CARDIAC ARRHYTHMIAS DETECTION WITH 12-LEAD ECG USING A MULTIMODAL NEURAL NETWORK

DIDIER JULIAN MORENO ORTIZ
KARLA VANESSA RUÍZ GONZÁLEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2025

CARDIAC ARRHYTHMIAS DETECTION WITH 12-LEAD ECG USING A MULTIMODAL NEURAL NETWORK

DIDIER JULIAN MORENO ORTIZ

KARLA VANESSA RUÍZ GONZÁLEZ

Thesis submitted in partial fulfillment of the requirements for the degree of Electronic Engineer

Advisor

MSc. (c) Harold H. Rodriguez

Co-advisors

MSc. Camilo Santos

Carlos A. Fajardo, Ph.D.

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE INGENIERÍAS FISICOMECÁNICAS

ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES

INGENIERÍA ELECTRÓNICA

BUCARAMANGA

2025

CONTENT

LIST OF FIGURES

LIST OF TABLES

p.

ANNEXES

RESUMEN

TÍTULO DETECCIÓN DE ARRITMIAS CARDÍACAS CON ECG DE 12 DERIVACIONES UTILIZANDO UNA RED NEURONAL MULTIMODAL [*]

AUTOR: DIDIER JULIAN MORENO ORTIZ, KARLA VANESSA RUÍZ GONZÁLEZ [**]

PALABRAS CLAVE: Arritmias cardíacas, ECG, Multimodal, Aprendizaje profundo, Interpretabilidad

DESCRIPCIÓN: Las enfermedades cardiovasculares son la principal causa de muerte a nivel mundial, siendo las arritmias cardíacas una de las condiciones más frecuentes y difíciles de diagnosticar. El diagnóstico se basa tradicionalmente en electrocardiogramas (ECG) de 12 derivaciones, los cuales generan grandes volúmenes de datos que requieren interpretación experta y pueden saturar al personal clínico en entornos de alta demanda. Este estudio propone una red neuronal multimodal para la detección de arritmias a partir de ECG de 12 derivaciones, incorporando interpretabilidad para identificar las derivaciones más influyentes. El modelo se entrenó inicialmente con datos de los desafíos PhysioNet/Computing in Cardiology 2020 y 2021, utilizando una red neuronal convolucional basada exclusivamente en la señal del ECG, y se analizó mediante Shapley Additive Explanations para jerarquizar la importancia de las derivaciones. Posteriormente, se reentrenó y evaluó empleando configuraciones de 6, 4 y 2 derivaciones, tanto en modelos basados únicamente en la señal del ECG como en configuraciones multimodales que combinan el ECG con características demográficas y temporales. Los resultados muestran que los modelos de 4 y 6 derivaciones alcanzaron puntajes F1 comparables o superiores a la línea base de 12 derivaciones, mientras que el modelo de 2 derivaciones mantuvo un rendimiento estable. La métrica oficial del PhysioNet/Computing in Cardiology Challenge alcanzó un valor máximo de 0.9074. El enfoque multimodal no produjo mejoras significativas, lo que resalta la contribución limitada de las características adicionales. Estos hallazgos demuestran la viabilidad de los modelos con derivaciones reducidas como alternativas eficientes y accesibles para la clasificación de arritmias.

---

[*] Tesis

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunucaciones. Director: MSc. (c) Harold H. Rodriguez. Codirectores: MSc. Camilo Santos; Carlos A. Fajardo, Ph.D. Electronic Engineer.

ABSTRACT

TITLE: CARDIAC ARRHYTHMIAS DETECTION WITH 12-LEAD ECG USING A MULTIMODAL NEURAL NETWORK *

AUTOR: DIDIER JULIAN MORENO ORTIZ, KARLA VANESSA RUÍZ GONZÁLEZ **

Keywords: Cardiac Arrhythmias, ECG, Multimodal, Deep Learning, Interpretability

Description: Cardiovascular diseases are the leading cause of death worldwide, with cardiac arrhythmias among the most prevalent and diagnostically challenging conditions. Diagnosis traditionally relies on 12-lead electrocardiograms (ECG), which produce large amounts of data requiring expert interpretation and may overwhelm clinicians in high-demand settings. This study proposes a multimodal neural network for arrhythmia detection from 12-lead ECGs, incorporating interpretability to identify the most influential leads. The model was initially trained on data from the PhysioNet/Computing in Cardiology Challenges 2020 and 2021 using a signal-only convolutional neural network and analyzed with Shapley Additive Explanations to rank lead importance. It was then retrained and evaluated using 6, 4, and 2 leads, in both signal-only and multimodal configurations combining ECGs with demographic and temporal features. Results show that 4- and 6-lead models achieved F1-scores comparable to or higher than the 12-lead baseline, while the 2-lead model maintained stable performance. The official PhysioNet/Computing in Cardiology Challenge metric reached 0.9074 at best. The multimodal approach did not yield significant improvements, underscoring the limited contribution of the additional features. These findings demonstrate the feasibility of reduced-lead models as efficient and accessible alternatives for arrhythmia classification.

* Thesis

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunucaciones. Advisor: MSc. (c) Harold H. Rodriguez. Co-advisors: MSc. Camilo Santos; Carlos A. Fajardo, Ph.D. Electronic Engineer.

INTRODUCTION

Cardiovascular diseases (CVD) are the leading cause of death worldwide, with 17.9 million deaths annually, according to the World Health Organization. Over 80 % of these deaths result from heart attacks and strokes [1]. Cardiac arrhythmias, a common condition of CVD, are characterized by irregular heart rhythms, typically detected using 12-lead electrocardiograms (ECGs) that capture the heart's full electrical activity [2]. However, this diagnostic process requires time-consuming cardiologist interpretation and faces practical limitations, as many healthcare facilities lack consistent access to 12-lead ECG systems [3].

Furthermore, the clinical complexity of arrhythmias may extend beyond what ECG signals alone can reveal, as metadata such as demographic information often influences their characteristics in ways that are not fully leveraged by traditional analysis [4]. Although medical staff take these aspects into account in practice, manual interpretation is constrained by time and workload, making it difficult to systematically incorporate such factors to reveal underlying features. This challenge is compoun-

---

[1] World Health Organization. *Cardiovascular diseases (CVDs)*. Website. Accessed: 2025-06-25. 2021.

