iCardo 3.0: ECG-based Prediction of Conduction Disturbances using Demographic Features

Nidhi Sinha[1*], Amit Joshi[2] and Saraju Mohanty[3]

[1*]Electronics and Communication Engineering, Malaviya National Institute of Technology, JLN Marg, Jaipur, India, 302017.

[2]Electronics and Communication Engineering, Malaviya National Institute of Technology, JLN Marg, Jaipur, India, 302017.

[3]Computer Science and Engineering, University of North Texas, Denton, USA, 76205.

*Corresponding author(s). E-mail(s): 2018rec9033@mnit.ac.in;
Contributing authors: amjoshi.ece@mnit.ac.in; saraju.mohanty@unt.edu;

Abstract

Cardiovascular Disease is one of the most contributing diseases to premature mortality around the globe. Low- and middle-income countries like India account for almost 80% of global CVD fatalities. Predicting cardiovascular diseases at an early stage can improve the quality of life. Electrocardiography (ECG) is one of the non-invasive methods to assess the disorders of heart functioning and hence CVDs. The paper presents the prediction of conduction disturbance or disorders (CD) through a 12-lead electrocardiogram (ECG) that leads to chronic heart failure or cardiac arrest. Three ensemble machine learning models i.e., Random Forest (RF), XG boost, and the Support Vector Machine (SVM), are used to classify the Conduction Disturbance subjects from the 'normal' subjects. In addition to this, the paper also presents a comparative study to show the effect of two demographic features, 'age' and 'sex' on the prediction of Conduction Disturbance's subjects. The performance of the classifiers' is measured in terms of Accuracy, Precision, Recall, and F1 score. 10-fold cross-validation is utilized, and the Receiver Operating Curve (ROC) is traced for each of the combinations of 10-fold cross-validation. The performance is measured with a confusion matrix for all three classifiers. The performance with Random Forest(RF) and XG boost performance is similar in terms of accuracy, whereas the total number of true predictions is higher in the case of RF. The proposed model would be useful for continuous monitoring and prediction conduction disturbance in the Smart Healthcare framework.

Keywords: Machine Learning, Cardiovascular Disease, Health Failure, Conduction Disturbance, Electrocardiography(ECG), Smart Healthcare

1 Introduction

According to WHO, Cardiovascular Disease (CVD) ranks as the world's primary reason of premature deaths. It is estimated that 31% of all fatalities worldwide—or around 17 million annually—are caused by CVD[1]. The World Heart Federation (WHF) published a report on May 20, 2023 stating that the number of fatalities attributable to cardiovascular disease increased from 12.1 million in 1990 to 20.5 million in 2021 [2]. The heart is one of the most crucial components of the circulatory system in humans. Several variables, including smoking, hypertension [3], and CVDs, cause a steady decline in heart function.

In India, cardiovascular problems are very common for people in the age group more than 45 years old [4]. Due to the absence of visual signs, early detection of CVD is difficult [5]. He re, we propose the prediction of Conduction Disturbance in the human heart using ECG, which can be acquired by wearable devices or Holter monitor, along with two demographic features. A simplified approach is depicted in Figure 1. The main contribution of the presented work is as follows:

- We are predicting the conduction disturbances of the heart using a non-invasive method, i.e., Electrocardiograph.
- Two demographic features that can be utilized with raw ECG to improve classification efficiency are also presented.
- A comparative analysis of three classifiers, i.e., Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Random Forest (RF)

Further, the organization of the paper is as follows: The background and work related to the prediction of conduction disturbances is discussed in section 2. Research gaps and novel contributions of the work are provided in section 3, and the application of the proposed work in smart healthcare is discussed in 4. While the detailed proposed methodology is discussed in section 5, that includes the information of the dataset in section 5.2, data preparation in section 5.3, and modelling in section 5.1. Following this, experimental results are discussed in section 6, including validation of the model in section 6.1 and comparison with state-of-the-art in section 6.2. Finally, the conclusion is provided in section 8.

2 Background and Related Work

In a human heart, conduction pathway is the path the electrical impulse takes that propels the heartbeat. Normally, the Sinoatrial (SA) node's impulse stimulates the atria. The left and right ventricles eventually contract concurrently due to the conduction routes, which continue to bundle the left and right bundle branches (LBB and RBB) [6]. One heartbeat is produced by one impulse, whereas a normal heart beats 60 to 100 times per minute, depending on the person's age. Figure 2 depicts the heart's conduction system.

A barrier in the conduction channels causes conduction disturbance (CD), a heart condition. Conduction problems lead to chronic heart failure. Conduction anomalies can be categorised into three fundamental categories: first degree, second degree, and third degree heart block [7]. The heart's atrioventricular (AV) node beats more slowly than usual in first-degree heart block, with no symptoms at all [5]. Only a portion of the electrical impulses go from the upper heart chamber, the atria, to the lower heart chamber, the ventricles, in a second-degree heart block. The heart may skip beats, beat irregularly, or beat slowly under this circumstance. It manifests as chest discomfort, breathlessness, and palpitations in the heart, among other symptoms. The electrical impulse cannot go from the heart's upper chamber to the lower chamber of the heart in a third-degree heart block, also known as a full heart block. However, the ventricles continue to contract and pump blood, albeit more slowly. Although the blood pumping is inefficient and the contraction is improper. The patient needs immediate assistance since cardiac arrest is a serious possibility. The ECG is a useful tool for predicting CD [8][9] because CD alters the heart's electrical impulse, which is reflected in the ECG. It can even predict the different types of heart block and arrhythmic risk in young patients with Kearns-Sayre syndrome (KSS) [10]. The following section discusses some CD-related researchers. The Electrocardiogram represents the heart's electrical activity using a time-voltage heartbeat chart. The ECG is an important part of clinical diagnosis and treatment as it offers crucial information. A reliable technique for early-stage CVD prediction is the electrocardiogram (ECG) or features of it [11]. The paper presents an ECG signal to predict the Conduction Disturbances in a human heart.

Several researchers have applied machine learning models to predict cardiovascular diseases (CVDs) [12]. The [13] offered an overview of cardiovascular disease risk prediction models and provided the traditional and machine learning-based approaches for CVD classification. To predict CVD, the Multilayer Perceptron (MLP) and K-Nearest Neighbours (KNN) models were tried

