Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 203 (2022) 231–238

Procedia Computer Science

www.elsevier.com/locate/procedia

The 19th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
August 9-11, 2022, Niagara Falls, Canada

Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability

Mohammad Moshawrab[a,c,*], Mehdi Adda[a], Abdenour Bouzouane[b], Hussein Ibrahim[c], Ali Raad[d]

[a]*Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, G5L 3A1, Québec, Canada*

[b]*Département D'informatique et de Mathématique, Université du Québec à Chicoutimi, Chicoutimi, 555 boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada*

[c]*Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, G4R 5B7, Québec, Canada*

[d]*Dean of the Faculty of Science and Arts, Islamic University of Lebanon, Wardaniyeh, Lebanon*

Abstract

Artificial Intelligence is exponentially evolving into a solution to many of humanity's complex problems. In this context, healthcare is benefiting from this technology and all its branches to improve the level of services offered, including cardiac health services. Cardiovascular diseases have always been among the most common and deadly diseases around the world, as studies have consistently shown. However, Artificial Intelligence services offer several tools to improve the diagnosis of these diseases and even predict their occurrence. In this study, four models are created and trained with "PhsyioNet Smart Health for Assessing the Risk of Events via ECG Database" to analyze the characteristics of heart rate variability and predict the occurrence of heart diseases and cerebrovascular events. The results obtained support the confidence in the use of Artificial Intelligence in cardiology, where Support Vector Machines, Deep Neural Networks, and XGBoost achieved an accuracy of 91.80%, 90.19%, and 89.10%, respectively.

© 2022 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the Conference Program Chairs.

Keywords: Artificial Intelligence, Machine Learning, Diseases Prediction, Cardiovascular Diseases, Heart Rate Variability

1. Introduction

Cardiovascular Diseases (CVDs) cause the most deaths and are therefore known as the most dangerous disease worldwide. According to the latest figures from the World Health Organization (WHO) in the field of cardiovascular diseases, the number of deaths caused by them increased from 12.1 million to 18.6 million between 1990 and 2019, with deaths accounting for 32% of global mortality in 2019. Moreover, cardiovascular diseases are not only a major cause of health conflict, but also of economic burden. According to "Medical Expenditure Panel Survey," the costs

* Corresponding author; Tel.: +1-581-624-9394
E-mail address: mohammad.moshawrab@uqar.ca

1877-0509 © 2022 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2022.07.030

due to CVDs were estimated to be $378.0 billion in the United States alone between 2017 and 2018, including $226.0 billion in expenditures and $151.8 billion in lost future productivity [1,2].

1.1. Artificial Intelligence in Healthcare: A New Cardiology Era

The deadly cardiovascular diseases urge to find efficient solutions that can help in the early diagnosis of these diseases and, if possible, even predict their occurrence. Traditional methods to detect these diseases include electrocardiogram, echocardiography, coronary angiography, stress test, magnetic resonance imaging or intracoronary ultrasound. However, technological developments, especially Information and Communication Technologies (ICT), and the rise of Artificial Intelligence (AI) and its variants are helping to improve the quality of healthcare services and thus facilitate the diagnosis of CVDs. Moreover, AI tools are considered the next revolution in cardiology as they help provide faster and more accurate patient care outcomes. Moreover, AI will soon transform the science of heart health, as its tools could outperform experts in diagnosing or even predicting CVDs [3,4].

1.2. Heart Rate Variability as a CVD Indicator

Recently, interest in the use of heart rate variability (HRV) as an indicator of cardiovascular diseases has increased, especially with the development of AI and the data analysis capabilities offered by its branches: Machine Learning and Deep Learning. Moreover, HRV is known as the beat-to-beat variation in heart rate or the duration of the RR peak interval, where R is a wave of the QRS complex extracted from a cardiac ECG signal. Knowing that changes in the autonomic regulation of the heart can be read from the temporal variations in heart rate, the parameters extracted from HRV data are divided into three main categories: Time domain, Frequency Domain and Non-Linear parameters. These categories are listed in Table 1 below [5].

