

# Cardiac Health Prediction using Electrocardiography



Pratishtha Agnihotri, Monika Jain

**Abstract:** Amongst various physiological signals, that can be collected from the human body, Electrocardiogram (ECG) is one widely used signal that gives an overview of individual's health non-invasively. Some prognostic tools, based on ECG, have already been introduced in the past. However, the diagnostic information contained in ECG is still under used. In the present study, we propose an algorithm that predicts the cardiac health (both present and future) by analyzing subject's ECG. The prediction is based on diagnostic information like Blood Pressure (BP), Arrhythmia and Heart Rate Variability (HRV), where BP and Arrhythmia are used to predict the present cardiac health, and Arrhythmia and HRV are used to predict the future cardiac health associated with an individual. To verify the algorithm, we use: (1) Linear Regression Model to extract BP based on parameters extracted from ECG; (2) Neural Network Pattern Recognition Application to detect Arrhythmia- Right and Left bundle branch block beat, Atrial premature contraction beat, Premature ventricular contraction beat and Premature or ectopic supraventricular beats, in any ECG signal; (3) Self-Organized Maps for HRV analysis using ECG. These models are used on ECG of 30 subjects chosen from an existing database. Based on the outputs of these models our algorithm predicts the present as well as the future cardiac health of 30 subjects under study. Our predictions are compared with the present and future cardiac health of these subjects already documented in the database. The prediction accuracy showed that present and future cardiac health risk of an individual can be satisfactorily determined using the proposed algorithm, which, in future, can be easily incorporated in any health monitoring device which can record ECG.

**Keywords:** Blood Pressure; Arrhythmia; Heart rate variability; Cardiac health monitoring.

## I. INTRODUCTION

Developing countries often face challenges like deforestation, pollution, urbanization and resource limitation. All these factors lead to a change in the socioeconomic conditions of the country. In developing countries like India, such rapidly changing socioeconomic conditions have forced the masses towards an insalubrious lifestyle. Consequently, changed habits and practices, directly or indirectly, affects the cardiovascular system and makes people prone to cardiovascular diseases (CVDs).

**Revised Manuscript Received on 30 July 2019.**

\* Correspondence Author

**Pratishtha Agnihotri\***, Department of Electronics And Communication, PSIT College of Engineering, Kanpur, (U.P.), India. E-mail: pratishtha.edu@gmail.com

**Monika Jain**, Department of Information Technology, IIITD, New Delhi, India. E-mail:monikaj@iiitd.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license [http://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Retrieval Number: B3284078219/2019@BEIESP

DOI: 10.35940/ijrte.B3284.078219

Journal Website: [www.ijrte.org](http://www.ijrte.org)

This made CVD one of the major causes of increase in morbidity and premature mortality in developing countries [1]. There are some CVDs which shows symptoms before actually posing any threat to life. In such cases, medical help can be provided to an individual according to the symptoms. But the situation worsens when an individual is carrying an asymptomatic CVD. In such cases, the subject may not even be aware of having a CVD and might ignore the medical help. About 1 to 2 million middle-aged men have asymptomatic but physiologically significant coronary artery disease, which puts them at increased risk of CVD events [2, 3]. Fortunately, indications of symptomatic as well as asymptomatic CVDs can be obtained by analysis of several physiological signals like Electrocardiogram (ECG), Photoplethysmogram (PPG), Phonocardiogram (PCG) and Impedance Cardiogram (ICG), without the help of trained medical practitioner. This algorithm can be incorporated in any physiological signal acquisition device which can record ECG. Since there are many low cost, non-invasive and compact ECG collecting devices already available in the market, no extra efforts are required to fabricate an entirely new device. Such device will make the subject aware about his cardiac health before any visible symptom occurs. Also, it would eliminate the requirement of a medical practitioner during the pre- and post-hospitalization care.

