Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n29.2023.07
pISSN: 1390-650X / eISSN: 1390-860X

Prediction of Arrhythmias and Acute Myocardial Infarctions using Machine Learning

Predicción de arritmias e infartos agudos de miocardio usando aprendizaje automático

Darwin Patiño[1,*], Jorge Medina[1], Ricardo Silva[2],
Alfonso Guijarro[1], José Rodríguez[1]

Received: 16-11-2022, Received after review: 06-12-2022, Accepted: 12-12-2022, Published: 01-01-2023

Abstract

Cardiovascular diseases such as Acute Myocardial Infarction is one of the 3 leading causes of death in the world according to WHO data, in the same way cardiac arrhythmias are very common diseases today, such as atrial fibrillation. The ECG electrocardiogram is the means of cardiac diagnosis that is used in a standardized way throughout the world. Machine learning models are very helpful in classification and prediction problems. Applied to the field of health, ANN, and CNN artificial and neural networks, added to tree-based models such as XGBoost, are of vital help in the prevention and control of heart disease. The present study aims to compare and evaluate learning based on ANN, CNN and XGBoost algorithms by using the Physionet MIT-BIH and PTB ECG databases, which provide ECGs classified with Arrhythmias and Acute Myocardial Infarctions respectively. The learning times and the percentage of Accuracy of the 3 algorithms in the 2 databases are compared separately, and finally the data are crossed to compare the validity and safety of the learning prediction.

Keywords: arrhythmias, acute myocardial, infarction, machine learning, artificial neural network, convolutional neural network, extreme gradient boosting

Resumen

Las enfermedades cardiovasculares, como el infarto agudo de miocardio, son una de las tres principales causas de muerte en el mundo según datos de la OMS. De forma similar, las arritmias cardíacas¸ como la fibrilación auricular, son enfermedades muy comunes en la actualidad. El electrocardiograma (ECG) es el medio de diagnóstico cardíaco que se utiliza de forma estandarizada en todo el mundo. Los modelos de aprendizaje automático son muy útiles en problemas de clasificación y predicción. Aplicadas al campo de la salud, las redes neuronales artificiales (ANN) y las redes neuronales convolucionales (CNN) en conjunto con modelos basados en árboles como XGBoost, son de vital ayuda en la prevención y control de enfermedades del corazón. El presente estudio tiene como objetivo comparar y evaluar el aprendizaje basado en los algoritmos ANN, CNN y XGBoost mediante el uso de las bases de datos de ECG Physionet MIT-BIH y PTB, que proporcionan ECG clasificados con arritmias e infartos agudos de miocardio, respectivamente. Se comparan por separado los tiempos de aprendizaje y el porcentaje de exactitud de los tres algoritmos en las dos bases de datos, y finalmente se cruzan los datos para comparar la validez y seguridad de la predicción.

Palabras clave: arritmias, infarto agudo miocardio, aprendizaje automático, red neuronal artificial, red neuronal convolucional, impulso del gradiente extremo

[1,*]Universidad de Guayaquil, Ecuador. Corresponding author ✉: darwin.patinop@ug.edu.ec.
[2]University of Villanova, Pensilvania, Estados Unidos.

Suggested citation: Patiño, D.; Medina, J.; Silva, R.; Guijarro, A. and Rodríguez, J. "Prediction of Arrhythmias and Acute Myocardial Infarctions using Machine Learning," *Ingenius, Revista de Ciencia y Tecnología*, N.° 29, pp. 79-89, 2023, DOI: https://doi.org/10.17163/ings.n29.2023.07.

1. Introduction

A multiplicity of devices (personal computers, smartphones, tablets, cell phones, etc.) are used today to accumulate and process bigdata about human behavior. This bigdata is available for a multiplicity of purpose including medicine [1,2]. Mobile Health (mHealth) and smart devices enable early detection and prompt intervention for patients with Atrial Fibrillation (AF). Single- and multi-lead ECG, photoplethysmography (PPG), and oscillometric, with validated diagnostic capability, can be integrated within the clinical practice to detect AF. Existing clinical practice guidelines suggest that pulse assessment with ECG screening for the high-risk population, for patients >65 years of age, is appropriate to reduce complications. However, easy-to-use, and affordable consumer health and smart devices may be a good alternative screening tool not only for the elderly population with comorbidities, but also for the general low-risk population with frequent monitoring [2–7].

mHealth devices that are capable of monitoring heart rate and/or heart rhythm come in multiple forms such as, smartphone apps, smart watches, rings, necklaces, wearable sensors, and patches [8–10]. Companies have created products capable of producing a point-of-care ECG registries, such as the AliveCor Kardia Monitor series of devices [8]. In a large cohort study conducted in Hong Kong, research on the Kardia single-lead ECG device found that a review cardiologist confirmed that 65% of AFs detected by the device were accurate. In this study of more than 10000 patients with a mean age of 78 years, the number needed to make an accurate new diagnosis of AF was 145 participants [8].36 The sensitivity and specificity of the Kardia monitor were found to be 99, 6% and 97.8%, respectively [11].

The term Ubiquitous Health (u-Health) defined by Weiser as the integration of computing into human actions and behaviors at “anytime” and “anywhere” has been gaining prominence [1,2]. The main attribute of u-Health is the capacity for interaction between individuals and devices in such a way that the technology is transparent to the user [12]. It is not clear what is the best algorithm to detect cardiovascular diseases by means of u-Health devices. The technology needs to be robust, reliable and with low computational cost so that it can be run directly in the devices even when offline. The goal of this paper is to evaluate the best alternative for arrhythmia detection with u-Health Devices.

This work is the product of an ongoing collaboration between the University of Guayaquil and the University of Villanova where multiple artificial intelligence strategies are being developed for real time detection of arrhythmias. The current work uses existing arrhythmia databases to validate the strategies, future work intends to process real time data from wearable devices. Results from this research are highly encouraging and will be further discussed throughout the article.