Table 1. Heart Rate Variability Parameters.

Group	Parameter	Unit	Description
Time Domain Features	Mean NN	(ms)	Mean of NN interval
	SDNN	(ms)	Standard deviation of NN intervals
	RMSSD	(ms)	Square root of the mean squared differences of successive NN intervals
	pNN50	(ms)	Proportion of interval differences of successive NN intervals greater than 50 ms
Frequency Domain Parameters	VLF	(ms^2)	Power in very low frequency range (0–0.04 Hz)
	LF	(ms^2)	Power in low frequency range (0.04–0.15 Hz)
	HF	(ms^2)	HF ms2 Power in high frequency range (0.15–0.4 Hz)
	LF/HF	(ratio)	Ratio of LF over HF
Non-Linear Parameters	SD1	(ms)	Standard deviation of points perpendicular to the axis of line of identity or the successive intervals scaled by $\sqrt{\frac{1}{2}}\ \sqrt{\frac{1}{2} var(RR_{\mathrm{n}} - RR_{\mathrm{n+1}})}$
	SD2	(ms)	Standard deviation of points along the axis of line of identity, or $\sqrt{2SDNN^2 - \frac{1}{2}SD1^2}$
	SD1/SD2	(ratio)	Ratio of SD1 over SD2

1.3. Prediction of CVDs with HRV; State of the Art

The number of studies on CVDs detection using HRV parameters is increasing rapidly. Researchers are using different AI models to analyze various HRV parameters, and AI has proven its efficiency and accuracy in this domain. For example, in[6], the authors used the Fast Fourier Transform (FFT) with the Blackman Harris window algorithm to build a model that analyzes various HRV features to predict the occurrence of Ventricular Tachycardia (VT) in the short term. In addition, the authors developed an Artificial Neural Networks (ANN) classifier in [7] and trained it with the "PhysioNet Spontaneous Ventricular Tachyarrhythmia Database" [8] to predict the occurrence of VT. They measured performance using several metrics, recording 76.60% 82.9% and 71.4% for accuracy, sensitivity, and specificity, respectively. In addition, the authors in [9] used Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machines (SVM) to predict cardiovascular risk. Their best model achieved 96.67% accuracy. In addition, the authors in [10] used SVM to develop a predictive model to predict cardiovascular risk after Myocardial Infarction, and their model accuracy was 89%.

Moreover, the authors of [11] created models to predict Sudden Cardiac Death (SCD) using the k-Nearest Neighbor and Multilayer Perceptron Neural Network (MLP) algorithms. They trained their models using the "PhysioNet Sudden Cardiac Death Holter database" [12] and the "PhysioNet Normal Sinus Rhythm database" [13], and their recorded performance measures were 99.73%, 96.52%, 90.37%, and 83.96% accuracy for the first, second, third, and fourth one-minute intervals, respectively. Furthermore, in [14], the authors did the same with SVM and Probabilistic Neural Network (PNN) to predict SCD two minutes before its onset. Similarly, the authors trained their models using the "PhysioNet Sudden Cardiac Death (SCD) Holter database" [12] and the "PhysioNet MIT Normal Sinus Rhythm database" [13] and SVM and PNN recorded prediction rates of 96.36% and 93.64%, respectively.

On the other hand, in [15], the authors targeted hypertensive patients and developed novel SVM, Tree-Based Classifier, Artificial Neural Network and Random Forest models to create an automated cardiovascular risk stratification model. The authors trained their data using the "Smart Health for Assessing the Risk of Events via ECG database" [16] and achieved a sensitivity of 71.4% and a specificity of 87.8%. Furthermore, in [17], the authors developed an Artificial Neural Networks model that analyzes respiratory rate in addition to HRV features to detect Ventricular Tachycardia one hour before its onset. The performance metrics of their model were 88%, 82%, and 93% for sensitivity, specificity, and area under the curve, respectively. In addition, the authors in [18] used a statistical model called MIL, to predict CVDs based on features of heart rate variability. Their model achieved high accuracy, as they mentioned. Finally, in [19], the authors created K Nearest Neighbor (k-NN), Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, Neural Network, and Vote and trained them with the "UCI Heart Diseases Repository" [20]. The models created were able to predict CVDs with 87.4% accuracy.
In this article, several artificial intelligence models were created to predict Cardiovascular Diseases and events. The models used are: Support Vector Machine (SVM), Deep Neural Networks (DNN), XGBoost, and Neural Oblivious Decision Ensembles (NODE). Section 2 below explains the dataset used in this study and the preprocessing steps used to prepare the data for the models. Section 3 explains the models created and the results obtained with these models are listed and discussed in Section 4.

