

Detecting False Arrhythmias Alarms in the ICU Using a Deep Learning Approach

Karol Milena Ardila Garnica y Germán Darío Moreno Jaimes

Director

Carlos Augusto Fajardo Ariza

Doctor en Ingeniería Electrónica

Codirector

Camilo Andrés Santos Ortiz

Ingeniero Electrónico

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Ingeniería Electrónica

Bucaramanga

2024

Dedicatoria

Cuando llegó, no paraba ni un instante. Una forma de ser algo revoltosa. Compartimos miles de momentos. De llegadas sonrientes a casa. De esos días que parece que nada te los puede arreglar y entonces apareces. No creo que nadie pueda imaginarse el cariño que nos teníamos. Aunque casi nadie nos entendía, nosotros sí. Una lealtad que algunas veces es difícil de encontrar en personas. Quiero dedicar este trabajo a mis mascotas, quienes son mucho más que un animal, son mis compañeros de vida. Y para Junior, Buenas noches, dulce príncipe. Gracias por tanto.

Germán Dario Moreno Jaimes

Desde el primer instante en que te vi, entendí que la vida me había regalado un ser maravilloso. Llenas de luz y amor cada cosa que haces. Me inspiras cada día más por lo valiente, fuerte y poderoso que eres, porque a pesar de cualquier adversidad le sonríes a la vida. Cada vez que miro tus ojitos bellos llenas de amor mi corazón, te amo con todas las fuerzas de mi ser, y reconozco que todo lo que he logrado en mi vida es gracias a ti. Cada paso, cada logro, es dedicado a ti mi amado hermano. Para Juandi con amor.

Karol Milena Ardila Garnica

Agradecimientos

Quiero expresar mi profundo agradecimiento a todas las personas que han sido parte fundamental de mi trayectoria académica y personal más especial hasta ahora. En primer lugar, a mi madre, cuyo apoyo inquebrantable me ha sostenido en los momentos de celebración y angustia a lo largo de esta carrera. A Rale, quién me heredó con amor, su pasión y responsabilidad por la academia y quién me enseñó que no todos los que deambulan, están perdidos. A mis amadas mascotas, que me han enseñado la paciencia y el amor incondicional, recordándome la importancia de disfrutar de las pequeñas alegrías de la vida en los paseos diarios. A mi familia y a todos los amigos que he tenido el privilegio de conocer durante este viaje, les agradezco por su apoyo constante, su amistad sincera y los invaluable momentos compartidos que han enriquecido mi vida. Finalmente, no puedo dejar pasar por alto a mi querida institución, la Universidad Industrial de Santander (UIS), mi alma mater, que ha sido el escenario en el que mi crecimiento no solo como profesional, sino como individuo desde que era un niño. Sus enseñanzas y aulas han sido testigos de todos aquellos momentos que han forjado mi carácter y mi ser hasta el día de hoy. Gracias totales.

Germán Dario Moreno Jaimes

En este momento tan gratificante de mi vida, deseo expresar mi profundo agradecimiento a todas las personas que han sido parte fundamental de esta etapa. En primer lugar, quiero dirigirme a mis padres, quienes han sido mi guía a lo largo de este proceso. A mi madre, cuyo amor incondicional y apoyo han sido mi luz en los días oscuros. A mi padre, fuente de inspiración y sabiduría, cuyo ejemplo me ha enseñado la importancia del esfuerzo y la dedicación. A mi querido hermano Juan Diego, quien ha sido mi mayor fuente de amor y motivación para cumplir mis

sueños. Agradezco de todo corazón a todos mis familiares y amigos que, con su apoyo incondicional y palabras de aliento, me impulsaron cada día a creer en mí misma y a tener la certeza de que podría superar cualquier desafío. Por otra parte, quiero extender mi gratitud a la Universidad Industrial de Santander por brindarme una formación académica impecable. Agradezco también a aquellos profesores que fueron verdaderas fuentes de luz e inspiración en mi trayectoria académica. A todos aquellos que formaron parte de este proyecto de grado, ya sea como colaboradores, asesores o contribuyentes, les estoy sinceramente agradecida por su tiempo, dedicación y aportes que hicieron posible la realización de este trabajo. Por último, quiero tomarme un momento para agradecerme a mí misma por cada noche de desvelo, cada sacrificio y cada esfuerzo realizado para llegar hasta aquí, por no rendirme en los momentos de dificultad, por mantener la determinación y la perseverancia a lo largo de este camino. Este logro es el resultado del amor, apoyo y esfuerzo de cada uno de ustedes, y por eso, desde lo más profundo de mi corazón, les agradezco por formar parte de este importante capítulo de mi vida.

Karol Milena Ardila Garnica

Content table

	Page.
Introduction.....	16
1. Objective	18
1.1. General Objective	18
1.2. Specific's Objectives	18
2. Addressing the Issue of False Alarms: A Rigorous Study of Predictive Approaches	19
2.1. Related work	19
2.2. Dataset.....	20
3. Methodology	22
3.1. Signal Organization	23
3.2. Data Preprocessing.....	25
3.3. False Alarm Identification	27
3.4. Model Evaluation.....	27
3.4.1. Scores.....	27
3.4.2 Confusion Matrix.....	29
3.5. Data Augmentation	29
4. Results.....	31
4.1 Model design.....	33
4.2 Metrics Evaluation	33
4.2.1 Model 1 Evaluation.....	35
4.2.1.1 Combination 3.....	35
4.2.1.2 Combination ALL.....	37

4.2.1.3 Analysis and Discussion	39
4.2.2 Model 2 Evaluation.....	39
4.2.2.1. Combination 3.....	40
4.2.2.2 Combination ALL.....	42
4.2.2.3 Analysis and Discussion	43
4.3 Data Augmentation	43
4.3.1 Model 1 Evaluation.....	44
4.3.1.1 Combination 3.....	44
4.3.1.2 Combination ALL.....	46
4.3.1.3 Analysis and Discussion	47
4.3.2 Model 2 Evaluation.....	47
4.3.2.1 Combination 3.....	47
4.3.2.2 Combination ALL.....	49
4.3.2.3 Analysis and Discussion	50
5. Conclusions.....	50
Bibliography	52
Appendix.....	54

Table List

Table 1 <i>Statistics of the dataset provided for the "PhysioNet/Computing in Cardiology Challenge" in 2015.</i>	20
Table 2 <i>Type of signal and quantity present in the 750 records of the database.</i>	24
Table 3 <i>Signal Configurations for Data Subsets</i>	25
Table 4. Lists of hyperparameters chosen for model 1 and 2 training	33
Table 5 <i>Score "Physionet Computing in Cardiology Challenge" 2015</i>	34
Table 6 <i>Model 1 evaluation metrics</i>	35
Table 7 <i>Model 2 evaluation metrics</i>	39
Table 8 <i>Model 1 evaluation metrics with data augmentation</i>	44
Table 9 <i>Model 2 evaluation metrics with data augmentation</i>	47

Figure List

Figure 1 <i>Visualization of Record "v1151" from the PhysioNet/Computing in Cardiology Challenge 2015 database reveals a steady signal in the fourth position of the record.</i>	21
Figure 2 <i>Block diagram illustrating the process developed for this research.</i>	22
Figure 3 <i>Record "v1151" Normalized with "Imputed" ABP Signal.....</i>	26
Figure 4 <i>Block description diagram for data augmentation algorithm with TSGM Library</i>	30
Figure 5 <i>Jiterring "ECG " created by TSGM Library algorithm.....</i>	31
Figure 6 <i>Model 1 proposed in the paper "A contrastive learning approach for ICU false arrhythmia alarm reduction."</i>	32
Figure 7 <i>Loss and Accuracy curves in training and validation for combination 3, Model 1</i>	36
Figure 8 <i>Confusion matrix and ROC curve for combination 3, Model 1.....</i>	37
Figure 9 <i>Loss and Accuracy curves in training and validation for combination ALL, Model 1</i>	37
Figure 10 <i>Confusion matrix and ROC curve for combination ALL, Model 1.....</i>	38
Figure 11 <i>Loss and Accuracy curves in training and validation for combination 3, Model 2 ...</i>	40
Figure 12 <i>Confusion matrix and ROC curve for combination 3, Model 2.....</i>	41
Figure 13 <i>Loss and Accuracy curves in training and validation for combination ALL, Model 2</i>	42
Figure 14 <i>Confusion matrix and ROC curve for combination ALL, Model 2.....</i>	43
Figure 15 <i>Loss and Accuracy curves in training and validation for combination 3, Model 1 trained with data augmentation.....</i>	44
Figure 16 <i>Confusion matrix and ROC curve for combination 3, Model 1 trained with data augmentation.....</i>	45
Figure 17 <i>Loss and Accuracy curves in training and validation for combination ALL, Model 1 trained with data augmentation.....</i>	46

Figure 18 <i>Confusion matrix and ROC curve for combination ALL, Model 1 with data augmentation.....</i>	46
Figure 19 <i>Loss and Accuracy curves in training and validation for combination 3, Model 2 trained with data augmentation.....</i>	48
Figure 20 <i>Confusion matrix and ROC curve for combination 3, Model 2 with data augmentation</i>	48
Figure 21 <i>Loss and Accuracy curves in training and validation for combination ALL, Model 2 trained with data augmentation</i>	49
Figure 22 <i>Confusion matrix and ROC curve for combination ALL, Model 2 with data augmentation.....</i>	49

Appendix's List

	Page
Appendix A. Github PMHS-DL-231 repository	54

Glossary

Ambulatory Blood Pressure (ABP): This waveform depicts blood pressure measurements in the body's arteries. Unlike the ECG signal, which focuses on the heart's electrical activity, ABP monitoring assesses the force exerted by blood against artery walls throughout the cardiac cycle. ABP monitoring is essential for gaining insights into cardiovascular function and evaluating a patient's overall health status.

Asystole: Absence of a QRS complex for at least 4 seconds, indicating a complete cessation of electrical activity in the heart, often considered a medical emergency requiring immediate intervention.

Cardiac Arrhythmias: Irregularity and inequality in heart contractions.

Deep Learning: A branch of machine learning that relies on artificial neural networks with multiple layers to learn and extract features from data.

End-to-End Learning Model (E2E): Refers to the training of a possibly complex learning system represented by a single model (specifically a deep neural network) that represents the entire target system, bypassing the intermediate layers that are usually present in traditional pipeline designs.

Extreme Bradycardia: Heart rate below 40 beats per minute (bpm) for 5 consecutive beats or fewer than 5 beats in 6 seconds, reflecting an abnormally slow heart rhythm that can lead to inadequate blood flow and tissue perfusion.

Extreme Tachycardia: Heart rate exceeding 140 bpm for 17 beats or more in 6.85 seconds, indicating an excessively rapid heart rhythm that may compromise cardiac function and lead to hemodynamic instability.

Photoplethysmogram (PPG): PPG is a non-invasive technique that detects variations in blood vessel volume using optical sensors. By illuminating the skin with light and measuring the intensity of reflected or transmitted light, PPG can capture blood flow changes in the body. The PPG waveform reflects these fluctuations and is synchronized with the cardiac cycle.

QRS Complex: In an ECG waveform signifies the depolarization of the ventricles, which is fundamental for initiating ventricular contraction and subsequent ejection of blood from the heart. It comprises three distinct components: the Q wave, R wave, and S wave.

Q Wave: The first negative wave of the QRS complex, with a duration of less than 0.04 seconds and a voltage not exceeding 0.2 mV, which should be less than 25% of the R wave.

Respiratory Signals (RESP): Respiratory activity has a significant impact on heart rate variability, a phenomenon known as Respiratory Sinus Arrhythmia (RSA). This relationship is evident in the ECG signal, where the influence of breathing on heart rhythm is reflected. Monitoring RESP signals provides valuable insights into the dynamic interplay between respiration and cardiac function.

Software Development: Refers to a set of computer activities dedicated to the process of creating, designing, deploying, and supporting software.

T Wave: Represents ventricular repolarization, usually with an amplitude lower than the preceding QRS complex. Its maximum amplitude is less than 5 mm in peripheral leads and 15 mm in precordial leads.

Ventricular Fibrillation: Fibrillation or chaotic electrical activity in the ventricles lasting for at least 4 seconds, resulting in the loss of coordinated ventricular contractions and severely compromised cardiac output, often leading to sudden cardiac arrest if not promptly treated with defibrillation.

Ventricular Tachycardia: Presence of 5 or more consecutive ventricular beats with a heart rate exceeding 100 bpm within a 2.4-second period, suggestive of abnormal electrical activity originating from the ventricles, which can lead to life-threatening arrhythmias.