## II. RELATED WORK

There are many concealed indications which a human body gives during every health advancement and deterioration. These indications can be analysed to obtain a reliable overview of an individual's cardiac health. One such widely considered foreteller of cardiac health is BP. Some studies have related high BP (Hypertension) with CVDs and have concluded that hypertensive adults are more likely to have CVDs [5, 6]. Consequently, using BP as one of the cardiac health prognosticators, in any study, can be conducive. Due to the high clinical importance of BP, in past few decades, a number of BP prediction methods have already been introduced, with the main agenda of predicting BP based on the parameters extracted from cardiac signals (ECG, PPG, Oscillometric waveform, ICG) [7-10]. Many researchers have proved that reliable BP estimation can be obtained using only one cardiac signal also [11-13]. Bussink *et al.* in [14] and Fernández-Lozano *et al.* in [15], conducted a study that focused on the relationship of right bundle branch block beats with the health of a person. Tabrizi *et al.* in [16] concluded that 1/3 of the heart failure patients suffers from left bundle branch block beat and requires hospitalization.



Published By:

Blue Eyes Intelligence Engineering  
& Sciences Publication

Such patients are associated with very high mortality. Murakoshi *et al.* in [17] and Binici *et al.* in [18] showed that supraventricular complexes are dangerous and give rise to atrial fibrillation (AF) and CVDs, with which an individual could end up having primary death or stroke. Juhl-Møller *et al.* in [19] showed that premature ventricular contraction in patients who already had myocardial infarction was found to predict death. Thus, the presence of arrhythmias indicates present as well as future cardiac health risks. Just like BP and Arrhythmia, another parameter that relates highly to an individual's cardiac health is HRV- beat to beat fluctuation of the Heart Rate (HR), which generally decreases under stress situations, and increases with rest [20]. In 1996, The Task Force of the European Society of Cardiology and The North American Society of Pacing Electrophysiology came up with HRV standards of measurement [21]. They found a significant relationship between the autonomic nervous system (ANS) and cardiovascular mortality. It is a standard method to study the mechanism of ANS on heart functionality. Several studies have shown, statistical analysis of HRV is a powerful tool for evaluation of cardiovascular health and an independent risk factor for cardiovascular events, for example: Dekker *et al.* in [20] concluded that low HRV is associated with increased risk of Coronary heart disease and hypothesized that it is also a marker of less favourable health. Binici *et al.* in [22] concluded that nocturnal HRV is a strong marker for stroke in healthy subjects and reduced parasympathetic activity may increase the risk of stroke by increasing the risk of Arrhythmias. Hillebrand *et al.* in [23] performed a meta-analysis, and concluded that low HRV is associated with 32-45% increased risk of first cardiovascular event in population without known CVD. Nolan *et al.* in [24] concluded a reduction in SDNN (deviation of RR intervals) identifies patients at high risk of death and is a better predictor than conventional clinical measurements. These studies suggest that HRV is highly associated with future cardiac risks and can be safely used to obtain the likelihood of having a CVD in near future. Based on these results, some research groups have also developed tools for prognosis of CVDs, particularly using HRV. Melillo *et al.* in [25] identified high-risk subjects, who experience a vascular event in 12 months follow-up, using HRV measures analysed by a predictive model based on random forest. Krüger *et al.* in [26] evaluated the effect of HRV parameters on the prognostic value in addition to established parameters with Congestive Heart Failure (CHF). They concluded that HRV enhances prediction of mortality or deterioration of CHF and improves risk stratification. Ebrahimzadeh *et al.* in [27] proposed an approach for prediction of sudden cardiac death, using HRV features, classified by K-nearest neighbour and Multilayer Perceptron Neural Networks. Ramirez-Villegas *et al.* in [28] also used Neural Networks (NN) and support vector machines for prognosis of cardiovascular risk based on HRV measures. Many researchers have shown combined analysis of BP and Arrhythmias. Sideris *et al.* in [29] reviewed the relationship between BP and Ventricular Arrhythmias and suggested that increase in BP by any means may induce ventricular Arrhythmias, both, experimentally and in patients with a history of ventricular ectopic beats.