2. Materials & Methods

2.1. Dataset

The dataset used in this study is the "PhysioNet Smart Health for Assessing the Risk of Events via ECG Database" (SHAREEDB) [16] that is offered by the PhysioNet online data repository. This dataset was collected to investigate the efficiency of classifying hypertensive patients at higher risk for cardiac and cerebrovascular events using heart rate variability characteristics. It consists of 139 records of 24-hour Electrocardiographic (ECG) Holter recordings. Each recording contains three ECG signals sampled at a rate of 128 samples per second with a precision of 8 bits. The population in which the data were collected consisted of 49 women and 90 men aged 55 years and older. They were followed up for 12 months to record the occurrence of serious cardiovascular and cerebrovascular events such as Coronary Revascularization, fatal or nonfatal Acute Coronary Syndromes, syncopal events, Myocardial Infarctions, fatal or nonfatal strokes, and Transient Ischemic Attacks. During the follow-up period, 17 patients experienced a cardiovascular event, including 11 Myocardial Infarctions, 3 strokes, and 3 syncopal events. In addition, the dataset contains some demographic and clinical information about the subjects, such as their age, sex, any vascular events, values of systolic and diastolic arterial pressure, and others.

2.2. Data Filtering & Preprocessing

The ECG signals provided by the SHAREEDB dataset are collected in laboratories and may be susceptible to a lot of noise that needs to be removed before the data is passed to the AI models. It is very important to clean the data and remove the noise to obtain high-quality ECG signals that are then analyzed by the models. The data cleaning and preparation steps used in this study are summarized below:

- **Filtering & Artifacts Removal [21]:** ECG recordings are susceptible to noise or interference from various signals, which can be divided into high and low-frequency noise sources. For example, noise can be caused by electrode interference, muscle motion interference, channel interference, baseline drift, or power line interference. Therefore, ECG signals can be cleaned by using the following filters:
  - **IIR Notch Filters:** remove motion artifacts and/or power line interference
  - **FIR Filters:** clean ECG data and are act on the range of ECG data that is between 1 and 100 hertz
- **R Peaks Detection [22,23]:** The ECG signal reflects the electrical activity of the myocardium and is divided into three distinct parts: the P wave, the QRS complex, and the T wave. However, the QRS complex is composed of Q-wave, R-wave and S-wave. The R-peak is the interval between the onset of the QRS complex and the peak of the R-wave and can be determined using various algorithms such as Hamilton, Christov, Engelse and Zeelenberg, Pan and Tompkins, Stationary Wavelet Transform and Two Moving Average. According to the results in [23], Engelse and Zeelenberg provided the best results in detecting R-peaks, which is why they were used in this study
- **Calculation of RR Intervals:** Heart Rate Variability is defined as the RR intervals or the difference between two consecutive R peaks, which are then calculated using the required equations

- **Outliers Removal:** After the RR intervals are detected, the outliers, defined as points that are extremely far from the mean, are removed and replaced with the mean value
- **Extract HRV features:** Finally, the HRV features were calculated using the appropriate mathematical formulas. In this study, 26 HRV features were calculated, and despite the high number of features calculated, the use of all features gave good results