2. Materials and Methods

2.1. Methodology

Two Physionet databases of ECG MIT electrocardiograms (arrhythmias) with 109444 records (normal 21891 and abnormal 87553) and PTDB (infarcts) with 14550 records (normal 4045 and abnormal 10505) were used. MIT database has 4 categories Normal “N” 0, Supraventricular “S” 1, Ventricular “V” 2, Ventricular Fibrillation “F” 3, Other unclassified “Q” 4. PTDB database has 2 categories “N” 0 and with cardiac problems “A” 1. The SMOTE was used to regularize the categories and avoid Overfitting and Underfitting.

The 80% of the records are used for training and 20% for tests, 20% of the 80% of training is taken again to evaluate the data prediction. This process was carried out separately for both the MIT and PTDB databases, using in both cases ANN artificial neural networks and CNN convolutional neural networks in addition to the decision tree-based algorithm called XGBoost or extreme gradient boosting, the 3 algorithms were evaluated with the 2 databases.

2.2. Cardiovascular Diseases and Artificial Intelligence

Cardiovascular diseases (CVD) are the leading cause of mortality worldwide, accounting for 31% of all deaths [13]. One of the main causes is acute myocardial infarction (AMI). There is an emerging need to study a wide range of cutting-edge techniques for its analysis and diagnosis of heart diseases. To judge the specific situation of the patient, doctors often look at the ECG (electrocardiograph) signal to get enough information to help them diagnose. Many researchers have applied Machine Learning algorithms to study the arrhythmia [14] classification problem. Advances in data processing, storage capacity and Machine Learning ML methods have been transforming the field of medicine, including cardiology [15].

AF is one of the most common types of arrhythmias which is characterized by a rapid and irregular heartbeat [**?**]. Ischemic heart disease (IHD) is a condition in which there is an inadequate supply of blood and oxygen to a part of the heart muscle [16]. This condition usually occurs when there is an imbalance between oxygen supply and demand to the heart muscle (myocardium), usually due to atherosclerotic heart disease [17]. Patients usually do not show typical signs and symptoms (asymptomatic) until ischemic heart disease manifests as angina, myocardial infarction, or sudden cardiac death [18].

Heartbeat classification, in ECG analysis, is the most common way to automate arrhythmia [19] diagnoses. The common machine learning-based ECG learning flow includes signal noise analysis, heartbeat recognition, feature extraction, and heartbeat classification. For deep learning, feature extraction can be replaced by storing fragments of beats from a complete ECG sequence [20]. Existing ECG signal identification algorithms in the literature incorporate three important points: preprocessing, classification, and imbalance of the data set. To analyze characteristics that can be directly related to physiological factors on the development of diseases, three Deep Learning algorithms were considered: CNN convolutional neural network, ANN artificial neural network, and the reinforced tree eXtreme Gradient Boost o XGBoost.

Artificial Neural Networks are learning algorithms that can identify complex relationships in data. ANNs are designed to mimic human nervous system. Typical ANN's are comprised of 3 layers: input, output, and hidden layers. Each layer is made up of neurons [21]. Convolutional neural networks (CNN), recurrent neural networks (RNN), and naïve Bayes (NB) are used as classifiers. There are techniques that use different combinations such as Discrete Wavelet Transform (DWT) and ANN combination were obtained for ECG arrhythmia classification [22]. When the number of features is greater than the number of samples, ANNs can handle multiple classes, there is no effect of large data sets on ANNs, and extensive memory is not required [23].

ANN-based study classifies IHD using heart rate variability (HRV) parameters along with a clinical data such as left ventricular ejection fraction (LVEF), age, and gender. A series of networks with different number of input nodes (varying between 7 and 15), hidden nodes (between 2 and 10) and two output nodes were tested. The training and test ranges were respectively 75% and 25% of the total amount of data [24]. Various investigators have also used artificial neutral network (ANN)-based approaches for diagnostic classification of ECG signals [16].

Modern Deep Neural Network DNN techniques are used to solve the problem of manual feature selection and extraction in conventional automatic systems for MI Image diagnostics [25]. The Neural Network backpropagation algorithms are used to train deep learning [26]. CNN is most applied to analyze visual images. Myocardial infarction is predicted using the characteristic images before and after the attack obtained as input image of a CNN [27]. Commonly used layers in CNN are convolution (Conv), rectified linear unit (ReLU), pooling, batch normalization, and fully connected layer [28]. An input matrix is fed into a detection model that is composed of CNNs and a bidirectional Long short-term memory network (bi-LSTM) with 5-fold cross-stratified validation [29].

DNN have shown success in several domains, including images, audio, and text [30]. In real-world applications, the most common data type is tabular data, which comprises sample (rows) with the same set of features (columns). Tabular data is used in many fields, including medicine, finance, manufacturing, climate science, and many others [31].

Traditional machine learning methods, such as gradient-powered decision trees (GBDT) [32], dominate tabular data modeling and show superior performance to deep learning. Despite their theoretical advantages [33–35], DNNs pose many challenges when applied to tabular data, such as lack of locality, sparseness of data (missing values), mixed feature types (numerical, ordinal, and categorical) and lack of prior knowledge about the structure of the data set (as opposed to text or images). Ensemble-of-trees algorithms, such as XGBoost, are considered the recommended choice for real-life tabular data problems [32], [36].

XGBoost has been used to classify Atrial Fibrillation [37]. One study proposes ECG signals classifier based on XGBoost and ensemble empirical mode decomposition (EEMD) that takes advantage of functions based on time, frequency, and morphological characteristics [38]. Another study proposes to create a set of five-dimensional morphological features regarding QRS complexes and RR intervals, as well as some wavelet coefficient features, to build the feature vector for highly efficient heartbeat classification [21]. Performance measures are trained to find features that are correctly classified and those that are not well classified, then their relationship is used to find the efficiency of the classifier. We can get a high ratio even if all the important classes are misclassified. To overcome this, the data must be properly balanced [39]. SMOTE, "Synthetic Minority Oversampling Technique" can overcome some classification disadvantages [40]. This method has been shown to be better than other mixtures of under-sampling and oversampling.