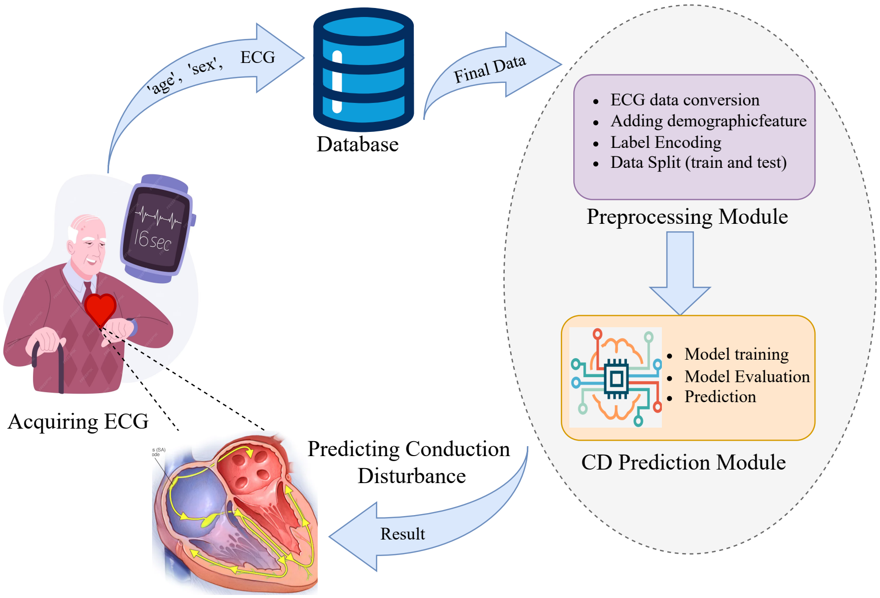

Fig. 1 Conceptual Diagram of Proposed Methodology

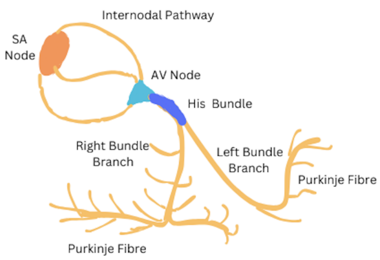

Fig. 2 Pathway of human heart conduction system

[14][15]. [16] has presented a study to identify the risk factors for predicting the complete heart block. The prediction of CVD is provided using a variety of supervised machine learning (ML) models, including Random Forest (RF), Decision Tree (DT), K-Nearest Neighbour (KNN), and Support Vector Machine (SVM) is presented by [17]. [18] predicted CVD using an adaptive boosting classifier based on UCI repository database of heart disease. While [19] predicted CVD employing SVM and Ensemble of Naive Bayes with memory-based learner Decision Tree Induction using Gini Index on Ricco and UCI repository data bases. While [15] provided a classifier for predicting cardiovascular disorders based on Random Forest. The authors of [20] investigated several deep learning machine learning algorithms and concluded that fusion model-based classifiers outperform separate classifiers. Numerous researches are emerging on the subject of CD prediction. It is noted that first and second-degree heart blockages don't present with peculiar symptoms that allow for accurate diagnosis. Mortality can be reduced via early CD prediction. After Transcatheter aortic valve replacement (TAVR), conductance disturbances are relatively prevalent [21] [22]. [23] compares machine learning algorithms and neural networks to predict atrioventricular block with a single lead ECG. An ML-based model for patient-specific monitoring was reported in [24] to properly predict the CD following TAVR. Conduction disturbances, atrial fibrillation, sudden coronary mortality, and device infection risk factors were evaluated by [25]. On the other hand, [26] studied the arrhythmia and conduction abnormalities connected to the septal defects. The other research was focused on systemic sclerosis patients' conduction disturbances [27]. Table 1 summarises some of the prior work.

3 Research Gap and Novel Contribution

3.1 Problem Statement

Most researchers have explored CD prediction or development of conduction abnormalities after TAVR. However, further research is required to investigate 12-lead ECG, a clinical standard for

Table 1 Summary of some prior work

Study	Type of CVD or Conduction Disturbance	Data Source	Classifier(s) included in study	Proposed Classifier/outcome
Al-Naami et-al, 2022 [28]	Complete LBBB	MIT-BIH dataset	Adaptive Neural Fuzzy Inference system (ANFIS)	ANFIS
Kammath et-al, 2022 [29]	MIT-BIH	Left Bundle Branch Block (LBBB) and Right Bundle Branch Block (RBBB)	SVM, KNN, LDA	KNN
Kirti Singh et-al, 2022 [23]	Atrioventricular block (AV block)	KURIAS-ECG database	Gaussian Naive Bayes, Random Forest, Neural Network	Random Forest classifier
Valeria Galli et-al, 2021 [24]	Right or Left Bundle Branch Block after TAVR	Private data of 151 cohort	SVC+SGD, RF, KNN, Gaussian Naive Bayes, Extreme gradient boosting, Decision Tree, Logistic Regression	K-Nearest Neighbour (KNN) classifier
Rasel et-al, 2020 [30]	1st degree AV block, LBBB, RBBB	Cardiology department of Chittogong Medical college (CMCH)	DT, RF, KNN, SVM	DT, RF
Matthew R. Williams et-al, 2018 [26]	Sinus and AV node conduction abnormality associated with Autism spectrum disorder(ASD)	PRAETORIAN trial patient data	NA	Analysed CD associated with ASD
Vincent Auffret et-al, 2017 [21]	Left Bundle Branch Block after TAVR	Private data	NA	Analysed the new onset of LBBB after TAVR to mitigate sudden cardiac death
Kora et-al, 2016 [31]	Bundle Branch Block (BBB)	MIT-BIH database	Adaptive Bacterial Foraging Optimization (ABFO)	ABFO and Levenberg Marquadt Neural Network (LMNN)
Hunifang Huang et-al, 2014 [32]	LBBB and RBBB	MIT-BIH Arrhythmia dataset	Minimum distance classifier, the Linear discriminant classifier, and the Linear SVM	Ensemble of the three classifiers

assessing the heart state without TAVR, for heart block or CD prediction.

3.2 Solution Proposed

Here, we propose a smart healthcare based framework to predict the conduction disturbance using ECG. The acquired ECG is combined with two demographic features i.e., 'age' and 'sex' to improve the prediction performance of the classifier.

3.3 Significance and Novelty of the Solution

The novel contribution of the presented work is:

1. A novel machine learning-based classifier is proposed to classify the conduction disturbance like Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Left anterior or posterior fascicular block, Atrioventricular (AV) block, on-specific intraventricular conduction block, complete or incomplete left or right bundle branch block, and Wolf-Parkinson-White syndrome.
2. The 12-lead raw ECG is utilised along with two ubiquitous demographic features, i.e., 'age' and 'sex,' to predict conduction abnormality in the heart.
3. It also presents a comparative analysis of three very famous machine learning classifiers

to show that ensemble machine learning is a better choice.

4. A smart healthcare framework is presented for predicting conduction disturbances or abnormalities in the human heart.

4 Smart Healthcare Framework for Prediction of Conduction Disturbances

The proposed algorithm's implementation is crucial in the modern smart healthcare system [33]. The clinical gold standard for the initial evaluation of CVD is the ECG, which is also economical. An IoT infrastructure was given by [15] that uses historical and empirical data to develop suggestions for real-time and remote health monitoring. While [34] presented a fog-based smart healthcare system that may notify medical professionals, caretakers, or respective hospital staff in advance about heart attack chances. An intelligent healthcare tracking system utilising ensemble deep learning and feature fusion to forecast heart disease is proposed by [35]. [36], [37], [38], [39], and [40] proposed a similar sort of cloud-based Smart Healthcare platform to identify cardiac problems. It employs a deep learning strategy in cloud networks using IoT. The smart healthcare framework for continuous glucose monitoring is presented in the paper [41] A Smart Healthcare framework for the early diagnosis of cardiac arrest circumstances was also given by [42], and for cardiovascular disease is given by [43]. Similarly, [44] suggests another Smart Healthcare framework to identify cardiac disease using machine learning and deep learning approaches.

