] Miguel Domingo, Alvaro Peris, and Francisco Casacuberta. Segment-based interactive-predictive machine translation. *Machine Translation*, 31:163–185, 2017.

[4] George Foster, Pierre Isabelle, and Pierre Plamondon. Target-text mediated interactive machine translation. *Machine Translation*, 12 (1):175–194, 1997.

[5] Guoping Huang, Lemao Liu, Xing Wang, Longyue Wang, Huayang Li, Zhaopeng Tu, Chengyan Huang, and Shuming Shi. Transmart: A practical interactive machine translation system. *arXiv preprint arXiv:2105.13072*, 2021.

[6] Philipp Koehn. europarl: A parallel corpus for statistical machine translation. In *Proceedings of machine translation summit x: papers*, pages 79–86, 2005.

[7] Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742, 2020.

[8] Lucia Specia and Kashif Shah. Machine translation quality estimation: Applications and future perspectives. *Translation quality assessment: from principles to practice*, pages 201–235, 2018.

[9] Jesús Tomás and Francisco Casacuberta. Statistical phrase-based models for interactive computer-assisted translation. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 835–841, 2006.

[10] Antonio Toral. Reassessing claims of human parity and super-human performance in machine translation at wmt 2019. *arXiv preprint arXiv:2005.05738*, 2020.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

[12] Qian Wang, Jiajun Zhang, Lemao Liu, Guoping Huang, and Chengqing Zong. Touch editing: A flexible one-time interaction approach for translation. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 1–11, 2020.

Electrocardiogram for Biometric Recognition: Collectability, Stability and Application Challenges

Teresa M.C. Pereira[123]
teresamcp@ua.pt
Raquel Sebastião[1]
raquelsebastiao@gmail.com
Raquel C. Conceição[2]
rcconceicao@ciencias.ulisboa.pt
Vitor Sencadas[3]
vsencadas@ua.pt

[1] IEETA, DETI, LASI, Universidade de Aveiro
Campus Universitário de Santiago, 3810-193 Aveiro
[2] IBEB, Faculdade de Ciências, Universidade de Lisboa
Campo Grande, 1749-016 Lisboa
[3] CICECO, DEMaC, Universidade de Aveiro
Campus Universitário de Santiago, 3810-193 Aveiro

INTRODUCTION

Biometric recognition has gained attention as traditional methods like passwords are no longer sufficient for protecting our personal data and belongings [2]. While fingerprints are commonly used, they can be easily circumvented by a skilled specialist. **Electrocardiogram (ECG)** has recently shown significant potential as a biometric trait mainly due to its uniqueness and hidden nature [1]. A biometric system can be divided into **data acquisition**, **feature extraction**, and **classification**.

Data Acquisition

Traditional ECG acquisition use Ag/AgCl gel electrodes due to their low cost and high signal quality. However, they suffer from practical limitations such as gel drying, skin irritation, and incompatibility with long-term wearables [3]. To address these issues, solid-state electrodes have emerged as a promising alternative, eliminating the need for conductive gels, enhancing comfort, reusability, and usability [4]. Nevertheless, ECG acquisitions still present some challenges:

- **Hardware and Physiological Variability:** Signal quality is influenced by electrode materials, cardiac conditions, posture, exercise, and emotional state.
- **Acquisition Protocols:** On-the-person setups (e.g., clinical electrodes) are inconvenient for rapid recognition. At the same time, off-the-person/wearable systems introduce noise from motion artifacts and variable skin-electrode impedance.

Feature Extraction

Regarding feature extraction, existing approaches fall into three categories, and there is still no consensus on the most effective approach.

- **Fiducial:** Relies on precise detection of ECG landmarks (e.g., QRS complexes, R-peaks), yielding high accuracy but requiring significant computational effort. These dominate in clinical systems.
- **Non-Fiducial:** Avoids landmark detection, reducing computational costs but often at the expense of performance, especially in noisy, off-the-person scenarios.
- **Partially-Fiducial:** Hybrid approaches usually rely on fiducial detection for segmentation purposes followed by non-fiducial techniques, which can be computationally more complex, justifying not being so commonly used in literature.

Classification

Classification can be performed using **machine learning models**, **distance-based methods**, or **deep learning**. Though SVM and kNN are widely adopted for their noise resilience, showing promising performances, the best classification approach still lacks consensus [5].

- SVM requires retraining for each new user, while kNN demands extensive template storage;
- Deep learning and distance-based methods have also shown potential but still lack in performance and generalization. Future work must optimize trade-offs between accuracy, adaptability and computational efficiency.

Goals

This project aims to investigate ECG-based biometrics using solid-state finger electrodes, focusing on robust feature extraction and classification under physiological variability (e.g., exercise, stress). The goal is to bridge gaps in real-world applications, by addressing noise resilience, scalability and user comfort. The following research questions guide this work:

1. **Electrode Feasibility:** Can polymeric-based dry electrodes achieve comparable or superior performance compared to conventional Ag/AgCl electrodes in ECG biometric systems, particularly in terms of signal quality, user comfort, and long-term stability?
2. **Protocol Design:** What acquisition conditions (e.g., posture, exercise, emotional state, session duration) are most critical to optimize for high-fidelity ECG biometric data collection?
3. **Variability and Performance:** How do intra-subject (e.g., heart rate variability, noise artifacts) and inter-subject (e.g., anatomical differences) factors influence recognition accuracy, and what mitigation strategies can improve robustness?
4. **Algorithm Optimization:** Which machine learning (e.g., SVM, kNN) or deep learning (e.g., CNNs) approaches are most effective for identification/authentication tasks under real-world variability? Which features optimize the performance of the system?

PROPOSED METHODOLOGY

The work begins with the design and fabrication of novel dry polymeric electrodes optimized for fingertip ECG acquisition. These electrodes will be engineered to combine high signal fidelity with practical advantages such as elimination of skin preparation requirements, enhanced biocompatibility to minimize irritation, and robust performance under varying environmental conditions including humidity and temperature fluctuations. Additionally, the project investigates how various physiological and psychological conditions affect ECG signals and biometric recognition performance, with a strong focus on intra- and inter-subject variability. Three acquisition protocols are designed: (1) multi-session recordings to evaluate signal stability over time, (2) ECG acquisition during physical activity to assess motion artifacts and postural effects, and (3) emotional elicitation via video stimuli to analyze changes induced by fear and happiness. Following data collection, raw ECG signals will undergo rigorous preprocessing to ensure analysis quality. This includes filtering to remove noises (baseline wander, powerline interference, and motion artifacts), signal normalization to account for amplitude variations, and outlier detection algorithms to identify and exclude corrupted segments. The cleaned signals will then be subjected to feature extraction pipelines employing fiducial, non-fiducial, and hybrid methodologies. Fiducial analysis will focus on precise detection of characteristic waveform components (P-QRS-T complexes) and subsequent measurement of temporal and amplitude-based features. Non-fiducial approaches will explore spectral and time-frequency domain characteristics through wavelet transforms and autocorrelation analysis, while the hybrid pipeline will strategically combine these approaches to potentially capture complementary discriminative information. The extracted feature sets will undergo dimensionality reduction techniques (such as PCA and LDA) and feature selection algorithms

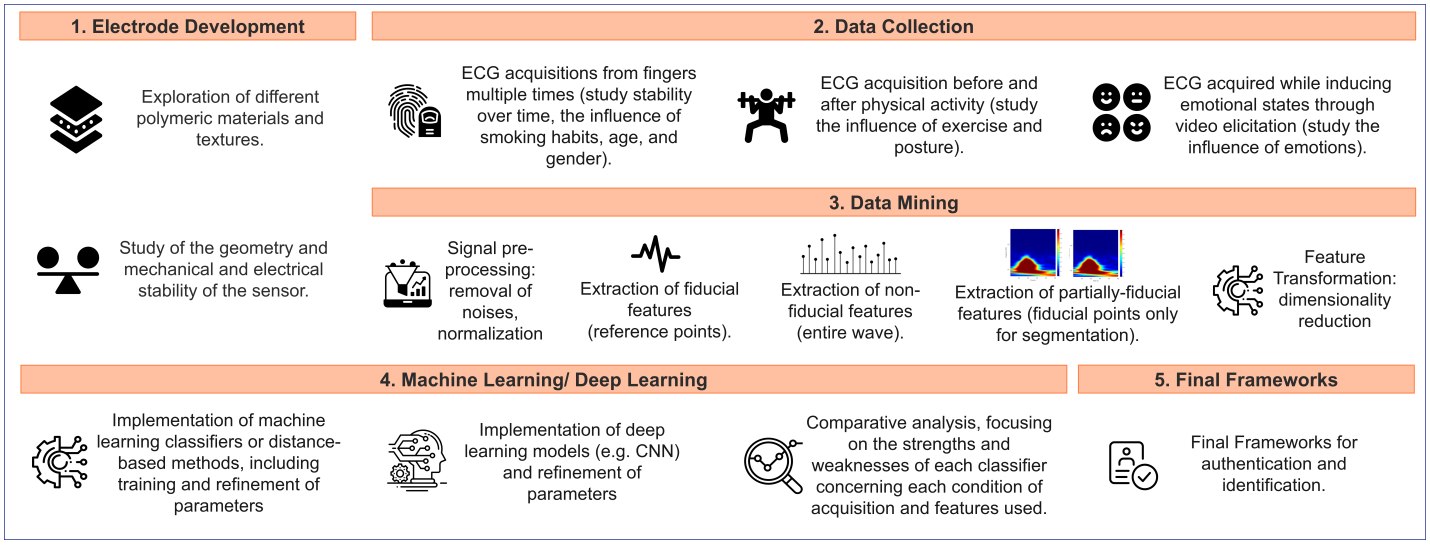


Figure 1: Graphical Abstract of the proposed PhD project.

to identify the most robust and computationally efficient features. These optimized features will be fed into various classification frameworks: machine learning approaches and deep learning frameworks. Model development will incorporate careful attention to the practical constraints of biometric systems, including computational efficiency and scalability for new user enrollment. Validation will occur through multiple comparative analyses: electrode performance will be benchmarked against clinical-grade Ag/AgCl systems across all experimental conditions; classification robustness will be assessed under varying physiological and psychological states; and long-term stability will be evaluated through the multi-session dataset. Performance metrics will include standard classification measures (accuracy, precision, recall, F1-score) as well as biometric-specific evaluations (equal error rate). The final phase will integrate these components into a unified biometric system prototype, with iterative refinement based on validation results to achieve optimal balance between accuracy, usability, and practical deployment considerations. Throughout this process, particular emphasis will be placed on reproducibility and generalizability, with all experimental protocols, algorithms, and validation procedures designed to facilitate future research and real-world implementation.

PRELIMINARY RESULTS

As this is an ongoing research project, we have obtained several promising preliminary findings that validate key aspects of our methodology. The fabrication process successfully produced dry solid-state ECG electrodes using screen printing technology, where conductive silver ink was deposited onto PDMS substrates. This elastomer was specifically chosen for its optimal combination of biocompatibility, mechanical flexibility, and permeability properties, making it ideal for long-term wearable applications. In our initial validation study involving 81 participants, we conducted simultaneous ECG recordings comparing the novel polymeric-based solid-state electrodes for the fingers against conventional wrist-placed Ag/AgCl electrodes. Dynamic Time Warping revealed strong waveform similarity between the two systems, with a root mean square error of 0.0174, mean absolute error of 0.0072, and an exceptionally high Pearson correlation coefficient of 0.9923. These results support the feasibility of our solid-state electrode design for high-quality ECG acquisition. For biometric recognition testing, we processed the finger-acquired ECG signals by segmenting them into individual heartbeats (600-sample windows centered on R-peaks) and transforming these into Gramian Angular Field (GAF) image representations. Our preliminary classification approach used a convolutional neural network (CNN) architecture that processed sequences of 10 GAF images (approximately 10 seconds of ECG data) per subject. The system demonstrated excellent same-session recognition performance, validating our core methodology. However, longitudinal analysis across four weekly acquisition sessions revealed two important findings: First, we observed a gradual decline in recognition accuracy as the time between enrollment and verification increased, highlighting the challenge of temporal variability in ECG biometrics. Second, our results indicated that multi-session training (incorporating data from different time points) yields superior performance compared to single-session enrollment, suggesting this as a potential strategy to enhance system robustness. The results to date confirm both the viability of our polymeric electrode design and the promise of our chosen processing pipeline while identifying key challenges for further investigation.