### III. METHODOLOGY

To prove our hypothesis, we designed an algorithm that predict present and future cardiac health of an individual using his BP, Arrhythmia and HRV, extracted from ECG signal. This algorithm uses the ECG signal of an individual for extracting his BP, detecting the type and frequency of the arrhythmias and analysing HRV. Based on the mentioned parameters, an individual is classified as healthy or unhealthy. The classifications done based on BP and Arrhythmia are used to predict the present cardiac health, whereas the classifications done based on Arrhythmia and HRV are used to predict the future cardiac health. The process flow of the algorithm is given in fig. 1 and each block of is subsequently discussed as follows:

#### A. ECG Pre-conditioning, Peak Detection and Feature Extraction

All ECG signals under study are pre-conditioned in MATLAB version 8.3.0.532 (R2014a) before using them in the analysis- DC component was removed, signals were normalized. Removal of baseline was done as mentioned in [30]. This was done to avoid any underlying noise component which could lead to faulty peak detection and hence faulty predictions dependent on detected peaks. For HRV analysis, the ECG signal are further fed into Kubios HRV software (version 2.2, University of Eastern Finland, Finland) [31]. For BP and arrhythmia, significant peaks like P, Q, R, S and T peaks of the ECG can be detected using any pre-existing optimized peak detection algorithm [32, 33]. These peaks were used to extract the parameters mentioned in Table I. All these parameters are used to for arrhythmia detection, whereas only three parameters marked green in Table I, viz. HR, QRS duration and PT duration are used for BP extraction using linear regression model. The models used for BP prediction, Arrhythmia detection and HRV analysis are discussed later in this paper. Based on the output of these models, 3-tier classification of an individual, whose ECG is under study, is done. The details of the classification are given under the following headings.

#### B. Classification Based on BP

There are a number of feature based BP prediction approaches, introduced recently. All the studies followed a common concept- Systolic and Diastolic BP prediction based on the parameters extracted from the cardiac signals like ECG and/or PPG [7-13, 33]. In our study, we predict beat to beat BP based on three parameters extracted from ECG- HR, QRS duration and PT duration. Once we have extracted beat to beat systolic and diastolic BP, depending on the value of this BP, each beat in the ECG is classified into one of the three categories- Healthy, Unhealthy and Very Unhealthy [34, 35], as shown in Table II.

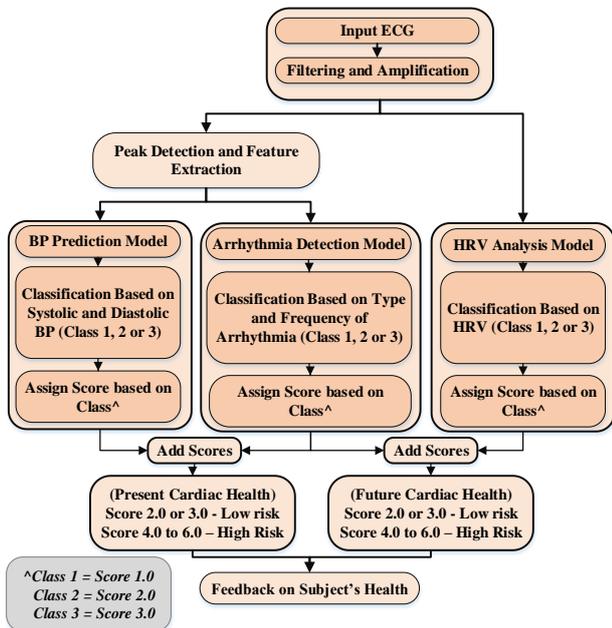


Figure 1. Block diagram showing process flow of the Proposed Algorithm

Table I. Parameters Extracted From Input ECG Signal

S.no	Parameters	Short Description
1	HR	Heart Rate
2	RR	Distance between two consecutive R-peaks
3	QRS	Distance between Q and S peak#
4	PT	Distance between P and T peak#
5	PR	Distance between P and R peak#
6	RT	Distance between R and T peak#
7	PS	Distance between P and S peak#
8	QT	Distance between Q and T peak#
9	ST	Distance between S and T peak#
10	PQ	Distance between P and Q peak#
11	P_AMP	Amplitude of P-peak
12	Q_AMP	Amplitude of Q-peak
13	R_AMP	Amplitude of R-peak
14	S_AMP	Amplitude of S-peak
15	T_AMP	Amplitude of T-peak