2.3. Artificial Intelligence Models

Cardiology is defined as the healthcare sector that takes care of heart health, and the use of AI in this field is growing briskly. AI has demonstrated high accuracy and efficiency in detecting CVDs, and sometimes it can go beyond professional diagnosis and even be used in predicting cardiovascular diseases instead of detecting them due to its high ability to analyze cardiac data [24,25]. In addition, AI is known for its various branches that are used in different areas of life around the world. For example, Machine Learning, Ensemble Learning and Deep Convolutional Neural Networks are AI branches that were used in this study:

- **Classical Machine Learning Algorithms[26]:** are algorithms that give computers learning potential by training them with experimental data and generating models based on these data, enabling them to make decisions in new situations such as: Support Vector Machines, Naïve Bayes, Logistic and Linear Regression and others.
- **Ensemble Learning [27]:** is a special branch of ML where its algorithms are based on merging predictions from different models. Some of these models are XGBoost, AdaBoost, GradientBoosting, LightGBM and others.
- **Deep Convolutional Neural Network (DCNNs) [26]:** are a type of Neural Networks that are used to analyze data with a grid-like structure. However, these networks are intended for analyzing multidimensional data such as images and videos. Using these networks to analyze tabular data may require transforming the data used. Nevertheless, there are several models that offer transformation of tabular data for use in DCNNs, such as TabNet, GrowNet, TreeEnsemble Layers, TabTransformers, Self Normalizing Neural Networks, Neural Oblivious Decision Ensembles (NODE), AutoInt, and Deep & Cross Neural Networks (DCNs) [28].

3. Construction of AI Models

In this study, different AI models were used to analyze HRV features to detect heart diseases and events. However, before passing the extracted features to the models, some data fitting steps should be performed, as explained below.

3.1. Data Adjustment

Considering that of the study population, 139 patients, only 17 developed a cardiovascular event in the 12-month follow up period, the extracted HRV features show an unbalanced identity, with the majority falling into the ”no cardiovascular event” class. Because the proportion of this class is 122 of 139, the performance of the prediction models may be negatively affected, suggesting the application of some data adjustments such as balancing and scaling:

- **Synthetic Minority Over-sampling Technique (SMOTE):** a data expansion in which new samples are drawn from existing ones to oversample the minority class
- **Preprocessing Standard Scaling:** the standardization of characteristics is achieved by removing the mean of the data and scaling it to a unit variance

3.2. Building the Models; hyperparameters to be considered

After applying the necessary data fitting steps to the extracted HRV features, they are then passed to the models created for fitting with the thresholds listed below:

3.2.1. Classical ML: Support Vector Machines

SVM is a supervised Machine Learning algorithm that is fed labeled training data to learn how to assign labels to objects based on examples, and then gain the ability to predict the category of new example(s) [26]. The performance of the SVM model is affected by the following hyperparameters [29]:

- **Kernel:** the function that converts the input data into the required form
- **Regularization:** denotes the misclassification or error term and is expressed as hyperparameter ”C”.
- **gamma:** interpret how far the effect of a single training sample extends
- **class weight:** used for imbalanced datasets and defines the weight of the classes to be predicted

3.2.2. Deep Learning: Deep Neural Networks

These networks are algorithms that mimic human brain cells called neurons. In general, these networks use brain simulations to improve their learning and increase the accuracy of the models. The structure of DNNs consists of more than two interconnected layers and is affected by the following hyperparameters [30]:

- **Number of layers:** input, output, and the hidden layers that define the structure of the network.
- **Units:** denotes the output of each layer.
- **Activation function:** also known as the "transfer function", which defines how the weighted sum of the input is converted into an output from one or more nodes in a layer of the network
- **Number of epochs:** a complete pass through all rows of the training data
- **Batch size:** samples that the model examines within each epoch before updating the weights.
- **Learning rates:** a variable that controls how the optimizer's learning rate changes over time
- **Momentum:** is the "delay" in learning the mean and variance

3.2.3. Ensemble Learning Algorithms: XGBoost

XGBoost is an Ensemble Learning algorithm that also belongs also to the Machine Learning AI Branch.