Resumen

Título: Identificación de Falsas Alarmas de Arritmias Cardiacas en Unidades de Cuidados Intensivos *

Autor: Karol Milena Ardila Garnica, Germán Darío Moreno Jaimes **

Director: Carlos Augusto Fajardo Ariza. Dr. Ingeniería Electrónica.

Co-director: Camilo Andrés Santos Ortiz. Ingeniero Electrónico.

Palabras Clave: Unidades de Cuidados Intensivos, Arritmias cardiacas, Falsas alarmas, Aprendizaje Profundo.

Descripción: En el entorno de la Unidad de Cuidados Intensivos (UCI), las falsas alarmas de arritmia se convierten en una perturbación constante que genera estrés tanto para el personal médico como para los pacientes. La elevada tasa de estos informes, unida a la incapacidad de los sistemas de monitorización para discernir con precisión entre señales cardiacas normales y arritmias, compromete la calidad de la atención y pone en riesgo la salud de los pacientes. El objetivo de esta investigación fue identificar falsas alarmas utilizando redes neuronales convolucionales (CNN) entrenadas con datos de la base de datos PhysioNet Computing in Cardiology Challenge 2015. Además, exploramos la eficacia de las técnicas de aumento de datos en el entrenamiento del modelo; las métricas de evaluación del rendimiento utilizadas incluyeron la tasa de verdaderos positivos (TPR), la tasa de verdaderos negativos (TNR), la puntuación F1, el área bajo la curva (ROC) y la puntuación final del PhysioNet Computing in Cardiology Challenge 2015. Nuestros resultados sugieren que el modelo propuesto genera una reducción de la tasa de falsos negativos (FN) al 5,9 % y revelan que las técnicas de aumento de datos no tuvieron un impacto significativo en la identificación de falsas alarmas.

* Trabajo de Grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Ingeniería Electrónica. Director: Carlos Augusto Fajardo Ariza. Doctor en Ingeniería Electrónica.

Abstract

Title: Detecting False Arrhythmias Alarms in the ICU Using a Deep Learning Approach *

Author(s): Karol Milena Ardila Garnica, Germán Dario Moreno Jaimes **

Director: Carlos Augusto Fajardo Ariza. Dr. Electronic Engineering.

Co-director: Camilo Andrés Santos Ortiz. Electronic Engineering.

Key Words: Intensive Care Units, Cardiac Arrhythmia, False Alarm, Deep Learning.

Description: In the Intensive Care Unit (ICU) setting, false arrhythmia alarms become a constant disturbance that generates stress for both medical staff and patients. The high rate of these reports, coupled with the inability of monitoring systems to accurately discern between normal cardiac signals and arrhythmias, compromises the quality of care and puts patients' health at risk. The aim of this research was to identify false alarms using convolutional neural networks (CNNs) trained with data from the PhysioNet Computing in Cardiology Challenge 2015 database. In addition, we explored the effectiveness of data augmentation techniques in training of the model; the performance evaluation metrics used included true positive rate (TPR), true negative rate (TNR), F1 score, area under the curve (ROC), and final score of the PhysioNet Computing in Cardiology Challenge 2015. Our results suggest that the proposed model generates reduction of the false negative rate (FN) to 5.9 % and reveal that data augmentation techniques did not have a significant impact on false alarm identification.

* Bachelor Thesis

**Faculty of Physical and Mechanical Engineering, School of Electrical, Electronic and Telecommunications Engineering. Electronic Engineering. Director: Carlos Augusto Fajardo Ariza. PhD. In Electronic Engineering.

Introduction

Intensive Care Units (ICUs) are specifically designed to deliver critical care to patients facing severe and life-threatening injuries or illnesses, one of their advantages is the employ of advanced bedside monitors equipped with alarm systems intended to alert healthcare providers when a patient's physiological signals deviate from predefined ranges such as the electrocardiogram (ECG) that serves as a vital tool in assessing the heart's electrical activity over time, offering a detailed depiction of various stages within the cardiac cycle, however, diagnosing cardiac arrhythmias requires more than just relying on Electrocardiogram (ECG) signals alone. Healthcare practitioners utilize a range of physiological signals too such as Ambulatory Blood Pressure (ABP) monitoring and Photoplethysmogram (PPG) to complement ECG analysis, providing valuable insights into cardiovascular dynamics. The first one tracks blood pressure fluctuations in arteries, offering essential information on cardiovascular function and overall health and the second one is a non-invasive technique that detects changes in blood vessel volume by analyzing light absorption or reflection on the skin's surface. Synchronized with the cardiac cycle, PPG captures blood flow variations. Additionally, to those measures, respiratory signals (RESP) uncover the intricate relationship between breathing patterns and heart rate variability, elucidating the influence of respiration on cardiac function and aiding in the diagnosis of respiratory sinus arrhythmia (RSA).

In this context, the high rate of false alarms, which can reach up to 88.8% in the ICU (Drew et al., 2014) is a considerable concern because it leads to a disruption of care, impacting both the patient and the clinical staff through noise disturbances (alarms produce sound intensities above 80 dB).

Among the many causes of this high false alarm rate in ICUs are various factors such as electromagnetic interference, voluntary and involuntary patient movements, and malfunctioning sensors, among others. In response to this issue, this project proposes to identify false alarms using deep learning algorithms, leveraging their ability to automatically learn physiological waveform characteristics and discriminate between true and false alarms. This detection will be carried out using the physiological signals mentioned, which will serve as the basis for training a CNN model. Unlike traditional methods based on signal processing, these models excel in capturing the complex patterns found in this data (Chambrin, 2001). The significance of this research lies in the need to enhance cardiac monitoring efficiency in critical environments, reducing medical staff fatigue, and improving patient care quality, at the end, this project aims to contribute not only to solving a technical problem but also to resource optimization and enhancing the well-being of those reliant on precise and timely medical care.

1. Objective

1.1. General Objective

Design and implement a deep learning algorithm that identifies false alarms of cardiac arrhythmias using the dataset from the "PhysioNet Computing in Cardiology Challenge 2015".

1.2. Specific's Objectives

- Preprocess the "PhysioNet Computing in Cardiology Challenge 2015" database. This preprocessing includes handling missing values, normalization, and data augmentation.
- Design a deep learning model capable of detecting false alarms with a level of precision comparable to that reported in the state-of-the-art literature.
- Evaluate the performance of the previously designed and trained model in terms of true positive rate (TPR), true negative rate (TNR), F1 score, and AUC.

2. Addressing the Issue of False Alarms: A Rigorous Study of Predictive Approaches

2.1. Related work

False alarms have consistently ranked as the primary concern in medical technology since 2012 (Keller, 2012) to confront this challenge, numerous studies have explored strategies based on machine learning, signal processing techniques, and deep learning methods to mitigate false alarms. Among the researches done, solutions employing rule-based algorithms have demonstrated their high effectiveness, like the approach developed by (Plesinger et al., 2016) which based on the study of the complex QRS and derived R-R information is able to identify and classify abnormal heart activity using associated medical rules. Similarly, (Daluwatte et al., 2015) developed an algorithm that leverages global detection of individual heartbeats from multiple physiological signals to apply criteria for arrhythmia detection and alarm classification. Another remarkable approach involves multimodal peak detection algorithms, as utilized by (Ansari et al., 2015), which combine the outcomes of various detection algorithms to create a more robust detection system.

Deep learning models represent another promising avenue, as demonstrated by (Mousavi et al., 2020), their proposal is a solution leveraging a network based on convolutional neural layers (CNN), augmented with an attention mechanism to enhance model performance by focusing on relevant input data. Complementarily, they integrate Long short-term memory (LSTM) neurons to capture short-term memory dependencies. Another notable contribution is from (Q Yu et al., 2020), who present an approach utilizing a multi-channel convolutional network to reduce false alarms. Their model incorporates raw signals from multiple input channels, enabling flexibility and specialization in detection, thereby improving classification effectiveness. Additionally,

(Zihlmann et al., 2017)) introduce two deep learning models for classifying variable-length Electrocardiogram signal records, incorporating both deep CNN architectures and short and long-term memory layers. They also employ data augmentation techniques, such as random resampling, to enrich dataset variability. Lastly, (Wu et al., 2019) propose an End-to-end (E2E) deep learning model for generating local representations of time series, preserving temporal and spatial relationships in the learned feature space through a local embedding loss optimization approach.

2.2. Dataset

For this project, the data used was the publicly available training dataset from the "PhysioNet/Computing in Cardiology Challenge" (Clifford et al., 2015). This dataset comprises 750 training records and an additional 500 records with similar characteristics, set aside for evaluation and internal scoring purposes within the competition. These 1250 records represent a robust collection of physiological signals derived from patients in Intensive Care Units (ICUs). The signals are sampled at a resolution of 12 bits and a frequency of 250 Hz, with FIR filtering applied to remove noise and each signal was truncated to precisely five minutes in duration. The data publication format consists of two files per record with MAT and HEA extensions. A statistical overview of the database content is provided in Table 1 a look of the register form of the signal can be found in Figure 1.

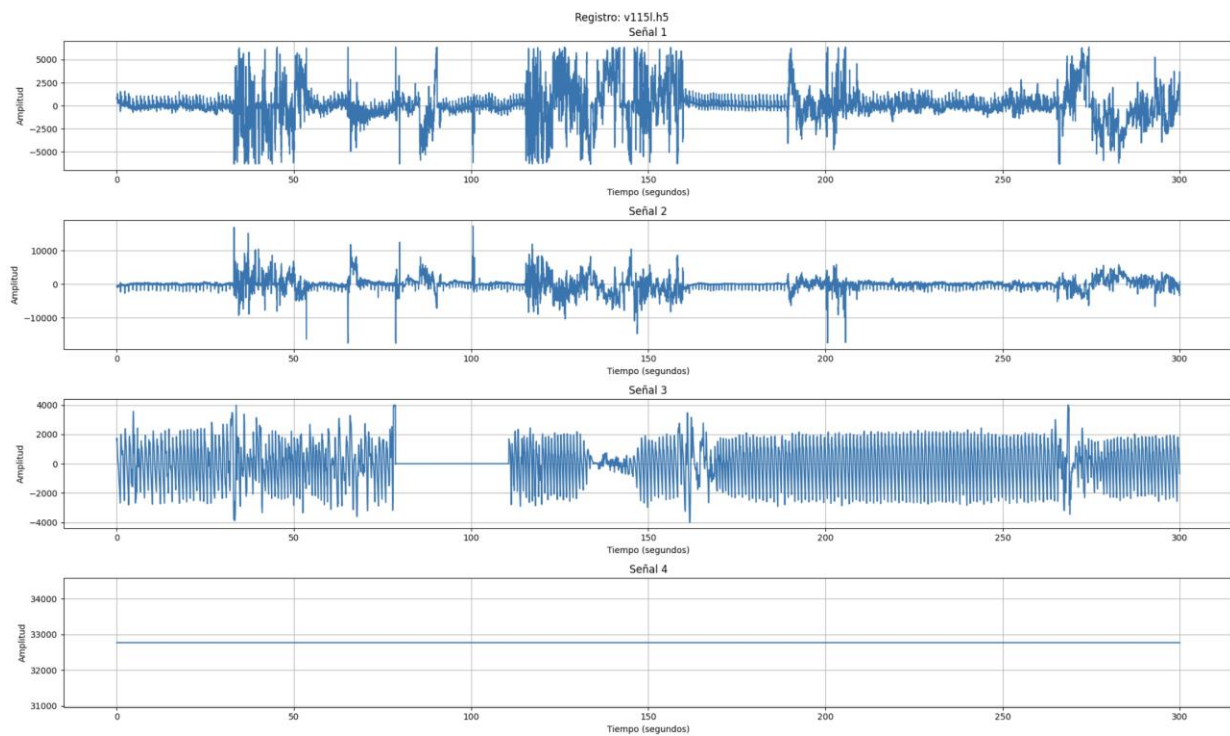
Table 1

Statistics of the dataset provided for the "PhysioNet/Computing in Cardiology Challenge" in 2015.

Tipo de Arritmia	Definición	Cantidad/Porcentaje	Alarmas Falsa	Alarmas Verdadera
VTA	> 5 pulsaciones ventriculares, FC > 100 ppm	341/47%	252	89
ETC	> 17 pulsaciones, FC > 140 ppm	140/17%	9	131
ASY	0 pulsaciones en 4s	122/17%	100	22
EBR	> 5 pulsaciones, FC < 40 ppm	89/11%	46	43
VFB	Ondas de Fibrilación en 4s	58/7%	52	6

Figure 1

Visualization of Record "v115l" from the PhysioNet/Computing in Cardiology Challenge 2015 database reveals a steady signal in the fourth position of the record.