#In the same cardiac cycle

Table II. Identifying Healthy and Unhealthy Beats Based on BP [34, 35]

Beats	Systolic BP (mmHg)	Diastolic BP (mmHg)
Healthy	90-120	60-80
Unhealthy	120-140	60-90
Very Unhealthy	>140 or <90	>90 or <60

Based on the percentage of healthy and unhealthy beats in the ECG of any subject, he is classified into one of the three health classes- Class 1, Class 2 or Class 3 (where Class 1 is comparatively healthier than Class 2 and 3, and Class 2 is healthier than Class 3). A detailed description of this classification is given as follows:

Assuming that in a given time frame, total number of beats in the ECG of any individual is N, then, based on the

percentage of Healthy beats (H), Unhealthy beats (U) and Very Unhealthy beats (V), he can be classified as Class 1, 2 or 3 as shown in Table III.

Table III. Classification Based on BP

Class	Conditions
Class1	If H = 100% Or If U < 10% and/or If V < 5%
Class 2	If (10% < U < 40%) And / Or (5% < V < 20%)
Class 3	if U > 40% And / Or V > 20%

C. Classification Based on Arrhythmia

In this study, the beats in any ECG signals are classified into Normal beat (NB), Right bundle branch block (RBBB) beat, Left bundle branch block (LBBB) beat, Atrial premature contraction (APC) beat, Premature ventricular contraction (PVC) beat and Premature or ectopic supraventricular beat (SVPB) in any ECG signal using Neural Network (details given in next section). These are some popularly explored Arrhythmias to study the cardiac health. Based on the frequency and type of these Arrhythmias present subject's ECG signal, he can be classified into three health classes-Class 1, Class 2 and Class 3 (meaning of each Class explained earlier). The criteria for classification is decided based on prior domain knowledge [14-19, 29, 36-38] which says that a limited number of bundle branch blocks (BBBs) is common in normal population and can be safely ignored, however, beyond a certain limit, BBBs cannot be counted as incidental findings and need medical attention. Somewhat same medical significance as BBBs is assigned to APC and PVC, with a constraint that frequent APCs and PVCs can be an indication of underlying heart disease. Slightly higher risk is associated with SVPBs, which are less common as compared to other mentioned Arrhythmias. Excessive SVPBs may trigger bursts of AF. Based on the already explored risks associated with different kinds of arrhythmia, a subject is considered to be in the following classes:

- 1) Class 1: If his ECG satisfies any one of the conditions given in Table IV
- 2) Class 2: If his ECG satisfies two or more conditions (amongst condition 2, 3, 4, 5, 6) given in Table IV.
- 3) Class 3: If his ECG satisfies any one or more than one conditions given in Table V

Table IV. Arrhythmia Based Classification Criteria For Class 1 and 2 Subjects [14-19, 29, 36-38]

S.no	Conditions
1	If all beats are normal
2	If RBBB < 50 beats/24 hours*
3	If LBBB < 50 beats/24 hours*
4	If APC < 50 beats/24 hours*
5	If PVC < 50 beats/24 hour*
6	If SVPB < 30 beats/24 hours*

All these classification criteria are generic and can be directly used for any other dataset. This is also true for classification based on BP and HRV analysis.