A study conducted in the year 2021 compared the performance of XGBoost and the DNN using the Adamax optimizer and binary cross-entropy loss function with four hidden layers. The results showed that the XGboost outperformed the DNN by achieving a learning accuracy of 100%, while its prediction accuracy was 95.60% and 93.08%, for the same phases [41]. The overall learning performance of the DNN model was 89.42% and 81.23%, while the prediction accuracy was 80.50% and 77.36%, respectively, for the same variables [41]. The goal of our study was to compare the algorithms to determine the most cost-effective solution for real-time arrhythmia detection.

Single-lead ECG monitors are frequently used because of their highly productive nature, short run time, and low cost [42]. However, single-lead ECG cannot capture all the information due to the great diversity of CVD features that can cause misdiagnosis [43].

2.3. Artificial Intelligence

The object of the work is to select an algorithm for the classification of cardiac alterations that can be executed in real time, while the electrocardiographic signal is being acquired. Given that the evaluated algorithms are based on the identification and classification of a single cardiac cycle, the ideal would be to have an algorithm capable of capturing and classifying the signals during the period between waves, better known as the T-P segment or interval, as shown in Figure 1.

To execute the algorithm in real time, the execution time should be less than the T-P interval or, in other words, less than 200 ms. We analyzed files, from the Physionet databases regarding spread and available for research of electrocardiograms ECG. Physionet was developed by the Beth Israel Hospital in Boston (now the Beth Israel Deaconess Medical Center) in conjunction with the Massachusetts Institute of Technology: (MIT-BIH) [44] and the Physikalisch - Technische Bundesanstalt, the National Metrology Institute of Germany: (PTB). The MIT-BIH Base has 109444 ECGs and the PTB Base has 14550 ECGs [45].

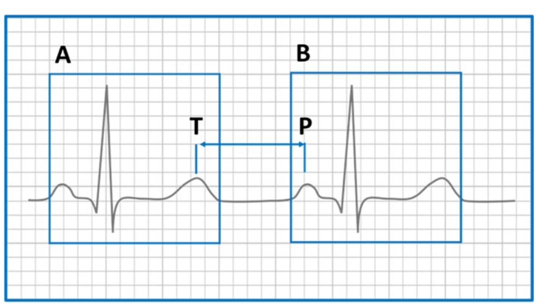

Figure 1. Two Independent Cardiac Cycles A and B within their respective detection window with the T-P interval identified

MIT-BIH is an arrhythmia database, so it has a classification labeled "N": 0, "S": 1, "V": 2, "F": 3, "Q": 4, where 0 is NORMAL and from 1 to 4 are arrhythmias which are classified as Bradyarrhythmia's and Tachyarrhythmias; subclassified as Supraventricular Tachyarrhythmias and Ventricular Tachyarrhythmias. In the PTB database, the classification is 0 NORMAL and 1 ABNORMAL, where severe heart disease such as myocardial infarction (mostly), heart failure and bundle branch block are considered. The csv files available in Kaggle have 187 columns that represent the ECG bio-signal and an additional 188 column that classifies the ECG. This field is available in both databases and allows the applicability of machine learning.

Three algorithms were considered: CNN convolutional neural network, ANN artificial neural network, and the reinforced tree eXtreme Gradient Boost o XGBoost. For the neural network algorithms, the works published by Premanand S, available at Analytics Vidhya, were taken as reference.

To avoid underfitting and overfitting problems in machine learning, the SMOTE function was applied to both databases separately, which creates new ECGs based on the original data and balances the categories.

A division by 80 is performed, -20-20 to both databases obtaining: 289878-72470-90587 ECGs and for PTB 13446-3362-4202 in Training, Test and Validation ECG data for MIT and PTB respectively. ECG samples from both databases and in the different categories of normal and abnormal have been plotted in Figures 2 and 3.

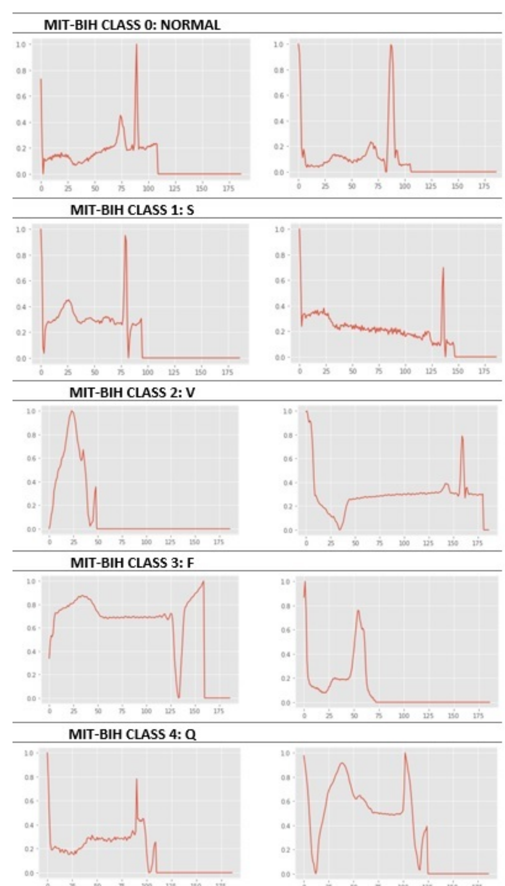

Figure 2. MIT-BIH classified signals

We have proceeded to train with ANN, CNN and XGBoost both the MIT Database and the PTB separately to make a comparison in terms of training times, and levels of precision in the prediction: accuracy, precision, recall, generating the required confusion matrices.

The training results were first validated using the test and validation data from both MIT and PTB separately. And then the prediction level is validated by crossing data between both databases. A CNN architecture proposes to select an optimal group of individual layers and the size of the filters. The following values were the chosen: 2 dense layers, layer size 128, number of 2D convolutional layers and MaxPooling 2D [12N] [46].