[2] Paul Kligfield et al. «Recommendations for the standardization and interpretation of the electrocardiogram: Part I: The electrocardiogram and its technology». En: *Journal of the American College of Cardiology* 49.10 (mar. de 2007), págs. 1109-1127. DOI: `10.1016/j.jacc.2007.01.024`.

[3] Mohammed B. Abubaker y Bilal Babayiğit. «Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods». En: *IEEE Transactions on Artificial Intelligence* 4.2 (2023), págs. 373-382. DOI: `10.1109/TAI.2022.3159505`.

[4] P. Uma Mageswari et al. «Decoding Heart Health: A Predictive Model for Demographic-Based Cardiovascular Risk». En: *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*. 2024, págs. 1584-1591. DOI: `10.1109/ICTACS62700.2024.10840610`.

ded by the extensive data volume generated by 12-lead ECGs, which can overwhelm healthcare systems, particularly in high-demand or resource-limited settings, leading to diagnostic delays and increased cardiovascular mortality. These limitations emphasize the need for diagnostic approaches that improve efficiency and accessibility by integrating complementary information sources.

Deep learning has emerged as a significant approach for ECG interpretation, demonstrating notable effectiveness in arrhythmia classification [5]. Among these techniques, Convolutional Neural Networks (CNNs) have shown considerable capability, as they learn both spatial patterns from multi-lead configurations and temporal features from cardiac rhythm sequences [6]. This dual capability enables CNNs to extract clinically relevant features directly from raw ECG data without depending on manual feature engineering, while identifying subtle diagnostic patterns that may elude traditional analysis methods.

This approach has since been extended through multimodal neural networks, which enrich ECG signals by incorporating patient details such as age, symptoms, and other relevant factors to gain a fuller understanding of cardiac conditions [7]. In parallel, interpretability has also become a key research focus, with explainable artificial intelligence techniques being employed to identify influential components of the data, guiding a more efficient use of information while also providing greater transparency

---

[5] Shijie Zhou et al. «Deep Learning Applied to Electrocardiogram Interpretation». En: *Canadian Journal of Cardiology* 37.1 (2021), págs. 17-18. DOI: `https://doi.org/10.1016/j.cjca.2020.03.035`.

[6] Zahra Ebrahimi et al. «A review on deep learning methods for ECG arrhythmia classification». En: *Expert Systems with Applications: X* 7 (2020), pág. 100033. DOI: `https://doi.org/10.1016/j.eswax.2020.100033`.

[7] Lukas Hilgendorf et al. «Fully Automated Diagnosis of Acute Myocardial Infarction Using Electrocardiograms and Multimodal Deep Learning». En: *JACC: Advances* 4.8 (2025), pág. 102011. DOI: `https://doi.org/10.1016/j.jacadv.2025.102011`.

in the decision-making process [8].

The present study aims to develop a multimodal neural network for the detection of cardiac arrhythmias using 12-lead ECG recordings, while also incorporating interpretability to identify the most influential leads. This approach validates performance through comprehensive metrics, addressing challenges of limited manual interpretation and lead availability to enhance healthcare delivery in diverse clinical settings.

---

8 Tanjila Alam Sathi et al. «An interpretable electrocardiogram-based model for predicting arrhythmia and ischemia in cardiovascular disease». En: *Results in Engineering* 24 (2024), pág. 103381. DOI: 10.1016/j.rineng.2024.103381.

1. OBJECTIVES

1.1. GENERAL OBJECTIVE

To develop a multimodal neural network for the detection of cardiac arrhythmias in 12-lead ECG recordings.

1.2. SPECIFIC OBJECTIVES

- To preprocess the 12-lead ECG data to enable the training, validation and testing of a multimodal neural network.
- To select a multimodal neural network model for cardiac arrhythmia detection based on the state of the art.
- To evaluate the multimodal neural network using cross-validation, sensitivity, specificity, F1-score, and the PhysioNet Challenge 2020 metric.

2. RELATED WORK

Several studies have applied one-dimensional Convolutional Neural Network (1D-CNN) architectures for arrhythmia classification using 12-lead ECG signals. Sowmiya et al. [9] proposed a 1D-CNN model with ten convolutional layers, achieving an accuracy of 97.18 % in the identification of different cardiac rhythms. In contrast, Kurniawan et al. [10] introduced a lighter 1D-CNN architecture with four convolutional layers, distinguished by a preprocessing step that extracts signal segments centered on the QRS complex, achieving an accuracy of 98.8 % in arrhythmia classification.
In the PhysioNet/Computing in Cardiology Challenge 2020 [11], the winning team led by Natarajan et al. proposed a multimodal Wide-and-Deep neural network, which combined convolutional and Transformer-based deep feature extraction with manually engineered features and demographic attributes, achieving a PhysioNet Challenge score of 0.533 on the hidden test set [12]. The team ranked second, led by Zhao et al., developed a modified ResNet architecture with Squeeze-and-Excitation

---

[9] M. Sowmiya et al. «Multi-class classification of arrhythmias with 12-lead ECG signals using one-dimensional convolutional neural network». En: *International Conference on Computer Vision and Internet of Things 2023 (ICCVIoT'23)*. Vol. 2023. 2023, págs. 90-96. DOI: `10.1049/icp.2023.2859`.

[10] Arief Kurniawan, Bayu Aditya Triwibowo y Dion Hayu Fandiantoro. «Classification of Arrhythmias 12-Lead ECG Signals Based on 1 Dimensional Convolutional Neural Networks». En: *2024 International Conference on Smart Computing, IoT and Machine Learning (SIML)*. 2024, págs. 220-225. DOI: `10.1109/SIML61815.2024.10578089`.

[11] Erick A Perez Alday et al. «Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020». En: *Physiological Measurement* 41.12 (2020), pág. 124003. DOI: `10.1088/1361-6579/abc960`.

[12] Annamalai Natarajan et al. «A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification». En: *2020 Computing in Cardiology*. 2020, págs. 1-4. DOI: `10.22489/CinC.2020.107`.

blocks and incorporated patient age and sex into the final layer, which yielded a score of 0.520 on the same evaluation [13].