ACKNOWLEDGEMENT

This work was funded by national funds through FCT — Fundação para a Ciência e a Tecnologia, I.P., under the PhD grant UI/BD/153605/2022 (https://doi.org/10.54499/UI/BD/153605/2022) (T.M.C.P.), under unit 00127-IEETA, and within the RD units UID/00645/2025 (IBEB), 2022.08973.PTDC, and CICECO-Aveiro Institute of Materials UIDB/50011/2020 (DOI 10.54499/UIDB/50011/2020), UIDP/50011/2020 (DOI 10.54499/UIDP/50011/2020) LA/P/0006/2020 (DOI 10.54499/LA/P/0006/2020), financed by national funds through the FCT/MCTES (PIDDAC).

REFERENCES

[1] André Lourenço, Hugo Silva, and Ana Fred. Unveiling the biometric potential of finger-based ecg signals. *Computational Intelligence and Neuroscience*, 2011(1):720971, 2011. doi: https://doi.org/10.1155/2011/720971. URL `https://onlinelibrary.wiley.com/doi/abs/10.1155/2011/720971`.

[2] Teresa M. C. Pereira, Raquel C. Conceição, and Raquel Sebastião. Initial study using electrocardiogram for authentication and identification. *Sensors*, 22(6), 2022. ISSN 1424-8220. doi: 10.3390/s22062202. URL `https://www.mdpi.com/1424-8220/22/6/2202`.

[3] Teresa M. C. Pereira, Raquel C. Conceição, Vitor Sencadas, and Raquel Sebastião. Biometric recognition: A systematic review on electrocardiogram data acquisition methods. *Sensors*, 23(3), 2023. ISSN 1424-8220. doi: 10.3390/s23031507. URL `https://www.mdpi.com/1424-8220/23/3/1507`.

[4] Teresa M. C. Pereira, Raquel Sebastião, Raquel C. Conceição, and Vitor Sencadas. A review on intelligent systems for ecg analysis: From flexible sensing technology to machine learning. *IEEE Journal of Biomedical and Health Informatics*, 29(5):3398–3413, 2025. doi: 10.1109/JBHI.2024.3508545.

[5] João Ribeiro Pinto, Jaime S. Cardoso, and André Lourenço. Evolution, current challenges, and future possibilities in ecg biometrics. *IEEE Access*, 6:34746–34776, 2018. doi: 10.1109/ACCESS.2018.2849870.

Graph-Imbalanced Regression for Rare Phenotypes

Brenda Nogueira
https://brendacnogueira.github.io/
Nuno Moniz
https://www.nunomoniz.co/
Nitesh Chawla
https://niteshchawla.nd.edu/

Lucy Family Institute for Data & Society
University of Notre Dame
Notre Dame, US

1 Introduction

Regression tasks are essential across many scientific domains, particularly in chemistry, where predictive models help prioritize compounds for experimental testing. However, the most transformative discoveries often lie in the rarest observations. In drug discovery, for instance, fewer than 0.1% of compounds demonstrate optimal binding properties [11], yet these rare compounds drive innovation.

Despite this, most current machine learning approaches are poorly suited for such extreme settings. They are typically optimized for average performance, systematically neglecting these high-impact, low-frequency cases [12, 13, 16]. Addressing this challenge requires tackling two key issues: scientific domains are inherently structured, and the distribution of values and domain preferences is highly imbalanced. While advances in graph neural networks (GNNs) [20] and imbalanced learning [21] have been significant, their intersection —**graph imbalanced regression**— remains largely unexplored.

My research addresses this gap by rethinking how we evaluate and optimize models in settings with non-uniform domain preferences, such as drug discovery. I aim to develop methods that prioritize accuracy in high-relevance regions. By doing so, we can better identify the most valuable compounds and extend these insights to other fields of natural and physical science, where similar imbalances persist.

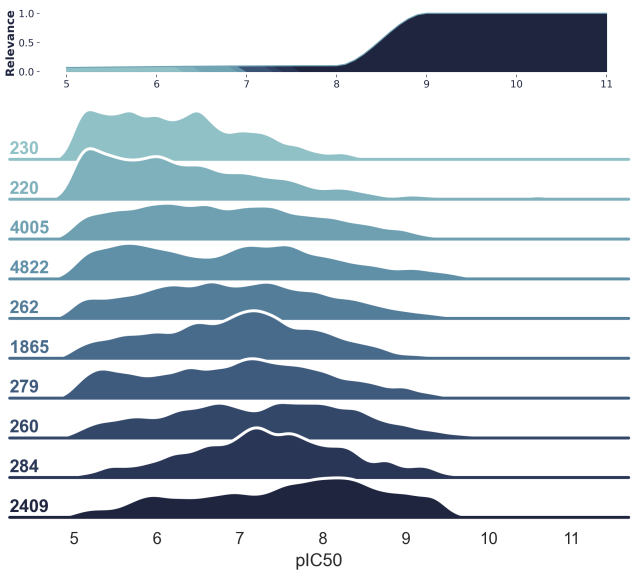


Figure 1: **Potency distributions and Relevance function.** The density plot illustrates the distributions of ten randomly selected activity classes representing ten distinct regression tasks. Data distribution is determined via kernel density estimation. The color of the densities is determined by the mean of each distribution, with lighter shades of blue representing lower means and darker shades representing higher means. Furthermore, the relevance of each pIC_{50} value is represented according to a Piecewise Cubic Hermite Interpolating Polynomial.

2 Literature Review

2.1 Imbalanced Learning Challenges

Imbalanced data problems have been extensively studied in classification contexts [3, 10], but remain underexplored for regression tasks [2]. Recent work on imbalanced regression includes SERA (Squared Error Relevance Area) [15] for traditional regression, and Deep Imbalanced Regression (DIR) [21] which introduces Label Distribution Smoothing (LDS) and Feature Distribution Smoothing (FDS). However, these approaches lack graph-awareness and cannot leverage structural relationships inherent in scientific data. Moreover, they face limitations in handling distribution shifts and may suffer from optimization instability when applied to complex structured data [19].

2.2 Graph Neural Networks for Scientific Discovery

Graph Neural Networks (GNNs) have transformed scientific machine learning through their ability to model complex relational data [4, 20]. In contrast, graph-aware architectures such as GraphSAGE [8], which samples and aggregates neighbor features for inductive node embeddings, and Graph Attention Networks (GAT) [18], which learn attention weights over edges, excel at node classification on balanced data but lack mechanisms to up-weight rare continuous outcomes. Further, recent developments in graph representation learning [5, 6, 9, 14, 22] have shown promise for general graph learning but are not tailored for regression or imbalanced learning contexts.

3 Preliminary results and ongoing work

In early work, we have shown that models which account for data imbalance and domain relevance can identify a greater number of unique and high-performing compounds compared to those trained with conventional techniques. While these models still capture key compounds detected by traditional approaches, they significantly improve predictive accuracy in high-relevance regions and better differentiate between critical and less critical cases.

Building on this, I propose to develop models that explicitly incorporate domain preferences by using domain-defined relevance functions to guide predictions toward high-value areas. Ongoing work includes:

- **Molecular property prediction for regression tasks (e.g., Lipo, ESOL, FreeSolv, Potency, Melting point):** We are adapting data augmentation strategies to reflect domain-specific preferences.
- **Imbalanced yield prediction:** Together with a lab colleague, I am developing a framework for predicting molecular reaction yields under imbalance.

4 Work Plan

1. **Multimodal Contribution Analysis** Assess the individual and combined effects of different molecular representations—SMILES (token-based), graph-based, and image-based—on predictive performance. Special attention will be given to how each modality contributes to accuracy in high-relevance (extreme) regions of the target distribution.

2. **Scalable, Domain-Relevant Augmentation Pipelines** Design and evaluate augmentation strategies tailored to each molecular modality, ensuring chemical validity and domain relevance, such as graph augmentation [1], loss function adaptation and image augmentation [17]. Also, we aim to build tools to extract and interpret visual molecular features [7].

5 Expected Impact

This work has meaningful real-world impact: more accurate predictive models can accelerate drug discovery, lower costs, and improve access to treatments for diseases. By improving virtual screening and reducing false negatives, my research helps identify promising compounds more efficiently.

Ultimately, my research bridges computational modeling and practical application, advancing techniques that align predictive performance with scientific relevance to support faster drug development, better environmental decisions, and more equitable AI deployment in imbalanced-data settings.

References

[1] Paula Branco, Luís Torgo, and Rita P. Ribeiro. Smogn: A pre-processing approach for imbalanced regression. In *First International Workshop on Learning with Imbalanced Domains: Theory and Applications*, pages 36–50, 2017.

[2] Paula Branco, Luís Torgo, and Rita P. Ribeiro. A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys*, 52(1):1–50, 2019.

[3] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[4] Gabriele Corso, Luca Cavalleri, Dominique Cantáro, Zhaoping Xiong, Alpha A Lee, Regina Barzilay, and Petar Veličković. Graph neural networks for scientific discovery. *Communications Materials*, 3(1):1–17, 2022.

[5] Gabriele Corso, Bohan Carderero, Robert Dürichen, Pim Andehekta, Zhihan Li, Anders Eriksson, Erik Klöppner, Du Huynh, Santiago Miret, David Charatan, et al. Message passing neural pde solvers. In *International Conference on Learning Representations*, 2023.

[6] Zheng Feng, Xiang Xu, Jiong Gao, Muhan Zhang, and Yu Wang. Grand: Graph neural diffusion. In *International Conference on Machine Learning*, pages 6233–6242. PMLR, 2022.

[7] Garrett B Goh, Charles Siegel, Abhinav Vishnu, Nathan O Hodas, and Nathan Baker. Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed qsar/qspr models. *arXiv preprint arXiv:1706.06689*, 2017.

[8] W. L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30:1024–1034, 2017.

[9] Yifan Han, Soheila Huang, Yingtao Ma, Huiyuan Chen, and Philip Yu. G-mixup: Graph data augmentation for graph classification. In *International Conference on Machine Learning*, pages 8230–8248. PMLR, 2022.

[10] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9): 1263–1284, 2009.

[11] James P. Hughes, Stephen Rees, S. Barrett Kalindjian, and Karen L. Philpott. Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249, 2011.

[12] Tiago Janela and Jürgen Bajorath. Rationalizing general limitations in assessing and comparing methods for compound potency prediction. *Scientific Reports*, 13(1):1–9, October 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-45086-3. URL `https://www.nature.com/articles/s41598-023-45086-3`. Number: 1 Publisher: Nature Publishing Group.

[13] Yihong Ma, Xiaobao Huang, Bozhao Nan, Nuno Moniz, Xiangliang Zhang, Olaf Wiest, and Nitesh V. Chawla. Are we making much progress? Revisiting chemical reaction yield prediction from an imbalanced regression perspective, February 2024. URL `http://arxiv.org/abs/2402.05971`. arXiv:2402.05971 [physics].

[14] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4602–4609, 2019.

[15] Rita P. Ribeiro and Nuno Moniz. Imbalanced regression and extreme value prediction. *Machine Learning*, 109:1803–1835, 2020.