Similar to the previous approaches, the proposed CD prediction machine learning model can also be deployed using cloud services. The raw ECG data would be acquired from smart wearable devices such as those presented by [45–49]. In the following step, the data preprocessing (i.e., noise filtering, data transformation, encoding, etc.) would be done at the ECG acquisition node or devices or on the cloud, depending upon the IoT infrastructure or healthcare service provider preference. After the preprocessing, data will be finally prepared for feeding the ML model, which will predict the presence or absence of any Conduction Disturbance in the subject. The prediction would be communicated further using Application Programming Interface (API) services.

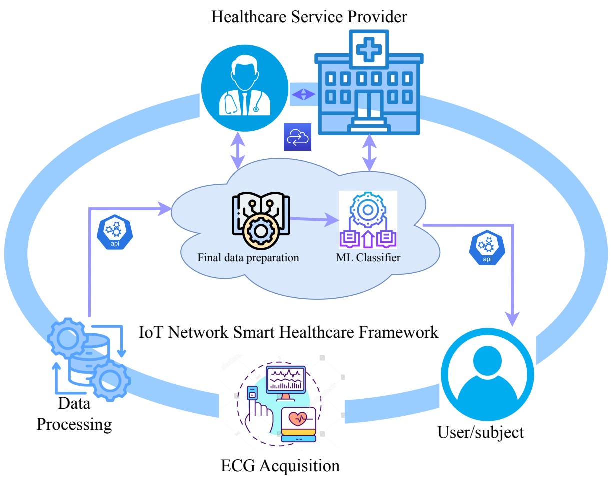

Fig. 3 Proposed Smart Healthcare Framework for CD prediction

5 iCardio 3.0: Proposed Methodology for Conduction Disturbance using Machine Learning method

The proposed methodology utilizes a 10-second 12-lead ECG collected from subjects of age group 0 to 95 years. The raw ECG is converted into an array and combined with the two demographic features, i.e., 'age' and 'sex.' Later, three famous classifiers, Support Vector Machine, XG Boost, and Random Forest, are utilized to predict the Conduction Disturbance in the heart. A comparison of the performance of all three classifiers' is done to find the best one in terms of accuracy. Further, a comparison of the performance of classifiers' before and after adding demographic features is also made, and all three classifiers have shown significant improvement in their performance after adding demographic features.

The next subsection discusses all the classifiers' estimated performance metrics, including SVM, RF, and XG Boost, and Figure 4 illustrates the process flow of the proposed work.

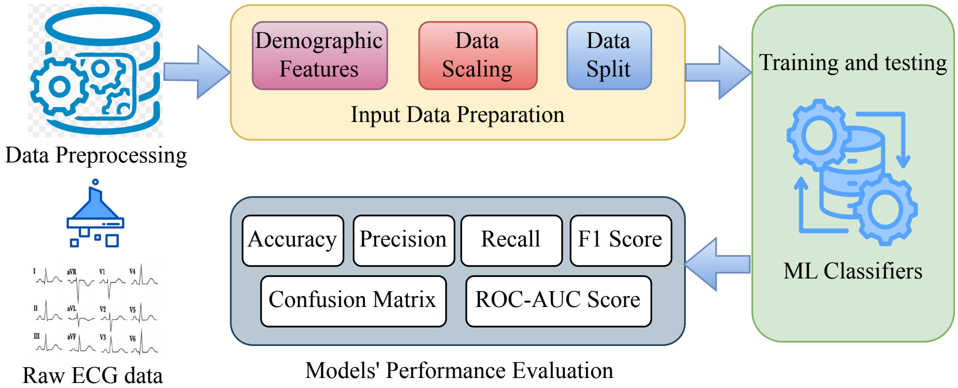

Fig. 4 Process Flow of the proposed work

5.1 Machine Learning Models

Ensemble machine learning models, i.e., Random Forest and XG Boost, are utilized to predict conduction disturbances. Performance and robustness are two other advantages of ensemble machine learning models for predictive modelling. An ensemble ML model can perform better than a single model and create more accurate predictions while also lowering the variance or dispersion of those predictions. The Support Vector Machine (SVM) classifier is also considered with the ensemble machine learning classifiers due to its discriminative power [50]. Here, in the presented work, SVM with 'rbf' (Radial Bias Function) kernel is used, which is given as:

$$k(x_a, x_b) = exp(-\frac{d_{ab}}{2\sigma^2}) \quad , \tag{1}$$

where x_a and x_b are two points of the sample data, and $_ab$ is the Euclidean Distance between x_a and x_b. The data, which cannot be separated linearly, is transformed by the kernel function into a higher-dimensional feature space where it can be separated linearly into two classes. However, RF and XGBoost both use the Decision Tree classifier at the backend and work on majority voting. One significant difference between them is that RF has parallel processing, and XGBoost has sequential. For selecting the attribute at the root node, the Decision Tree of the presented RF model uses the Gini Index, which can be calculated as follows:

$$GiniIndex = 1 - \sum_{i=1}^{n}(P_i)^2 \quad , \tag{2}$$

where P_i is the probability of a data point to being classified in a particular class. Here n=2 as we are assuming two classes NORM and CD. As XGBoost is a sequential process, it tries to minimise the loss at each step, which is calculated as follows:

$$L(f) = \sum_{i=1}^{n}(L(y_i, (f(x_i)) \quad . \tag{3}$$

The above mentioned three classifiers—Random Forest (RF), XG Boost (XGB), and Support Vector Machine (SVM) are set up with their default hyper-parameters and trained using the training dataset [51]. The models performance is assessed using the testing dataset after training. Performance metrics included for comparing the performance of the classifiers are accuracy, precision, recall, and f1-score [52]. Although accuracy is the most used performance metric for classification models, a brief description of the performance mentioned above measures is provided below:

- Accuracy: It provides us with the proportion of overall correct forecasts to overall predictions (see equation 1).
- Precision: It gives the percentage of True Positives compared to all of the Positive Predictions (refer to equation 2).
- Recall: The ratio of rightly predicted CVD class cases to actual CVD incidence is provided. The formula to calculate Recall is given in equation 3.
- F1 score: It is the harmonic mean of precision and recall, it is calculated as per the formula given below (refer to equation 4):

The overall process can be summarized in four stages. First is data processing, which includes noise removal, transformation, and normalization. The second step is to prepare the data for input to a machine-learning model; in this step, two sets of data were prepared, one containing only 12-lead ECG recording and the labels and another that has two demographic features 'age' and 'sex' together with 12-lead ECG and labels. Both the datasets are later split into train and test set in 80:20. In the third step, the three models (i.e., SVM, RF, and XGB) are trained with the training datasets, and finally, they are tested on testing datasets, and the performance metrics are compared. The models are validated in the fourth and final step using 10-fold cross-validation. Mean accuracy, Receiver Operating Curve (ROC), and Area Under Curve (AOC) scores for each model are calculated using 10-fold cross-validation. The overall process flow is depicted in Figure 4.