Building on these directions, Ponemash et al. [14] proposed a hybrid approach that integrates handcrafted time-domain features, with deep representations obtained through CNN and ResNet18 architectures applied to raw ECG signals. Similarly, Kiladze et al. [15] developed a multimodal neural network in which 12-lead ECG signals are processed by an LSTM branch in parallel with patient metadata such as age and sex through a linear network.

Other studies have focused on feature extraction through architectural designs tailored to the structure of ECG data. Jiang et al. [16] proposed a dual-branch CNN that processes limb and precordial leads separately, complemented by a domain-informed attention mechanism that leverages RR intervals to emphasize dynamic characteristics of the signal. In a related effort, Ayano et al. [17] emphasized interpretability by applying post-hoc methods such as Gradient-weighted Class Activation

---

[13] Zhibin Zhao et al. «Adaptive Lead Weighted ResNet Trained With Different Duration Signals for Classifying 12-lead ECGs». En: *2020 Computing in Cardiology*. 2020, págs. 1-4. DOI: `10.22489/CinC.2020.112`.

[14] Oleksandr Ponemash et al. «Hybrid Feature Extraction for 12-Lead Ecg Classification by Integrating Handcrafted and Deep Learning Techniques». En: *2025 17th International Conference on Computer and Automation Engineering (ICCAE)*. 2025, págs. 53-58. DOI: `10.1109/ICCAE64891.2025.10980605`.

[15] Mariya R. Kiladze et al. «Multimodal Neural Network for Recognition of Cardiac Arrhythmias Based on 12-Load Electrocardiogram Signals». En: *IEEE Access* 11 (2023), págs. 133744-133754. DOI: `10.1109/ACCESS.2023.3335176`.

[16] Rucheng Jiang et al. «A dual-branch convolutional neural network with domain-informed attention for arrhythmia classification of 12-lead electrocardiograms». En: *Engineering Applications of Artificial Intelligence* 139 (2025), pág. 109480. DOI: `https://doi.org/10.1016/j.engappai.2024.109480`.

[17] Yehualashet Megersa Ayano et al. «Interpretable Hybrid Multichannel Deep Learning Model for Heart Disease Classification Using 12-Lead ECG Signal». En: *IEEE Access* 12 (2024), págs. 94055-94080.

Mapping Plus Plus (Grad-CAM++) and SHapley Additive exPlanations (SHAP) [18] to analyze model decisions, enabling the visualization of both ECG leads and temporal segments that most influenced predictions.

Finally, a recent study combined handcrafted features with deep learning architectures by evaluating raw 12-lead ECG signals alongside Heart Rate Variability (HRV) descriptors extracted using Python biosignal processing libraries. While the hybrid design did not surpass the raw-signal models, the study incorporated explainability through SHAP, which identified the most influential leads and temporal segments for each arrhythmia class [19].

Previous studies have either focused on multimodal integration without interpretability or on interpretability applied in isolation. The most recent work combined both perspectives, but only at the class level and using ECG-derived features alone. In contrast, our approach integrates signal features with patient information within a multimodal framework, while incorporating interpretability from a global perspective to strengthen transparency and clinical applicability.

[18] Scott M Lundberg y Su-In Lee. «A Unified Approach to Interpreting Model Predictions». En: *Advances in Neural Information Processing Systems*. Ed. por I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017.

[19] Emmanuel C. Chukwu y Pedro A. Moreno-Sánchez. «Enhancing Arrhythmia Diagnosis with Data-Driven Methods: A 12-Lead ECG-Based Explainable AI Model». En: *Digital Health and Wireless Solutions*. Ed. por Mariella Särestöniemi et al. Cham: Springer Nature Switzerland, 2024, págs. 242-259.

3. METHODOLOGY

3.1. DATASET

This study utilizes publicly available data from the PhysioNet/Computing in Cardiology Challenges of 2020 and 2021 [20]. The data originates from four distinct sources, PTB-XL across various European countries, Georgia in the United States, Ningbo in China, and Chapman-Shaoxing in China, ensuring broad demographic representation. It comprises 12-lead ECG recordings sampled at 500 Hz with a duration of 10 seconds, accompanied by patient metadata and diagnoses.

As the study focuses specifically on cardiac arrhythmias, we initially filter records to include only arrhythmic events and normal rhythm. From this subset, we select six arrhythmia classes: atrial fibrillation (AF), atrial flutter (AFL), sinus arrhythmia (SA), sinus bradycardia (SB), sinus tachycardia (STach), and normal sinus rhythm (NSR), based on their sufficient sample sizes to ensure robust model training. Additionally, we exclude records containing multiple arrhythmias from the selected classes to support the study's emphasis on single-label classification, though this excludes complex real-world cases that could affect generalizability. We also discard samples lacking patient demographic data, resulting in a final database of 64,801 records, each corresponding to a unique patient. Table 1 shows the distribution of these records across the four contributing databases. For clarity, the Ningbo and Chapman-Shaoxing collections were grouped into a single column, as both were acquired in China and follow similar acquisition protocols. The percentage column indicates the relative proportion of each class within the database.

---

[20] Matthew A. Reyna et al. «Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021». En: *2021 Computing in Cardiology (CinC)*. Brno, Czech Republic, 2021, págs. 1-4. DOI: 10.23919/CinC53138.2021.9662687.

Table 1. Distribution of normal rhythm and arrhythmia classes across PhysioNet Challenge databases.

Dx	**PTB-XL**	**Georgia**	**Ningbo/Chap.**	**Total**	**Percentage**
AF	1437	509	1780	3726	5.75 %
AFL	46	155	7440	7641	11.79 %
SA	338	325	1531	2194	3.39 %
SB	276	1435	15283	16994	26.22 %
STach	557	1142	6498	8197	12.65 %
NSR	16652	1715	7682	26049	40.20 %
Total	19306	5281	40214	64801	100 %

3.2. FEATURE EXTRACTION

From the same ECG recordings, we extracted additional physiological parameters using the open-source NeuroKit2 Python library [21]. This tool supports signal preprocessing by removing noise and correcting baseline wander, detects key physiological events such as R-peaks to calculate instantaneous heart rate, and provides metrics related to HRV. Due to the 10-second signal duration, we prioritized temporal domain features for their reliability, with specific metrics detailed in Table 2, while non-linear, fractal, and entropy-based measures were excluded. Additionally, variables such as age and gender, obtained from the demographic information in the database, were included to provide clinical context.