- **XGBoost [31]:** eXtreme Gradient Boosting package is a scalable implementation of the gradient boosting framework built with an efficient linear model solver and a tree learning algorithm with hyperparameters:
  - **Booster:** the type of model to run at each iteration
  - **Learning Rate:** is the step size shrinkage used during the update to prevent overfitting
  - **Gamma:** specifies the minimum loss reduction required to perform splitting
  - **Max Depth:** the parameter used to control overfitting
  - **Min Child Weight:** defines the minimum sum of weights of all observations required in a child
  - **Max Delta Step:** helps to make the update step more conservative
  - **Sub Sample:** denotes the fraction of observations that are randomly selected for each tree
  - **Lambdas:** is used to handle the regularization part
  - **Alpha:**is used in case of very high dimensionality to make the algorithm run faster during implementation
  - **Tree Method:** Algorithm for tree construction
  - **Scale Weight:** controls the balance of positive and negative weights
  - **Objective:** defines the loss function to be minimized

3.2.4. Deep Convolutional Neural Networks: Neural Oblivious Decision Ensembles

In this study, the following model was used to apply DCNN to the SHAREEDB tabular data:

- **Neural Oblivious Decision Ensembles (NODE)[32]:** a model with a layered structure built from differentiable oblivious trees, which are decision tables that decompose the data along dd-splitting features and compare each feature to a learned threshold. It was trained in an end-to-end manner using backpropagation and is affected by the following hyperparameters:
  - **Number of Layers:** Number of layers forming the Neural Network
  - **Number of Trees:** Number of trees in each layer
  - **Depth:** Depth of the tree
  - **Learning Rate:** is the shrinkage step size used in the update to prevent overfitting.

3.3. Wrapping Up, Training, Prediction, and Optimization

Once the models were created, they were trained with the fitted version of the extracted HRV features. The obtained results are explained and discussed in detail in Section 4. Figure 1 shows the overall architecture of the data preparation steps and the models created in this study.

4. Results & Discussion

The created models were trained with the HRV features. The SVM, DNN, XGBoost, and NODE models were evaluated with the metrics of Accuracy, Precision, Recall, Specificity, Negative Predictive Value NPV, and F1 Score. For better measurement, Repeated K-fold Cross Validation [33] was implemented with 10 folds and repeated 5 times. The results are shown in Table 2 below, and the values of accuracy, precision, recall, specificity, negative predictive value, and F1 score are denoted as AC, PR, RE, SP, NPV, and F1 , respectively. In addition, the values of the hyperparameters used are listed in the table. Figure 2 below also shows a graphical representation of the performance of the models created in this study.

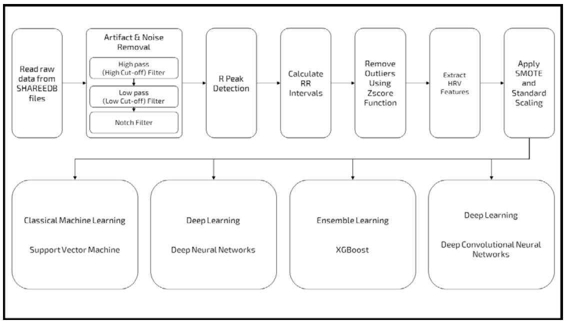

Fig. 1. Overall Architecture Followed in this Study

Table 2. AI Models Evaluation Metrics.

Model	Hyper Parameters		Evaluation Metrics					
	Parameter	Value	AC	PR	RE	SP	NPV	F1
SVM	Training(Testing)	0.75(0.25)	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
	Kernel	rbf						
	Regularization(C)	2.66						
	Gamma	0.141						
DNN	Training(Testing)	0.79(0.21)	90.19%	85.18%	95.83%	85.18%	95.83%	90.19%
	Layers	Input/3 Hidden/Output						
	Units	512/256/128/64/1						
	Activation Function	tanh/tanh/tanh/sigmoid						
	Dropout	Before Output Layer 0.2						
	Optimizer	SGD						
	Epochs	6850						
	Batch Size	250						
	Learning Rate	0.005						
	Momentum	default						
XGBoost	Training(Testing)	0.79(0.21)	89.10%	86.00%	93.80%	85.10%	92.50%	89.10%
	Booster	gbtree						
	Learning Rate	0.01						
	Gamma	0.1						
	Maximum Depth	10						
	Minimum Child Weight	0.01						
	Max Delta Step	0						
	Sub Sample	0.75						
	Lambda	1						
	Alpa	0.01						
	Tree Method	Auto						
NODE	Training(Testing)	0.71(0.29)	76.92%	77.77%	73.68%	80%	76.19%	75.67%
	Number of Layers	5						
	Depth	10						
	Number of Trees	1						
	Learning Rate	0.1						
	Batch Size	26						