[16] Jannik P. Roth and Jürgen Bajorath. Relationship between prediction accuracy and uncertainty in compound potency prediction using deep neural networks and control models. *Scientific Reports*, 14(1):6536, March 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-57135-6. URL `https://www.nature.com/articles/s41598-024-57135-6`. Publisher: Nature Publishing Group.

[17] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

[18] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[19] Ziyan Wang and Hao Wang. Variational imbalanced regression: Fair uncertainty quantification via probabilistic smoothing. *Advances in Neural Information Processing Systems*, 36:30429–30452, 2023.

[20] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2022.

[21] Yuzhe Yang, Kaibin Zha, Yuchen Chen, Hao Wang, and Dina Katabi. Delving into deep imbalanced regression. In *International Conference on Machine Learning*. PMLR, 2021.

[22] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. In *Advances in Neural Information Processing Systems*, volume 33, pages 5812–5823, 2020.

Image and Video-Based Automatic Body and Gait Biomarker Computation for Turner Syndrome Diagnosis

Maria del Mar Coch-Alcina
mariadelmar.coch@salle.url.edu
Xavier Sevillano
xavier.sevillano@salle.url.edu

Human-Environment Research Group
La Salle, Universitat Ramon Llull
Barcelona, Spain

INTRODUCTION

Turner Syndrome (TS) is a rare genetic disorder caused by the complete or partial absence of one X chromosome, affecting approximately 1 in every 2,000 to 2,500 live-born females worldwide [2]. It presents with diverse clinical manifestations, including short stature, skeletal disproportion, cardiovascular anomalies, neurodevelopmental alterations, and motor coordination challenges. Diagnosis is often delayed due to clinical variability, limited awareness of TS phenotypes, and restricted access to confirmatory testing such as karyotyping [8, 11]. In many cases, diagnosis only occurs after the onset of delayed puberty or infertility, long after early signs such as short stature or subtle dysmorphic features have been overlooked or misinterpreted.

The BeNeXT project [5] addresses these challenges through a multi-omic framework that integrates phenomics, genomics, and machine learning to improve diagnostic accuracy for TS. The approach of the project explicitly tackles biases in current diagnostic models, which have been developed primarily using data from individuals of European descent. By studying populations from Spain and Latin America, BeNeXT aims to expand representation and ensure that the developed tools are effective across diverse ancestral backgrounds [9], aiming also to be applicable to diagnose other rare genetic conditions.

Within the BeNeXT framework, this doctoral thesis investigates the potential of body proportions and gait characteristics as phenomic biomarkers for TS. It proposes a computer vision-based pipeline for extracting morphometric and kinematic parameters from RGB images and video using 3D model fitting and human pose estimation. These predictions will be validated against synchronously recorded motion capture data serving as ground truth. By identifying body-based features that distinguish TS from control populations, this thesis aims to contribute scalable, non-invasive tools to the diagnostic pipeline developed within the BeNeXT project.

RESEARCH PROBLEM

The central aim of this research is to develop low-cost computational tools capable of extracting body and gait parameters from images and video that can be quantitatively compared between TS individuals and matched controls. Although body proportion differences and motor coordination issues are frequently reported in TS [6], these traits are rarely evaluated systematically or used with diagnostic purposes. The lack of standardized, scalable methods for assessing full-body phenomic features presents a barrier to earlier and more equitable diagnosis, particularly in populations with limited access to genetic testing [8].

By enabling reproducible, low-cost analysis of morphometric and kinematic patterns extraction from image and video data, this thesis seeks to support the identification of phenomic biomarkers that could assist in the early screening and diagnosis of TS.

RESEARCH QUESTIONS

- Are there statistically significant differences in body structure and gait parameters between TS and control populations that can be measured computationally?
- Can video-based human pose estimation and 3D model fitting provide reliable estimates of these parameters when compared with ground truth data?
- Which specific biomechanical and morphological markers can be used as diagnostic features for TS, and how accurately can they be extracted from low-cost video systems?

PROPOSED METHODOLOGY

The proposed methodology is based on the collection of synchronized motion capture and RGB pictures and video recordings to extract full-body morphological and motor features in individuals with Turner Syndrome and matched control participants. Marker-based optoelectronic motion capture data will serve as ground truth to define key morphometric and kinematic parameters, including (but not limited to) limb proportions, full-body joint trajectories, range of motion, and gait symmetry indicators [1, 7]. All participants will be recorded following a standardized physical task protocol developed in collaboration with physiotherapists and medical experts from the BeNeXT project, specifically designed to elicit relevant motor and postural features associated with TS [10].

The kinematic data acquisition will take place at the motion capture laboratory of La Salle - Universitat Ramon Llull, equipped with 8 Vero 2.2 cameras working at 100 Hz sample rate, and the Nexus 2.16.0 software. Reflective markers will be placed on the participants body following the Plug-in Gait full-body template, composed by 39 marker locations on the pelvis, trunk, limbs, and head [12]. Simultaneously, synchronized RGB video will be captured from multiple viewpoints to enable downstream computer vision analysis. The video data will be processed using human pose estimation and 3D model fitting techniques to extract parameters comparable to those obtained from motion capture. Prior work on gait and identity analysis has demonstrated the effectivity of pose-based models such as OpenPose, HRNet, and PoseGait for extracting kinematic features from video data [3, 4].

The parameters estimated from video-based models will be quantitatively compared to their motion capture counterparts using evaluation metrics including joint position error (JPE), limb length deviation, and gait cycle consistency. This comparative analysis will be used to identify and adapt the most accurate and robust models for estimating diagnostic physical features. The resulting dataset will also support classification and clustering tasks aimed at identifying group-level differences, with the long-term objective of informing low-cost, reproducible screening tools that can be integrated into BeNeXT's diagnostic pipeline.

REFERENCES

[1] Gema Beltrán-Serrano, Alba Llauró, Álvaro Heredia-Lidón, Aroa Casado, Laura Menés-Fernández, Esther Esteban, Neus Martínez-Abadías, and Xavier Sevillano. Biomechanical manifestations of skeletal alterations in Turner syndrome: a study of full-body gait dynamics. In *Proceedings of the 2025 Computer Methods in Biomechanics and Biomedical Engineering Symposium (CMBBE)*, 2025.

[2] Agnethe Berglund, Kirstine Stochholm, and Claus Højbjerg Gravholt. The epidemiology of sex chromosome abnormalities. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, volume 184, pages 202–215. Wiley Online Library, 2020.

[3] Francisco M Castro, Rubén Delgado-Escaño, Ruber Hernández-García, Manuel J Marín-Jiménez, and Nicolás Guil. Attengait: Gait recognition with attention and rich modalities. *Pattern Recognition*, 148:110171, 2024.

[4] Nicolás Cubero, Francisco M Castro, Julián R Cózar, Nicolas Guil, and Manuel J Marín-Jiménez. Empirical study of human pose representations for gait recognition. *Expert Systems with Applications*, 275:126946, 2025.

[5] Marc Freixes, Xavier Sevillano, Esther Esteban, Aroa Casado, Carmen Garrido, Alejandro González, Alvaro Heredia-Lidón, Jordi Malé, Joan Claudi Socoró, Luis Joglar-Ongay, et al. BeNeXT project: Biomarker enhanced diagnostic and prognostic tools for

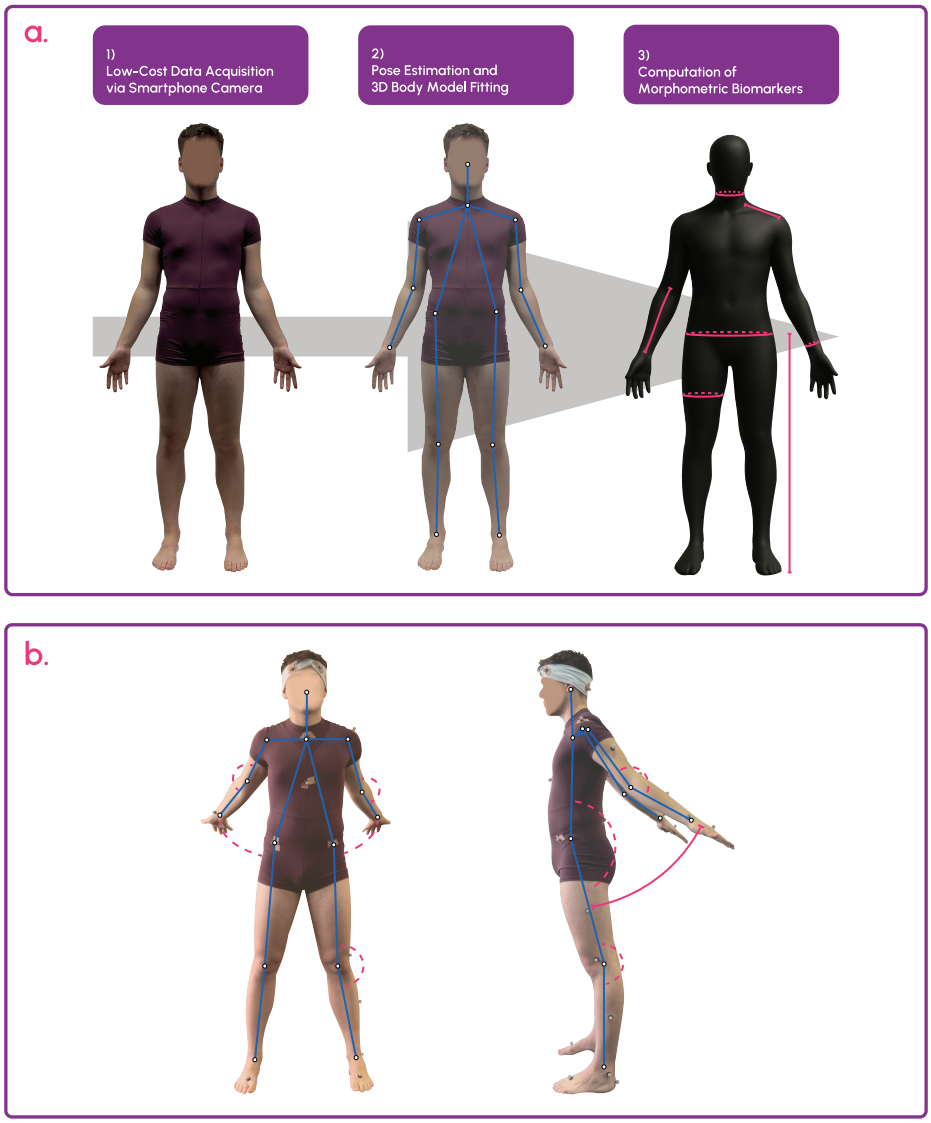


Figure 1: Overview of the image-based pipeline for extracting morphometric biomarkers relevant to Turner Syndrome diagnosis. (a) The general workflow: low-cost data acquisition via smartphone (left), pose estimation and 3D body model fitting (middle), and computation of morphological biomarkers (right). (b) Example keyframes from motor task trials captured from two viewpoints, used for pose estimation and extraction of diagnostic features.

rare disorders–using X-chromosome alterations in Turner syndrome as a model. In *Proc. IberSPEECH 2024*, pages 281–284, 2024.

[6] Claus H Gravholt, Niels H Andersen, Gerard S Conway, Olaf M Dekkers, Mitchell E Geffner, Karen O Klein, Angela E Lin, Nelly Mauras, Charmian A Quigley, Karen Rubin, et al. Clinical practice guidelines for the care of girls and women with Turner syndrome: Proceedings from the 2016 Cincinnati International Turner Syndrome Meeting. *European Journal of Endocrinology*, 177(3): G1–G70, 2017.