3.3. DATA PREPROCESSING

To optimize computational efficiency, we downsampled the ECG recordings from 500 Hz to 250 Hz, following prior research demonstrating preserved diagnostic informa-

[21] Dominique Makowski et al. «NeuroKit2: A Python toolbox for neurophysiological signal processing». En: *Behavior Research Methods* 53 (2021), págs. 1689-1696. DOI: 10.3758/s13428-020-01516-y.

Table 2. Selected time-domain HRV features extracted from ECG signals.

Feature	Description
ECG_Rate_Mean	Average heart rate (beats per minute)
HRV_MeanNN	Mean of NN intervals (R-R intervals)
HRV_RMSSD	Root mean square of successive differences
HRV_SDSD	Standard deviation of successive differences
HRV_MedianNN	Median of NN intervals
HRV_MinNN	Minimum NN interval
HRV_MaxNN	Maximum NN interval
HRV_pNN20	Percentage of NN intervals $>$ 20 ms apart
HRV_SD1	Non-linear measure from Poincaré plot

tion at lower frequencies [22] [23]. Additionally, we applied z-score normalization to both the ECG signals and tabular features to standardize the data, ensuring a mean of 0 and a standard deviation of 1. For ECG signals, normalization was performed across all leads, while for tabular features, each continuous feature was standardized, facilitating model training and enhancing robustness across modalities.

3.4. MODEL DESCRIPTION

Figure 1 shows the multimodal architecture used in this work. The ECG branch was derived from a previously proposed convolutional model for arrhythmia classification [24] and was subsequently adapted to the characteristics and requirements of this

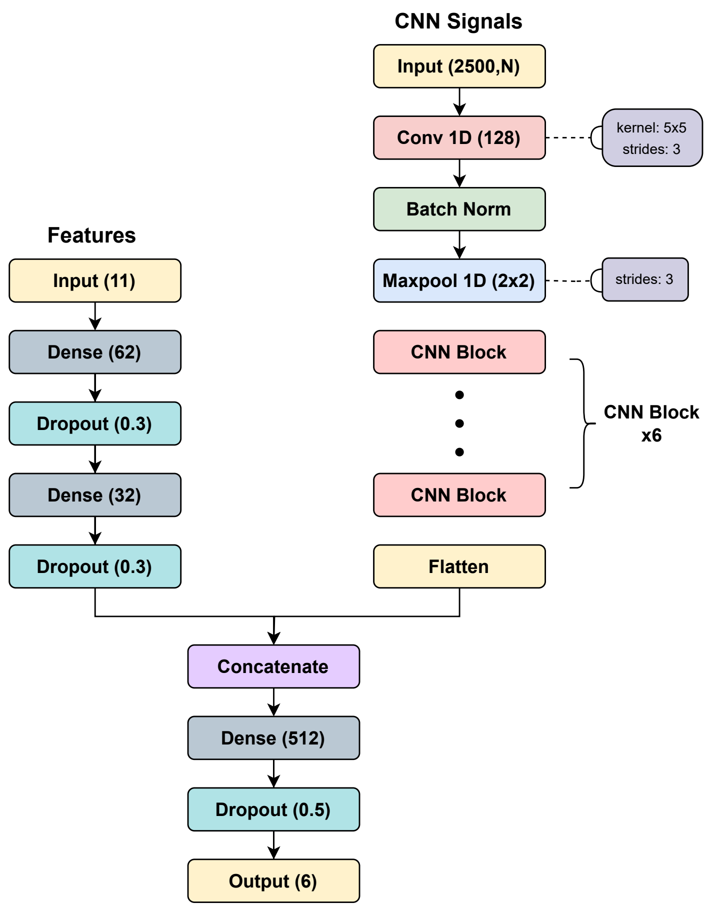
[22] Bochao Zhao et al. «DRL-ECG-HF: Deep reinforcement learning for enhanced automated diagnosis of heart failure with imbalanced ECG data». En: *Biomedical Signal Processing and Control* 107 (2025), pág. 107680. DOI: `https://doi.org/10.1016/j.bspc.2025.107680`.

[23] Yufeng Wei et al. «Bimodal Masked Autoencoders with internal representation connections for electrocardiogram classification». En: *Pattern Recognition* 161 (2025), pág. 111311. DOI: `https://doi.org/10.1016/j.patcog.2024.111311`.

[24] Özal Yıldırım et al. «Arrhythmia detection using deep convolutional neural network with long duration ECG signals». En: *Computers in Biology and Medicine* 102 (2018), págs. 411-420. DOI: `10.1016/j.compbiomed.2018.09.009`.

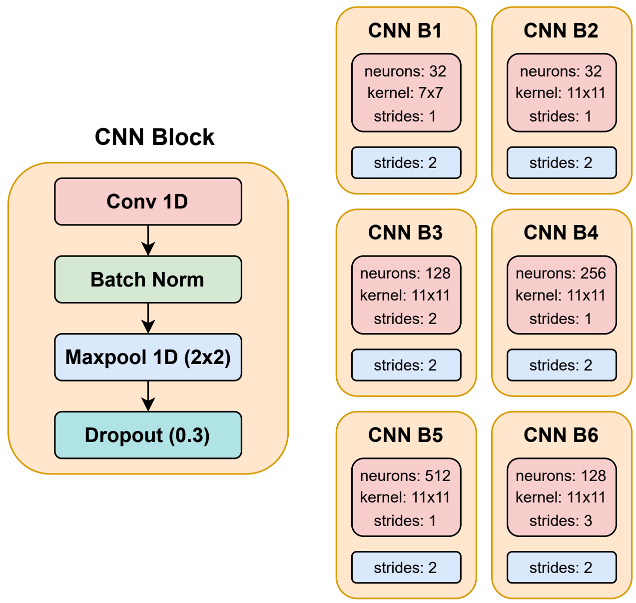
study. It consists of an initial 1D convolutional layer, followed by batch normalization and a max pooling layer to extract features. This is followed by six subsequent CNN blocks, each comprising a 1D convolutional layer, batch normalization, a max pooling layer, and a dropout layer, with ReLU activation and He uniform initialization, as detailed in Figure 2. The output of this branch is then passed through a flatten layer.