4.1. Discussion

In this study, several models were created to analyze HRV characteristics to detect cardiovascular risks. The results obtained demonstrate the high efficiency of AI models in predicting cardiovascular disease. However, the results obtained in this study outperformed previous implementations.

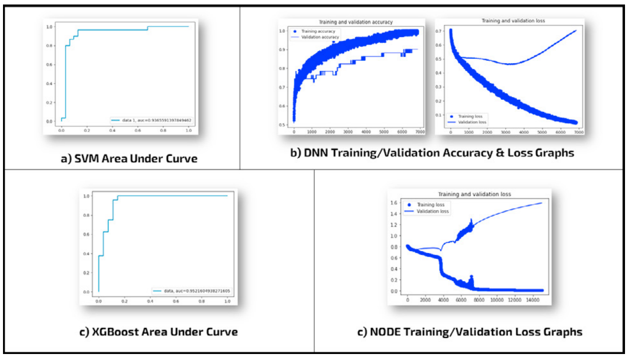

Fig. 2. Models Performance Graphical Representation

First, the authors in [15]applied similar models to the same dataset. Nevertheless, the results obtained in this study exceeded their results. For example, their SVM model recorded accuracy, recall and specificity results were 89.00%, 86.30% and 91.80% respectively, whereas our results are 91.80%, 96.66% and 87.09% for the same performance metrics. In addition, the performance metrics of their Multi Layer Perceptron (MLP) model were Accuracy: 78.10%, Recall: 86.30%, Specificity: 69.90% and our model recorded 90.19%, 95.83% and 85.18% for the same metrics.
In addition, our SVM model achieved 91.80% accuracy, the highest performance among all previous implementations.

For example, the SVM models in [9] recorded an accuracy of 88.64%, 82.95% and 82.58% for the Linear, Polynomial and RBF kernels, respectively. Moreover, the accuracy of SVM in [10,15,19] was 79.81%, 89.00% and 85.19%, respectively. Even though the accuracy is close, other metrics such as Precision and Recall clearly outperform the previous results by a large margin. Knowing that Recall measures how a model correctly classifies True Positives, the models presented in this study are more accurate in predicting whether a person will have a cardiovascular disease in future. The high recall for SVM, DNN, and XGBoost, which are 96.66%, 95.83% and 93.80%, respectively, reflects the highest ability of all implementations to correctly predict that a person is in the cardiovascular risk zone.
Likewise, the DNN model presented in this article also outperforms all previous implementations. The accuracy of this model is 90.19%, whereas the multilayer perceptron in [9,15] is 86.67% and 78.10%, and the accuracy of artificial neural networks in [7,17] is 76.60% and 85.30%. Moreover, precision and recall are significantly higher than the previous implementations, which also reflects a higher capability in cardiovascular risk detection. Table 3 provides a detailed comparison between the results of the models presented in this article and the previous implementations. The sybmoles of the performance metrics used in this table are similar to those in Table 2, and a ”NA” symbol indicates that the corresponding metric was not mentioned in the associated study.
On the other hand, none of the previous implementations used XGBoost , which also outperformed the previous implementations with an accuracy of 98.10% and a recall of 94.60%, reflecting high efficiency in predicting cardiovascular risk, in contrast to the implementation of NODE, which achieved an accuracy of 76.92%, which is not comparable with the previous implementations.
Finally, the SVM, DNN, and XGBoost models discussed in this study can be considered the most accurate models for predicting cardiac disease and events. Even the implementations in [11,14] had higher accuracy and relatively higher recall, but their models were developed to detect sudden cardiac death (SCD) only minutes before its occurrence. For example, the model mentioned in [11] achieved 99.73% accuracy in predicting sudden cardiac death one minute before its onset, but the performance drops to 83.93% when the event is predicted four minutes before its occurrence. However, the models presented here are able to predict cardiovascular disease 12 months before its onset, demonstrating high efficiency in predicting cardiovascular disease and cardiac events long before their onset, thus increasing confidence in the use of AI in detecting and predicting cardiac disease and related events.