[7] Claus H Gravholt, Mette Viuff, Jesper Just, Kristian Sandahl, Sara Brun, Janielle van der Velden, Niels H Andersen, and Anne Skakkebaek. The changing face of Turner syndrome. *Endocrine Reviews*, 44(1):33–69, 2023.

[8] Daniel F Gunther, Erica Eugster, Anthony J Zagar, Constance G Bryant, Marsha L Davenport, and Charmian A Quigley. Ascertainment bias in Turner syndrome: new insights from girls who were diagnosed incidentally in prenatal life. *Pediatrics*, 114(3):640–644, 2004.

[9] Stéphanie Nguengang Wakap, Deborah M Lambert, Annie Olry, Charlotte Rodwell, Charlotte Gueydan, Valérie Lanneau, Daniel Murphy, Yann Le Cam, and Ana Rath. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics*, 28(2):165–173, 2020.

[10] Uma Kaimal Saikia, Dipti Sarma, and Yogesh Yadav. Delayed presentation of Turner syndrome: challenge to optimal management. *Journal of Human Reproductive Sciences*, 10(4):297–301, 2017.

[11] Arrigo Schieppati, Jan-Inge Henter, Erica Daina, and Anita Aperia. Why rare diseases are an important medical and social issue. *The Lancet*, 371(9629):2039–2041, 2008.

[12] Vicon Motion Systems Ltd. *Plug-in Gait Reference Guide*, n.d. URL `https://help.vicon.com/space/Nexus216/11607059/Plug-in+Gait+Reference+Guide`. Accessed: 2025-05-22.

Modeling Music: Explorations in Computation, Language and Recognition

Aitana Menárguez-Box
amenbox@prhlt.upv.es
Enrique Vidal
evidal@prhlt.upv.es
Alejandro H. Toselli
ahector@prhlt.upv.es

PRHLT Research Center
Universitat Politècnica de València,
Valencia, Spain
http://www.prhlt.upv.es

INTRODUCTION

Music can be considered one of the most universal forms of human expression, and its preservation has taken many shapes throughout history – from oral tradition and handwritten manuscripts to printed scores and audio recordings. Yet beyond the medium of preservation, music is also a deeply interpretative practice. Understanding it involves more than decoding symbols; it requires engaging with its structure, meaning, and context, and furthermore connecting with its artistic expression and with emotions.

In recent decades, the digitization of music has opened up new frontiers. Musical data is now more accessible and diverse than ever before, and computational methods have become central to interacting with music at scale. Automatic Music Recognition (AMR) systems are a cornerstone of this development, enabling a wide range of applications such as supporting for digital archiving, facilitating musicological research and offering powerful tools for artistic exploration.

This doctoral research, currently in its early stages, engages with these possibilities by combining computational, cognitive, and linguistic perspectives. It aims to deepen our understanding of music as both a symbolic and expressive system, while also contributing to practical applications.

THE RESEARCH

The starting point of this research has been treating with ancient *handwritten* sheet music for automatic *recognition* and information *retrieval*. Although these manuscripts may appear simpler than other repertoires due the nature of the music inside them (monophonic tunes without rhythm annotations), they present real and underexplored challenges for automatic recognition and retrieval systems – challenges that are, in fact, common across many different styles and notations from different time periods [2, 4, 7, 10]. Their relative visual and structural simplicity also allows for isolating core aspects of musical representation, making them ideal for early-stage experimentation.

Problem

Working with handwritten sheet music, whether ancient or modern, remains a difficult task due to several interrelated factors: high variability in handwriting, historical changes in notation standards, differences in musical contents (such as clef changes, instrumentation, or stylistic conventions), among others. Notably, one of the most significant and often overlooked challenges is the limited understanding of what current models actually "learn" about music. While recent machine learning approaches have achieved impressive results, models used often operate as black boxes, offering limited insight into how musical structures are encoded and interpreted.

By treating music as a structured and interpretable system – akin to natural language – it becomes possible to analyze and formalize the syntax, semantics, and rules for its construction. This opens the door to more accurate (and "conscious") recognition and richer interpretation. Moreover, this is also beneficial for retrieval tasks, as it enables the better use of probabilistic approaches.

Questions

This work is *initially* guided by the following research questions: 1) How can the structure and patterns of musical language in handwritten sheet music be mathematically and linguistically modeled? 2) In what ways can the mathematical and linguistic modeling of musical notation improve the accuracy of both music recognition and retrieval systems? 3) How can the methods developed for ancient handwritten manuscripts be generalized to recognize and retrieve music information with more complex notations, such as modern polyphonic music, rhythmic annotations, and other advanced musical elements? 4) What role does music theory play in refining the design of computational models for automatic music recognition and retrieval?

OBJECTIVES

The central objective of this research is to explore mathematical and linguistic modeling for the recognition and representation of music, with ancient handwritten sheet music as the starting point. The aim is not only to improve music transcription but also to enable more effective information retrieval, digital preservation, and creative use of music data.

To achieve this, the research will first focus on the analysis of historical manuscripts which will support the development of a formal representation of music. Beyond the immediate scope of ancient notation, the research will explore how the proposed models and methods can be extended to more complex repertoires in modern music – particularly those involving rhythm, polyphony and diverse layout conventions. This includes both printed and handwritten sources, aiming to build systems that are robust across different notational standards and time periods.

Moreover, a key part of this research involves Music Information Retrieval (MIR) technique. Leveraging probabilistic strategies, such as those explored with Probabilistic Indexing (PrIx) [3, 11], the work aims to go beyond traditional best-hypothesis matching [1, 6], facilitating search and access in large digital archives.

Finally, while the initial focus is on symbolic music in handwritten form, this work is envisioned as a step toward integrating these insights into other modalities such as audio. In the long term, the goal is to support cross-modal tasks like score-audio alignment, transcription, or creative applications.

Work Plan

To accomplish the stated objectives, the research has been structured into the following interrelated points:

- **Bibliographic and Theoretical Grounding:** The initial phase focuses on a review of the existing literature in several key areas: optical music recognition (OMR) [4, 10], automatic music transcription and retrieval [2, 7], formal modeling of musical notation [5, 9, 12]. This will serve for identifying the state of the art existing gaps, and selecting the most relevant theories, tools, and datasets. This revision will be present throughout the entire project, continuously incorporating new publications and discoveries relevant for the research.

- **Dataset Preparation and Exploratory Analysis:** This phase involves the selection and preprocessing of relevant datasets – initially focusing on historical handwritten sheet music from sources such as the *Cantus Database* [8], and later incorporating modern samples.

- **Theoretical Modeling:** Parallel to the analysis, the research will focus on developing formal mathematical representations of music. This will involve identifying rule-based structures, abstract symbols, and their relationships, supported by linguistic concepts such as grammar, syntax, and semantics.

- **Model Integration and OMR Systems Development:** Once the theoretical models are mature, they will be integrated into a computational system for optical music recognition. Machine learning methods, especially those capable of handling sequence modeling and visual variability will be employed.
- **Evaluation and Retrieval Testing:** The performance of the system will be evaluated quantitatively using standard metrics (e.g., recognition accuracy, character error rate) and qualitatively by examining the interpretability of the results. Retrieval experiments will also be conducted to assess the model's utility in information access scenarios.
- **Creative and Cross-domain Applications:** Broader applications and adaptation to modern notational systems with added complexity (e.g., polyphony, rhythm, instrumentation) will be studied.
- **Writing and Dissemination:** Throughout all phases, findings will be documented and communicated through academic writing, conference participation, and collaborative discussions within the research community.

Thin plan ensures on one hand, the fulfillment of scientific and technical goals and on the other it supports the theoretical framework through empirical feedback and creative exploration.

EXPECTED CONTRIBUTIONS

This doctoral research aims to contribute both theoretical and practical tools to the field of automatic music recognition and music information retrieval. The expected contributions include:

- Formal models of musical notation grounded in mathematical and linguistic theory, enabling structured representation and analysis of symbolic music.
- New methodologies for the recognition of handwritten sheet music, especially in historical contexts, addressing challenges such as variability in notation and handwriting.
- Extension of recognition and retrieval methods to broader musical contexts, including printed modern scores with rhythm and polyphony, and potentially toward audio-based music analysis.
- Improved music information retrieval through probabilistic indexing and structured representations, supporting more flexible and accurate search in digital music archives.
- A theoretical foundation for treating music as a formal, interpretable language, aimed to interdisciplinary connections between musicology, linguistics, and AI.
- Empirical evaluation and resources, such as datasets, annotated corpora, and open-source tools, to support reproducibility and future research in the community.

These contributions are intended to support the development of more robust, interpretable, and creative systems for music recognition, analysis, and interaction.

REFERENCES

[1] Laurent Pugin Jason Hockman John Ashley and Burgoyne Ichiro Fujinaga. Gamera versus aruspix two optical music recognition approaches. *ISMIR 2008*, page 139, 2008.

[2] Donald Byrd and Tim Crawford. Problems of music information retrieval in the real world. *Information Processing & Management*, 38(2):249–272, 2002.

[3] Jorge Calvo-Zaragoza, Alejandro H. Toselli, Enrique Vidal, and Joan Andreu Sánchez. Music Symbol Sequence Indexing in Medieval Plainchant Manuscripts. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 882–887, 2019.

[4] Jorge Calvo-Zaragoza, Juan C. Martinez-Sevilla, Carlos Penarrubia, and Antonio Rios-Vila. Optical Music Recognition: Recent Advances, Current Challenges, and Future Directions. In Mickael Coustaty and Alicia Fornés, editors, *Document Analysis and Recognition – ICDAR 2023 Workshops*, pages 94–104, Cham, 2023. Springer Nature Switzerland.

[5] Keiji Hirata, Satoshi Tojo, and Masatoshi Hamanaka. *Music, mathematics and language: The new horizon of computational musicology opened by information science*. Springer, Singapore, 2023.

[6] Yu-Hui Huang, Xuanli Chen, Serafina Beck, David Burn, and Luc Van Gool. Automatic handwritten mensural notation interpreter: From manuscript to MIDI performance. In *ISMIR*, pages 79–85, 2015.

[7] Marius Kaminskas and Francesco Ricci. Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, 6(2):89–119, 2012.

[8] Debra Lacoste, Terence Bailey, Ruth Steiner, and Jan Koláček. Cantus: A database for latin ecclesiastical chant, 2011. URL `https://cantusdatabase.org`. Funded through the Digital Analysis of Chant Transmission project at Dalhousie University, Halifax, Nova Scotia, Canada (SSHRC 895-2023-1002).

[9] Meinard Muller. *Fundamentals of music processing: Audio, analysis, algorithms, applications*. Springer International Publishing, Cham, Switzerland, 2016.

[10] Ana Rebelo, Ichiro Fujinaga, Filipe Paszkiewicz, Andre RS Marcal, Carlos Guedes, and Jaime S Cardoso. Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1:173–190, 2012.

[11] Alejandro Héctor Toselli, Joan Puigcerver, and Enrique Vidal. *Probabilistic indexing for information search and retrieval in large collections of handwritten text images*, volume 49 of *The Information Retrieval Series*. Springer Cham, Switzerland, 2024. ISBN 978-3-031-55389-9.

[12] Damiáan H. Zanette. Zipf's law and the creation of musical context. *Musicae Scientiae*, 10(1):3–18, 2006.

On the Use of Implicit Representations for Deepfake Detection

Miguel Leão
https://visteam.isr.uc.pt/team/120/
Nuno Gonçalves
https://visteam.isr.uc.pt/team/nuno-goncalves-2/

Institute of Systems and Robotics
University of Coimbra
Portugal

INTRODUCTION

The developments in home computers, united with the thousands upon thousands of images/videos of individuals present on the Internet, allowed for the proliferation of deepfaked media affecting the lives of private individuals and the dangerous spread of misinformation. Current state-of-the art detection methods show impressive results. However the development of improved generation methods overcomes them, as there are generalization difficulties.