Figure 1. Multimodal architecture for arrhythmia classification. The left-hand side shows the feature branch, and the right-hand side illustrates the ECG branch with an input of 2500×N (where N is the number of leads used).



In parallel, the tabular branch was incorporated to process the extracted physiological and demographic features, allowing the integration of complementary information

Figure 2. CNN block in the ECG branch, where the diagrams on the right specify the parameters of each block, with colors indicating their corresponding layer type.



and enabling the development of a multimodal architecture. This branch processes the input through two fully connected layers, each followed by a dropout layer.

The outputs of both branches are concatenated to form a joint representation, which is passed through a dense layer and a dropout layer before the final classification layer. The output layer uses the softmax activation function with Glorot normal initialization to perform multiclass classification across the six categories.

The model was optimized during training with the Adam optimizer and categorical cross-entropy loss function, using a batch size of 256. To prevent overfitting, L2 regularization and Dropout at 30 % and 50 % rates are integrated, reducing model complexity and randomly deactivating neurons during training, respectively, while early stopping halts the process when validation loss stabilizes for 20 epochs, learning rate reduction adjusts the rate by a factor of 0.2 after 10 epochs of no improvement with a minimum of 0.0001, and model checkpointing saves the best model based on validation performance.

3.5. MODEL EVALUATION

The dataset was split into 85 % for training and 15 % for independent testing to ensure rigorous evaluation. Model performance was assessed through a 10-fold stratified cross-validation approach on the training data, where the best model was selected based on the F1-score, valued for its ability to robustly handle potential class imbalances across the six arrhythmia categories. This method enhances the models' generalizability and robustness by testing across diverse data subsets, minimizing the risk of bias from specific partitions. Subsequently, the selected model underwent further evaluation on the independent test set. The evaluation metrics include accuracy, precision, recall, F1-score, and the PhysioNet Challenge score, where this last metric provides partial credit for misclassifications that would lead to similar treatments or outcomes, while more harmful errors are heavily penalized. In this way, the score reflects the clinical relevance of different types of mistakes, offering a more nuanced assessment of model performance.

3.6. MODEL INTERPRETATION

In this study, SHAP was used to identify the most influential ECG leads in arrhythmia classification, providing interpretability and insights into the model's decision-making process. The SHAP DeepExplainer was applied to the model trained using only the ECG signal branch with all 12 leads, using background and test samples randomly selected from its training set. The background samples serve as a baseline to compare the model's predictions, while the test samples help evaluate the feature contributions. The average absolute SHAP values were calculated for each lead across all arrhythmia classes, revealing the leads that most contributed to the model's predictions.

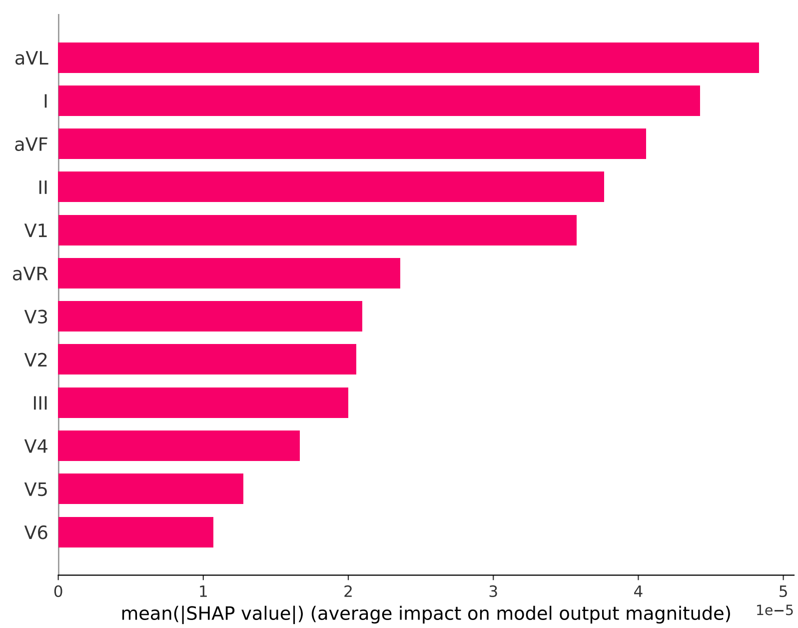
This analysis enabled a reduction in the number of leads, beginning with the full set and progressively focusing on the most relevant ones. The ECG model was retrained and evaluated with 6, 4, and 2 leads, assessing how lead reduction impacted performance. The same ranked configurations were then used to train and evaluate the multimodal model, ensuring a consistent comparison across modalities. By emphasizing the most important leads, this method aims to optimize the classification process, resulting in a more interpretable and efficient solution.

For full reproducibility, the complete code used in this work, including data preprocessing, feature extraction, model training, and evaluation, can be found in the GitHub repository listed in Annex1.

4. RESULTS

Based on the unimodal model, an interpretability analysis was conducted using SHAP values to determine the relative importance of each ECG lead in the classification process. Figure 3 presents the results of this analysis, where the mean absolute SHAP values were calculated across the six considered classes. The leads were ordered from highest to lowest relevance, with aVL, I, aVF, and II consistently emerging as the most influential in the model's decisions. From this information, four input configurations were defined to evaluate the impact of the number of leads on model performance: the first included all 12 leads, the second used the 6 most relevant, the third considered only the 4 most representative, and finally, the fourth was limited to the 2 most influential leads.

Figure 3. Average relative importance of ECG leads in the classification task, calculated as the mean absolute SHAP value averaged across the six considered arrhythmia classes. The X-axis represents the unitless mean contribution of each lead, and the Y-axis displays the leads ordered from most to least important.



Based on the defined configurations, the data were evaluated with the multimodal model and, complementarily, reevaluated with the unimodal model. Table 3 presents the results obtained on the training subset using a 10-fold cross-validation scheme. Four performance metrics are reported—accuracy, precision, recall, and F1-score—evaluated under the 12-, 6-, 4-, and 2-lead configurations for both models. Each metric is presented together with its corresponding standard deviation, allowing for an assessment of the model's stability in each scenario. In this study, the focus is primarily placed on the F1-score, given its particular relevance in the clinical domain and its ability to adequately reflect model behavior in scenarios with class imbalance.