Table V. Arrhythmia Based Classification



Criteria for Class 3 Subjects [14-19, 29, 36-38]

S.no	Conditions
1	If RBBB > 50 beats/24 hours*
2	If LBBB > 50 beats/24 hours*
3	If APC > 50 beats/24 hours*
4	If PVC > 50 beats/24 hour*
5	If SVPB > 30 beats/24 hours*

D. Classification Based on HRV

ECG signals from different subjects are taken from multiple databases (explained in detail in Section IV (C)), to form a single dataset, which is a mix of healthy and unhealthy subjects. Further, using Kubios software, HRV parameters are extracted from ECG signal. It is well identified that time domain and frequency domain parameters of HRV decreases progressively with aging [39, 40]. Thus to ensure correct risk classification, all subjects are divided into three categories according to their age, and analysis is done separately for three categories. Subjects in their 20s and 30s were grouped in Category A (Young). The 40s, 50s and 60s aged subjects were grouped in Category B (Middle) while subjects age in 70s and 80s were kept in Category C (Old). Fig. 2 shows the boxplot of subjects in all three categories. We used various online available databases to train our model. Different database helped to create a diverse dataset of healthy and diseased subjects aged 20-85 years. To classify subjects on the basis of HRV analysis, we have used Self-Organizing Maps (SOM) on the dataset, which clusters healthy and unhealthy subjects separately as it contains diverse subjects with a different state of health and age. SOM is a type of Artificial Neural Network (ANN), used for unsupervised learning. They learn to cluster data on the basis of similarity and topology of the input vector. We have used default batch train algorithm in MATLAB. Through SOM, we have tried to cluster similar subjects into a group and ranked those groups with relative health. In our method, a total of six neurons are used, where neighboring neurons in the SOM learn to recognize neighboring sections of input space. As input to SOM, nine parameters (8 HRV parameters & age) are used. Three different SOMs are trained for subjects in Category A, B & C. Subjects are further classified into six classes, as in Fig. 3, shown for Category C subjects. Moreover, they are further reduced into three classes, by combining the classes that fall in the same range of total power – sum of the power in VLF, LF and HF. The new classes are formed by combining classes from SOM, represented in black rounded-corner boxes (fig. 3). Thus, the number of classes are reduced from six to three, as shown in fig. 4. Fig. 3 and 4 show the box-plot of one of the HRV parameter: Total power (ms<sup>2</sup>), compared in different classes obtained from SOM.

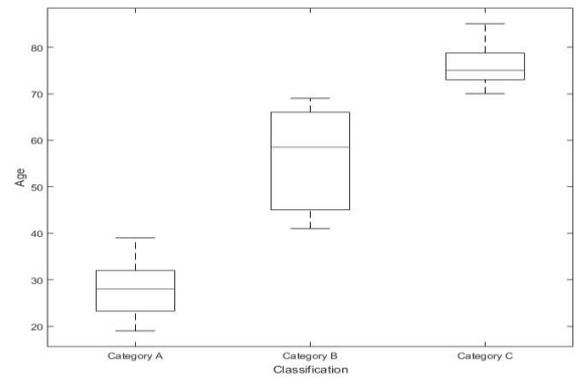


Fig 2. Age-based subject classification

As discussed in Section II, a low HRV is hypothesized as a marker of less favorable health, hence subjects classified in Class 1 possess higher HRV values as compared to other two classes, and therefore, are at lesser risk relative to subjects in Class 2 and Class 3. Similarly, subjects in Class 3 have lower HRV values and hence are at a higher risk than subjects in first two classes.