Following the logic of forensic approaches in both the color and frequency space, this work investigates the use of the implicit space in the problem of deepfake detection. Implicit representations have recently offered new research avenues for image analysis, translating a scene usually in a coordinates-based representation, that allows for detailed reconstructions of the original. Using Sinusoidal Representation Networks (SIRENs) [9] the video frames of the Deepfake Detection Challenge Dataset (DFDC) [2] were translated to the implicit space and analyzed.

This work uses Fréchet Video Distance (FVD) [10] between the original DFDC videos and their respective SIREN reconstruction, to show a significant difference in the average FVDs of the bonafide and deepfake pairs. We expect that this work might open new avenues of research for the deepfake detection problem.

METHOD

Implicit representation

An image is represented as a function $I : \Omega \subset R^2 \rightarrow C$, where Ω is the image's domain and C is the color space. The image is then parameterized with a coordinate-based neural network $I_\theta : R^2 \rightarrow C$ with parameters θ. To train the neural image I_θ so that it approximates I, the model optimizes the following objective:

$$\int_\Omega (I - I_\theta)^2 \, dx.$$

The coordinate-based network is a sinusoidal multilayer perceptron (MLP) $f_\theta(p) : R^n \rightarrow R^m$, defined as a composition of d sinusoidal layers:

$$f_\theta(x) = W_d \circ f_{d-1} \circ \cdots \circ f_0(x) + b_d,$$

where each layer $f_i(x_i) = \sin(W_i x_i + b_i) = x_{i+1}$, with $W_i \in R^{n_{i+1} \times n_i}$ being the weight matrices, and $b_i \in R^{n_{i+1}}$ being the biases. The collection of these parameters defines θ. The integer d denotes the depth of the network, and n_i refers to the width of the layers.

With the neural image defined by θ, the RGB values for any pixel of a reconstructed image are given by the value of f_θ at x coordinates.

Through the method used in [8], the neural images of each frame of the subject's face is obtained. The individual frames are then joined into a reconstructed video.

Distance between original and reconstructed videos

This article proposes to show that there is a difference between how reliable the neural reconstruction of a video is for bonafide and deepfake video cases, so that it can be used to detect the latter. This is measured through Fréchet Video Distance (FVD).

FVD is proposed as an improvement on common video analysis approaches such as Peak Signal-to-Noise-Ratio (PSNR) or Structural Similarity (SSIM) [11] claiming that these lack for the temporal coherence of the video, aside from the video quality itself. It is based on the principal of Fréchet Inception Distance (FID) [4], commonly used for image analysis, where the distance between the real world data distribution P_R and the distribution defined by the generative model P_G is defined by:

$$d(P_R, P_G) = min_{X,Y} E|X - Y|^2$$

where the minimization is over all random variables X and Y with distributions P_R and P_G respectively. With the data distribution being represented as a multivariate Gaussian using a suitable feature space, the previous equation can be solved as:

$$d(P_R, P_G) = |\mu_R - \mu_G|^2 + Tr(\Sigma_R + \Sigma_G - 2(\Sigma_R \Sigma_G)^{\frac{1}{2}})$$

where μ_R and μ_G are the means and Σ_R and Σ_G are the co-variance matrices of P_R and P_G. This representation is obtained from an Inflated 3D ConvNet (I3D) [1], and the distance between videos is obtained. In our work, we obtained the FVD through the implementation used in [3].

EXPERIMENTS AND RESULTS

Dataset

The Deepfake Detection Challenge (DFDC) [2] is a self-designated third generation dataset featuring 23,654 videos from 960 actors hired for this purpose, from which 104,500 fake videos are created using various deepfake creation methods.

These include a Deepfake Auto Encoder (DFAE) model with a shared encoder but two isolated decoders, one for each identity, and a Neural Talking Heads (NTH) [12] model comprised of a metalearning stage and a fine-tuning stage.

It also includes deepfakes generated from FSGAN [6] which applies an adversarial loss to generators for reenactment and inpainting, and trains additional generators for face segmentation and Poisson blending and StyleGAN [5] which is modified to produce a face swap between a given fixed identity descriptor onto a video by projecting this descriptor on the latent face space. Finally, certain videos from the previous categories are processed with a sharpening filter to improve the quality of the final video and certain videos receive vocal deepfakes as presented in [7].

SIREN reconstructions

The SIREN models were trained for 1000 epochs for each frame, resulting in a reconstruction that shows no differences to the naked eye, for both deepfake and bonafide videos, even for the ones scoring the highest FVD scores, as shown in figures 1.

Figure 1: Comparison between an original frame (left) from a video, it's SIREN reconstruction (center) and their difference (right), for bonafide cases in green and deepfake cases in red.

Although the reconstructions do not show visible differences when analyzed, it is possible to find the areas in the image where the reconstructions is not perfect. Analyzing these areas together with additional information from the scene can give insights into the problem.

This would greatly benefit from a labeling effort on the dataset to properly analyze if and how different conditions affect the SIREN representation for bonafide and deepfake cases.

Testing the hypothesis

The average FVD scores obtained show that SIREN reconstructions for the bonafide videos have lower fidelity to their original video than in the deepfake videos, achieving higher FVD scores, as shown in figure 1. Higher FVD scores mean higher distance between video pairs i.e. worse reconstructions.

To confirm that the data shows that SIREN reconstructions could be used for deepfake detection, specifically that the neural reconstructions of bonafide material have lower fidelity than deepfake reconstructions, a one tail significance test is conducted. First the null hypothesis is defined as there is no difference between the distribution of FVD scores for the original and SIREN reconstruction of bonafide videos and deepfake videos, or that the FVD scores for bonafide videos are lower then the deepfake scores, i.e. SIRENS achieve better reconstructions on bonafide videos than deepfake ones:

$$H_0 : \mu FVD_{bonafide} <= \mu FVD_{deepfake}$$

and present the alternative hypothesis that the bonafide video pairs score higher FVDs, therefore have worse fidelity, than the deepfake video pairs:

$$H_a : \mu FVD_{bonafide} > \mu FVD_{deepfake}$$

conducting the test with a significance level of $\alpha = 0.01$, the p-value result is equal to $1.145e-5$ giving $p < \alpha$, thus rejecting the null hypothesis.

Discussion

The data shows that SIREN reconstructions bonafide videos have lower fidelity than the reconstructions of deepfake videos. This could suggest that bonafide videos contain richer information, which is lost during manipulation.

The fact that the standard deviation for deepfake scores is lower, may also indicates a process of homogenization of the information. The DFDC is a large dataset, so as previously mentioned, its full translation into neural representations would, at this pace, take a not practical amount of time. However by expanding this research over more videos from the dataset, it would give a clearer idea of how different attributes from these videos might affect the neural representation, or how certain deepfake generative methods impact the image.

This result could contribute to the development of a system for detecting deepfakes by learning how to distinguish between bonafide videos and deepfake videos by analyzing the original video and its neural reconstruction.

However, there is still work to be done in this area, particularly in understanding the influence of certain factors like resolution, the context of the video (e.g. how much of the frame does the face occupy), among other elements.

CONCLUSION AND FUTURE WORK

Conclusion

This article presented the hypothesis of using implicit representations of facial videos to distinguish between bonafide and deepfake videos. Carrying out this first analysis with videos reconstructed from the SIREN representation, the FVD value between the original videos and their reconstructions was measured. These values were use to test the hypothesis that the bonafide reconstructions have lower fidelity to their original material when compared to the reconstruction of deepfake videos. Carrying out a significance test at a significance level of 99%, we were able to show that the null hypothesis was rejected. Although these are initial results, the hypothesis that we can use implicit representations to detect deepfakes seems promising.

Future work

Having reached these conclusions, it is necessary to consider how to proceed. The end result of this research is expected to achieve state-of-the-art deepfake detection. There are still a number of obstacles to overcome, with problems such as data volume. It is still required to test if all frames from a video are required to achieve satisfactory results. This, among a battery of ablation tests, will be conducted as to conceive the "ideal" conditions to proceed with research.

While this paper revolves around videos reconstructed from their SIREN representations, it is to show a discernible distinction between deepfakes and bonafide material. Future work will be conducted as much as possible with the implicit representation itself.

REFERENCES

[1] João Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4733, 2017. doi: 10.1109/CVPR.2017.502.

[2] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton-Ferrer. The deepfake detection challenge dataset. *ArXiv*, abs/2006.07397, 2020.

[3] Songwei Ge, Aniruddha Mahapatra, Gaurav Parmar, Jun-Yan Zhu, and Jia-Bin Huang. On the Content Bias in Fréchet Video Distance . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7277–7288, Los Alamitos, CA, USA, June 2024. IEEE Computer Society. doi: 10.1109/CVPR52733.2024.00695.

[4] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[5] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405, 2019. doi: 10.1109/CVPR.2019.00453.

[6] Yuval Nirkin, Yosi Keller, and Tal Hassner. Fsgan: Subject agnostic face swapping and reenactment. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7183–7192, 2019. doi: 10.1109/ICCV.2019.00728.

[7] Adam Polyak, Lior Wolf, and Yaniv Taigman. Tts skins: Speaker conversion via asr. In *Interspeech*, 2019.

[8] Guilherme Schardong, Tiago Novello, Hallison Paz, Iurii Medvedev, Vinicius Da Silva, Luiz Velho, and Nuno Goncalves. Neural Implicit Morphing of Face Images . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7321–7330, Los Alamitos, CA, USA, June 2024. IEEE Computer Society. doi: 10.1109/CVPR52733.2024.00699.

[9] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 7462–7473. Curran Associates, Inc., 2020.

[10] Thomas Unterthiner, Sjoerd van Steenkiste, Karol Kurach, Raphaël Marinier, Marcin Michalski, and Sylvain Gelly. Fvd: A new metric for video generation. In *DGS@ICLR*, 2019.

[11] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

[12] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. Few-shot adversarial learning of realistic neural talking head models. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9458–9467, 2019. doi: 10.1109/ICCV.2019.00955.

Remote Sensing and AI-based Land Coverage Analysis for Wildfire Prevention and Planning

Matheus F. Kovaleski[1]
matheus.kovaleski@isr.uc.pt
João Ruivo Paulo[1]
jpaulo@isr.uc.pt
Cristiano Premebida[1]
cpremebida@isr.uc.pt
Jérôme Mendes[2]
jerome.mendes@uc.pt

[1] University of Coimbra
Institute of Systems and Robotics
Coimbra, PT

[2] University of Coimbra, CEMMPRE, ARISE, Department of Mechanical Engineering, Coimbra, PT

Abstract

Wildfires pose an increasing threat to ecosystems, humans, and global climate stability. With rising temperatures and changing precipitation patterns, fire-prone landscapes are expanding, demanding more precise data-driven approaches for wildfire prevention and real-time monitoring. Traditional fire risk assessment methods are heavily based on historical fire records, meteorological models, and ground-based surveys, often lacking the spatial and temporal resolution needed for proactive management. On the other hand, remote sensing on active fire mapping still lacks multimodal data and temporal resolution to mitigate smoke and cloud cover penetration and accurate and timely wildfire spread modeling limitations. To address these challenges, this work aims to develop advanced wildfire management tools based on machine learning approaches applied to multispectral satellite drone-based data. The main expected contributions of the work are: (i) identification of preventive actions for assessment of wildfire risk through land coverage analyses, and (ii) low latency and accurate active perimeter mapping for improved wildfire spread predictions.