5.2 Dataset

This study is employed on the PTB-XL publicly accessible large electrocardiography data collection obtained from [53, 54]. It comprises 12 lead ECG recordings, each lasting 10 seconds, from 21837 records obtained from 18885 people (I, II, III, aVL, aVR, aVF, V1, V2, V3, V4, V5, and V6). The ECG data was annotated by two cardiologists, making it a multi-label data set. It was later combined as a diagnostic superclass and subclass. The five superclasses are Myocardial Infarction (MI), Normal ECG (NORM), Hypertrophy (HYP), Conduction Disturbance (CD)and ST/T change (STTC). Only the classes NORM and CD are taken into consideration in this case. The following diseases are included in conduction disturbance or disorder (CD):(i)(LAFB/LPFB) Left anterior/left posterior fascicular block, (ii) (IRBBB) left anterior/left posterior fascicular block, (iii) (ILBBB) incomplete left bundle branch block, (iv) (CLBBB) complete left bundle branch block, (v) (CRBBB) complete right bundle branch block, (vi) (AVB) AV block, (vii) (IVCB) on-specific intraventricular conduction lock or disturbance and (viii) (WPW) Wolf-Parkinson-White syndrome.

The data inclusion and exclusion process for the presented work is shown in Figure 5.

5.3 Data Preparation

Here, in this presented work, CD classification is modelled as a binary classification problem, where class 1 corresponds to normal (NORM) subjects, and class 2 corresponds to CD patients. In total there were 10,777 samples left out of which 9,069 were of NORM class and 1,708 were of CD class. Using the Numpy library, the 12-lead ECG data for all 10,777 individuals has been transformed into a 3-D array. Later, it was flattened into a 2-D array and transformed into a Pandas data frame. After that, two more features, 'age' and 'sex' were added to the data. The data is split into 80:20 ratios for training and testing purposes. Further, using Standard Scalar from the Sklearn package, the training data is scaled or standardized. Three distinct classifiers are trained using training data and evaluated through testing. The machine learning models are used to predict conduction disturbances and are discussed in the next section.

6 Experimental Results

The data included for the work has 10,777 samples, which are split into training and testing sets. The machine learning classifiers, i.e., SVM, RF, and XGB, are first trained and tested with raw ECG, and the performance metrics are recorded for each one of them. Later, we added two demographic features, 'age' and 'sex,' to raw ECG and trained and tested the same three algorithms (i.e., SVM, RF, and XGB) for classification. Both cases' performance metrics are listed in Table 2 and Table 3. The class-wise performance metrics are also provided in Table 6 for the second case where demographic features are added. Further, the confusion matrix for the same is provided in Figure 6.

As reported in Table 2, the maximum accuracy with raw ECG only is 84% with SVM and XGB classifier. However, the best value for precision is with the RF that is 74 %. At the same time, the best values for recall and F1 score are 84 % and 77 %, respectively.

A significant improvement in the performance of the classifiers' can be observed after adding the demographic features, as seen in Table 3. The accuracy of all the classifiers improved significantly, i.e., 90 % for the RF and XGB and 87 % for

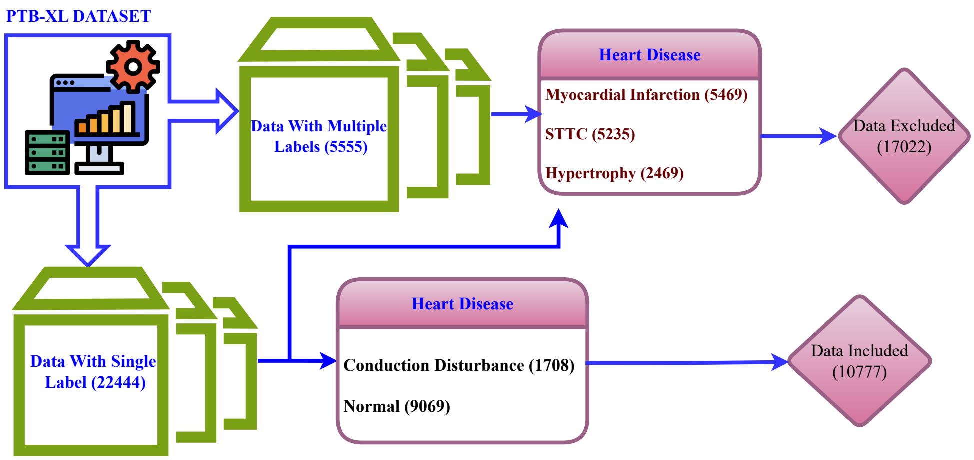

Fig. 5 Data inclusion and exclusion process

Table 2 Performance metrics of SVM, XG Boost, and RF before adding 'age' and 'sex' features

ML Classifiers	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
Support Vector Machine (SVM)	70	84	77	84
XG Boost	72	84	76	84
Random Forest (RF)	74	83	77	83

Table 3 Performance metrics of SVM, XG Boost, and RF after adding 'age' and 'sex' features

ML Classifiers	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
Support Vector Machine (SVM)	89	87	83	87
XG Boost	89	90	88	90
Random Forest (RF)	90	90	89	90

SVM. Precision for RF also significantly enhanced, i.e., 90%. At the same time, recall and F1 score for the RF classifier is 90% and 89%, respectively.

6.1 Validation

To further validate the classifier's result, 10-fold cross-validation is performed, and performance metrics like the average accuracy score, ROC-AUC value, and confusion matrix are obtained. The 10-fold cross-validation process assures that the proposed classifier is unbiased. The accuracy and ROC-AUC score in each fold of 10-fold cross-validation is listed in Table 5.

It is evident from the 10-fold cross-validation scores that the Random Forest classifier exceeds the SVM and XG Boost classifier's performance. However, the accuracy of XG Boost is comparable. Further, the cumulative values of the confusion matrix of 10-fold cross-validation are provided in Figure 7 to better assess the classifiers' performance.