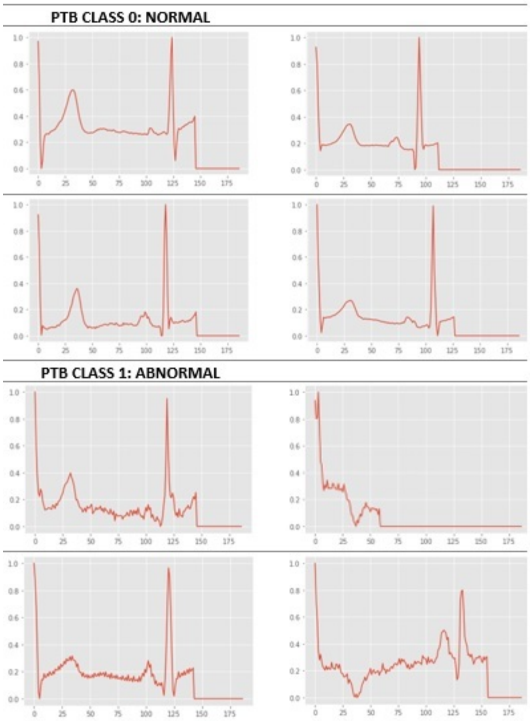

Figure 3. MIT-BIH classified signals

3. Results and discussion

We proceeded to train with ANN, CNN and XGBoost both the MIT Database and the PTB separately to make a comparison in terms of Training times, and levels of precision in the prediction: Accuracy, Precision, Recall, are presented in the required confusion matrices, Figures 4 and 5.

XGBOOST

	precision	recall	f1-score	support
0,0	0,90	0,97	0,93	2091
0,1	0,97	0,89	0,93	2111
accuracy			0,93	4202
macro avg	0,93	0,93	0,93	4202
weighted avg	0,93	0,93	0,93	4202
[[2026 65]				
[236 1875]]				

ANN

	precision	recall	f1-score	support
0,0	0,95	0,99	0,97	2091
0,1	0,99	0,95	0,97	2111
accuracy			0,97	4202
macro avg	0,97	0,97	0,97	4202
weighted av	0,97	0,97	0,97	4202
[[2061 30]				
[104 2007]]				

CNN

	precision	recall	f1-score	support
0,0	0,99	1,00	0,99	2076
0,1	1,00	0,99	0,99	2126
accuracy			0,99	4202
macro avg	0,99	0,99	0,99	4202
weighted avg	0,99	0,99	0,99	4202
[[2074 2]				
[29 2097]]				

Figure 4. PTB classified results

XGBOOST

	precision	recall	f1-score	support
0,00	0,81	0,85	0,83	18109
1,00	0,92	0,87	0,89	17952
2,00	0,93	0,91	0,92	18228
3,00	0,90	0,94	0,92	18037
4,00	0,97	0,97	0,97	18261
accuracy			0,91	90587
macro avg	0,91	0,91	0,91	90587
weighted avg	0,91	0,91	0,91	90587
[[15396 1016 545 859 293]				
[2004 15539 132 233 44]				
[735 117 16534 659 183]				
[503 112 468 16927 27]				
[354 52 150 27 17678]]				

ANN

	precision	recall	f1-score	support
0,00	0,92	0,96	0,94	18109
1,00	0,97	0,95	0,96	17952
2,00	0,99	0,96	0,98	18228
3,00	0,98	0,99	0,99	18037
4,00	0,99	0,99	0,99	18261
accuracy			0,97	90587
macro avg	0,97	0,97	0,97	90587
weighted av	0,97	0,97	0,97	90587
[[17433 393 132 102 49]				
[906 16971 28 19 20]				
[416 28 17563 204 17]				
[75 16 25 17920 1]				
[101 13 18 6 18123]]				

CNN

	precision	recall	f1-score	support
0,00	1,00	0,98	0,99	18026
1,00	0,99	1,00	0,99	17939
2,00	1,00	0,99	0,99	18271
3,00	0,99	1,00	0,99	18032
4,00	1,00	1,00	1,00	18319
accuracy			0,99	90587
macro avg	0,99	0,99	0,99	90587
weighted avg	0,99	0,99	0,99	90587
[[17742 151 43 60 30]				
[42 17893 3 1 0]				
[19 8 18150 90 4]				
[19 0 24 17989 0]				
[2 4 10 2 18301]]				

Figure 5. Resultados de clasificación sobre la MIT-BIH

Accuracy *vs.* Loss for each of the algorithms for the PTB and MIT databases are presented in Figures 6 and 7.

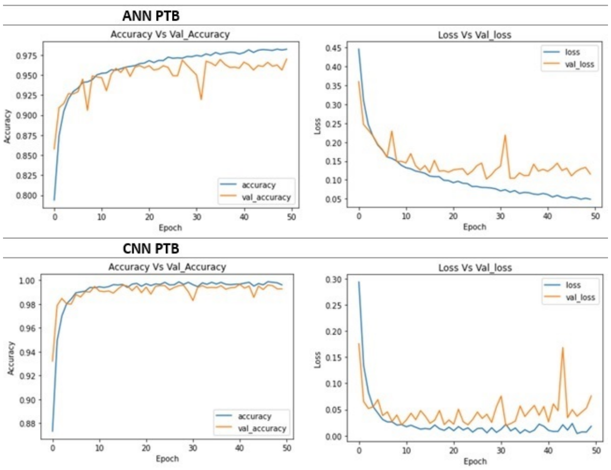

Figure 6. Accuracy *vs.* Loss for each of the algorithms using PTB dataset

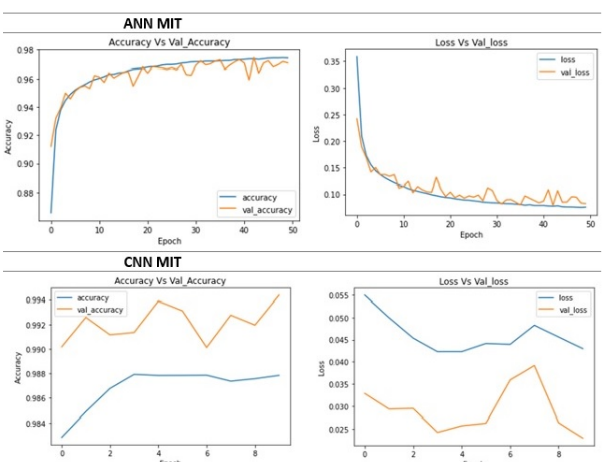

Figure 7. Accuracy *vs.* Loss for each of the algorithms using MIT dataset

FIT Times and Accuracy obtained for each of the databases is presented in Figure 8.