1 Introduction

Remote sensing has emerged as a valuable technology for wildfire prevention and planning. On the prevention side it has helped to automate field surveys which are labor-intensive and not feasible for large-scale or frequent assessments [5, 8]. Remote sensing data and machine learning (ML) can be used to estimate fuel load, moisture, stress, and vegetation health, which are critical for predicting wildfire behavior and severity. On the other hand, during active wildfires, remote sensing can provide wildfire perimeter mapping, which allows accurate fire spread modeling for adequate planning.

Looking at the literature on fuel mapping, there are many relevant parameters to consider. These are the type of vegetation, the different height layers of the fuel (canopy, surface, vegetation), and moisture, where two major categories of fuel are considered, live and dead fuel [11]. These parameters are fundamental to provide an accurate model of the fuel in a given area. Previous works on remote sensing focused on the use of MODIS and MSG-SEVIRI data to overcome limitations from other ground observations for retrievals of fuel parameters [2, 6, 7]. Unmanned aerial vehicles (UAVs) are also used to collect vegetation indexes, using both infrared and optical cameras [1, 10]. ML has been used in these works, mainly convolutional neural networks (CNNs) for image processing. Despite these advancements, several challenges persist. These include technological constraints that hinder optimal fuel load mapping, the scarcity of ground truth data for algorithm calibration, and the difficulty in mapping understory vegetation and surface fuels.

When it comes to active fire mapping, recent advancements have significantly reduced data latency, providing both global and regional coverage [9]. However, the works found in the literature are mainly focused on fire detection. The works that focus on fire perimeter delineation [3, 4] use specific physics based algorithms like convex hull and concave to fit the wildfire's perimeter.

There is still a need for ultra real time (URT) and more advanced ML algorithms to deal with multispectral data for segmentation approaches of wildfire perimeter mapping.

Considering the previous analysis, this work aims to innovate and contribute to the state-of-the-art by:

- Combining multiple satellite band imagery for an improved accuracy in fuel analysis;
- Fusing satellite and drone images for a richer and more complete fuel analysis, compensating the limitations of each technology;
- Proposing ML-based wildfire risk assessment and preventive actions identification, such as targeted fuel cleaning, controlled burns, and defining containment lines. This effectively bridges the gap between remote sensing analytics and practical, on-ground wildfire mitigation efforts;
- Improving fire mapping using multispectral satellite data and advanced deep learning models;
- Combining complementary satellite platforms, including Sentinel-2, MODIS, and VIIRS, to achieve enhanced temporal resolution.

1.1 Objectives

This PhD work aims to advance wildfire management by leveraging remote sensing and machine learning. It addresses the following key research questions:

- How can multispectral satellite imagery and drone-based data be effectively combined to enhance vegetation analysis for wildfire prevention? This explores integrating diverse data sources to improve mapping of understory and surface fuels, critical for accurate vegetation classification and fuel load estimation.
- How can remote sensing data be translated into actionable preventive measures for wildfire risk mitigation? This focuses on converting data insights into practical tools that guide wildfire prevention strategies based on vegetation and risk analysis.
- What ML techniques can achieve near-real-time mapping of active wildfire perimeters using satellite data? This question targets the development of algorithms to process satellite data swiftly, enabling timely and precise delineation of fire perimeters.

Hence, the objectives of this work are:

- Develop a Multimodal Vegetation Analysis Model: Build an ML model that integrates satellite data (for example, Sentinel-2, MODIS) with drone imagery to classify vegetation and estimate fuel load.
- Create a Preventive Action Decision Support Tool: Design a tool that assesses wildfire risk using vegetation analysis and recommends preventive actions.
- Implement Near-Real-Time Wildfire Mapping: Create a system to map active wildfire perimeters using satellite data, targeting a temporal resolution of 10 minutes post-data acquisition.

2 Methodology

An overview of the methodology is shown in Fig. 1. The key blocks (modules) are described in the following sections.

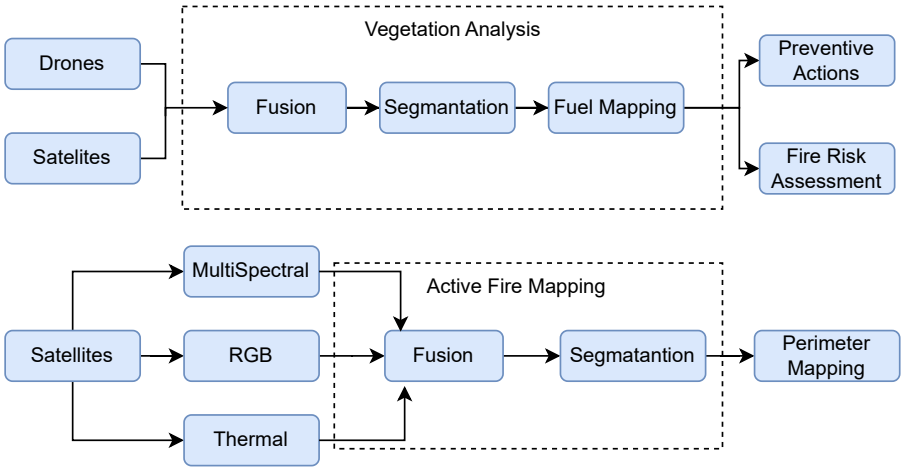


Figure 1: Diagram showing the use of drones and satellites for wildfire management operations.

2.1 Dataset Collection

The first step involves selecting multispectral satellite data (e.g. Sentinel-2) and high-resolution drone imagery for vegetation analysis, creating a multimodal dataset from national case studies. For active fire monitoring, satellite constellations with ultra-real-time feedback (e.g., GOES-R) are chosen, and a dataset is collected during controlled burns in the CEIF Lous field laboratory with annotated ground truths.

2.2 Vegetation Analysis

This task focuses on processing satellite imagery to extract NDVI and textural features, using convolutional neural networks for classification. Drone data are processed into orthomosaics and 3D point clouds, applying object detection and semantic segmentation to assess flammability. Data fusion aligns datasets, using multimodal neural networks to enhance fuel characteristic estimation, and regression models map these to preventive actions (e.g., firebreaks) for a decision support system.

2.3 Active Wildfire Mapping

This task relies on high-temporal-resolution satellite data: Access the satellite constellations with high temporal resolution and thermal infrared capabilities to detect active fire signatures effectively, and establish protocols for rapid data acquisition, leveraging ultra-real-time (URT) feedback systems to ensure timeliness. Data will be preprocessed to extract key fire-detection features, such as brightness temperature anomalies and spectral indices (e.g., Normalized Burn Ratio), to enhance mapping precision. The next step is the use of machine learning methodologies for accurate and timely mapping of active wildfire perimeters using multispectral data. Particularly, segmentation models to identify and outline active fire fronts from preprocessed satellite data. This task will consider an optimized computational pipeline to process satellite data swiftly and apply trained models.

Acknowledgments

This research is partially sponsored by national funds through FCT under projects UID/00285 - Centre for Mechanical Engineering, Materials and Processes, LA/P/0112/2020, UIDB/00048/2020 (https://doi.org/10.54499/UIDB/00048/2020), UIDP/00048/2020 (https://doi.org/10.54499/UIDP/00048/2020), UIDP&B/4950/2020.

References

[1] Turkia Al-Moustafa, Richard P. Armitage, and F. Mark Danson. Mapping fuel moisture content in upland vegetation using airborne hyperspectral imagery. *Remote Sensing of Environment*, 127:74–83, 2012. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2012.08.034.

[2] Alexandra Bjånes, Rodrigo De La Fuente, and Pablo Mena. A deep learning ensemble model for wildfire susceptibility mapping. *Ecological Informatics*, 65:101397, 2021. ISSN 1574-9541. doi:https://doi.org/10.1016/j.ecoinf.2021.101397.

[3] Carlos Ivan Briones-Herrera, Daniel José Vega-Nieva, Jaime Briseño-Reyes, Norma Angélica Monjarás-Vega, Pablito Marcelo López-Serrano, José Javier Corral-Rivas, Ernesto Alvarado, Stéfano Arellano-Pérez, Enrique J. Jardel Peláez, Diego Rafael Pérez Salicrup, and William Matthew Jolly. Fuel-specific aggregation of active fire detections for rapid mapping of forest fire perimeters in mexico. *Forests*, 13(1), 2022. ISSN 1999-4907. doi:https://doi.org/10.3390/f13010124.

[4] Yang Chen, Stijn Hantson, Niels Andela, Shane Coffield, Casey Graff, Douglas Morton, Lesley Ott, Efi Foufoula-Georgiou, Padhraic Smyth, Michael Goulden, and James Randerson. California wildfire spread derived using viirs satellite observations and an object-based tracking system. *Scientific Data*, 9, 05 2022. doi:https://doi.org/10.1038/s41597-022-01343-0.

[5] Matthew G. Gale, Geoffrey J. Cary, Albert I.J.M. Van Dijk, and Marta Yebra. Forest fire fuel through the lens of remote sensing: Review of approaches, challenges and future directions in the remote sensing of biotic determinants of fire behaviour. *Remote Sensing of Environment*, 255:112282, 2021. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2020.112282.

[6] Yizhi Han, Xiaojing Bai, Wei Shao, and Jie Wang. Retrieval of soil moisture by integrating sentinel-1a and modis data over agricultural fields. *Water*, 12(6), 2020. ISSN 2073-4441. doi:https://doi.org/10.3390/w12061726.

[7] Xingwen Quan, Marta Yebra, David Riaño, Binbin He, Gengke Lai, and Xiangzhuo Liu. Global fuel moisture content mapping from modis. *International Journal of Applied Earth Observation and Geoinformation*, 101:102354, 2021. ISSN 1569-8432. doi:https://doi.org/10.1016/j.jag.2021.102354.

[8] John S. Schreck, William Petzke, Pedro A. Jimenez, Thomas Brummet, Jason C. Knievel, Eric James, Branko Kosovic, and David John Gagne. Machine learning and viirs satellite retrievals for skillful fuel moisture content monitoring in wildfire management, 2023. URL https://doi.org/10.48550/arXiv.2305.11910.

[9] Joseph M. Smith. FIRMS Adds Ultra Real-Time Data from MODIS and VIIRS | NASA Earthdata — earthdata.nasa.gov. URL https://shorturl.at/xjHq4. [Accessed 13-05-2025].

[10] Jian Xing, Chaoyong Wang, Ying Liu, Zibo Chao, Jiabo Guo, Haitao Wang, and Xinfang Chang. Uav multispectral imagery predicts dead fuel moisture content. *Forests*, 14(9), 2023. ISSN 1999-4907. doi:https://doi.org/10.3390/f14091724.

[11] Marta Yebra, Philip E. Dennison, Emilio Chuvieco, David Riaño, Philip Zylstra, E. Raymond Hunt, F. Mark Danson, Yi Qi, and Sara Jurdao. A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products. *Remote Sensing of Environment*, 136:455–468, 2013. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2013.05.029.

Robustness of Deep Learning Based Face Recognition Under Morphing Attacks

Iurii Medvedev
https://visteam.isr.uc.pt/team/iurii-medvedev/

Institute of Systems and Robotics,
University of Coimbra,
Portugal

Nuno Gonçalves
https://visteam.isr.uc.pt/team/nuno-goncalves-2/

Portuguese Mint and Official Printing Office (INCM),
Lisbon,
Portugal

INTRODUCTION

Last decades with the development of deep learning techniques the evident advances have been reached in the field of face recognition. However at the same time more evolved and sophisticated techniques for performing the presentation attacks continue to appear (see Fig. 1), which require the development of new protection solutions [1].

One of such face image manipulating methods is **Face Morphing**. Image morphing techniques are used to combine information from two (or more) images into one image. Over the past decade, it has gained significant attention and has been more thoroughly investigated. As awareness of the problem has grown, numerous counterfeit documents employing face morphing techniques have been uncovered at control gates [12] (an example is presented at Fig. 2).