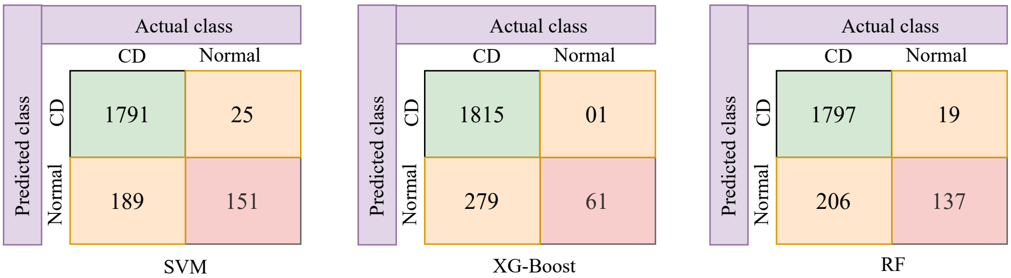

Fig. 6 Confusion matrix of SVM, XG Boost, and RF respectively before adding demographic feature

Table 4 Class-wise performance metrics of each classifier after adding 'age' and 'sex' features

Performance Metrics	SVM		XG Boost		RF	
	NORM	CD	NORM	CD	NORM	CD
Precision	0.87	0.98	0.90	0.88	0.90	0.86
Recall	1.00	0.18	0.99	0.39	0.99	0.44
F1 Score	0.93	0.30	0.94	0.54	0.94	0.59

Table 5 Accuracy and ROC-AUC score of the classifiers' in each fold of 10-fold cross-validation

Fold	Accuracy			ROC-AUC		
	RF	XG-Boost	SVM	RF	XG-Boost	SVM
1-Fold	89.14	88.49	85.25	0.89	0.90	0.76
2-Fold	88.68	88.31	85.52	0.86	0.85	0.72
3-Fold	90.35	89.61	86.82	0.90	0.90	0.74
4-Fold	90.44	89.33	86.17	0.87	0.86	0.74
5-Fold	89.98	90.25	86.08	0.88	0.87	0.71
6-Fold	90.16	90.53	85.89	0.88	0.91	0.75
7-Fold	89.33	89.98	86.36	0.86	0.86	0.72
8-Fold	90.71	90.62	85.88	0.86	0.86	0.74
9-Fold	90.15	90.06	85.42	0.87	0.87	0.71
10-Fold	92.20	91.17	86.53	0.88	0.87	0.72
Average	**90.11**	89.83	85.99	**0.87**	0.87	0.73

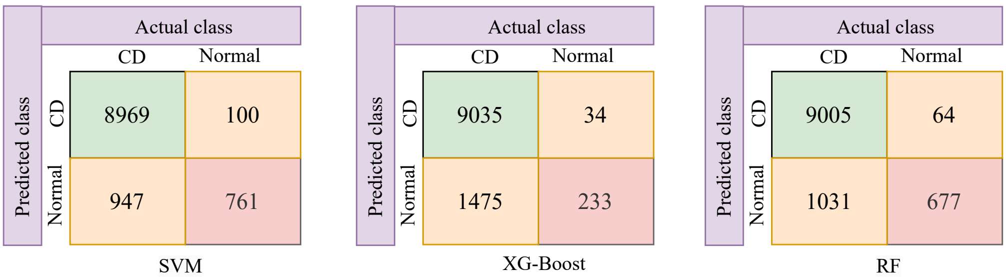

Fig. 7 10-fold cross-validation Confusion Matrix after adding demographic features

To ensure the unbiased performance of the presented classifier, their Receiver Operating Curve (ROC)is plotted, and the Area Under the Curve (AUC) is also calculated. Figure 8, Figure 9 and Figure 10 show the ROC-AUC score for all three classifiers. ROC curve is drawn on the 45-degree diagonal, i.e., y=x. Ten curves are plotted in all three graphs, each representing the corresponding K-Fold test data. Mean and standard deviations are also plotted along with the 10-fold cross-validation receiver operating curves. Mean indicates the average of the 10-fold receiver operating curves, while standard deviation measures the dispersion of data with the mean value.

6.2 Comparison with State of the Art Methods

The existing methods are compared with the proposed method to predict the conduction abnormalities in the human heart, which is presented in Table 6. Most of the methods are tested on ECG features that require an additional step before the final prediction through the final classification model. Wherever the presented method requires no feature extraction step, the raw ECG, along with two common demographic features, i.e., 'age' and 'sex', can be directly be applied to the classifier to predict conduction disturbances in the human heart. Apart from this, the number of subjects considered in the existing state-of-the-art methods is comparatively less; they are hundreds, while in this presented work, it is more than 10 thousand. This makes the presented method more promising and reliable. Subsequently, a 10-fold cross-validation score is presented, proving the classifier is robust and accurate. The area under the curve of the receiver operating curve confirms the unbiased prediction by the classifier.

7 Discussion

It is evident from the result that adding two demographic features, i.e., 'age' and 'sex,' has significantly improved the performance measures (that includes precision, recall, F1 score, and accuracy) of all three classifiers. However, if we dig deeper into the model and look into the classifiers' class-wise performance measure, we observe that even after adding the demographic features, recall for all three classifiers is below 0.50 for the CD class, whereas it is 0.99 and above for the NORM class. The low recall value for the CD class is due to the imbalanced data of 9069 samples of the NORM category and only 1708 samples of the CD class. In the future, this problem can be attempted to solve by applying the techniques of data balancing [43]. The relation between Diabetes and CVD is well established [60, 61]. People with Diabetes have a greater prevalence rate of cardiovascular disease (CVD) than people without diabetes [62–64]. Moreover, people with Diabetes mellitus experience higher rates of morbidity and death due to cardiovascular disease mention yao2023age, einarson2018prevalence. So, to lower the chance of dying from cardiovascular disease, diabetes management is crucial [65]. Similarly, BMI (Body Mass Index) is directly related to CVD. Higher BMI increases the risk of CVD [66–68]. Several studies [69–71] have been done to showcase the relationship between obesity and CVD. Hence, including Diabetes and BMI with other demographic features, viz. 'age' and 'sex,' shall increase the prediction of CVD. Still, unfortunately, the subjects' Diabetes information is not available in the data. In the future, the impact of adding features like 'BMI' and 'Diabetes' on predicting CVD can be done.

8 Conclusion

The ensemble ML algorithms perform better than the SVM, whereas RF uses the bagging approach and works in parallel. The XG Boost works on the concept of boosting, which is a sequential process. It has been observed that RF is better regarding overall performance metrics. The performance measures for each class are listed for a better understanding of the classifiers' behaviour. However, the accuracy and ROC-AUC score are recorded good results in 10-fold cross-validation. The Random Forest classifier is better in terms of accuracy and ROC-AUC score. However, the mean of the 10-fold cross-validation score of ROC-AUC score is the same for the Random Forest classifier and XG boost classifier; Random Forest offers a standard deviation of 0.01 only, while XG Boost classifiers have a standard deviation of 0.02. In future work, the resource requirement of a neural network-based classifier for predicting CVD using 12-lead ECG can be studied, and performance can be compared. In addition to that, Optuna optimization would also be performed