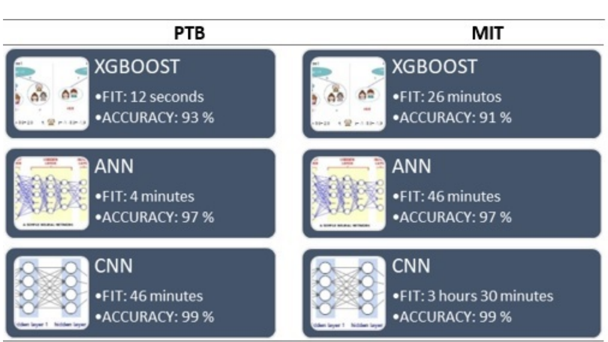

Figure 8. Fit and Accuracy for PTB and MIT datasets

After evaluating the training results separately prediction algorithms level was cross validated by exchanging data between both databases, as shown in Figure 9. For validation purposes, the PTB database was categorized as 0 = normal and 1 = abnormal and processed by the learning models based on XGBoost and ANN.

ANN Normal ECGs (0)

	precision	recall	f1-score	support
0,00	1,00	0,86	0,93	4045
1,00	0,00	0,00	0,00	0
2,00	0,00	0,00	0,00	0
3,00	0,00	0,00	0,00	0
4,00	0,00	0,00	0,00	0
accuracy			0,86	4045
macro avg	0,20	0,17	0,19	4045
weighted av	1,00	0,86	0,93	4045
[[3484 370 163 20 8]				
[0 0 0 0 0]				
[0 0 0 0 0]				
[0 0 0 0 0]				
[0 0 0 0 0]]				

ANN Abnormal ECGs (1)

	precision	recall	f1-score	support
0,00	0,00	0,00	0,00	0
1,00	1,00	0,15	0,27	10505
2,00	0,00	0,00	0,00	0
3,00	0,00	0,00	0,00	0
4,00	0,00	0,00	0,00	0
accuracy			0,15	10505
macro avg	0,20	0,03	0,05	10505
weighted av	1,00	0,15	0,27	10505
[0 0 0 0 0]				
[7247 1616 856 180 606]				
[0 0 0 0 0]				
[0 0 0 0 0]				
[0 0 0 0 0]]				

XBOOST Normal ECGs (0)

	precision	recall	f1-score	support
0,00	1,00	0,85	0,92	4045
1,00	0,00	0,00	0,00	0
2,00	0,00	0,00	0,00	0
3,00	0,00	0,00	0,00	0
4,00	0,00	0,00	0,00	0
accuracy			0,85	4045
macro avg	0,20	0,17	0,18	4045
weighted av	1,00	0,85	0,92	4045
[3439 545 49 6 6]				
[0 0 0 0 0]				
[0 0 0 0 0]				
[0 0 0 0 0]				
[0 0 0 0 0]]				

XBOOST Abnormal ECGs (1)

	precision	recall	f1-score	support
0,00	0,00	0,00	0,00	0
1,00	1,00	0,11	0,20	10505
2,00	0,00	0,00	0,00	0
3,00	0,00	0,00	0,00	0
4,00	0,00	0,00	0,00	0
accuracy			0,11	10505
macro avg	0,20	0,02	0,04	10505
weighted av	1,00	0,11	0,2	10505
[[0 0 0 0 0]				
[7292 1176 641 302 1094]				
[0 0 0 0 0]				
[0 0 0 0 0]				
[0 0 0 0 0]]				

Figure 9. PTB data Cross Validated on MIT dataset using ANN and XGBoost

The training phase was completed through trial-and-error definitions, the hyperparameters were properly configured according to Table 1 so that they reach the expected level of accuracy.

ECGs from the PTB database were validated for prediction to machine learning models based on ECGs from the MIT-BIH database, obtaining Accuracy levels of 85% and 86% for normal ECGs. Regarding the validation of abnormalities, XGBoost had an Accuracy of 11% and ANN 15%. This is because they handle different heart diseases MIT-BIH arrhythmias and PTB acute myocardial infarctions.

Table 1. Hyper-Parameter

Classifier Type	Hyper-Parameter	MIT	PTDB
XGBoost	Max deph	6	6
	Learning rate	0,1	0,1
	Optimum number - estimators	100	100
	Random state	42	42
ANN	Activation function - hidden layers	ReLU	ReLU
	Activation function - output layers	Softmax	Softmax
	Number of epochs	50 epochs	50 epochs
	Batch_size	10	10
	Optimization method	adam	adam
	Layer hidden	3	3
	Learning rate	0,0001	0,0001
CNN	Activation function - hidden layers	ReLU	ReLU
	Activation function - output layers	Softmax	Softmax
	Number of epochs	10 epochs	10 epochs
	Batch_size	10	10
	Optimization method	adam	adam
	Layer hidden	7	7
	Learning rate	0,0001	0,0001

This proves that normal signals can be recognized cross platform, but abnormal ECG data is not interoperable between one database and another.

Finally, using supervised learning, the 2 artificial neural network models and the XGBoost algorithm for prediction by classification, were compared using a weight matrix, considering criteria such as prediction accuracy, sensitivity of medical data (false positive/-false negative), the learning time, the prediction time, as shown in Table 2.

Table 2. PTB data Cross Validated on MIT dataset using ANN and XGBoost

Criterio	ANN	Value	CNN	Valor	XgBoost	Value	Weight	ANN	CNN	XgBoost
Training time	46 min	4	3 h, 30 min	2	26 min	5	3	12	6	15
Resources required for training	Mid	4	High	2	Low	5	3	12	6	15
Resources required for operation	Mid	5	Mid	5	Mid	5	3	15	15	15
Predictive capacuty	Hgh	4	Low	2	High	5	3	12	6	15
Sensitivity	2631	3	504	5	8477	2	5	15	25	10
Mean prediction time	0:00:04.14 2830	3	0:01:22.14 2794	2	0:00:02,02 8421	4	5	15	10	20
Prediction accuracy	97%	3	99 %	5	91 %	2	5	15	25	10
Total							110	96	93	100
							100	87,3	84,5	90,9

As previously defined, the object of the work is to select an algorithm for the classification of cardiac alterations that can be executed in real time, while the electrocardiographic signal is being acquired. Given that the results from Table 2 a point comparison can be obtained to determine what is the best algorithm for a u-health solution (Figure 10).