Figure 1: Example of various face morphing techniques.

Figure 2: Real face morphing example. a) - ID image of accomplice (requester of a document); b) Face image on counterfeited ID Document; c) Face image of a person that used a counterfeited document.

Given the importance of reducing vulnerabilities in modern face recognition systems and the significant risks posed by presentation attacks, this thesis aims to contribute with the tools for combating the face morphing problem involving deep learning algorithms.

FACE MORPHING ATTACKS PROBLEM

The potential for morphing attacks to compromise identification systems was first explored in [3]. This study shows how ICAO-compliant morphed images could effectively bypass both human and automated border control checks.

The typical pipeline for creating a morphed identity document for impersonation involves several coordinated steps designed to exploit face recognition systems. First, a wanted individual collaborates with a complicit accomplice to generate a synthetic morphed face image with facial features from both parties. This morphed image is made to be realistic and compliant with official ID photo standards. Next, the accomplice uses this image to apply for and obtain a legitimate identification document, such as a passport or national ID card. Once issued, the document (though legally tied to the accomplice) can also successfully match the facial characteristics of the wanted person. As a result, the wanted individual is able to use the authentic document to travel or access secure services while evading detection.

RESEARCH OBJECTIVES AND QUESTIONS

The primary objective of this thesis is to contribute to enhancing the robustness of face recognition systems against presentation attacks, particularly those involving face morphing [10, 11]. Our research will focus on two key areas: improving the detection of morphing attacks and strengthening the robustness of deep facial feature representations against morphing.

In this context, we can outline our specific objectives as follows:

• Study data collection for face recognition, including morphed face generation.

• Improve deep learning methods for morphing detection and vulnerability analysis in ID enrollment.

• Develop morph-resistant face recognition strategies for secure document applications.

• Evaluate the effectiveness of proposed methods with custom protocols and bublic benchmarks.

MORPHING ATTACK DETECTION

Morphing Attack detection is a straightforward approach to combat their risks for facial biometric systems and it is typically categorized into two processing pipelines based on the availability of reference data: *no-reference* and *differential* approaches.

In the *no-reference* or *Single Morphing Attack Detection (SMAD)* scenario, the algorithm receives only a single face image without any corresponding trusted reference and must determine whether the image is morphed (for instance, in Enrollment pipelines).

Differential morphing attack detection (DMAD) methods rely on the availability of a trusted reference image typically captured during a live interaction with the facial biometric system allowing the comparison between the live capture and the enrolled (potentially morphed) image (for instance in Automated Border Control).

For morphing attack detection, we proposed advanced detection strategies based on multitask learning frameworks with sophisticated morph sample labeling (see Fig. 3). In our setup, images are processed by two parallel feature extractors, and their outputs are compared to evaluate whether they belong to the same identity. The underlying principle is that genuine (non-morphed) face pairs will produce highly similar features, while morphed images will result in dissimilar outputs.

On practice this implies the designing of a complex multitask learning problem and we will call such concept as *Fused Classification* further in the work.

Our methods achieved state-of-the-art performance in publicly available benchmarks for both Single-image Morphing Attack Detection (SMAD) [4] and Differential Morphing Attack Detection (DMAD) [8] scenarios.

INCREASING THE ROBUSTNESS OF FACIAL BIOMETRIC TEMPLATES TO MAD

Beyond detection, alternative strategies exist. For instance in can be approached by increasing the *Robustness* the face feature templates to morphing attacks. This approach shifts the emphasis from detecting morphs

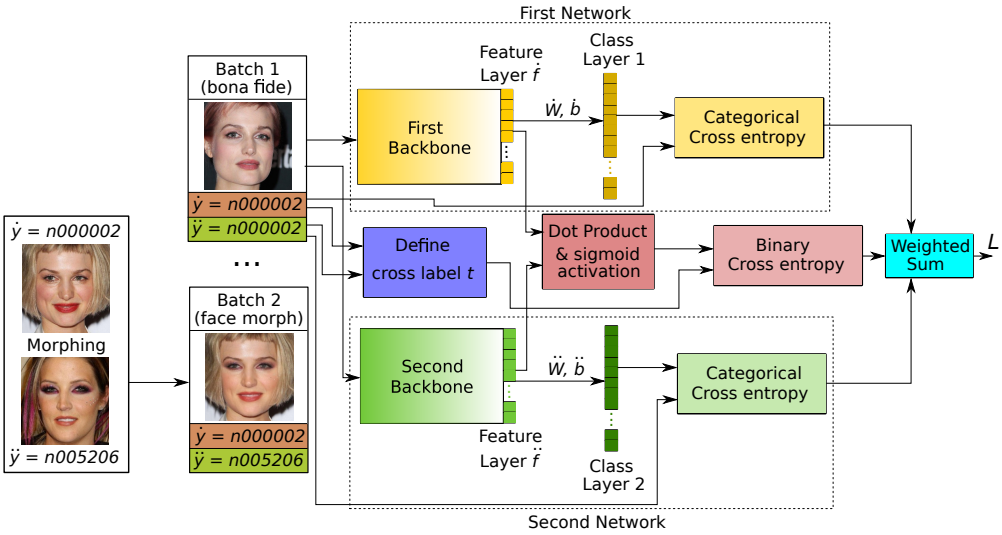


Figure 3: Schematic of the Fused Classificartion. For clarity in visualization, the presented batch contains only a single image. Labels $\dot{y}$ and $\ddot{y}$ are displayed with descriptive names for better understanding. In the actual implementation, these labels are represented by their numerical index values, which are subsequently encoded into one-hot vectors for processing.

to mitigating their impact by reducing the likelihood that morph samples can successfully match legitimate face templates.

Achieving this level of robustness requires modifications to the core of the face recognition system. Namely the target is to modify the face feature extraction mechanism, which is used for generating biometric templates. Such robustness based strategies focus on improving the discriminatory power of the templates themselves by modifying the deep face feature domain.

Our initial approach here addresses contrastive learning methods by introducing a dedicated branch for morph samples, allowing explicit control over their feature distribution [2]. Additionally, we refine traditional classification strategies through a carefully designed softmax-based margin loss, which intentionally disbalabce morph samples from bona fide ones [5].

THESIS CONTRIBUTIONS

In this thesis we approached the problem of face recognition robustness to morphing attacks from multiple angles, introducing new methods for morphing attacks detection and face image templates robustness.

In the work, many collateral contributions were presented, which appeared mainly in the process of data curation and performance assessment. This include novel datasets and benchmarks tailored to face recognition tasks [9]. These resources have been made available to the academic community and are already in use for new research projects.

This also include the developed data filtering techniques and resulting metadata for public academic face datasets [13] [6], which can facilitate more refined data preparation in future studies.

Our contributions include the creation of a dedicated benchmarks for face morphing detection [4][8] and evaluating robustness to morphing attacks [7], which is a resource that, to our knowledge, currently has no publicly available alternative and is designed to scale with future developments .

Summing up in this thesis we provided significant contributions that support the broader academic community in advancing research in face recognition and presentation attack detection.

REFERENCES

[1] Zahid Akhtar, Dipankar Dasgupta, and Bonny Banerjee. Face Authenticity: An Overview of Face Manipulation Generation, Detection and Recognition. *SSRN Electronic Journal*, 01 2019. doi: 10.2139/ssrn.3419272.

[2] W. Chen, X. Chen, J. Zhang, and K. Huang. Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1320–1329, 2017. doi: 10.1109/CVPR.2017.145.

[3] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. The magic passport. *IJCB 2014 - 2014 IEEE/IAPR International Joint Conference on Biometrics*, 12 2014. doi: 10.1109/BTAS.2014.6996240.

[4] I. Medvedev, F. Shadmand, and N. Gonçalves. Mordeephy: Face morphing detection via fused classification. In *Proceedings of ICPRAM*, pages 193–204. SciTePress, 2023. ISBN 978-989-758-626-2. doi: 10.5220/0011606100003411.

[5] Iurii Medvedev and Nuno Goncalves. Morphguard: Morph specific margin loss for enhancing robustness to face morphing attacks, 2025. URL https://arxiv.org/abs/2505.10497.

[6] Iurii Medvedev and Nuno Gonçalves. Improving performance of facial biometrics with quality-driven dataset filtering. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–8, 2023. doi: 10.1109/FG57933.2023.10042579.

[7] Iurii Medvedev and Nuno Gonçalves. Morfacing: A benchmark for estimation face recognition robustness to face morphing attacks. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10, 2024. doi: 10.1109/IJCB62174.2024.10744449.

[8] Iurii Medvedev, Joana Alves Pimenta, and Nuno Gonçalves. Fused classification for differential face morphing detection. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, pages 1043–1050, 2024. doi: 10.1109/WACVW60836.2024.00114.

[9] Iurii Medvedev, Farhad Shadmand, and Nuno Gonçalves. Young labeled faces in the wild (ylfw): A dataset for children faces recognition. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–10, 2024. doi: 10.1109/FG59268.2024.10582021.

[10] Raghavendra Ramachandra and Christoph Busch. Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey. *ACM Comput. Surv.*, 50(1), March 2017. ISSN 0360-0300. doi: 10.1145/3038924. URL https://doi.org/10.1145/3038924.

[11] U. Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch. On the vulnerability of face recognition systems towards morphed face attacks. In *2017 5th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6, 2017. doi: 10.1109/IWBF.2017.7935088.

[12] Matjaž Torkar. Morphing cases in slovenia. NIST IFPS, 2022. Ministry of the Interior Police, Slovenia.

[13] João Tremoço, Iurii Medvedev, and Nuno Gonçalves. QualFace: Adapting Deep Learning Face Recognition for ID and Travel Documents with Quality Assessment. In *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6, 2021.

Towards Power-Efficient Bayesian Causal Spiking Neural Networks

Dylan Perdigão
dgp@dei.uc.pt
Francisco Antunes
fnibau@dei.uc.pt
Catarina Silva
catarina@dei.uc.pt
Bernardete Ribeiro
bribeiro@dei.uc.pt

University of Coimbra
CISUC/LASI – Centre for Informatics and Systems
of the University of Coimbra
Department of Informatics Engineering
Coimbra (PT)

INTRODUCTION

The brain implements causal inference to solve complex problems like object recognition through our eyes or smell identification with our nose. Many decisions are based on assumptions about unknown causes from ambiguous and noisy observations. The brain's ability to efficiently implement these inferences is critical because the number of potential sensory stimuli is enormous, and stimuli from different classes are often remarkably similar [11]. For instance, distinguishing the smell of coffee from tea involves recognizing a specific mixture of chemicals from countless possible combinations. Neuromorphic (or brain-inspired) algorithms emerged to solve computational problems efficiently by advancing Artificial Intelligence (AI). More specifically, neuroscience has inspired different modeling of Artificial Neural Networks (ANNs) through different neuron models, led to the current Spiking Neural Networks (SNNs) pioneered by Maass in 1997 [10]. Unlike traditional artificial neural networks that use continuous activation functions, SNNs use discrete spikes for reasoning, mimicking how biological neurons operate. Moreover, SNNs become significantly power-efficient when implemented on neuromorphic hardware, making this difference point to a singularity, where computers reach human performance levels [9]. While the brain uses spiking neurons for learning, most theoretical research has focused on non-spiking networks. The nature of spike-based algorithms that achieve complex computations, such as object probabilistic inference, has been largely unknown until recent advancements. For instance, Moreno-Bote demonstrated that a family of high-dimensional quadratic optimization problems with non-negativity constraints can be solved precisely and efficiently by a spiking neuron network [11]. This network naturally imposes the non-negativity of causal contributions fundamental to causal inference and employs simple operations like linear synapses and neural spike generation. The spiking networks are robust against internal and external variability and can dynamically implement explaining away through spike-based, tuned inhibition. This robustness and efficiency highlight the potential of SNNs in neuromorphic computing, where power efficiency and computational performance are essential. Thus, understanding and leveraging the mechanisms of causal spiking networks could shorten the gap between artificial and biological intelligence, leading to more robust and powerful AI systems.