Table 3. Comparison of model performance metrics for unimodal and multimodal approaches using 12-, 6-, 4-, and 2-lead configurations on the training set, evaluated through 10-fold cross-validation. Standard deviation is reported for each metric, and the best F1-score values are highlighted in bold.

UNIMODAL MODEL				
Leads	**Accuracy**	**Precision**	**Recall**	**F1-score**
12-Leads	0.9354 ± 0.0023	0.9110 ± 0.0059	0.8269 ± 0.0048	0.8493 ± 0.0056
6-Leads	0.9362 ± 0.0025	0.9107 ± 0.0048	0.8298 ± 0.0085	**0.8512 ± 0.0073**
4-Leads	0.9364 ± 0.0030	0.9105 ± 0.0064	0.8330 ± 0.0052	**0.8536 ± 0.0059**
2-Leads	0.9306 ± 0.0031	0.9062 ± 0.0064	0.8155 ± 0.0064	0.8340 ± 0.0064

MULTIMODAL MODEL				
Leads	**Accuracy**	**Precision**	**Recall**	**F1-score**
12-Leads	0.9353 ± 0.0029	0.9061 ± 0.0068	0.8313 ± 0.0089	**0.8517 ± 0.0074**
6-Leads	0.9359 ± 0.0021	0.9128 ± 0.0084	0.8261 ± 0.0124	0.8495 ± 0.0079
4-Leads	0.9350 ± 0.0021	0.9076 ± 0.0085	0.8276 ± 0.0091	0.8498 ± 0.0067
2-Leads	0.9302 ± 0.0019	0.9104 ± 0.0052	0.8095 ± 0.0058	0.8312 ± 0.0051

Similarly, Table 4 presents the results obtained on the test subset for the unimodal and multimodal models. The model was selected through a 10-fold cross-validation process applied to the training set; among the ten models generated, the one with the best performance in terms of F1-score was chosen and subsequently evaluated on the test set. The table reports the four previously described performance metrics and, additionally, includes the metric used in the PhysioNet Computing in Cardiology Challenge 2020, evaluated under dataset configurations corresponding to ECG recordings with 12, 6, 4, and 2 leads.

Table 4. Comparison of model performance metrics for unimodal and multimodal approaches using 12-, 6-, 4-, and 2-lead configurations on the test set, evaluated with the best model obtained through 10-fold cross-validation on the training set. The official Challenge metric is also included.

UNIMODAL MODEL					
Leads	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Challenge**
12-Leads	0.9364	0.9153	0.8301	0.8536	0.9064
6-Leads	0.9368	0.9155	0.8337	0.8561	0.9065
4-Leads	0.9366	0.9163	0.8334	0.8569	0.9058
2-Leads	0.9329	0.9168	0.8188	0.8422	0.9014

MULTIMODAL MODEL					
Leads	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Challenge**
12-Leads	0.9371	0.9104	0.8382	0.8581	0.9074
6-Leads	0.9331	0.8997	0.8377	0.8535	0.9015
4-Leads	0.9337	0.9112	0.8253	0.8501	0.9017
2-Leads	0.9315	0.9105	0.8193	0.8413	0.8980

5. DISCUSSION

Table 3 shows that, in the unimodal model, the 4-lead and 6-lead configurations achieved the highest average F1-scores, exceeding 85 % and, in some cases, even surpassing the 12-lead configuration. This finding indicates that it is possible to maintain, or even slightly improve, model performance using a reduced subset of leads. The 12-lead configuration, in turn, exhibited the lowest standard deviation, reflecting greater stability and consistency across the cross-validation folds.

In contrast, the 2-lead configuration showed a minimal reduction of 1.84 % compared to the 12-lead setup, indicating that the model maintains consistent behavior despite having less information available. The 4-lead configuration slightly outperformed the 6- and 12-lead configurations, suggesting that the selected leads (aVL, I, aVF, and II) contain most of the discriminative information. By excluding less relevant or redundant leads, noise is reduced, pattern recognition is facilitated, and the risk of overfitting decreases. However, this technical efficiency poses a clinical challenge: reducing the number of leads could lower sensitivity to arrhythmias with subtle manifestations, increasing the likelihood of missing clinically relevant patterns.

In the multimodal model, the validation set results reflect performance comparable to that of the unimodal model. The 12-lead configuration achieved the highest F1-score, exceeding its unimodal counterpart by only 0.28 %; however, this difference is not statistically significant. The 6-, 4-, and 2-lead configurations showed slightly lower performance than in the unimodal model, accompanied, in some cases, by an increase in standard deviation, particularly in the 6- and 4-lead setups.

The test set analysis of the unimodal model confirms consistency with the validation results (see Table 4). The 12-, 6-, and 4-lead configurations exceeded 85 % in F1-score, with statistically equivalent values among them, while the 2-lead configuration reached over 84 %, approaching the performance of the other setups. This finding highlights the model's ability to generalize using only one-sixth of the information, representing a significant advantage for efficient arrhythmia classification.

Consistently, the test set analysis shows that the multimodal approach does not offer significant improvements over the unimodal model. This indicates that the incorporation of additional information—age, sex, and temporal features—does not increase the model's discriminative capacity. Reduced configurations of 6 or 4 leads achieve performance comparable to that of 12 leads, reinforcing the robustness of the unimodal model and its potential for applications in resource-limited settings.

Regarding the Challenge metric, no significant difference is observed between the unimodal and multimodal models, as both present very similar values. However, it is particularly noteworthy that, in the best case, the metric reached a value of 0.9074, whereas the winning teams of the Challenge reported results below 0.6. This discrepancy is explained by the fact that, of the 27 original classes proposed in the competition, only 6 were used in this study. By reducing the number of classes, the model makes fewer errors and achieves better performance on the selected categories, which justifies the high value obtained. It is worth noting that this metric does not simply measure overall accuracy but more precisely reflects the model's ability to correctly identify the most relevant arrhythmias defined by the Challenge. Moreover, it considers the clinical importance of misclassifications, imposing heavier penalties on those with more severe implications. In this way, it provides a more comprehensive and nuanced evaluation of performance than traditional metrics such as the F1-score.