Table 3. Comparison with Previous Implementations.

Study	Model	AC	PR	RE	SP	NPV	F1
Our Study	Support Vector Machines	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
	DNN	90.19%	85.18%	95.83%	85.18%	95.83%	90.19%
	XGBoost	89.10%	86.00%	93.80%	85.10%	92.50%	89.10%
	NODE	76.92%	77.77%	73.68%	80.00%	76.00%	75.67%
[7]	Artificial Neural Network	76.60%	70.70%	82.90%	71.40%	NA	NA
[9]	Support Vector Machines (Linear Kernel)	88.64%	90.84%	86.36%	90.91%	86.96%	NA
	Support Vector Machines (Polynomial Kernel)	82.95%	80.85%	79.55%	86.36%	85.37%	NA
	Support Vector Machines (RBF Kernel)	82.58%	79.45%	77.27%	87.88%	86.44%	NA
	Multi Layer Perceptron (Top 15 Features)	86.67%	100%	73.33%	100%	78.95%	NA
[10]	Support Vector Machines	79.81%	21.15%	91.67%	79.08%	99.36%	NA
[11]	MLP (A Minute Before the SCD Event)	99.73%	NA	NA	NA	NA	NA
	K-NN (A Minute Before the SCD Event)	98.32%	NA	NA	NA	NA	NA
[14]	SVM (2 minutes before VF Event)	96.36%	NA	NA	NA	NA	NA
	Penalized Neural Network	93.64%	NA	NA	NA	NA	NA
[15]	Support Vector Machines	89.00%	NA	86.30%	91.80%	NA	NA
	Multi Layer Perceptron	78.10%	NA	86.30%	69.90%	NA	NA
[17]	Artificial Neural Network	85.30%	83.30%	88.20%	82.40%	87.50%	NA
[18]	MIL Statisitcs Algorithm	85.47%	92.11%	86.42%	83.33%	NA	NA
[19]	Vote	87.41%	NA	NA	NA	NA	NA
	Naïve Bayes	84.81%	NA	NA	NA	NA	NA
	Support Vector Machines	85.19%	NA	NA	NA	NA	NA

5. Conclusion

AI will one day be destiny, some have said. But what we are witnessing today through the use of these technologies in many areas of life confirms that they have become a reality and that their use is increasing day by day. Moreover, AI is expected to evolve the concepts of cardiology and the mechanisms to diagnose its diseases, and even use them to predict the occurrence of these diseases in the future. In this study, we have presented a group of models capable of predicting the occurrence of heart diseases or events with high accuracy, which increases the confidence in AI and its branches in the health field. Furthermore, adapting these models to work in real time will certainly help create personalized and continuous monitoring that can be used to track patients' heart health or even monitor the health of workers who work in stressful environments or for extremely long periods of time.

References

[1] Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., ... & GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. Journal of the American College of Cardiology, 76(25), 2982-3021.

[2] Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., ... & American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2022). Heart Disease and Stroke Statistics—2022 Update: A Report From the American Heart Association. Circulation, 145(8), e153-e639.

[3] National Heart Lung and Blood Institute. (2005). In Brief: Your guide to living well with heart disease (NIH Publication 06-5716). Washington, DC: National Institutes of Health.

[4] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. Journal of the American College of Cardiology, 71(23), 2668-2679.

[5] Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. Medical and biological engineering and computing, 44(12), 1031-1051.