SPIKING NEURAL NETWORKS

Biologically, the brain comprises the fundamental nervous system cells called neurons [4]. Neurons communicate via electrochemical impulses known as action potentials (or spikes), which travel along the cell structures. The neuron body is a cell surrounded by a pored membrane that separates the inside from the outside. These pores will regulate the concentration of sodium and potassium ions. For instance, when the electrical charge of the sodium ions passes the threshold of the neuron's membrane, the ions are quickly replaced with the absorption of potassium ions [8]. This rapid variation in the electrochemical forces and the resultant equilibrium potentials are fundamental to generating and transmitting electrical signals between neurons, which are the action potentials [6]. In spiking neural networks, the exchanges of chemical ions are simulated as an electrical circuit with the Leaky Integrate-and-Fire (LIF) neuron model [5], which consists of a differential equation system, with a capacitor in parallel and a resistor driven by a current intensity [3]. This system can be approximated using the forward Euler method for compatibility with sequential networks. In the simplified model, the input current combines the weighted sum of inputs of the neural network with the electrical potential of the circuit. The raw input is converted into a matrix of spikes, where their superposition on each layer of the SNN creates the current. This current is translated into the potential of the membrane, generating spikes for the next layer. The reset is defined by subtracting the threshold via the Heaviside step function of the last recorded spike, as shown in Figure 1.

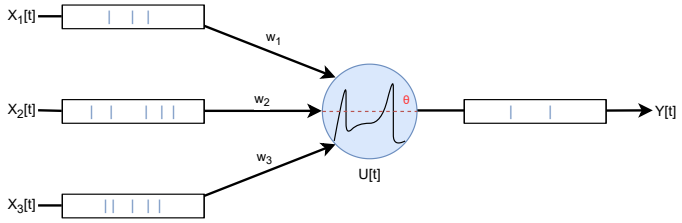


Figure 1: Spiking Neural Network with a LIF neuron.

Causal spiking neural networks can combine the temporal dynamics of spiking neurons with causal inference principles to create models that can better emulate the brain's processing of information. Causal networks can capture and explain the probabilistic dependencies between neurons and their spike timings, enabling a deeper understanding of the underlying causal mechanisms. However, there is a gap in this domain, where only a few works address this area.

NEUROMORPHIC COMPUTING

The popularity of SNNs increased drastically since they are energetically efficient on neuromorphic hardware, making them suitable to mitigate the high energy costs of AI that make performant models inaccessible for production purposes [2]. Neuromorphic hardware solutions are inspired by the human brain's ability to function on approximately 20W of power. As a comparison, a traditional computer vision algorithm can spend up to 250W for object recognition tasks [17] on a Graphical Processing Unit (GPU). Our current research is focused on Synsense's Speck [16] hardware, which is an ideal candidate for developing sustainable AI systems that align both energy efficiency goals and the fairness needed in model deployment. The advantage of using SNNs as opposed to the traditional ANNs or other Machine learning (ML) algorithms lies in their temporal and spatial sparsity, enabling the reduction of on/off activation on neuromorphic hardware, resulting in energy-efficient algorithms.

RESEARCH APPROACH

The current research, described in Figure 2, focuses on understanding the behavior of the current state-of-the-art SNNs. This understanding allows a first phase for the development of diverse SNN architectures that are fair and performant in the context of constrained financial fraud detection [7]. SNN architectures, which are very sensitive to their hyperparameters, are optimized through multi-objective Bayesian optimization process. This optimization is challenging due to its constraint of identifying the best True Positive Rate (TPR) at a specific point of the Receiver Operating Characteristic (ROC) curve, which is 5% of False Positive Rate (FPR) while attending to the fairness of the model towards sensitive attributes [1, 14]. This enables the analysis of the disparities of fraud, for instance, between the people's age, income, or employment status [18]. The second phase involves integrating causal inference mechanisms into the spiking

neuron to improve its performance and fairness. Some strategies studied are estimating the causal effect when the neuron is close to the spike and reconsidering the neuron's spiking decision. Another approach consists of evaluating the causality inside the neuron's temporal steps, for instance with Granger causality. This enables a more realistic modeling of the neuron, which is closer to the biological brain. To validate this approach, a dataset will be developed with a priori known relations between the features to observe the behavior of the causal spiking neuron. Finally, we will propose an approach that evaluates a spiking neural network with a causal inference framework. The idea is to use causality in the spiking neural network while giving some insights into its decision-making. The causal inference framework will be divided into two approaches. The first is a data-driven approach, where the data will be analyzed with causality to understand the relation of a causal effect between features. The second, is a model-driven approach, where the model will be analyzed to dive insights into the decision process of the spiking neural network. The idea of the framework is to enable a generalization of the analysis, not just to the financial sector.

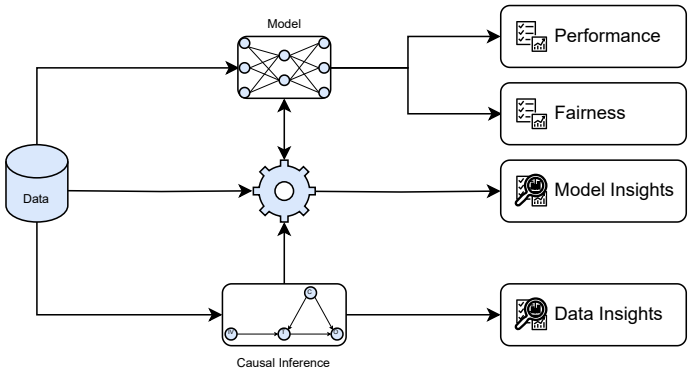


Figure 2: Pipeline towards Bayesian Causal Spiking Neural Networks.

CONCLUSION

The integration of Bayesian causal inference with spiking neural networks presents a promising avenue for enhancing the explainability, fairness, and performance of AI models. By exploring the biological neuron mechanisms of SNNs and combining them with Bayesian inference, this research aims to bridge the gap between artificial and biological intelligence. The key steps of this research work involve optimizing SNN architectures for specific real-world applications, such as financial fraud detection, which has optimization constraints in a highly imbalanced context, and then integrating causal inference mechanisms to improve these models further while developing a framework to give more insights on the decision-making process of SNNs.

The current work focused on exploring different optimization strategies for SNNs, specifically applied to the highly imbalanced task of financial fraud detection [13, 15]. Bayesian optimization was used, alongside other approaches such as population coding mechanisms [12]. Additionally, GPU-based power consumption was evaluated as a baseline for energy efficiency. Future work will focus on integrating our SNN models into the Speck neuromorphic hardware, which has already demonstrated competitive performance and represents a promising step towards a more sustainable AI.

REFERENCES

[1] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic Decision Making and the Cost of Fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 797–806, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 978-1-4503-4887-4. doi: 10.1145/3097983.3098095.

[2] Alex de Vries. The growing energy footprint of artificial intelligence. *Joule*, 7(10):2191–2194, October 2023. ISSN 2542-4351. doi: 10.1016/j.joule.2023.09.004.

[3] Sangya Dutta, Vinay Kumar, Aditya Shukla, Nihar R. Mohapatra, and Udayan Ganguly. Leaky Integrate and Fire Neuron by Charge-Discharge Dynamics in Floating-Body MOSFET. *Scientific Reports*, 7(1):8257, 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-07418-y.

[4] Michael Foster and Charles Scott Sherrington. *A Textbook of Physiology*, volume 3. Macmillan, London, 7th ed edition, 1897. ISBN 978-1-4325-1085-5.

[5] Wulfram Gerstner and Werner M. Kistler. *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press, Cambridge, U.K. ; New York, 2002. ISBN 978-0-521-81384-6 978-0-521-89079-3.

[6] Michael H. Grider, Rishita Jessu, and Rian Kabir. Physiology, Action Potential. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2024.

[7] Sérgio Jesus, José Pombal, Duarte Alves, André F Cruz, Pedro Saleiro, Rita P Ribeiro, João Gama, and Pedro Bizarro. Turning the Tables: Biased, Imbalanced, Dynamic Tabular Datasets for ML Evaluation. In *36th Conference on Neural Information Processing Systems Datasets and Benchmark Track*, 2022. doi: 10.48550/arXiv.2211.13358.

[8] Louis Lapicque. Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *Journal de physiologie et de pathologie générale*, pages 1–16, 1907.

[9] Richard Liu and Fredrik Bixo. *Analysing the Energy Efficiency of Training Spiking Neural Networks*. First Cycle Project, KTH Royal Institute of Technology, 2022.

[10] Wolfgang Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997. ISSN 0893-6080. doi: 10.1016/S0893-6080(97)00011-7.

[11] Rubén Moreno-Bote and Jan Drugowitsch. Causal Inference and Explaining Away in a Spiking Network. *Scientific Reports*, 5(1):17531, December 2015. ISSN 2045-2322. doi: 10.1038/srep17531.

[12] Dylan Perdigão, Francisco Antunes, Catarina Silva, and Bernardete Ribeiro. Exploring Neural Joint Activity in Spiking Neural Networks for Fraud Detection. In Ruber Hernández-García, Ricardo J. Barrientos, and Sergio A. Velastin, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 45–59, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-76604-6. doi: 10.1007/978-3-031-76604-6_4.

[13] Dylan Perdigão, Francisco Antunes, Catarina Silva, and Bernardete Ribeiro. Improving Fraud Detection with 1D-Convolutional Spiking Neural Networks Through Bayesian Optimization. In Manuel Filipe Santos, José Machado, Paulo Novais, Paulo Cortez, and Pedro Miguel Moreira, editors, *Progress in Artificial Intelligence*, pages 127–138, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-73503-5. doi: 10.1007/978-3-031-73503-5_11.

[14] José Pombal, André F. Cruz, João Bravo, Pedro Saleiro, Mário A. T. Figueiredo, and Pedro Bizarro. Understanding Unfairness in Fraud Detection through Model and Data Bias Interactions, 2022.

[15] Bernardete Ribeiro, Francisco Antunes, Dylan Perdigão, and Catarina Silva. Convolutional Spiking Neural Networks targeting learning and inference in highly imbalanced datasets. *Pattern Recognition Letters*, 189:241–247, March 2025. ISSN 0167-8655. doi: 10.1016/j.patrec.2024.08.002.

[16] Ole Richter, Yannan Xing, Michele De Marchi, Carsten Nielsen, Merkourios Katsimpris, Roberto Cattaneo, Yudi Ren, Yalun Hu, Qian Liu, Sadique Sheik, Tugba Demirci, and Ning Qiao. Speck: A Smart event-based Vision Sensor with a low latency 327K Neuron Convolutional Neuronal Network Processing Pipeline, May 2024. arXiv:2304.06793.

[17] Kaushik Roy, Akhilesh Jaiswal, and Priyadarshini Panda. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784):607–617, November 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1677-2. Publisher: Nature Publishing Group.

[18] Pedro Saleiro, Benedict Kuester, Loren Hinkson, Jesse London, Abby Stevens, Ari Anisfeld, Kit T. Rodolfa, and Rayid Ghani. Aequitas: A Bias and Fairness Audit Toolkit, 2018.

IbPRIA 2025

INSTITUTO DE SISTEMAS E ROBÓTICA
UNIVERSIDADE DE COIMBRA

Book of Extended Abstracts

Bernadete Ribeiro,
Catarina Silva,
Nuno Gonçalves and
Gustavo Bongiovi (Eds.)