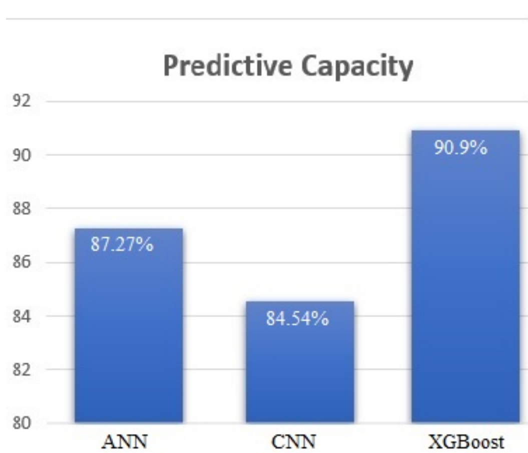

Figure 10. Point comparison between the three algorithms

For the learning phase of the models, the 123994 electrocardiographic records were used, among the predictive values obtained, the model that provides the highest prediction accuracy was determined; Since artificial intelligence has managed to learn how to properly classify conditions with cardiac pathologies, the most suitable and applicable model for the diagnosis of people or population groups was selected. In matters of prevention, the prediction of risk for the general population is of vital importance since it can reduce the impact of deaths due to cardiac pathologies, as well as reduce the costs related to these cases.

4. Conclusions

From the discussion, a weight matrix was used to compare the quality of the 3 prediction algorithms. Based on such results we conclude that CNN (convolutional neural networks) are much more accurate than other algorithms (99%), however, training time is high (in terms of hours), when compared to the XGBoost training that is obtained within minutes. Since we are dealing with Human Health, precision and accuracy in prediction have more weight than speed in training. As an intermediate we have the artificial neural network ANN that with 97% accuracy is very acceptable. XGBoost, given the tabular nature of the data, is the best choice as seen from Figure 10.

Prior conclusion indicates that it is possible to obtain information about arrhythmia within the RR interval. Since the goal of the project was to process data real time, the results are highly encouraging. For future work we intend to use ECG data from smart watches that are being generated as part of a doctoral research. Arrhythmia detection from smart watches would be a great tool for early detection of potential life-threatening events such as fibrillation. False positives however need to be reduced and since we could process data in real time, the joint probability distribution can be used in future work to increase the predictive nature of the algorithm. All in all, this is a significant contribution to the field of real-time arrhythmia detection.

References

[1] K.-Y. Chin, K.-F. Lee, and Y.-L. Chen, "Using an interactive ubiquitous learning system to enhance authentic learning experiences in a cultural heritage course," *Interactive Learning Environments*, vol. 26, no. 4, pp. 444–459, 2018. [Online]. Available: https://doi.org/10.1080/10494820.2017.1341939

[2] F. P. Mota, F. P. d eTôledo, V. Kwecko, S. Devincenzi, P. Núñez, and S. S. da C. Botelho, "Ubiquitous learning: Asystematic review," in *2019 IEEE Frontiers in Education Conference (FIE)*, 2019, pp. 1–9. [Online]. Available: https://doi.org/10.1109/FIE43999.2019.9028361

[3] Y. Guo and G. Y. H. Lip, "Beyond atrial fibrillation detection: how digital tools impact the care of patients with atrial fibrillation," *European Journal of Internal Medicine*, vol. 93, pp. 117–118, 2021. [Online]. Available: https://doi.org/10.1016/j.ejim.2021.08.026

[4] Y. Guo, H. Wang, H. Zhang, T. Liu, Z. Liang, Y. Xia, L. Yan, Y. Xing, H. Shi, S. Li, Y. Liu, F. Liu, M. Feng, Y. Chen, G. Y. H. Lip, and M.A.F.A. II Investigators, "Mobile photoplethysmographic technology to detect atrial fibrillation," *Journal of the American College of Cardiology*, vol. 74, no. 19, pp. 2365–2375, Sep. 2019. [Online]. Available: https://doi.org/10.1016/j.jacc.2019.08.019

[5] M. V. Perez, K. W. Mahaffey, H. Hedlin, J. S. Rumsfeld, A. Garcia, T. Ferris, V. Balasubramanian, A. M. Russo, A. Rajmane, L. Cheung, G. Hung, J. Lee, P. Kowey, N. Talati, D. Nag, S. E. Gummidipundi, A. Beatty, M. T. Hills, S. Desai, C. B. Granger, M. Desai, and M. P. Turakhia, "Large-scale assessment of a smartwatch to identify atrial fibrillation," *New England Journal of Medicine*, vol. 381, no. 20, pp. 1909–1917, 2019, pMID: 31722151. [Online]. Available: https://doi.org/10.1056/NEJMoa1901183

[6] G. Boriani, R. B. Schnabel, J. S. Healey, R. D. Lopes, N. Verbiest-van Gurp, T. Lobban, J. A. Camm, and B. Freedman, "Consumer-led screening for atrial fibrillation using consumer-facing wearables, devices and apps: A survey of health care professionals by af-screen international collaboration," *European Journal of Internal Medicine*, vol. 82, pp. 97–104, 2020. [Online]. Available: https://doi.org/10.1016/j.ejim.2020.09.005

[7] G. H. Mairesse and H. Heidbüchel, "Consumer-led screening for atrial fibrillation: What is the next step?" *European Journal of Internal Medicine*, vol. 90, pp. 16–18, 2021. [Online]. Available: https://doi.org/10.1016/j.ejim.2021.05.030

[8] J. R. Baman, D. T. Mathew, M. Jiang, and . Passman, R, "Mobile health for arrhythmia diagnosis and management," *Journal of General Internal Medicine*, no. 37, pp. 188–197, 2022. [Online]. Available: https://doi.org/10.1007/s11606-021-07007-w