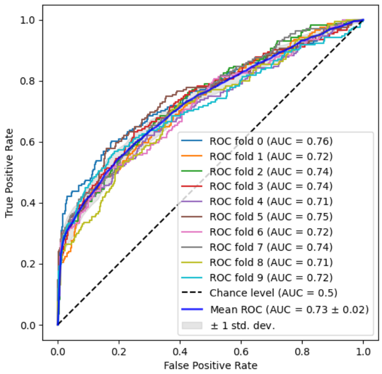

Fig. 8 Receiver Operating Curve of SVM classifier

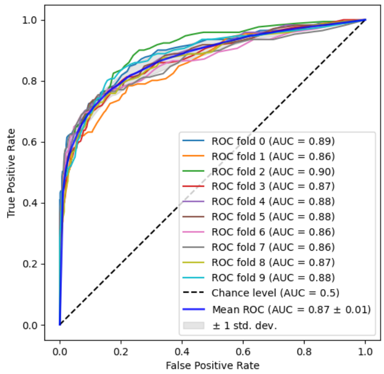

Fig. 9 Receiver Operating Curve of RF classifier

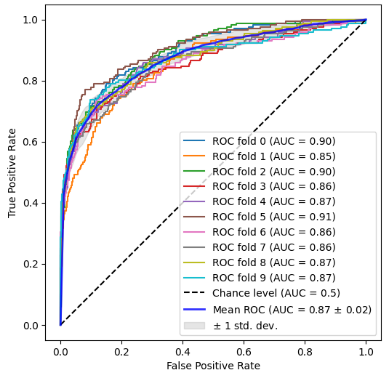

Fig. 10 Receiver Operating Curve of RF classifier

Table 6 Comparative Analysis of the Methods of prediction of Conduction Disturbances

Reference	Conduction Abnormality Included	Data	No. of subjects	Input data/Features	Cross- validation	Model	Accuracy	Area under curve
Kora et-al, 2015 [55]	LBBB, RBBB	MIT-BIH		ECG features	NA	BFPSO	↑ 90.0%	NA
Rasel et-al, 2019 [56]	1st AV block, LBBB, RBBB	Chittagong Medical College (CHCH)	208	ECG Features (32 attributes)	NA	DT and RF	↑ 90.0%	NA
Yang et-al, 2020 [57]	Strict LBBB, and LBBB	MADIT-CRT clinical trial data	602	ECG Features (14 attributes)	NA	Neural Network	↓ 90.0%	NA
Kaya et-al, 2021 [58]	LBBB, RBBB	MIT-BIH	45	ECG Features	NA	BE+ KNN	↑ 90.0%	NA
Aldave et-al, 2023 [59]	LBBB	Physiological signal challenge (ICBED 2018)	300	ECG Features	NA	RF	↓ 90.0%	NA
Galli et-al, 2021 [24]	Patient specific conduction abnormality after TAVR	ECG data nine European Centres	151	Preoperative multi-slice computed tomography (MSCT) features	5-fold	KNN	↓ 90.0%	0.84
Proposed	ILBBB, CLBBB, CRBBB, AVB, IVCB, WPW	PTB-XL	10,777	Raw ECG with 'age', and 'sex'	10-fold	RF	↑ 90.0%	0.87

on the presented ML models to improve their performance further.

9 Compliance with Ethical Standards

The authors declare that they have no conflict of interest and there was no human or animal testing or participation involved in this research. All data were obtained from public domain sources.

10 Acknowledgment

This article is an extended version of our previous conference paper presented at [72].

11 Declarations

11.1 Competing Interests

All the authors declare that they have no competing interest.

11.2 Funding Information

Not Applicable

11.3 Author contribution

All the authors contributed equally.

11.4 Data Availability Statement

The data that support the findings of this study are openly available in Physionet at https://physionet.org/content/ptb-xl/1.0.3/.

11.5 Research Involving Human and/or Animals

The presented research does not involve any human or animal.

11.6 Informed Consent

Not Applicable

References

[1] WHO: Cardiovascular Diseases. https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1/. Online; accessed 28-June-2023 (2023)

[2] WHF: Deaths From Cardiovascular Disease. https://world-heart-federation.org/news/deaths-from-cardiovascular-disease-surged-/60-globally-over-the-last-30-years-report/. Online; accessed 21-Sept-2023 (2023)

[3] Diamond, J.A., Phillips, R.A.: Hypertensive heart disease. Hypertension research **28**(3), 191–202 (2005)

[4] Bhatia, M., Dixit, P., Kumar, M., Dwivedi, L.K.: Impending epidemic of cardiovascular diseases among lower socioeconomic groups in india. The Lancet Healthy Longevity **2**(6), 314–315 (2021)

[5] Sinha, N., Jangid, T., Joshi, A.M., Mohanty, S.P.: icardo: A machine learning based smart healthcare framework for cardiovascular disease prediction. arXiv preprint arXiv:2212.08022 (2022)

[6] Sinha, N., Joshi, A.M., Mohanty, S.P.: icardo 2.0: A smart healthcare framework for cardiovascular disease accurate prediction by using t-wave morphology of ecg. In: 2022 IEEE International Symposium on Smart Electronic Systems (iSES), pp. 343–348 (2022). IEEE

[7] Heart Conduction Disorders. https://www.heart.org/en/health-topics/arrhythmia/about-arrhythmia/conduction-disorders. Accessed: 2023-05-14

[8] Meghna, S., Joshi, A.M.: Real-time ecg based authentication using hardware accelerated implementation on zynq soc. In: 2019 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS), pp. 99–104 (2019). IEEE

[9] Silva, R.M., Souza Maciel, A.: Conduction disorders: The value of surface ecg. Current Cardiology Reviews **17**(2), 173–181 (2021)

[10] Di Mambro, C., Tamborrino, P.P., Silvetti, M.S., Yammine, M.L., Marcolin, C., Righi, D., Baban, A., Martinelli, D., Dionisi Vici, C., Drago, F.: Progressive involvement of cardiac conduction system in paediatric patients with kearns–sayre syndrome: how to predict occurrence of complete heart block and sudden cardiac death? EP Europace **23**(6), 948–957 (2021)

[11] Goldberger, A.: Goldberger's clinical electrocardiography. Elsevier (2018)

[12] Oyeleye, M., Chen, T., Titarenko, S., Antoniou, G.: A predictive analysis of heart rates using machine learning techniques. International Journal of Environmental Research and Public Health **19**(4), 2417 (2022)

[13] Faizal, A.S.M., Thevarajah, T.M., Khor,

S.M., Chang, S.-W.: A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach. Computer methods and programs in biomedicine **207**, 106190 (2021)

[14] Pal, M., Parija, S., Panda, G., Dhama, K., Mohapatra, R.K.: Risk prediction of cardiovascular disease using machine learning classifiers. Open Medicine **17**(1), 1100–1113 (2022)

[15] Kaur, P., Kumar, R., Kumar, M.: A healthcare monitoring system using random forest and internet of things (iot). Multimedia Tools and Applications **78**, 19905–19916 (2019)

[16] Karimianpour, A., Heizer, J., Leaphart, D., Rier, J.D., Shaji, S., Ramakrishnan, V., Nielsen, C.D., Fernandes, V.L., Gold, M.R.: Predicting complete heart block after alcohol septal ablation for hypertrophic cardiomyopathy using a risk stratification model and clinical tool. Catheterization and Cardiovascular Interventions **98**(2), 393–400 (2021)

[17] Gupta, J.: The accuracy of supervised machine learning algorithms in predicting cardiovascular disease. In: 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), pp. 234–239 (2021). https://doi.org/10.1109/ICAICST53116.2021.9497837

[18] Miao, K.H., Miao, J.H., Miao, G.J.: Diagnosing coronary heart disease using ensemble machine learning. International Journal of Advanced Computer Science and Applications **7**(10) (2016)