Also, there is a HEA file that provides relevant information regarding the cardiac signal, specifying parameters such as sampling frequency, signal length, comments, and signal names. The dataset includes an additional file named "ALARMS" in TXT format, which contains the

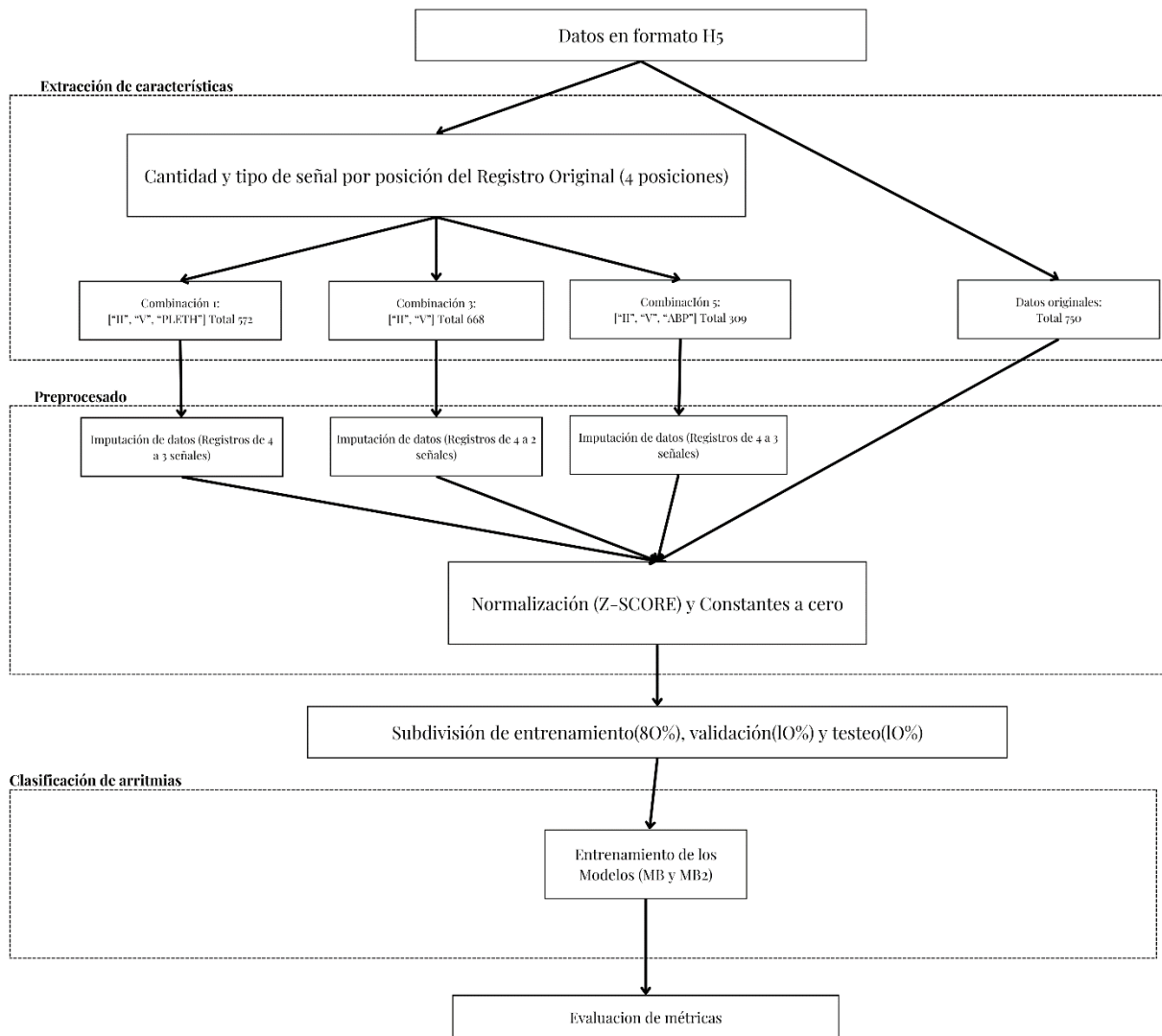
signal name, associated arrhythmia type, and a label indicating whether the alarm is false (0) or true (1). All this information serves as labels and is associated with five arrhythmia events: asystole (ASY), extreme bradycardia (EBR), extreme tachycardia (ETC), ventricular fibrillation or flutter (VFB), and ventricular tachycardia (VT).

3. Methodology

The proposed methodology to address this problem of false alarm identification consists of five main stages which are, signal organization step, data preprocessing, false alarm identification, metrics evaluation and data augmentation. In the first two stages, the physiological signals collected from database are organized and prepared for further analysis with processes to ensure uniform formatting and consistency. The third stage is proposed to demonstrate the remarkable abilities in feature extraction from their training data by CNN as mentioned by (Mousavi et al., 2020); in the metrics evaluation step its performance is evaluated using various metrics such as accuracy, precision, recall, F1 score, and area under the ROC curve (AUC). These metrics provide insight into the model's effectiveness in correctly identifying false alarms. Finally, as part of the final stage, a data augmentation technique is employed to increase the diversity and size of the training dataset and evaluate the effectiveness of this technique in time series signals. The details of the process mentioned are shown in Figure 2.

Figure 2

Block diagram illustrating the process developed for this research.



3.1. Signal Organization

In this dataset, there was a significant diversity in waveform shapes and record compositions, alongside instances of missing signals. Each medical record could have between three or four signals, consistently comprising two ECG leads and at least one pulsatile waveform, such as PPG or ABP. In certain instances, both pulsatile waveforms were present, and one of them was accompanied by a respiratory signal (RESP) (Clifford et al., 2015); this diverse data landscape poses an intriguing challenge for artificial intelligence models' classification and encouraged a

debate when formulating adaptive and specific strategies to harness the waveform variability for this project, within this context, a signal organization by specific characteristics represents a meticulous approach to discerning patterns and key relationships within the dataset. Through this process, the aim is to identify subsets sharing characteristics that contribute to enhancing the efficiency of the proposed machine learning model and provide a more insightful perspective on the interactions within the dataset. To achieve this, an exploratory data analysis was conducted to ascertain the most prevalent signal types at each record position. Signals lacking a fourth signal underwent an imputation process, where a signal with a value of 0 (constant) was added to ensure that each record contained four signals, irrespective of its original configuration. By position, the following counts were obtained in the analysis.

Table 2

Type of signal and quantity present in the 750 records of the database

1st Position (Signal: Count/Perc)	2nd Position (Signal: Count/Perc)	3rd Position (Signal: Count/Perc)	4th Position (Signal: Count/Perc)
ECG - II: 724 / 96.5%	ECG – V: 684 / 91.2%	PPG: 627 / 83.6%	RESP: 278 / 37.1%
ECG – I: 13 / 1.7%	ECG – III: 27 / 3.6%	ABP: 123 / 16.4%	ABP: 220 / 29.3%
ECG – III: 12 / 1.6%	ECG – aVF: 2 / 0.3%		Data in zero: 23.3%
ECG – aVF: 1 / 0.1%	ECG – MCL: 28 / 3.7%		
	ECG – II: 4 / 0.5%		
	ECG – aVL: 2 / 0.3%		
	ECG – aVR: 3 / 0.4%		
Total: 750 Signals per-position			

Table 3 showcases signal configurations for various subsets of data utilized in the analysis. Each combination is denoted by a unique number, outlining the composition of corresponding signals alongside the total instances in each subset. A dedicated row is allocated for the original dataset

("ALL"), representing the comprehensive data provided by Physionet. Signal representations include "II" for the second derivation, "V" for the fifth derivation, "PLETH" for plethysmograph, "RESP" for respiration, and "ABP" for arterial blood pressure. The objective is to discern signal combinations that could offer more pertinent information to the machine learning model and avoid the missing data that the original dataset had.

Table 3

Signal Configurations for Data Subsets

Combination	Signal composition	TOTAL
1	["II", "V", "PLETH"]	572
2	["II", "V", "PLETH", "RESP"]	167
3	["II", "V"]	668
4	["II", "V", "PLETH", "ABP"]	213
5	["II", "V", "ABP"]	309
6	["II", "V", "ABP", "RESP"]	76
7	["PLETH", "ABP"]	220
8	["PLETH", "RESP"]	178
ALL	Original Data	750

3.2. Data Preprocessing

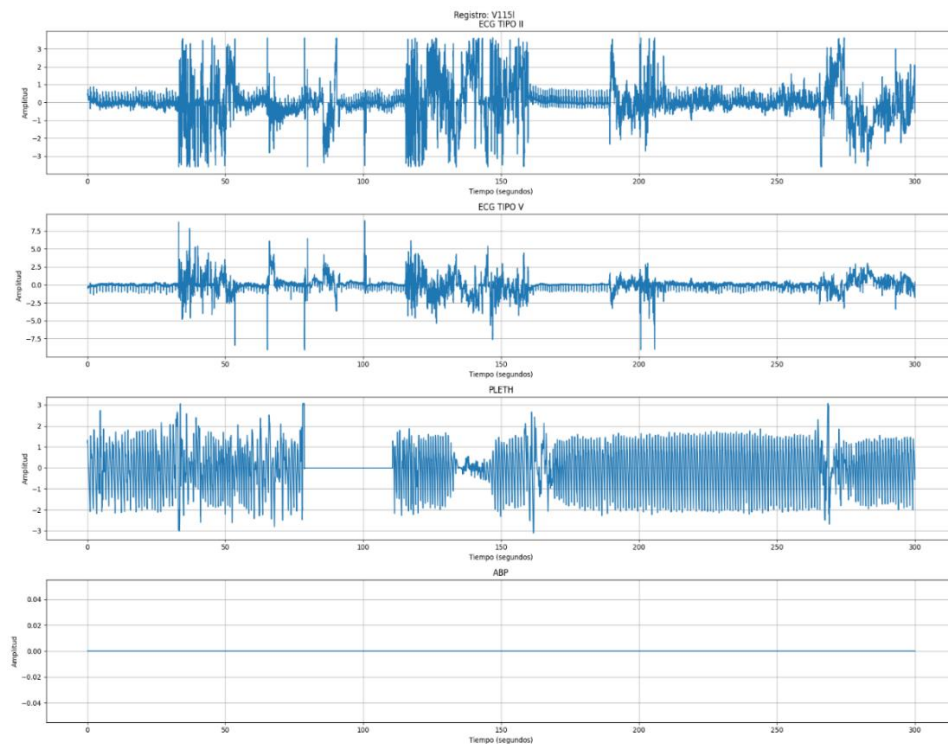
Given the variability observed in the original database, the attention was directed towards optimizing the preprocessing of signals after the subdivision proposed in Table 3 for model training. The objective was twofold: to streamline training time while preserving the temporal and structural integrity of signal waveforms. To achieve this, a data standardization technique was employed—specifically, the use of the standard score, often referred to as the 'z-score.'

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Where x is the raw score, μ the mean and σ the standard deviation as it shown in equation (1). Implementing the 'z-score' transformation facilitates the conversion of signal data into a standard normal distribution. This normalization process is pivotal for ensuring uniformity and comparability across signals, thus laying the groundwork for effective preprocessing and subsequent model training. The Figure 3 shows a register after the preprocessed step. In the context of this study, where signals exhibit diverse characteristics, the 'z-score' offers a valuable means to ascertain the likelihood of a signal's adherence to a normal distribution. Additionally, it furnishes a standardized framework for comparing scores originating from disparate normal distributions, a common occurrence in the signals comprising our dataset (Al-Faiz et al., 2019).

Figure 3

Record "v1151" Normalized with "Imputed" ABP Signal



3.3. False Alarm Identification

To get an optimal model training, an evaluation of each hyperparameter was undertaken to discern those who has the greatest impact on performance and generalization capability of the model. The outcomes of this "fine tuning" revealed that, for optimal training, the learning rate was fixed at 0.001 to regulate model weight adjustments, while the batch size was set at 128 to manage the processing of training examples. The Adam optimizer was chosen for its efficacy in optimizing neural networks, and the binary cross-entropy loss function was employed for binary classification problems. Additionally, L2 regularization and an 85% dropout rate were applied to mitigate overfitting, complemented by the utilization of callbacks such as early stopping, learning rate reduction, and model checkpointing to enhance training efficiency and performance. Finally, the He uniform weight initializer was implemented to ensure the proper initialization of neural network weights.

Moreover, all these techniques were implemented using the well-established cross-validation method to ensure the robustness and reliability of the results obtained during model training. All the tests performed to come with these results can be found in Appendix A.

3.4. Model Evaluation

3.4.1. Scores

For testing, a selection of metrics specialized in assessing the model's performance in classifying cardiac arrhythmias was chosen.