It is surprising that the multimodal model does not outperform its unimodal counterpart, despite the inclusion of additional variables that could, in principle, enhance performance. This lack of advantage may be due to several reasons: first, the incorporated information is limited, as the 11 features were extracted exclusively from lead II, which restricts their discriminative contribution compared to what is already captured by the full set of 12 ECG leads. Second, although age and sex are clinically relevant risk factors, they do not appear to contribute significantly to arrhythmia prediction in the instantaneous classification of short-duration ECGs. In future work, it would be interesting to explore the inclusion of laboratory variables, which could complement the information provided to the network and potentially improve its discriminative capacity.

Finally, it is worth noting that the leads included in the 4- and 6-lead configurations were selected using a SHAP-based importance analysis, supporting their value as highly informative signals for arrhythmia classification. This analysis was reviewed with Professor Ximena Ramos, a healthcare professional, who confirmed that the leads selected in both configurations are clinically appropriate and anatomically relevant for arrhythmia detection. She also noted that these subsets include lead II, which is commonly used to evaluate cardiac rhythm, thereby reinforcing the physiological plausibility and clinical coherence of the model's predictions. Overall, the results indicate that a well-justified reduction in the number of leads can maintain, or even improve, model performance, provided that highly informative signals are selected. Nevertheless, before considering these reduced configurations as substitutes for a full electrocardiogram, it is essential to validate the approach in real clinical settings, such as hospitals or diagnostic centers, to assess its effectiveness across different patients, acquisition conditions, and diagnostic scenarios. This will determine whether reduced-lead models can maintain reliable and clinically meaningful performance for the arrhythmias addressed in this study.

6. CONCLUSIONS

The results show that reducing the number of ECG leads does not compromise the model's performance. The 4- and 6-lead configurations achieve F1-scores similar to or even higher than the 12-lead model, indicating that a selected subset of leads captures most of the discriminative information. The model trained with only 2 leads, despite a slight decrease in performance, achieves a comparable F1-score and maintains consistent behavior, suggesting the feasibility of simplified solutions in resource-limited settings or portable devices.

The performance of the multimodal model is comparable to that of the unimodal one, and the incorporation of additional variables—such as age, sex, and temporal features—does not significantly improve discriminative capacity, likely due to the limited added information and the short duration of the signals. The high Challenge metric, reaching 0.9074 in the best case, demonstrates that the model is highly effective in detecting the arrhythmias considered in the selected classes, and this metric remains stable even with reduced information.

Finally, the selection of leads through SHAP analysis supports their informative value and shows that SHAP is an appropriate interpretability tool for this type of data, as it enables the identification of key leads and supports decision-making based on the relevance of each signal. Taken together, these findings support the development of more accessible, efficient, and easy-to-implement automated diagnostic systems without significantly compromising accuracy, which is particularly valuable in remote environments or outpatient monitoring scenarios.

BIBLIOGRAPHY

Abousaber, Inam, Hany El-Ghaish y Haitham F. Abdallah. «A novel explainable ECG classification with spatio-temporal transformers and hybrid loss optimization». En: *Biomedical Signal Processing and Control* 110 (2025), pág. 108142. DOI: 10.1016/j.bspc.2025.108142.

Abubaker, Mohammed B. y Bilal Babayiğit. «Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods». En: *IEEE Transactions on Artificial Intelligence* 4.2 (2023), págs. 373-382. DOI: 10.1109/TAI.2022.3159505 (vid. pág. 9).

Alamatsaz, Negin et al. «A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection». En: *Biomedical Signal Processing and Control* 90 (2024), pág. 105884. DOI: 10.1016/j.bspc.2023.105884.

Ayano, Yehualashet Megersa et al. «Interpretable Hybrid Multichannel Deep Learning Model for Heart Disease Classification Using 12-Lead ECG Signal». En: *IEEE Access* 12 (2024), págs. 94055-94080 (vid. pág. 14).

Chukwu, Emmanuel C. y Pedro A. Moreno-Sánchez. «Enhancing Arrhythmia Diagnosis with Data-Driven Methods: A 12-Lead ECG-Based Explainable AI Model». En: *Digital Health and Wireless Solutions*. Ed. por Mariella Särestöniemi et al. Cham: Springer Nature Switzerland, 2024, págs. 242-259 (vid. pág. 15).

Ebrahimi, Zahra et al. «A review on deep learning methods for ECG arrhythmia classification». En: *Expert Systems with Applications: X* 7 (2020), pág. 100033. DOI: https://doi.org/10.1016/j.eswax.2020.100033 (vid. pág. 10).

EPMoghaddam, Dorsa et al. «Reconstructing 12-lead ECG from reduced lead sets using an encoder–decoder convolutional neural network». En: *Biomedical Signal Processing and Control* 104 (2025), pág. 107486. DOI: https://doi.org/10.1016/j.bspc.2024.107486.

Hilgendorf, Lukas et al. «Fully Automated Diagnosis of Acute Myocardial Infarction Using Electrocardiograms and Multimodal Deep Learning». En: *JACC: Advances* 4.8 (2025), pág. 102011. DOI: https://doi.org/10.1016/j.jacadv.2025.102011 (vid. pág. 10).

Jiang, Rucheng et al. «A dual-branch convolutional neural network with domain-informed attention for arrhythmia classification of 12-lead electrocardiograms». En: *Engineering Applications of Artificial Intelligence* 139 (2025), pág. 109480. DOI: https://doi.org/10.1016/j.engappai.2024.109480 (vid. pág. 14).

Kiladze, Mariya R. et al. «Multimodal Neural Network for Recognition of Cardiac Arrhythmias Based on 12-Load Electrocardiogram Signals». En: *IEEE Access* 11 (2023), págs. 133744-133754. DOI: 10.1109/ACCESS.2023.3335176 (vid. pág. 14).

Kligfield, Paul et al. «Recommendations for the standardization and interpretation of the electrocardiogram: Part I: The electrocardiogram and its technology». En: *Journal of the American College of Cardiology* 49.10 (mar. de 2007), págs. 1109-1127. DOI: 10.1016/j.jacc.2007.01.024 (vid. pág. 9).