[19] Bashir, S., Qamar, U., Khan, F.H., Javed, M.Y.: Mv5: a clinical decision support framework for heart disease prediction using majority vote based classifier ensemble. Arabian Journal for Science and Engineering **39**, 7771–7783 (2014)

[20] Subramani, S., Varshney, N., Anand, M.V., Soudagar, M.E.M., Al-Keridis, L.A., Upadhyay, T.K., Alshammari, N., Saeed, M., Subramanian, K., Anbarasu, K., *et al.*: Cardiovascular diseases prediction by machine learning incorporation with deep learning. Frontiers in Medicine **10**, 1150933 (2023)

[21] Auffret, V., Puri, R., Urena, M., Chamandi, C., Rodriguez-Gabella, T., Philippon, F., Rodes-Cabau, J.: Conduction disturbances after transcatheter aortic valve replacement: current status and future perspectives. Circulation **136**(11), 1049–1069 (2017)

[22] Sammour, Y., Kapadia, S.R.: Conduction Disturbance After TAVR. https://citoday.com/articles/2022-mar-apr/conduction-disturbance-after-tavr Accessed 2023-05-14

[23] Kirti Singh, V.N., et-al: Machine learning algorithms for atrioventricular conduction defects prediction using ecg: A comparative study. In: 2022 IEEE Delhi Section Conference (DELCON), pp. 1–5 (2022). https://doi.org/10.1109/DELCON54057.2022.9753488

[24] Galli V, L.F., et-al: Towards patient-specific prediction of conduction abnormalities induced by transcatheter aortic valve implantation: a combined mechanistic modelling and machine learning approach. Eur Heart J Digit Health, 606–615 (2021) https://doi.org/10.1093/ehjdh/ztab063

[25] Crea, F.: Novel risk factors for atrial fibrillation, conduction disturbances, sudden coronary death, and device infection. European Heart Journal **43**(47), 4853–4857 (2022) https://doi.org/10.1093/eurheartj/ehac734

[26] Williams, M.R., Perry, J.C.: Arrhythmias and conduction disorders associated with atrial septal defects. Journal of Thoracic Disease **10**(Suppl 24) (2018)

[27] Vrancianu, C.A., Gheorghiu, A.M., Popa, D.E., Chan, J.S.K., Satti, D.I., Lee, Y.H.A., Hui, J.M.H., Tse, G., Ancuta, I., Ciobanu, A., *et al.*: Arrhythmias and conduction disturbances in patients with systemic sclerosis—a systematic literature review. International Journal of Molecular Sciences **23**(21), 12963 (2022)

[28] Al-Naami, B., Fraihat, H., Owida, H.A., Al-Hamad, K., De Fazio, R., Visconti, P.: Automated detection of left bundle branch block from ecg signal utilizing the maximal overlap discrete wavelet transform with anfis. Computers **11**(6), 93 (2022)

[29] Kammath, P.S., Gopal, V.V., Kuriakose, J.: Detection of bundle branch blocks using machine learning techniques. Indonesian Journal of Electrical Engineering and Informatics (IJEEI) **10**(3), 559–566 (2022)

[30] Rasel, R.I., Sultana, N., Meesad, P., Chowdhury, A., Hossain, M.: 1st-degree atrioventricular (av-block) and bundle branch block prediction using machine learning. In: Proceedings of the 11th International Conference on Advances in Information Technology, pp. 1–6 (2020)

[31] Kora, P., Kalva, S.R.K.: Detection of bundle branch block using adaptive bacterial foraging optimization and neural network. Egyptian Informatics Journal **18**(1), 67–74 (2017)

[32] Huang, H., Liu, J., Zhu, Q., Wang, R., Hu, G.: Detection of inter-patient left and right bundle branch block heartbeats in ecg using ensemble classifiers. Biomedical engineering online **13**, 1–22 (2014)

[33] Jain, P., Joshi, A.M., Agrawal, N., Mohanty, S.: iglu 2.0: A new non-invasive, accurate serum glucometer for smart healthcare. arXiv preprint arXiv:2001.09182 (2020)

[34] Nancy, A.A., Ravindran, D., Vincent, D.R., Srinivasan, K., Chang, C.-Y.: Fog-based smart cardiovascular disease prediction system powered by modified gated recurrent unit. Diagnostics **13**(12), 2071 (2023)

[35] Ali, F., El-Sappagh, S., Islam, S.R., Kwak, D., Ali, A., Imran, M., Kwak, K.-S.: A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Information Fusion **63**, 208–222 (2020)

[36] Nancy, A.A., Ravindran, D., Raj Vincent, P.D., Srinivasan, K., Gutierrez Reina, D.: Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. Electronics **11**(15), 2292 (2022)

[37] Kumar, A., Satyanarayana Reddy, S.S., Mahommad, G.B., Khan, B., Sharma, R., et al.: Smart healthcare: Disease prediction using the cuckoo-enabled deep classifier in iot framework. Scientific Programming **2022** (2022)

[38] Raju, K.B., Dara, S., Vidyarthi, A., Gupta, V.M., Khan, B., et al.: Smart heart disease prediction system with iot and fog computing sectors enabled by cascaded deep learning model. Computational Intelligence and Neuroscience **2022** (2022)

[39] Sundararajan, S.C.M., Bharathi, G., Loganathan, U., Vadivel, S.: Improved smart healthcare system of cloud-based iot framework for the prediction of heart disease. Information Technology and Control **52**(2), 529–540 (2023)

[40] Patro, S.P., Padhy, N.: An rhmiot framework for cardiovascular disease prediction and severity level using machine learning and deep learning algorithms. International Journal of Ambient Computing and Intelligence (IJACI) **13**(1), 1–37 (2022)

[41] Jain, P., Joshi, A.M., Mohanty, S.P.: iglu 1.0: An accurate non-invasive near-infrared dual short wavelengths spectroscopy based glucometer for smart healthcare. arXiv preprint arXiv:1911.04471 (2019)

[42] Shah, A., Ahirrao, S., Pandya, S., Kotecha, K., Rathod, S.: Smart cardiac framework for an early detection of cardiac arrest condition and risk. Frontiers in Public Health **9**, 762303 (2021)

[43] Sinha, N., Kumar, M.G., Joshi, A.M., Cenkeramaddi, L.R.: Dasmcc: Data augmented smote multi-class classifier for prediction of cardiovascular diseases using time series features. IEEE Access (2023)

[44] Elwahsh, H., El-Shafeiy, E., Alanazi, S.,

Tawfeek, M.A.: A new smart healthcare framework for real-time heart disease detection based on deep and machine learning. PeerJ Computer Science **7**, 646 (2021)

[45] Romagnoli, S., Ripanti, F., Morettini, M., Burattini, L., Sbrollini, A.: Wearable and portable devices for acquisition of cardiac signals while practicing sport: A scoping review. Sensors **23**(6), 3350 (2023)

[46] Neri, L., Oberdier, M.T., Abeelen, K.C., Menghini, L., Tumarkin, E., Tripathi, H., Jaipalli, S., Orro, A., Paolocci, N., Gallelli, I., *et al.*: Electrocardiogram monitoring wearable devices and artificial-intelligence-enabled diagnostic capabilities: A review. Sensors **23**(10), 4805 (2023)