- **PhysioNet Challenge Score.** This metric assesses the model's performance based on four key variables: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). The final score is computed using the formula (2).

$$Score = \frac{TP + TN}{TP + TN + FP + 5 * FN} \quad (2)$$

This formula weights FN more than FP, aiming to minimize errors in cardiac arrhythmia diagnosis (Clifford et al., 2015)

- Accuracy. It measures the proportion of correct predictions made by the model. It is calculated by dividing the number of correct predictions (TP + TN) by the total number of predictions made.
- F1 Score. The F1 Score is a measure that combines the precision and recall of the model into a single metric. It is especially useful when classes are imbalanced, as it considers both false positives and false negatives. It's calculated using the formula (3)

$$F1\ Score = \frac{2}{\left(\frac{1}{Precision} + \frac{1}{Recall}\right)} \quad (3)$$

- Precision. Represents the proportion of true positive instances correctly classified as positive relative to all instances classified as positive by the model. It is calculated using the following formula (4)

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

- Recall. Denotes the proportion of true positive instances correctly classified as positive relative to all positive instances present in the actual data. It's calculated using the formula (5)

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

- **ROC Score.** This metric evaluates the model's discriminative ability based on the Receiver Operating Characteristic (ROC) curve. It corresponds to the area under this curve, depicting the true positive rate versus the false positive rate at different classification thresholds.
- **Specificity.** In the context of binary classification, specificity measures the ability of a model to accurately identify true negative instances among all actual negative instances. It quantifies the proportion of negative instances that are correctly classified as negative by the model.

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

3.4.2 Confusion Matrix

Another crucial metric utilized is the confusion matrix, offering both a visual representation of the model's accuracy by comparing its predictions with the true labels and a direct relationship with the PhysioNet score. As noted, it employs TP, FP, TN, and FN values to calculate a score reflecting the quality of the model's predictions.

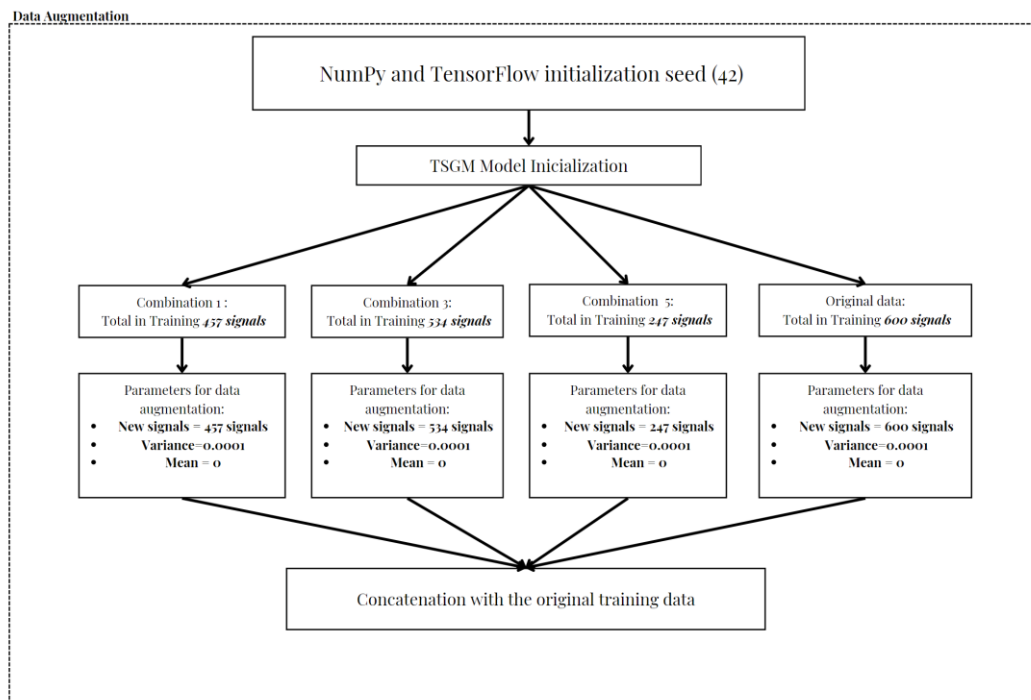
3.5. Data Augmentation

Expanding the dataset is crucial, especially when employing Convolutional Neural Networks (CNNs) for identifying time series signals. By increasing the data volume, CNN has a broader range of examples to learn from, improving its ability to detect patterns and generalize to

unseen data. Generating new samples from existing ones by applying transformations like scaling, rotation, or adding noise helps alleviate overfitting. In this context, the process followed to create signals based in the database is shown in Figure 4, and it was created by the TSGM (Time Series Generative Modeling) library (Nikitin et al., 2023) adding gaussian noise to original time series signals, thereby introducing subtle variations while trying to preserve essential data characteristics. The process was designed to duplicate training data by adding an equal amount of original training data with a variance of 0.0001 and a mean of 0 to the signal.

Figure 4

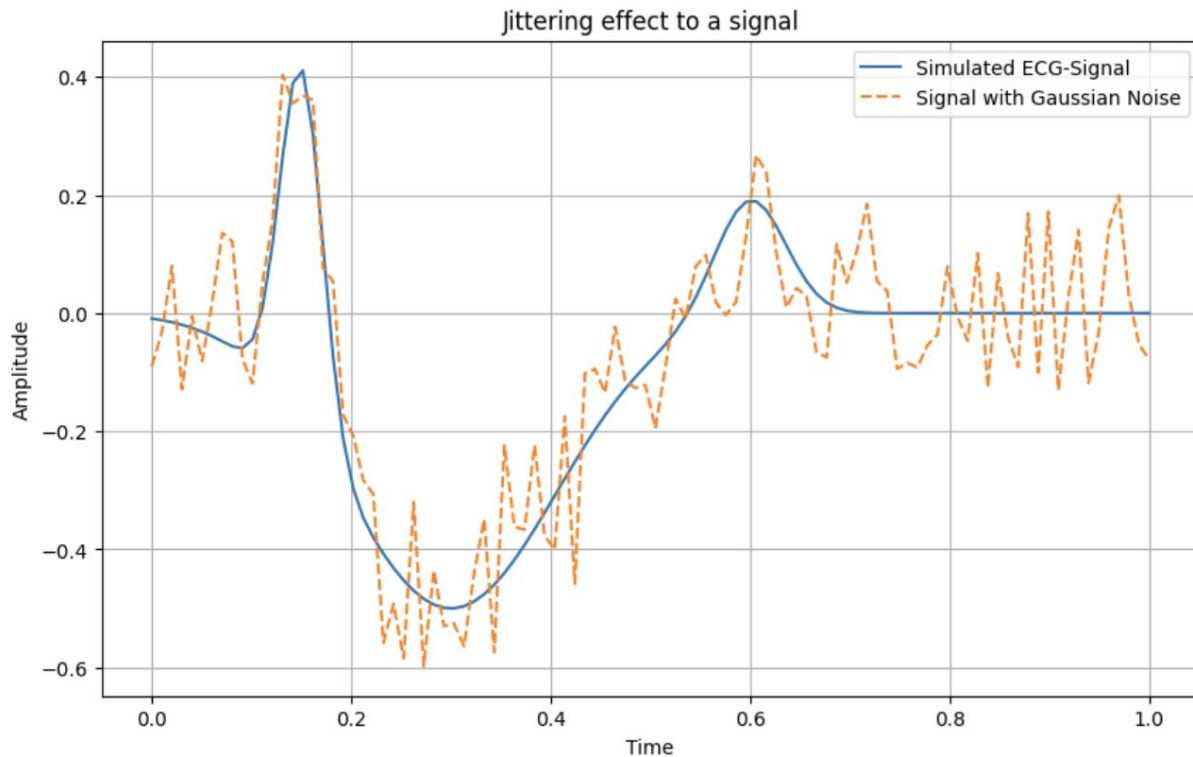
Block description diagram for data augmentation algorithm with TSGM Library



A figure shown the augmentation signal created can be shown:

Figure 5

Jittering "ECG " created by TSGM Library algorithm



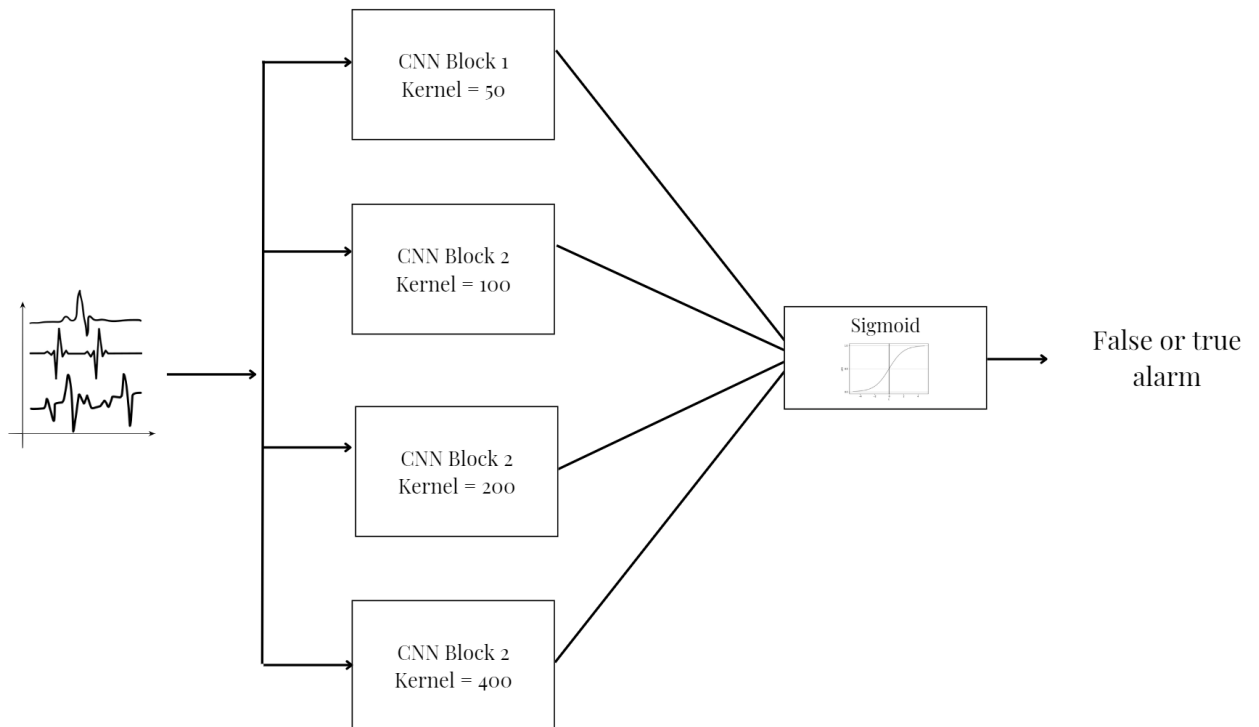
4. Results

The experiment employed two distinct models, both designed based on the "signal encoder" proposed by (Zhou et al., 2022); Model 1 and Model 2 share the same architecture that consists of four parallel convolutional blocks, each using a different kernel size. Additionally, a sigmoid activation function is integrated into the network's output layer, its purpose is to process the neural network's output, converting it into a probability ranging between 0 and 1, where it represents the network's confidence in its determination, particularly in distinguishing false or true alarms. The imbalanced volume of the data provided by the PhysioNet database presents an additional challenge for model identification, for this research, the rate of false negatives is

extremely important and it's crucial to clarify that our aim is to maximize both true positive and true negative rates while minimizing false positive and especially, false negative rates.

Figure 6

Model 1 proposed in the paper "A contrastive learning approach for ICU false arrhythmia alarm reduction."



To explore and further enhance the model's performance, it was decided to subject the architecture to a filter optimization process, the goal is to develop a CNN architecture that maximizes accuracy while minimizing the number of parameters, that's why the Convolutional Neural Network in 1 dimension (CNN 1D) based on its documented efficiency for time series data as it mentions (Castillo et al., 2020) represents a pivotal role in this fine tuning of kernels. It was identified that

filter sizes of 11, 75, 150, and 40 proved to be highly effective in capturing the relevant characteristics of the signals. This new approach, incorporating the optimized filter sizes, is referred to as 'model 2'.

4.1 Model design

Various combinations of hyperparameters were explored to find the optimal model for our false alarm detection task. The determination of these hyperparameters involved a total of 6 trials, assessing 4 different signal combinations detailed in Appendix A. Following these trials, the best hyperparameters for the model were identified, as presented in Table 4.

Table 4.