[9] B. Freedman, J. Camm, H. Calkins, J. S. Healey, M. Rosenqvist, J. Wang, C. M. Albert, C. S. Anderson, S. Antoniou, E. J. Benjamin, G. Boriani, J. Brachmann, A. Brandes, T.-F. Chao, D. Conen, J. Engdahl, L. Fauchier, D. A. Fitzmaurice, L. Friberg, B. J. Gersh, D. J. Gladstone, T. V. Glotzer, K. Gwynne, G. J. Hankey, J. Harbison, G. S. Hillis, M. T. Hills, H. Kamel, P. Kirchhof, P. R. Kowey, D. Krieger, V. W. Y. Lee, L.-A. Levin, G. Y. H. Lip, T. Lobban, N. Lowres, G. H. Mairesse, C. Martinez, L. Neubeck, J. Orchard, J. P. Piccini, K. Poppe, T. S. Potpara, H. Puererfellner, M. Rienstra, R. K. Sandhu, R. B. Schnabel, C.-W. Siu, S. Steinhubl, J. H. Svendsen, E. Svennberg, S. Themistoclakis, R. G. Tieleman, M. P. Turakhia, A. Tveit, S. B. Uittenbogaart, I. C. V. Gelder, A. Verma, R. Wachter, B. P. Yan, A. A. Awwad, F. Al-Kalili, T. Berge, G. Breithardt, G. Bury, W. Caorsi, N. Chan, S. Chen, I. Christophersen, S. Connolly, H. Crijns, S. Davis, U. Dixen, R. Doughty, X. Du, M. Ezekowitz, M. Fay, V. Frykman, M. Geanta, H. Gray, N. Grubb, A. Guerra, J. Halcox, R. Hatala, H. Heidbuchel, R. Jackson, L. Johnson, S. Kaab, K. Keane, Y. Kim, G. Kollios, M. Lochen, C. Ma, J. Mant, M. Martinek, I. Marzona, K. Matsumoto, D. McManus, P. Moran, N. Naik, T. Ngarmukos, D. Prabhakaran, D. Reidpath, A. Ribeiro, A. Rudd, I. Savalieva, R. Schilling, M. Sinner, S. Stewart, N. Suwanwela, N. Takahashi, E. Topol, S. Ushiyama, N. V. van Gurp, N. Walker, and T. Wijeratne, "Screening for atrial fibrillation," *Circulation*, vol. 135, no. 19, pp. 1851–1867, 2017. [Online]. Available: https://doi.org/10.1161/CIRCULATIONAHA.116.026693

[10] M. V. McConnell, M. P. Turakhia, R. A. Harrington, A. C. King, and E. A. Ashley, "Mobile

health advances in physical activity, fitness, and atrial fibrillation: Moving hearts," *Journal of the American College of Cardiology*, vol. 71, no. 23, pp. 2691–2701, 2018. [Online]. Available: https://doi.org/10.1016/j.jacc.2018.04.030

[11] N. Brasier, C. J. Raichle, M. Dörr, A. Becke, V. Nohturfft, S. Weber, F. Bulacher, L. Salomon, T. Noah, R. Birkemeyer, and J. Eckstein, "Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO)," *EP Europace*, vol. 21, no. 1, pp. 41–47, 2018. [Online]. Available: https://doi.org/10.1093/europace/euy176

[12] M. Weiser, "The computer for the 21st century," *ScientificAmericanUbicompPaperafterSciAmediting*, vol. 265, no. 3, pp. 94–104, 2011. [Online]. Available: https://bit.ly/3uYsmiU

[13] X. Ye, Y. Huang, and Q. Lu, "Explainable prediction of cardiac arrhythmia using machine learning," in *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2021, pp. 1–5. [Online]. Available: https://doi.org/10.1109/CISP-BMEI53629.2021.9624213

[14] K. Mc Namara, H. Alzubaidi, and J. K. Jackson, "Cardiovascular disease as a leading cause of death: how are pharmacists getting involved?" *Integrated pharmacy research & practice*, vol. 8, pp. 1–11, Feb. 2019. [Online]. Available: https://doi.org/10.2147/iprp.s133088

[15] J. Bao, "Multi-features based arrhythmia diagnosis algorithm using xgboost," in *2020 International Conference on Computing and Data Science (CDS)*, 2020, pp. 454–457. [Online]. Available: https://doi.org/10.1109/CDS49703.2020.00095

[16] G. Silveri, M. Merlo, L. Restivo, B. De Paola, A. Miladinović, M. Ajčević, G. Sinagra, and A. Accardo, "Identification of ischemic heart disease by using machine learning technique based on parameters measuring heart rate variability," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1309–1312. [Online]. Available: https://doi.org/10.23919/Eusipco47968.2020.9287800

[17] X. Wu, Y. Zheng, C.-H. Chu, and Z. He, "Extracting deep features from short ecg signals for early atrial fibrillation detection," *Artificial Intelligence in Medicine*, vol. 109, p. 101896, 2020. [Online]. Available: https://doi.org/10.1016/j.artmed.2020.101896

[18] D. Kasper, A. Fauci, S. Hauser, D. Longo, J. Jameson, and J. Loscalzo, *Harrison's principles of internal medicine, 19th ed.* Mc Graw Hill, 2014. [Online]. Available: https://bit.ly/3hqHin8

[19] I. Goldenberg, R. Goldkorn, N. Shlomo, M. Einhorn, J. Levitan, R. Kuperstein, R. Klempfner, and B. Johnson, "Heart rate variability for risk assessment of myocardial ischemia in patients without known coronary artery disease: The HRV-DETECT (heart rate variability for the detection of myocardial ischemia) study," *Journal of the American Heart Association*, vol. 8, no. 24, p. e014540, Dec. 2019. [Online]. Available: https://doi.org/10.1161/jaha.119.014540

[20] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144–164, 2016. [Online]. Available: https://doi.org/10.1016/j.cmpb.2015.12.008

[21] H. Zhu, Y. Zhao, Y. Pan, H. Xie, F. Wu, and R. Huan, "Robust heartbeat classification for wearable Single-Lead ECG via extreme gradient boosting," *Sensors (Basel)*, vol. 21, no. 16, Aug. 2021. [Online]. Available: https://doi.org/10.3390/s21165290