[47] Lee, S.-Y., Hung, Y.-W., Su, P.-H., Lee, I.-P., Chen, J.-Y.: Biosignal monitoring clothing system for the acquisition of ecg and respiratory signals. IEEE Access **10**, 66083–66097 (2022)

[48] Nawawi, M.M.M., Sidek, K.A., Dafhalla, A.K., Azman, A.W.: Review on data acquisition of electrocardiogram biometric recognition in wearable smart textile shirts. In: Journal of Physics: Conference Series, vol. 1900, p. 012019 (2021). IOP Publishing

[49] Lin, B.-S., Wong, A.M., Tseng, K.C.: Community-based ecg monitoring system for patients with cardiovascular diseases. Journal of medical systems **40**, 1–12 (2016)

[50] Youn-Jung Son, H.-G.K. RN, et-al: Application of support vector machine for prediction of medication adherence in heart failure patients. Healthc Inform Res **16**(4), 253–259 (2010)

[51] Pancholi, S., Joshi, A.M., Joshi, D.: A robust and accurate deep learning based pattern recognition framework for upper limb prosthesis using semg. arXiv preprint arXiv:2106.02463 (2021)

[52] Sharma, G., Joshi, A.M., Gupta, R., Cenkeramaddi, L.R.: Depcap: A smart healthcare framework for eeg based depression detection using time-frequency response and deep neural network. IEEE Access (2023)

[53] Patrick Wagner, N.S., et-al: Ptb-xl, a large publicly available electrocardiography dataset. Scientific Data **7**(1), 154 (2020) https://doi.org/10.1038/s41597-020-0495-6

[54] Ary L. Goldberger, L.A.N.A., et-al: Physiobank, physiotoolkit, and physionet. Circulation **101**(23), 215–220 (2000) https://doi.org/10.1161/01.CIR.101.23.e215

[55] Kora, P., Kalva, S.R.: Hybrid bacterial foraging and particle swarm optimization for detecting bundle branch block. SpringerPlus **4**(1), 481 (2015)

[56] Rasel, R.I., Meesad, P., Chowdhury, A., Hossain, M.: Heart block prediction using data mining and machine learning. 8th Joint Symposium Computational Intelligence (JSCI), Bangkok Thailand (2020)

[57] Yang, T., Gregg, R.E., Babaeizadeh, S.: Detection of strict left bundle branch block by neural network and a method to test detection consistency. Physiological Measurement **41**(2), 025005 (2020)

[58] Kaya, Y.: Detection of bundle branch block using higher order statistics and temporal features. Int. Arab J. Inf. Technol. **18**(3), 279–285 (2021)

[59] Alessandra Aldave, P.M.P.R.S.T. Huaylinos J, T, S.: Addressing detection of left bundle branch block by means of electrocardiogram's signals using machine learning algorithms. In: INTERCON 2023 (2023). IEEE

[60] Grundy, S.M., Benjamin, I.J., Burke, G.L., Chait, A., Eckel, R.H., Howard, B.V., Mitch, W., Smith Jr, S.C., Sowers, J.R.: Diabetes and cardiovascular disease: a statement for healthcare professionals from the american heart association. Circulation **100**(10), 1134–1146 (1999)

[61] Sharma, A., Mittal, S., Aggarwal, R.,

Chauhan, M.K.: Diabetes and cardiovascular disease: inter-relation of risk factors and treatment. Future Journal of Pharmaceutical Sciences **6**(1), 1–19 (2020)

[62] Yusuf, S., Joseph, P., Rangarajan, S., Islam, S., Mente, A., Hystad, P., Brauer, M., Kutty, V.R., Gupta, R., Wielgosz, A., *et al.*: Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (pure): a prospective cohort study. The Lancet **395**(10226), 795–808 (2020)

[63] Wannamethee, S.G., Shaper, A.G., Whincup, P.H., Lennon, L., Sattar, N.: Impact of diabetes on cardiovascular disease risk and all-cause mortality in older men: influence of age at onset, diabetes duration, and established and novel risk factors. Archives of internal medicine **171**(5), 404–410 (2011)

[64] Jong, M., Woodward, M., Peters, S.A.: Duration of diabetes and the risk of major cardiovascular events in women and men: A prospective cohort study of uk biobank participants. Diabetes Research and Clinical Practice **188**, 109899 (2022)

[65] Hirakawa, Y., Ninomiya, T., Kiyohara, Y., Murakami, Y., Saitoh, S., Nakagawa, H., Okayama, A., Tamakoshi, A., Sakata, K., Miura, K., *et al.*: Age-specific impact of diabetes mellitus on the risk of cardiovascular mortality: An overview from the evidence for cardiovascular prevention from observational cohorts in the japan research group (epoch-japan). Journal of epidemiology **27**(3), 123–129 (2017)

[66] Khan, S.S., Ning, H., Wilkins, J.T., Allen, N., Carnethon, M., Berry, J.D., Sweis, R.N., Lloyd-Jones, D.M.: Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. JAMA cardiology **3**(4), 280–287 (2018)

[67] Held, C., Hadziosmanovic, N., Aylward, P.E., Hagström, E., Hochman, J.S., Stewart, R.A., White, H.D., Wallentin, L.: Body mass index and association with cardiovascular outcomes in patients with stable coronary heart disease–a stability substudy. Journal of the American Heart Association **11**(3), 023667 (2022)

[68] Chen, Y., Copeland, W.K., Vedanthan, R., Grant, E., Lee, J.E., Gu, D., Gupta, P.C., Ramadas, K., Inoue, M., Tsugane, S., et al.: Association between body mass index and cardiovascular disease mortality in east asians and south asians: pooled analysis of prospective data from the asia cohort consortium. Bmj **347** (2013)

[69] Iyen, B., Weng, S., Vinogradova, Y., Akyea, R.K., Qureshi, N., Kai, J.: Long-term body mass index changes in overweight and obese adults and the risk of heart failure, cardiovascular disease and mortality: a cohort study of over 260,000 adults in the uk. BMC Public Health **21**(1), 1–13 (2021)

[70] Powell-Wiley, T.M., Poirier, P., Burke, L.E., Després, J.-P., Gordon-Larsen, P., Lavie, C.J., Lear, S.A., Ndumele, C.E., Neeland, I.J., Sanders, P., *et al.*: Obesity and cardiovascular disease: a scientific statement from the american heart association. Circulation **143**(21), 984–1010 (2021)

[71] Dikaiou, P., Björck, L., Adiels, M., Lundberg, C.E., Mandalenakis, Z., Manhem, K., Rosengren, A.: Obesity, overweight and risk for cardiovascular disease and mortality in young women. European journal of preventive cardiology **28**(12), 1351–1359 (2021)

[72] Nidhi Sinha, A.J., Mohanty, S.P.: 4th International Conference on Data Science and Applications (ICDSA), Jaipur, India (2023). Malaviya National Institute of Technology