Lists of hyperparameters chosen for model 1 and 2 training

Hyperparameters	Description
Learning rate	0.001
Batch Size	128
Optimizer	Adam
Loss function	Binary Cross entropy

However, in this document, we primarily focus on two combinations: Combination 3, comprising two variations of ECG signals, and Combination ALL, consisting of ECG, plethysmographic, blood pressure, and respiratory signals. It is important to note that in Combination ALL, the position and variation of the signals may vary per record, as explained in detail in Section 2.

4.2 Metrics Evaluation

The models underwent evaluation using the 5-Fold cross-validation technique, a method chosen to ensure balanced class distribution which is crucial as it divides the data into five groups

Fifth	0.6901	0.60919	0.5242	0.466019
MEAN	0.6524	0.53412	0.41093	0.434366
STANDART DEVIATION	0.08891	0.06236	0.0647	0.046282

4.2.1 Model 1 Evaluation

Results in Table 6 display the mean alongside the standard deviation and it summarizes the findings, offering an overview of the model's performance across various metrics

Table 6

Model 1 evaluation metrics

Metrics	COMBINATION 3	COMBINATION ALL
Accuracy Test	0.77910 +- 0.0356	0.64800 +- 0.04648
Precisión	0.81409 +- 0.1101	0.55412 +- 0.07846
Recall	0.53333 +- 0.16329	0.47586 +- 0.07363
F1 score	0.61931 +- 0.10338	0.51017 +- 0.06951
ROC SCORE	0.79748 +- 0.04659	0.65592 +- 0.05526
Physionet	0.65240 +- 0.08891	0.41093 +- 0.06476

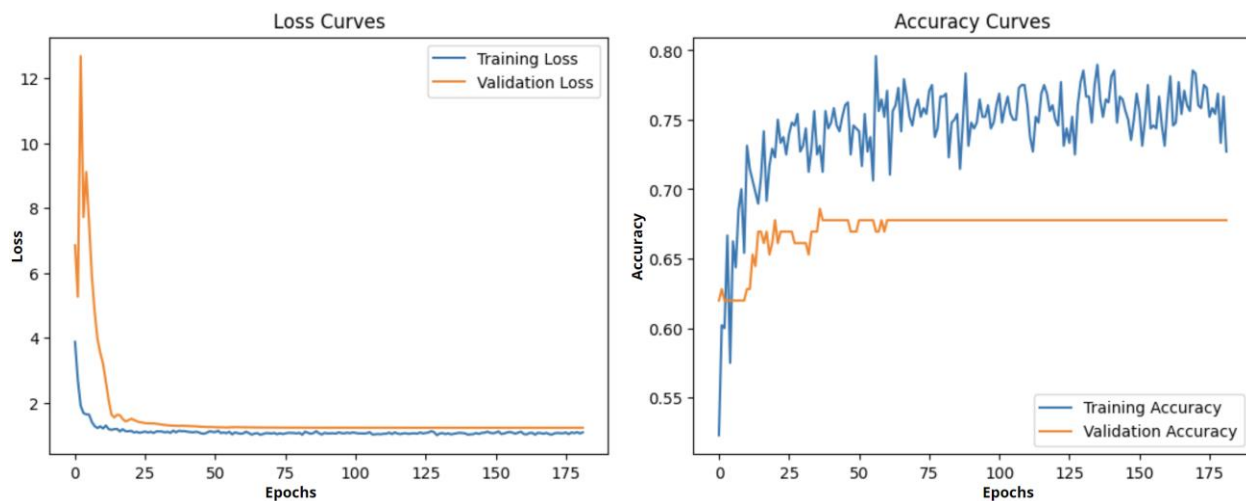
4.2.1.1 Combination 3

The combination 3 achieved a precision level of 77.9%; however, in imbalanced datasets like these, this metric can be deceptive as it tends to favor the prediction of the majority class. Therefore, for a more comprehensive evaluation of the model's performance, it is essential to

analyze the loss and validation curves, these graphs offer a more detailed insight into the model's training and validation process, enabling a better understanding of its ability to fit the data and generalize to new examples. The corresponding graphs for a thorough evaluation are presented below.

Figure 7

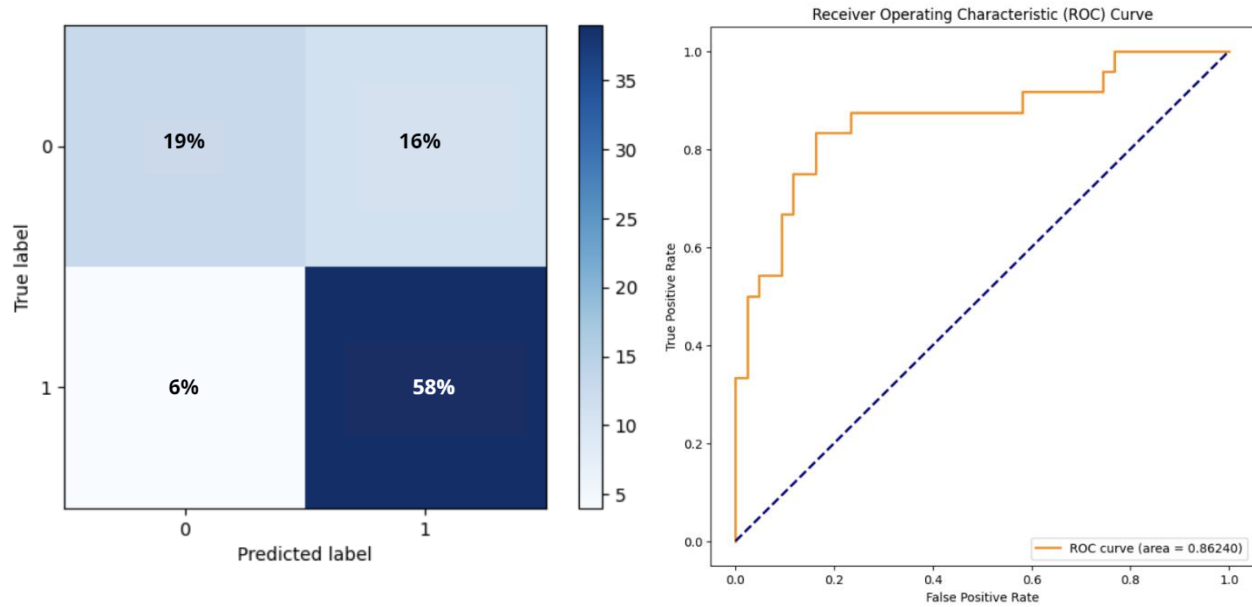
Loss and Accuracy curves in training and validation for combination 3, Model 1



The remaining metrics were extracted from the confusion matrix depicted in the Figure 8 which provides a visual representation of the model's accuracy in classifying each class. Particularly notable is a false negative rate of 10.44% in the test data.

Figure 8

Confusion matrix and ROC curve for combination 3, Model 1

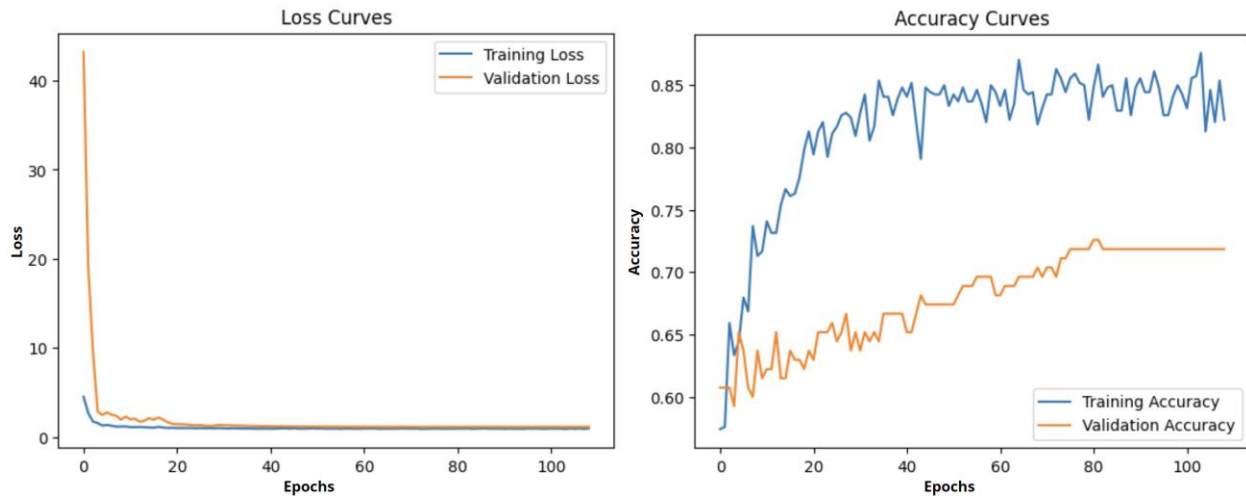


4.2.1.2 Combination ALL

The full dataset achieved a moderate accuracy in classifying the test samples, with a value of 0.648 ± 0.046 , this indicates that the model demonstrates acceptable capability in accurately predicting the labels of the test samples, the learning curves of the model during training are shown below

Figure 9

Loss and Accuracy curves in training and validation for combination ALL, Model 1

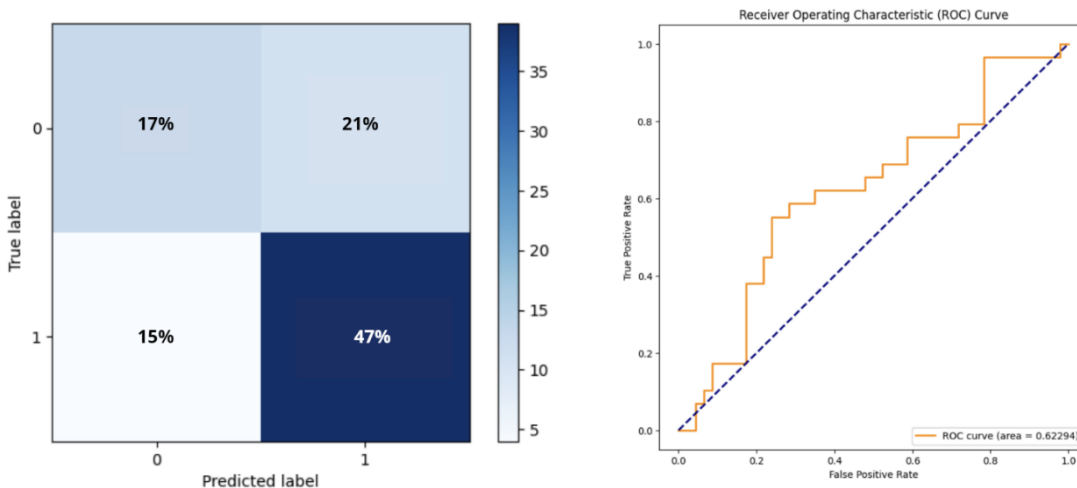


While Figure 9 depicts an acceptable loss during training, upon comparing the outcomes of all metrics with those of Combination 3, it becomes evident that they are lower than combination 3 To corroborate this observation, the confusion matrix depicted in

Figure 10 shows a notable increase in the percentage of false negatives, which rises to 14.6%. This suggests that the combined dataset encounters challenges in correctly identifying positive samples, thereby adversely affecting all evaluation metrics when compared to Combination 3.

Figure 10

Confusion matrix and ROC curve for combination ALL, Model 1



4.2.1.3 Analysis and Discussion

Combination 3 (ECG, ECG) stands out for its high precision in predicting true arrhythmia alarms, as evidenced by the Accuracy and Precision metrics. Furthermore, when considering the F1 score, which strikes a balance between precision and recall, this combination shows proficiency in both correctly identifying true arrhythmia alarms and minimizing false alarms. However, the total combination exhibits relatively lower performance compared to Combination 3, particularly in terms of accuracy and ROC score metrics. Moreover, upon scrutinizing its loss and precision curves Figure 7 it becomes apparent that this model struggles with generalizing data, which consequently affects its efficacy in detecting false alarms.