[6] Baumert, M., Wessel, N., Schirdewan, A., Voss, A., & Abbott, D. (2007). Forecasting of ventricular tachycardia using scaling characteristics and entropy of heart rate time series. In World Congress on Medical Physics and Biomedical Engineering 2006 (pp. 1001-1004). Springer, Berlin, Heidelberg.

[7] Joo, S., Choi, K. J., & Huh, S. J. (2010, September). Prediction of ventricular tachycardia by a neural network using parameters of heart rate variability. In 2010 Computing in Cardiology (pp. 585-588). IEEE.

[8] Spontaneous Ventricular Tachyarrhythmia Database v1.0. (2007, May 2). PhysioNet. https://physionet.org/content/mvtdb/1.0/

[9] Ramirez-Villegas, J. F., Lam-Espinosa, E., Ramirez-Moreno, D. F., Calvo-Echeverry, P. C., & Agredo-Rodriguez, W. (2011). Heart rate variability dynamics for the prognosis of cardiovascular risk. PloS one, 6(2), e17060.

[10] Song, T., Qu, X. F., Zhang, Y. T., Cao, W., Han, B. H., Li, Y., ... & Da Cheng, H. (2014). Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. BMC cardiovascular disorders, 14(1), 1-8.

[11] Ebrahimzadeh, E., Pooyan, M., & Bijar, A. (2014). A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. PloS one, 9(2), e81896.

[12] Sudden Cardiac Death Holter Database v1.0.0. (2004, July 2). PhysioNet. https://physionet.org/content/sddb/1.0.0/

[13] MIT-BIH Normal Sinus Rhythm Database v1.0.0. (1999, August 3). PhysioNet. https://physionet.org/content/nsrdb/1.0.0/

[14] Murukesan, L., Murugappan, M., Iqbal, M., & Saravanan, K. (2014). Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features. Journal of Medical Imaging and Health Informatics, 4(4), 521-532.

[15] Melillo, P., Izzo, R., Orrico, A., Scala, P., Attanasio, M., Mirra, M., ... & Pecchia, L. (2015). Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. PloS one, 10(3), e0118504.

[16] Smart Health for Assessing the Risk of Events via ECG Database. (2015, May 19). PhysioNet. Retrieved November 1, 2021, from https://physionet.org/content/shareedb/1.0.0/

[17] Lee, H., Shin, S. Y., Seo, M., Nam, G. B., & Joo, S. (2016). Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks. Scientific reports, 6(1), 1-7.

[18] Lan, K. C., Raknim, P., Kao, W. F., & Huang, J. H. (2018). Toward hypertension prediction based on PPG-derived HRV signals: A feasibility study. Journal of medical systems, 42(6), 1-7.

[19] Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, 82-93.

[20] UCI Machine Learning Repository: Heart Disease Data Set. (n.d.-a). UCI Machine Learning Repository: Heart Disease Data Set. https://archive.ics.uci.edu/ml/datasets/heart+disease.

[21] Luo, S., & Johnston, P. (2010). A review of electrocardiogram filtering. Journal of electrocardiology, 43(6), 486-496.

[22] Ashley, E. A., & Niebauer, J. (2004). Cardiology explained.

[23] Eilers, J., Chromik, J., & Arnrich, B. (2021). Choosing the Appropriate QRS Detector. In BIOSIGNALS (pp. 50-59).

[24] Mathur, P., Srivastava, S., Xu, X., & Mehta, J. L. (2020). Artificial intelligence, machine learning, and cardiovascular disease. Clinical Medicine Insights: Cardiology, 14, 1179546820927404.

[25] Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?. Heart, 104(14), 1156-1164.

[26] Ertel, W. (2018). Introduction to artificial intelligence. Springer.

[27] Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14(2), 241-258.

[28] Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems, 34.

[29] Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[30] Anderson, J. A. (1995). An introduction to neural networks. MIT press.

[31] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

[32] Popov, S., Morozov, S., & Babenko, A. (2019). Neural oblivious decision ensembles for deep learning on tabular data. arXiv preprint arXiv:1909.06312.

[33] Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, 21(2), 137-146.