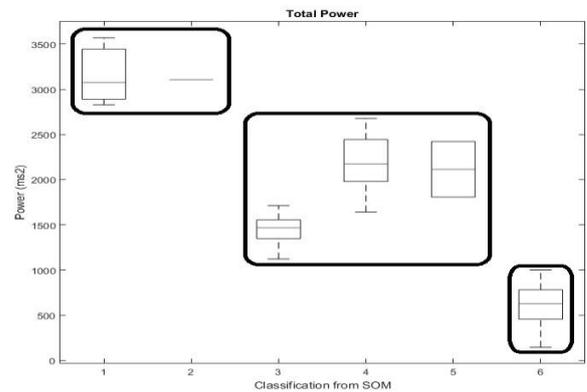


Figure 3. Category C – SOM based classification

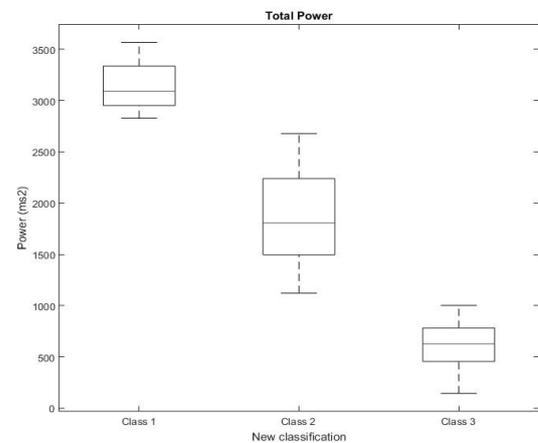


Figure 4. Category C – new classification

Present Cardiac health prediction

To predict present cardiac health, we use the classification from BP and Arrhythmia. Each subject is assigned a particular score, based on which Class they fall in.



For subjects in Class 1, a score of 1.0 is given, for Class 2 - 2.0 and Class 3 - 3.0. Hence, each subject gets two separate scores on the basis of BP and Arrhythmia classification, which are both added. The present cardiac health is predicted based on the value of this score (which lies between 2.0 to 6.0). A score of 2.0 and 3.0 is considered as low risk, and a score 4.0 to 6.0 is considered as high risk. The interpretation drawn from risk based on score are as follows:

- Low risk: The BP of the subject is in normal ranges, with no or very fewer Arrhythmias. No medical attention is required at this stage.
- High risk: The BP of the subject is beyond the acceptable ranges, with many irregular beats. The subject is advised seek medical help.

**E. Future cardiac health prediction**

In a similar way, as done for present cardiac health prediction, we use the classification done for arrhythmia and HRV for future cardiac health prediction. The score for the subject is calculated by adding scores obtained through classification based on arrhythmia and HRV. A score of 2.0 and 3.0 is considered as low risk, and a score 4.0 to 6.0 is high risk. The interpretation drawn from risk based on score are as follows:

- Low risk: The HRV of the subject is normal with almost no irregular beats. The chances of having any major CVDs in near future are very less. No medical attention is required at the present stage.
- High risk: The HRV is beyond acceptable with many irregular beats. There are chances of having any major CVDs in near future. The subject is advised to seek medical help.

**VI. EXPERIMENTAL SET-UP AND RESULTS**

The proposed algorithm is independent of the BP prediction model, Arrhythmia detection model and HRV analysis model. In the end, the complete algorithm is tested on 30 subjects chosen from SHAREE database [4]. The details of each model are discussed as follows:

**A. Blood Pressure Prediction**

This section elaborates on development and validation of an efficient BP prediction model. ECG and measured systolic/diastolic BP of 72 subjects was chosen from two online available databases given below:

1) *The MIMIC II Clinical Database:* It contains clinical data from bedside workstations as well as hospital archives. The ECG signals sampled at 125 Hz and are available freely via PhysioNet ATM [4, 41]. In this study, ECG and BP of 40 subjects were chosen from this database.

2) *The University of Queensland Vital Signs Database:* It contains a wide range of patient monitoring data and vital signs that were recorded during 32 surgical cases where patients underwent anesthesia at the Royal Adelaide Hospital [42]. ECG signals were recorded at 100 Hz. ECG and BP of all the subjects from this database were used in this study. HR, QRS duration and PT duration was extracted from all the ECG signals under study. A database of 72 subjects was created using these extracted parameters and corresponding measured BP. The entire database is divided randomly into three equal subsets, containing the data of 24 subjects each. Linear regression was employed to model the relationship between extracted parameters and BP by using any two subsets of data (48 subjects), and testing was performed over remaining 24 test subjects. This process was

repeated thrice until each subset was used as testing subset once (three-fold cross validation). This developed regression model is capable of predicting beat to beat BP based on the parameters extracted from each beat. The accuracy of the model during each fold was measured in terms of Normalized mean square error (NMSE), Mean absolute error (MAE) and Error standard deviation (ESD) with respect to the standard BP already mentioned in the used databases (measured BP), as shown in Table VI.