4.2.2 Model 2 Evaluation

Considering the insights gained from Model 1, an experiment was conducted with Model 2 to assess whether false alarm detection could be enhanced through kernel adjustments. The training process for Model 2 followed the same methodology as implemented in Model 1, their results of this experiment are summarized in the following table:

Table 7

Model 2 evaluation metrics

Metrics	COMBINATION 3	COMBINATION ALL
Accuracy Test	0.77910 ± 0.04648	0.67466 ± 0.02109
Precision	0.69041 ± 0.04677	0.58536 ± 0.05197
Recall	0.71666 ± 0.09646	0.52413 ± 0.14367
F1 score	0.69700 ± 0.03680	0.54492 ± 0.09989

ROC SCORE	0.82655 ± 0.03838	0.66416 ± 0.04952
Physionet	0.53412 ± 0.06236	0.43436 ± 0.04628

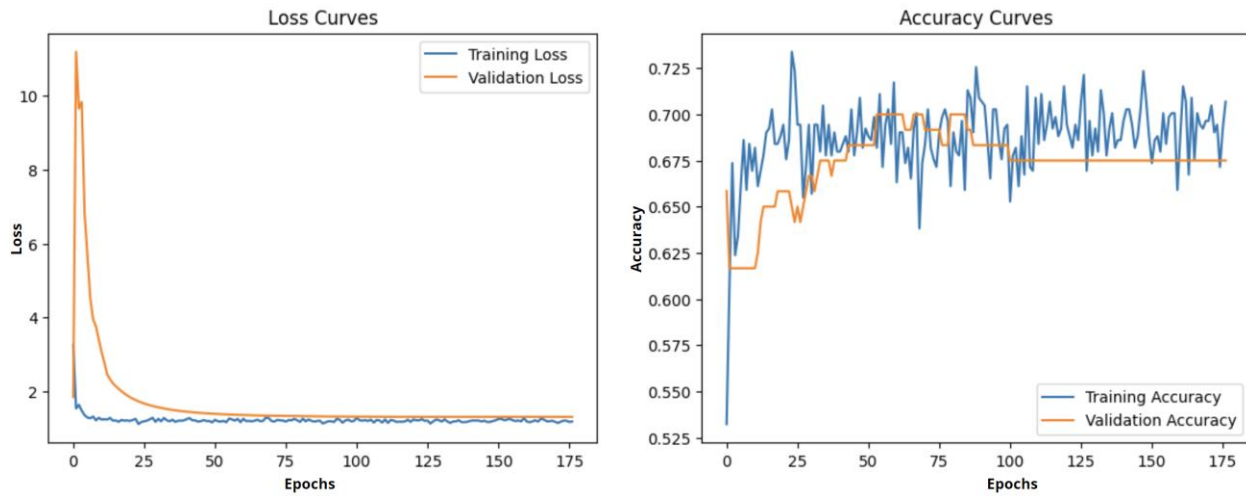
These results suggest that while Model 2 exhibits some improvements in performance metrics compared to Model 1, especially in terms of accuracy and precision, it still faces challenges in false alarm detection, as indicated by the Physionet metric. Further investigation is warranted to elucidate the underlying factors contributing to these results and to refine the model for improved false alarm detection capabilities.

4.2.2.1. Combination 3

For the analysis of Combination 3 in Model 2, we observe that it achieved an accuracy of 0.77910 ± 0.0464 , maintaining a consistent result with the previous model. This suggests a reliable capability of the model to classify test samples accurately. When considering the loss and accuracy curves depicted in Figure 11 it can be observed how the model learned during training. The loss curve shows a gradual decrease in the model's loss as training progresses, indicating an improvement in its ability to make precise predictions. On the other hand, the accuracy curve illustrates how the model's accuracy increases over time, supporting the effectiveness of the model in sample classification.

Figure 11

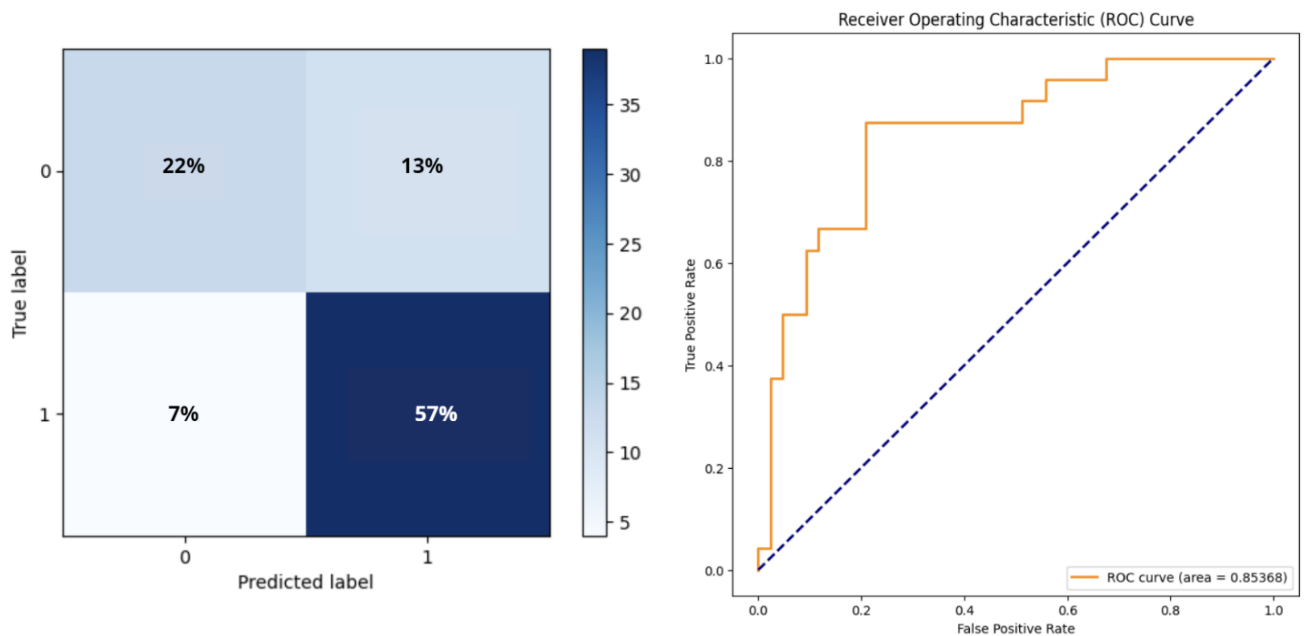
Loss and Accuracy curves in training and validation for combination 3, Model 2



Additionally, metrics such as precision, recall, F1-score, among others, were computed from the confusion matrix presented in Figure 12, particularly noteworthy is a false negative rate of 7,462 % in the test data. This finding underscores the importance of looking beyond overall accuracy to understand the model's performance in each class of interest.

Figure 12

Confusion matrix and ROC curve for combination 3, Model 2

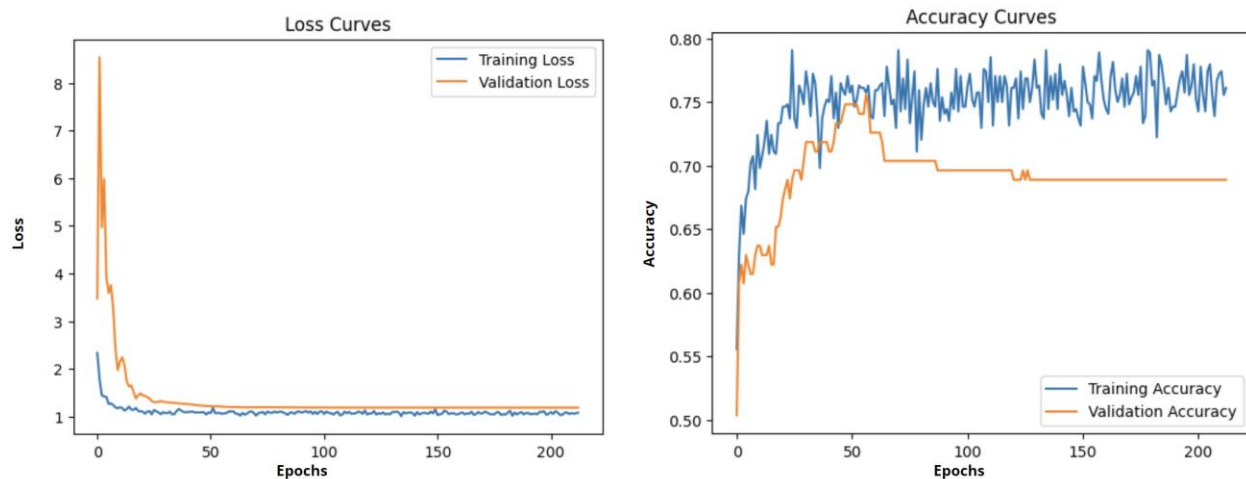


4.2.2.2 Combination ALL

For this combination, higher results were obtained for all metrics compared to Model 1, including the PhysioNet metric. Therefore, this model with different kernels achieves a higher percentage in false alarm detection. For a better understanding of this, training graphs are provided shown below

Figure 13

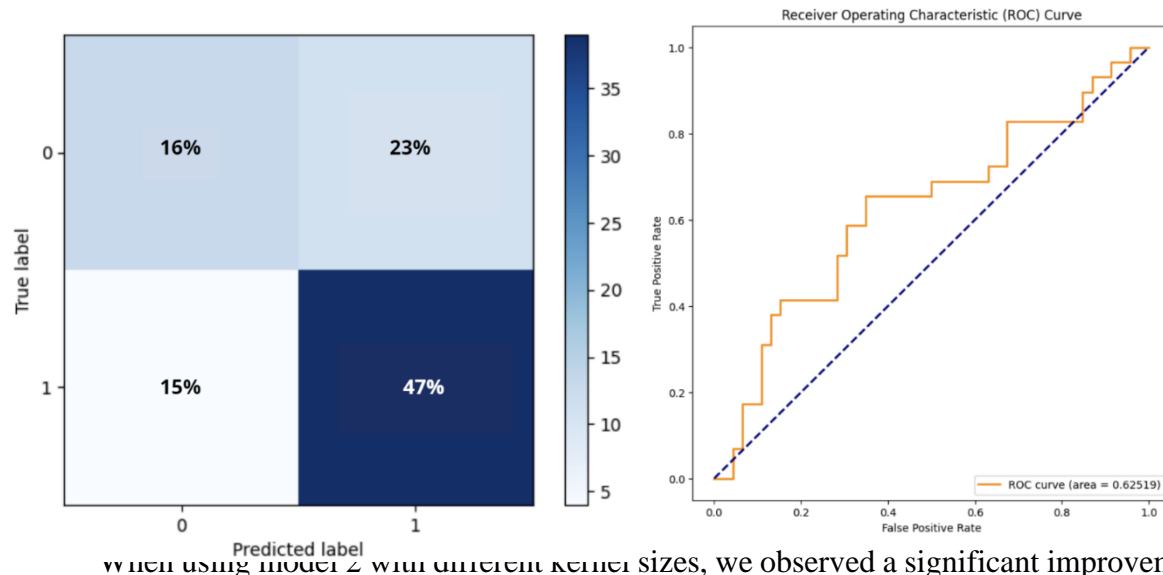
Loss and Accuracy curves in training and validation for combination ALL, Model 2



On the other hand, the metrics have experienced an increase compared to model one, resulting in an improvement in the false negative percentage of 14,66 %. However, this value is lower compared to the results of combination 3. Additionally, when comparing all metrics with the results of combination 3, a decrease is observed in all of them, suggesting a relatively inferior performance in identifying arrhythmia alarms. This phenomenon is also reflected in the confusion matrix shown below, which provides a detailed insight into how the model classifies samples and how classification errors are distributed.

Figure 14

Confusion matrix and ROC curve for combination ALL, Model 2



WHEN USING MODEL 2 WITH DIFFERENT KERNEL SIZES, we observed a significant improvement in most metrics across all evaluated combinations. Once again, combination 3 stands out by achieving the best results in metrics such as precision, accuracy, F1 score, and Physionet Score, highlighting not only its excellent performance in classifying arrhythmia alarms but also its ability to detect false alarms.

4.3 Data Augmentation

Traditionally, data augmentation, a technique in machine learning for increasing the amount of training data by generating additional instances from existing data, is not commonly considered in the medical field. This is because generating synthetic data does not adequately capture the complexity and diversity of real medical conditions, which could lead to unreliable and biased models, however, in this study, the use of this technique has been explored to see how a false alarm detection model for cardiac arrhythmias performs with data augmentation, specifically using a technique known as Gaussian noise, described in the 3.5. Data Augmentation section.

4.3.1 Model 1 Evaluation

Results in Table 8 display the mean alongside the standard deviation and it summarizes the findings, offering an overview of the model's performance across various metrics.