[22] S. Bhalerao, I. A. Ansari, and A. Kumar, "Reversible ecg data hiding: Analysis and comparison of ann, regression svm and random forest regression," in *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 0667–0671. [Online]. Available: https://doi.org/10.1109/ICCSP48568.2020.9182219

[23] M. Manjula and A. Sarma, "Comparison of empirical mode decomposition and wavelet based classification of power quality events," *Energy Procedia*, vol. 14, pp. 1156–1162, 2012. [Online]. Available: https://doi.org/10.1016/j.egypro.2011.12.1069

[24] S. Murawwat, H. M. Asif, S. Ijaz, M. Imran Malik, and K. Raahemifar, "Denoising and classification of arrhythmia using memd and ann," *Alexandria Engineering Journal*, vol. 61, no. 4, pp. 2807–2823, 2022. [Online]. Available: https://doi.org/10.1016/j.aej.2021.08.014

[25] M. Chandra Gaddam and S. Pattnaik, "An ann ensemble based ecg signal classification approach for accurate arrhythmia detection," *International Journal of Emerging Technology and Advanced Engineering*, vol. 10, pp. 57–61, 2020. [Online]. Available: http://dx.doi.org/10.46338/IJETAE0820_08

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[27] T. Kanai, N. Tanabe, Y. Miyagi, and J. Aoyama, "Cnn-type myocardial infarction prediction based on cardiac cycle determination," in *2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2021, pp. 1–2. [Online]. Available: https://doi.org/10.1109/ISPACS51563.2021.9651000

[28] A. Escontrela. (2020) Convolutional neural networks from the ground up. [Online]. Available: https://bit.ly/2EXtsnf

[29] M. Dey, N. Omar, and M. A. Ullah, "Temporal feature-based classification into myocardial infarction and other cvds merging cnn and bi-lstm from ecg signal," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21 688–21 695, 2021. [Online]. Available: https://doi.org/10.1109/JSEN.2021.3079241

[30] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *NAACL*, vol. abs/1810.04805, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1810.04805

[31] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *CoRR*, vol. abs/2106.03253, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2106.03253

[32] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. abs/1603.02754, 2016. [Online]. Available: https://doi.org/10.1145/2939672.2939785

[33] R. Shwartz-Ziv, A. Painsky, and N. Tishby, "Representation compression and generalization in deep neural networks," in *ICLR 2019 Conference Blind Submission*, 2018. [Online]. Available: https://bit.ly/3YjzJz0

[34] T. Poggio, A. Banburski, and Q. Liao, "Theoretical issues in deep networks," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 039–30 045, 2020. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1907369117

[35] Z. Piran, R. Shwartz-Ziv, and N. Tishby, "The dual information bottleneck," *CoRR*, vol. abs/2006.04641, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2006.04641

[36] A. V. Dorogush, A. Gulin, G. Gusev, N. Kazeev, L. O. Prokhorenkova, and A. Vorobev, "Fighting biases with dynamic boosting," *CoRR*, vol. abs/1706.09516, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1706.09516

[37] H. Shi, H. Wang, Y. Huang, L. Zhao, C. Qin, and C. Liu, "A hierarchical method based on weighted extreme gradient boosting in ecg heartbeat classification," *Computer Methods and Programs in Biomedicine*, vol. 171, pp. 1–10, 2019. [Online]. Available: https://doi.org/10.1016/j.cmpb.2019.02.005

[38] Z. Yue and Z. Jinjing, "Atrial fibrillation detection based on eemd and xgboost," *Journal of Physics: Conference Series*, vol. 1229, no. 1, p. 012074, 2019. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1229/1/012074

[39] B. R. Manju and A. R. Nair, "Classification of cardiac arrhythmia of 12 lead ecg using combination of smoteenn, xgboost and machine learning algorithms," in *2019 9th International Symposium on Embedded Computing and System Design (ISED)*, 2019, pp. 1–7. [Online]. Available: https://doi.org/10.1109/ISED48680.2019.9096244

[40] N. V. Chawla, K. W. Bowyer, L. O.Hall, and W. P. Kegelmeyer, "Smote: synthetic minority oversampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. [Online]. Available: https://doi.org/10.1613/jair.953

[41] F. Giannakas, C. Troussas, A. Krouska, C. Sgouropoulou, and I. Voyiatzis, "Xgboost and deep neural network comparison: The case of teams' performance," in *Intelligent Tutoring Systems*, A. I. Cristea and C. Troussas, Eds. Springer International Publishing, 2021, pp. 343–349. [Online]. Available: https://doi.org/10.1007/978-3-030-80421-3_37

[42] P. D. Arini and E. R. Valverde, "Beat-to-beat electrocardiographic analysis of ventricular repolarization variability in patients after myocardial infarction," *Journal of electrocardiology*, vol. 49, no. 2, pp. 206–213, Dec. 2015. [Online]. Available: https://doi.org/10.1016/j.jelectrocard.2015.12.003

[43] W. Liu, F. Wang, Q. Huang, S. Chang, H. Wang, and J. He, "MFB-CBRNN: A hybrid network for MI detection using 12-lead ECGs," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 503–514, Apr. 2019. [Online]. Available: https://doi.org/10.1109/jbhi.2019.2910082

[44] M. R. Rajeshwari and K. S. Kavitha, "Arrhythmia ventricular fibrillation classification on ECG signal using ensemble feature selection and deep

neural network," *Cluster Computing*, vol. 25, no. 5, pp. 3085–3102, 2022. [Online]. Available: https://doi.org/10.1007/s10586-022-03547-w

[45] H. M. Rai and K. Chatterjee, "Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data," *Applied Intelligence*, vol. 52, no. 5, pp. 5366–5384, 2022. [Online]. Available: https://doi.org/10.1007/s10489-021-02696-6

[46] B. Król-Józaga, "Atrial fibrillation detection using convolutional neural networks on 2-dimensional representation of ecg signal," *Biomedical Signal Processing and Control*, vol. 74, p. 103470, 2022. [Online]. Available: https://doi.org/10.1016/j.bspc.2021.103470