Kolliyil, Jibin Joy y Melissa C. Brindise. «Automated detection of arrhythmias using a novel interpretable feature set extracted from 12-lead electrocardiogram». En: *Computers in Biology and Medicine* 189 (2025), pág. 109957. DOI: 10.1016/j.compbiomed.2025.109957.

Kurniawan, Arief, Bayu Aditya Triwibowo y Dion Hayu Fandiantoro. «Classification of Arrhythmias 12-Lead ECG Signals Based on 1 Dimensional Convolutional Neural Networks». En: *2024 International Conference on Smart Computing, IoT and Machine Learning (SIML)*. 2024, págs. 220-225. DOI: 10.1109/SIML61815.2024.10578089 (vid. pág. 13).

Lundberg, Scott M y Su-In Lee. «A Unified Approach to Interpreting Model Predictions». En: *Advances in Neural Information Processing Systems*. Ed. por I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017 (vid. pág. 15).

Mageswari, P. Uma et al. «Decoding Heart Health: A Predictive Model for Demographic-Based Cardiovascular Risk». En: *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*. 2024, págs. 1584-1591. DOI: 10.1109/ICTACS62700.2024.10840610 (vid. pág. 9).

Makowski, Dominique et al. «NeuroKit2: A Python toolbox for neurophysiological signal processing». En: *Behavior Research Methods* 53 (2021), págs. 1689-1696. DOI: 10.3758/s13428-020-01516-y (vid. pág. 17).

Natarajan, Annamalai et al. «A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification». En: *2020 Computing in Cardiology*. 2020, págs. 1-4. DOI: 10.22489/CinC.2020.107 (vid. pág. 13).

Perez Alday, Erick A et al. «Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020». En: *Physiological Measurement* 41.12 (2020), pág. 124003. DOI: 10.1088/1361-6579/abc960 (vid. pág. 13).

Ponemash, Oleksandr et al. «Hybrid Feature Extraction for 12-Lead Ecg Classification by Integrating Handcrafted and Deep Learning Techniques». En: *2025 17th*

International Conference on Computer and Automation Engineering (ICCAE). 2025, págs. 53-58. DOI: 10.1109/ICCAE64891.2025.10980605 (vid. pág. 14).

Reyna, Matthew A. et al. «Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021». En: *2021 Computing in Cardiology (CinC)*. Brno, Czech Republic, 2021, págs. 1-4. DOI: 10.23919/CinC53138.2021.9662687 (vid. pág. 16).

Reznichenko, Serhii et al. «Comparing ECG Lead Subsets for Heart Arrhythmia/ECG Pattern Classification: Convolutional Neural Networks and Random Forest». En: *CJC Open* 7 (2025), págs. 176-186. DOI: 10.1016/j.cjco.2024.10.012.

Sadeghi, Alireza et al. «3DECG-Net: ECG fusion network for multi-label cardiac arrhythmia detection». En: *Computers in Biology and Medicine* 182 (2024), pág. 109126. DOI: 10.1016/j.compbiomed.2024.109126.

Sathi, Tanjila Alam et al. «An interpretable electrocardiogram-based model for predicting arrhythmia and ischemia in cardiovascular disease». En: *Results in Engineering* 24 (2024), pág. 103381. DOI: 10.1016/j.rineng.2024.103381 (vid. pág. 11).

Sowmiya, M. et al. «Multi-class classification of arrhythmias with 12-lead ECG signals using one-dimensional convolutional neural network». En: *International Conference on Computer Vision and Internet of Things 2023 (ICCVIoT'23)*. Vol. 2023. 2023, págs. 90-96. DOI: 10.1049/icp.2023.2859 (vid. pág. 13).

Tao, Rui et al. «IM-ECG: An interpretable framework for arrhythmia detection using multi-lead ECG». En: *Expert Systems With Applications* 237 (2024), pág. 121497. DOI: 10.1016/j.eswa.2023.121497.

Wei, Yufeng et al. «Bimodal Masked Autoencoders with internal representation connections for electrocardiogram classification». En: *Pattern Recognition* 161 (2025), pág. 111311. DOI: https://doi.org/10.1016/j.patcog.2024.111311 (vid. pág. 18).

World Health Organization. *Cardiovascular diseases (CVDs)*. Website. Accessed: 2025-06-25. 2021 (vid. pág. 9).

Yıldırım, Özal et al. «Arrhythmia detection using deep convolutional neural network with long duration ECG signals». En: *Computers in Biology and Medicine* 102 (2018), págs. 411-420. DOI: 10.1016/j.compbiomed.2018.09.009 (vid. pág. 18).

Zeng, Wei et al. «Advancing cardiac diagnostics: Exceptional accuracy in abnormal ECG signal classification with cascading deep learning and explainability analysis». En: *Applied Soft Computing Journal* 165 (2024), pág. 112056. DOI: 10.1016/j.asoc.2024.112056.

Zhang, Dongdong et al. «Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram». En: *iScience* 24 (2021), pág. 102373. DOI: 10.1016/j.isci.2021.102373.

Zhao, Bochao et al. «DRL-ECG-HF: Deep reinforcement learning for enhanced automated diagnosis of heart failure with imbalanced ECG data». En: *Biomedical Signal Processing and Control* 107 (2025), pág. 107680. DOI: https://doi.org/10.1016/j.bspc.2025.107680 (vid. pág. 18).

Zhao, Zhibin et al. «Adaptive Lead Weighted ResNet Trained With Different Duration Signals for Classifying 12-lead ECGs». En: *2020 Computing in Cardiology*. 2020, págs. 1-4. DOI: 10.22489/CinC.2020.112 (vid. pág. 14).

Zhou, Shijie et al. «Deep Learning Applied to Electrocardiogram Interpretation». En: *Canadian Journal of Cardiology* 37.1 (2021), págs. 17-18. DOI: https://doi.org/10.1016/j.cjca.2020.03.035 (vid. pág. 10).

ANNEXES

Annex A. GitHub Repository

The complete code developed for this study is available in the public GitHub repository: `https://github.com/21DJMO/12ECG`