Table 8

Model 1 evaluation metrics with data augmentation

Metrics	COMBINATION 3	COMBINATION ALL
Accuracy Test	0.76119 +- 0.02110	0.65066 +- 0.01959
Precision	0.67990 +- 0.03880	0.56294 +- 0.03836
Recall	0.64166 +- 0.09354	0.46896 +- 0.05160
F1 score	0.65517 +- 0.04549	0.50811 +- 0.02509
ROC SCORE	0.78003 +- 0.02234	0.64422 +- 0.02519
Physionet	0.53184 +- 0.04778	0.41698 +- 0.04775

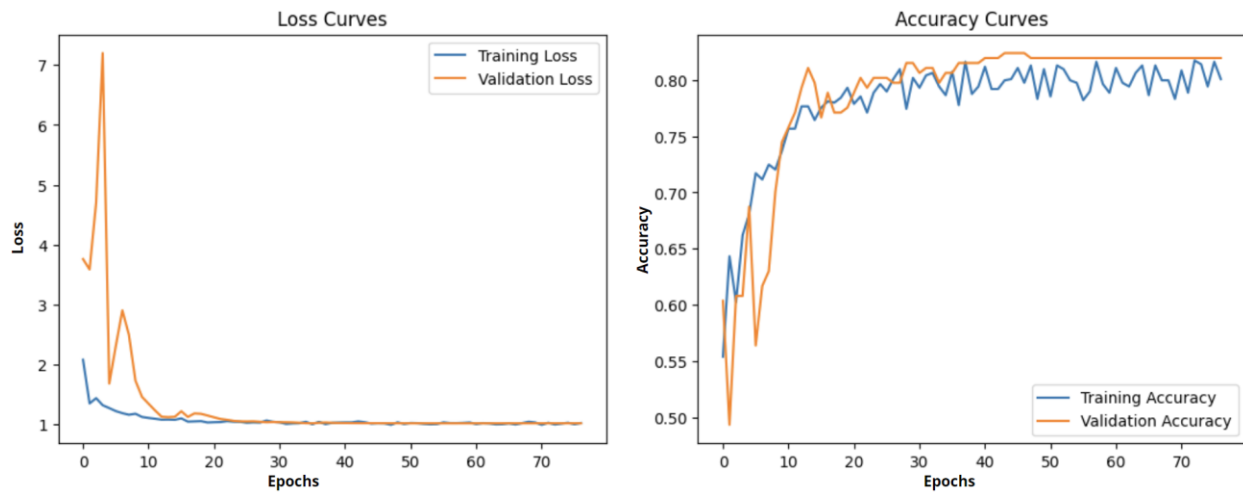
4.3.1.1 Combination 3

With data augmentation in combination 3 of model 1, we observe in the **Table 8** that most of the metrics remain within their ranges, except for the Physionet metric, which decreased by approximately 18,46 % compared to the experiment without data augmentation. Particularly, the recall increased from 0.5345 without augmentation to a value of 0.6416 with data augmentation, suggesting a significant improvement in the model's ability to capture positive samples.

Figure 15

Loss and Accuracy curves in training and validation for combination 3, Model 1 trained with

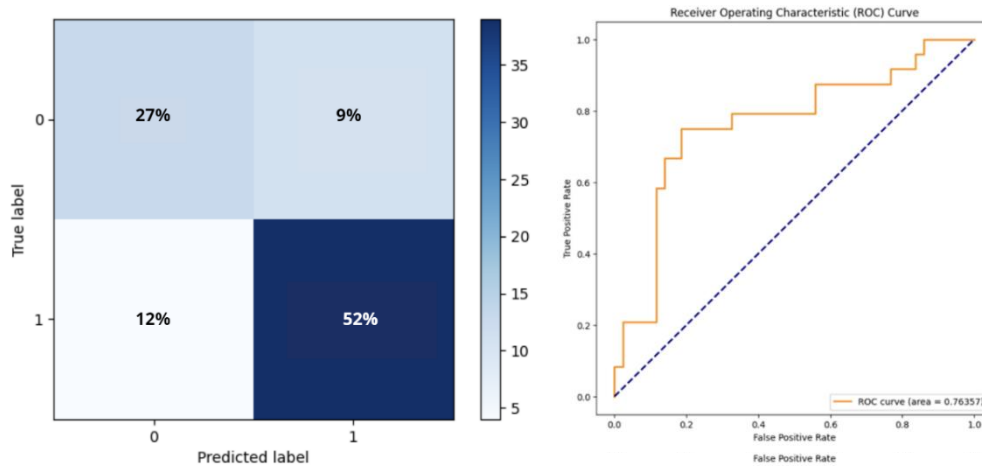
data augmentation



For a more detailed understanding of this combination, we turn to the confusion matrix, where we observe a false negative rate of, indicating the proportion of positive samples incorrectly classified as negative. This provides a more precise insight into how the model is performing in identifying false arrhythmia alarms. On the other hand, with data augmentation in combination 3 of model 1, we observe a significant improvement in the model’s ability to generalize complex patterns in the data, resulting in a notable reduction in the overfitting phenomenon.

Figure 16

Confusion matrix and ROC curve for combination 3, Model 1 trained with data augmentation



4.3.1.2 Combination ALL

In the case of the total combination, an improvement is evident when compared to the original data. However, it still does not match the performance achieved by combination 3, emphasizing the crucial role of data augmentation in enhancing model efficacy. The respective graphics depicting these results are provided below

Figure 17

Loss and Accuracy curves in training and validation for combination ALL, Model 1 trained with data augmentation

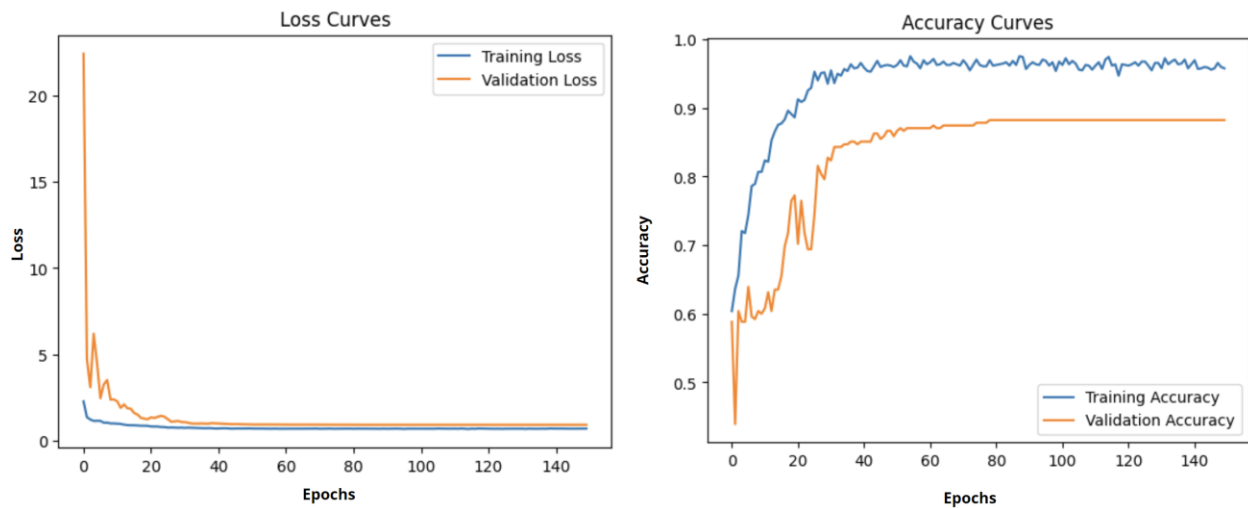
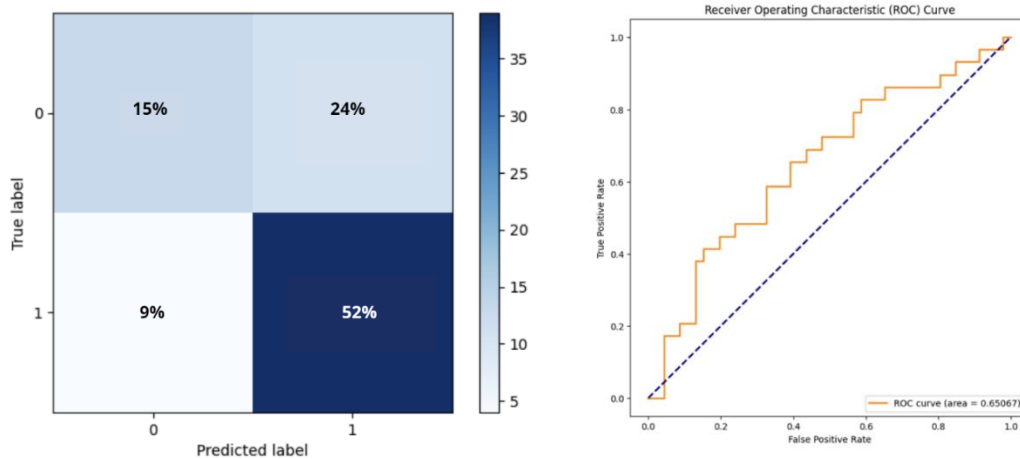


Figure 18

Confusion matrix and ROC curve for combination ALL, Model 1 with data augmentation



4.3.1.3 Analysis and Discussion

In this case, data augmentation has allowed the model to be exposed to a greater variety of examples, making it more robust and less prone to memorizing the training data. As a result, there is a decrease in the discrepancy between the performance on the training data and on the test data, indicating a better ability of the model to generalize to previously unseen data.

4.3.2 Model 2 Evaluation

Results shown below display the mean alongside the standard deviation and it summarizes the findings, offering an overview of the model's performance across various metrics.

Table 9

Model 2 evaluation metrics with data augmentation

Metrics	COMBINATION 3	COMBINATION ALL
Accuracy Test	0.74626 +- 0.01334	0.59733 +- 0.00997
Precision	0.61987 +- 0.02594	0.48161 +- 0.01131
Recall	0.76666 +- 0.04249	0.54482 +- 0.02580
F1 score	0.68395 +- 0.00581	0.51111 +- 0.01641
ROC SCORE	0.81821+- 0.01033	0.63568 +- 0.01749
Physionet	0.44647 +- 0.03691	0.31342 +- 0.00903

4.3.2.1 Combination 3

With the change in kernels, combination 3 maintains its metrics within ranges very close to those obtained without data augmentation. However, an important point to highlight is the metric of the Physionet Challenge, where without data augmentation a value of 0.53412 ± 0.06236

was obtained, while with data augmentation this value decreased to 0.44647 ± 0.03691 , as seen in the **Table 9**; this suggests that the model reduced its performance in identifying false alarms, but, on the contrary, increased its generalization capacity once again.

Figure 19

Loss and Accuracy curves in training and validation for combination 3, Model 2 trained with data augmentation

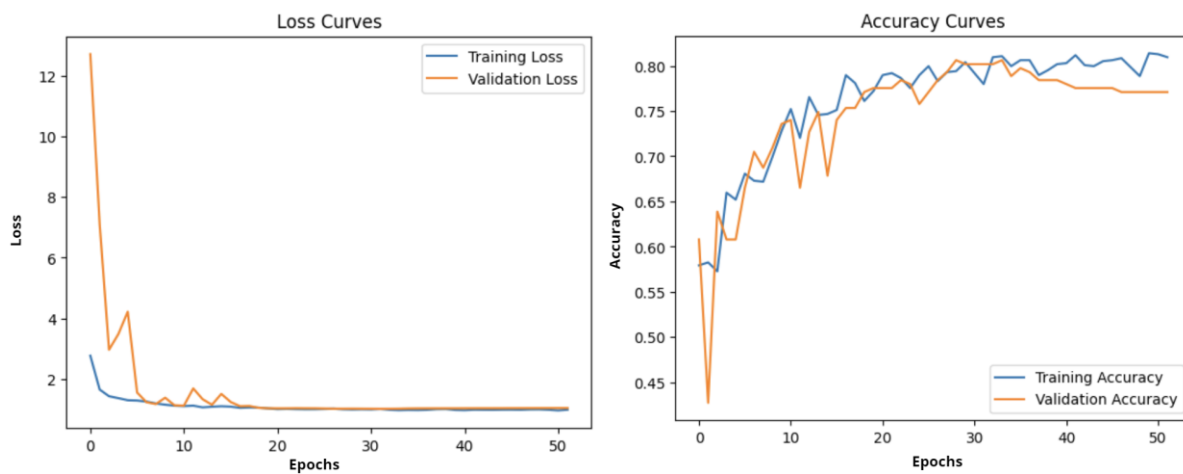
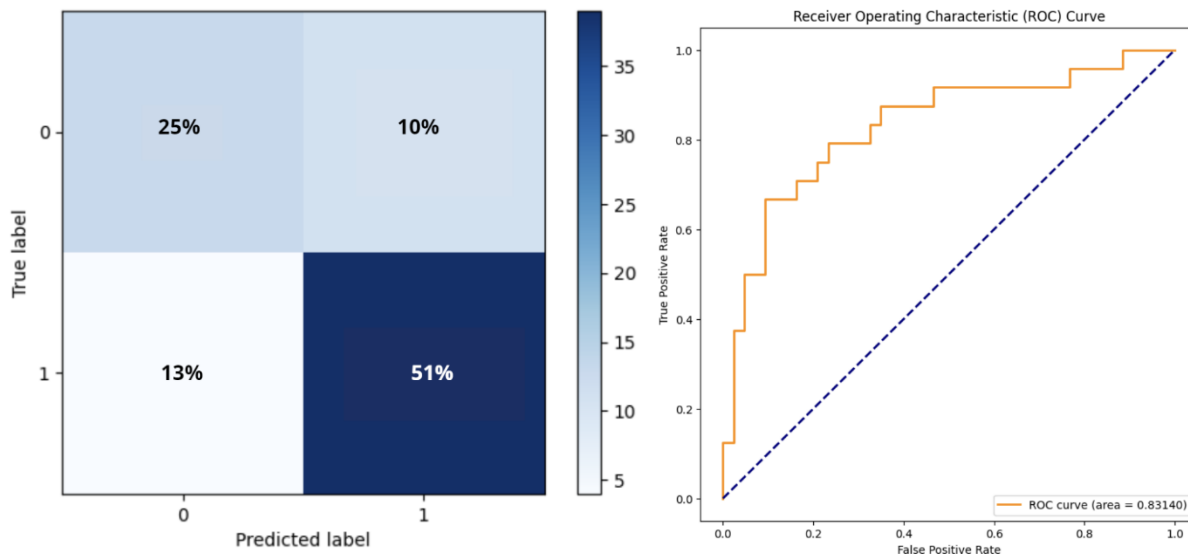


Figure 20

Confusion matrix and ROC curve for combination 3, Model 2 with data augmentation



4.3.2.2 Combination ALL

Similarly to the previous cases, the total combination showed a reduction in the Physionet metric, with a value of 0.31342 ± 0.00903 , as shown in the Table 9. This result is much lower than that of the model without data augmentation. This suggests that the use of data augmentation in the medical field, especially for identifying false alarms, may not be an effective strategy.

Figure 21

Loss and Accuracy curves in training and validation for combination ALL, Model 2 trained with data augmentation

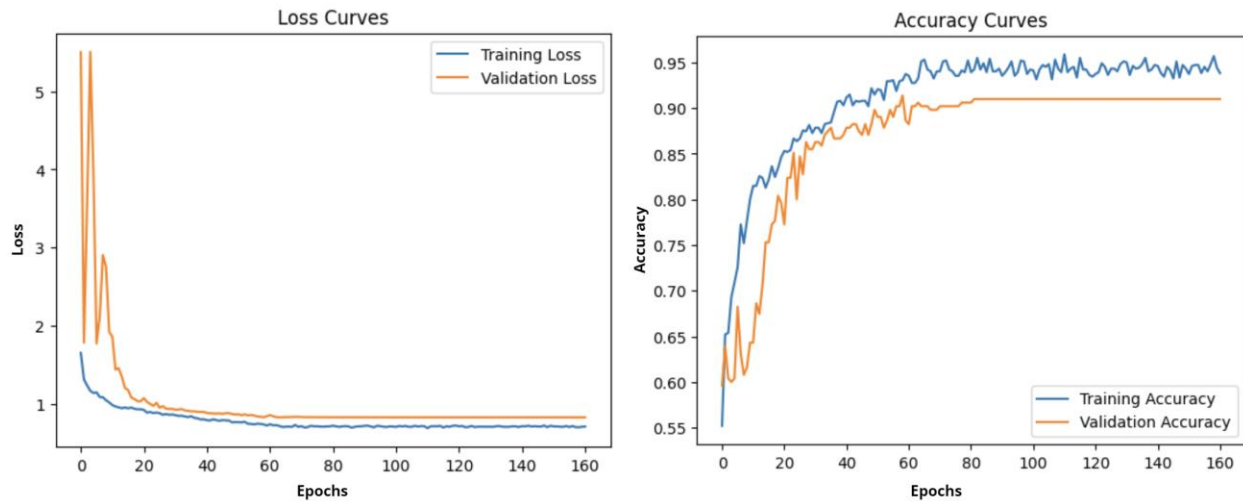
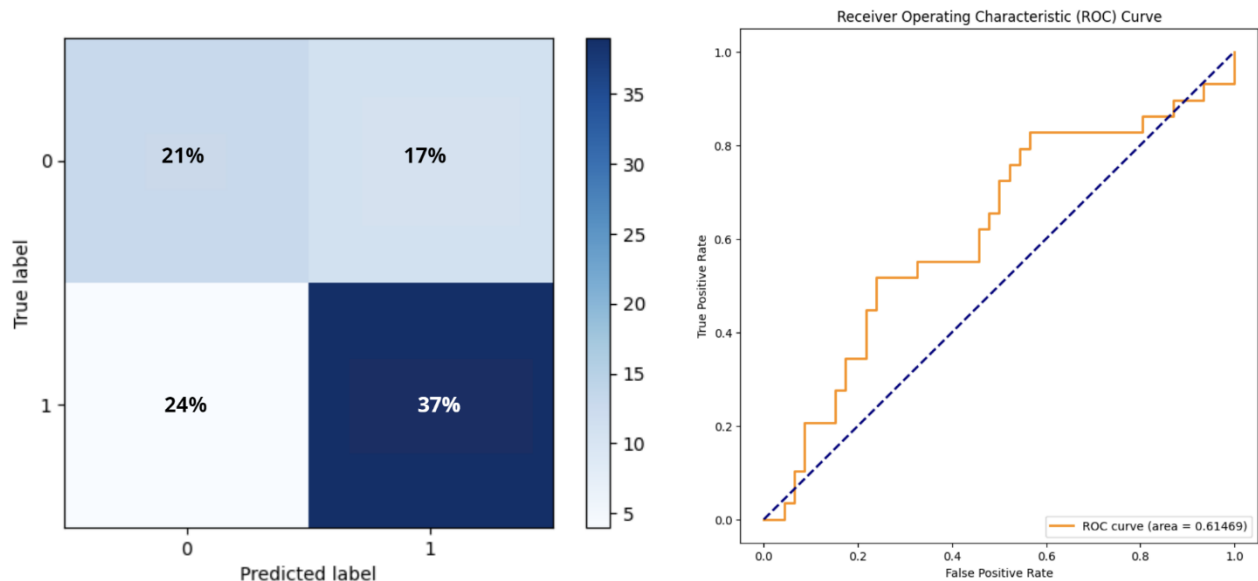


Figure 22

Confusion matrix and ROC curve for combination ALL, Model 2 with data augmentation



4.3.2.3 Analysis and Discussion

Although data augmentation enhances the model's generalization ability by providing greater diversity during training, our findings indicate that its application in false alarm detection with our data combinations did not yield the expected results. Despite optimizing training and adding a substantial amount of additional data, we did not observe a significant improvement in the model's performance for identifying false alarms. These results suggest that, in our specific context, the model trained without using data augmentation might be more effective. It is possible that data augmentation introduced additional noise in the dataset or unbalanced the class distribution, which could make it difficult for the model to distinguish between real and false alarms.

5. Conclusions

- The preprocessing steps to prepare the PhysioNet Computing in Cardiology Challenge 2015 database for model training were effectively implemented. Our analysis highlighted the importance of the composition of dataset to improve model's performance. Specifically, Combination 3 outperformed Combination ALL due to the diverse waveform shapes present in this dataset's third and fourth signals, which affected model generalization during training.
- Through the preprocessing stages, it was observed that various signals in the dataset contained outliers, which negatively affected the quality of the records. However, by implementing combinations, the occurrence of these outliers was diminished, thereby enhancing the model's performance during the evaluation stages.

- During the testing stages, the variation of kernels between models, while not improving the PhysioNet score, does significantly enhance other metrics evaluated such as Accuracy, Precision, Recall, F1 Score and ROC score.
- The performance of the developed deep learning model was evaluated comprehensively using metrics such as true positive rate (TPR), true negative rate (TNR), F1 score, and area under the curve (AUC). This evaluation provided insights into the model's ability to accurately detect false alarms while minimizing false positives and false negatives, meeting the objective of assessing precision comparable to state-of-the-art literature, achieving a performance of 65.24 % on the PhysioNet challenge metric using deep learning exclusively.
- The trained deep learning model exhibited robust detection capabilities, achieving high levels of precision in identifying false alarms of cardiac arrhythmias. By optimizing model parameters and leveraging advanced training techniques, the model demonstrated competitive performance, highlighting its potential for real-world application in clinical settings.
- The data augmentation technique was explored to enhance the training process of the deep learning model for false alarm identification in cardiac arrhythmias. Despite its potential to increase the diversity and quantity of training data, our findings suggest that data augmentation was not as effective as expected in improving the model's performance as it shows on the reduction of the FN from 5.9% without data augmentation to 12% with it.

Bibliography

- Ansari, S., Belle, A., & Najarian, K. (2015). Multi-modal integrated approach towards reducing false arrhythmia alarms during continuous patient monitoring: The Physionet Challenge 2015. 1181–1184. <https://doi.org/10.1109/CIC.2015.7411127>
- Castillo, J. A., Granados, Y. C., & Fajardo, C. A. (2020). Patient-Specific Detection of Atrial Fibrillation in Segments of ECG Signals using Deep Neural Networks. *Ciencia e IngenierÃ-a Neogranadina*, 30, 45–58.
- Chambrin, M. C. (2001). Alarms in the intensive care unit: How can the number of false alarms be reduced? *Critical Care (London, England)*, 5(4), 184–188. <https://doi.org/10.1186/cc1021>
- Clifford, G. D., Silva, I., Moody, B., Li, Q., Kella, D., Shahin, A., Kooistra, T., Perry, D., & Mark, R. G. (2015). The PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU. *Computing in Cardiology*, 2015, 273–276. <https://doi.org/10.1109/CIC.2015.7408639>
- Daluwatte, C., Johannesen, L., Vicente, J., Scully, C., Galeotti, L., & Strauss, D. (2015). Heartbeat fusion algorithm to reduce false alarms for arrhythmias. 745–748. <https://doi.org/10.1109/CIC.2015.7411018>
- Drew, B., Harris, P., Zègre-Hemsey, J., Mammone, T., Schindler, D., Salas-Boni, R., Bai, Y., Tinoco, A., Ding, Q., & Hu, X. (2014). Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients. *PloS one*, 9, e110274. <https://doi.org/10.1371/journal.pone.0110274>

- Keller, J. (2012). Clinical alarm hazards: A top ten health technology safety concern. *Journal of electrocardiology*, 45. <https://doi.org/10.1016/j.jelectrocard.2012.08.050>
- Mousavi, S., Fotoohinasab, A., & Afghah, F. (2020). Single-modal and multi-modal false arrhythmia alarm reduction using attention-based convolutional and recurrent neural networks. *PloS One*, 15(1), e0226990. <https://doi.org/10.1371/journal.pone.0226990>
- Nikitin, A., Iannucci, L., & Kaski, S. (2023). TSGM: A Flexible Framework for Generative Modeling of Synthetic Time Series.
- Plesinger, F., Klimes, P., Halamek, J., & Jurak, P. (2016). Taming of the monitors: Reducing false alarms in intensive care units. *Physiological Measurement*, 37(8), 1313–1325. <https://doi.org/10.1088/0967-3334/37/8/1313>
- Q Yu, Q. Y., Wang, C., Xi, J., Chen, Y., Li, W., Ge, Y., & Huang, X. (2020). ICU False Alarm Identification Based on Convolution Neural Network. <https://doi.org/10.21203/rs.3.rs-130985/v1>
- Wu, X., Kimura, A., Iwana, B., Uchida, S., & Kashino, K. (2019). Deep Dynamic Time Warping: End-to-End Local Representation Learning for Online Signature Verification (p. 1110). <https://doi.org/10.1109/ICDAR.2019.00179>
- Zhou, Y., Zhao, G., Li, J., Sun, G., Qian, X., Moody, B., Mark, R. G., & Lehman, L.-W. H. (2022). A contrastive learning approach for ICU false arrhythmia alarm reduction. *Scientific Reports*, 12(1), 4689. <https://doi.org/10.1038/s41598-022-07761-9>
- Zihlmann, M., Perekrestenko, D., & Tschannen, M. (2017). Convolutional recurrent neural networks for electrocardiogram classification. *2017 Computing in Cardiology (CinC)*, 1–4. <https://doi.org/10.22489/CinC.2017.070-060>

Appendix

Appendix A. Github PMHS-DL-231 repository

In the following link it can be consulted for free the code used for the develop of the Project,
posted in Github.

<https://github.com/German0715/PMHS-DL-231>

This repository has a small section with instructions for their understanding and use.