**Table VI. Blood Pressure Prediction Error**

Fold	Systolic BP			Diastolic BP		
	NMSE (%)	MAE (mmHg)	ESD (mmHg)	NMSE (%)	MAE (mmHg)	ESD (mmHg)
I	4.19	4.14	5.13	7.25	4.58	5.50
II	4.24	4.12	5.17	4.83	2.90	3.82
III	6.09	5.57	6.96	6.52	4.08	5.22
Mean	4.84	4.61	5.75	6.20	3.85	4.85

The computed error for BP prediction falls under the standard allowable error mentioned by Association for the Advancement of Medical Instrumentation, that is, MAE and ESD to be less than or equal to 5 mmHg and 8 mmHg respectively [43]. Three fold cross validation verified that the model trained with the data of 48 test subjects can be safely used for BP prediction.

**B. Arrhythmia Detection**

Literature survey presented in Section II reflects that irregular heartbeats can certainly indicate the presence of underlying heart disease and health conditions. In this section, NN is used to classify each beat as NB, RBBB, LBBB, APC, PVC and SVPB.

The training of the NN was done using the following databases:

1) *The MIT-BIH Arrhythmia Database:* It consists of continuous ECG signals recorded for slightly more than 30 minutes. Each ECG file consists of samples recorded from two leads at 360 Hz. [4, 44].

2) *The MIT-BIH Malignant Ventricular Arrhythmia Database:* It consist of 22 recordings of thirty-five minute each, obtained from Holter tapes of 16 subjects [4, 44]. Each signal file consists of signals recorded from two leads at 250 Hz. Although, this database consists of annotations with respect to rhythm changes, we have used this data to study normal beats only.

3) *The MIT-BIH Supraventricular Arrhythmia Database:* It is used to study supraventricular arrhythmias. The records in this database were obtained from Holter tapes, and each file consist of two signals recorded at 128 Hz [4]. From the databases mentioned above, 15 parameters (shown in Table I) were extracted corresponding to each beat of the pre-conditioned ECG signal under study. We were able to extract the parameters of 155067 NBs, 5319 RBBBs, 4476 LBBBs, 2433 APCs, 9384 PVCs and 4728 SVPBs. Each type of beat was divided into three equal parts, where each part consisted of 51689 NBs, 1773 RBBBs, 1492 LBBBs, 811 APCs, 3128 PVCs and 1576 SVPBs. These parts were used for three-fold cross validation.

For classification, we have used NN pattern recognition



app. [45, 46], which uses a two-layer Feed Forward Network, with sigmoid hidden and softmax output neurons. It can classify vectors arbitrarily well, given enough neurons in its hidden layer. The network was trained with scaled conjugate gradient back-propagation. The input to NN was the 15 parameters extracted from the each beat, and the output was the type of Arrhythmia (or category) to which the beat belonged. The prediction accuracy during each fold is given in Table VII.

**Table VII. arrhythmia prediction accuracy of nn during each fold**

Type of Beat	Detection Accuracy (%)		
	Fold I	Fold II	Fold III
NB	100	100	100
RBBB	94.75	93.06	92.49
LBBB	95.84	93.83	94.50
APC	88.77	94.94	92.47
PVC	96.54	95.90	97.50
SVPB	92.00	93.90	93.27

Based on the abnormal beats possessed, an individual is assigned a Health Class which reflects the level of cardiac health. Details of the risk classification based on Arrhythmia are already discussed.

### C. HRV Analysis

As discussed in Section II, HRV is variation in the time interval between heart beats, a reliable parameter for future cardiac health prediction. The following sections describes the process of HRV analysis done, to obtain final classification. In order to form a single dataset, various online databases from physionet.org [4] were chosen, which are as follows: “*Fantasia Database*” contains 40 healthy screened subjects, with ECG signal recorded in the awake state. “*MIT-BIH Normal Sinus Rhythm Database*” includes 18 long-term ECG recordings of subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital. “*Data for Development and Evaluation of ECG-based APNEA detector*” database contains 10 controlled records, obtained from healthy volunteers during sleep, which contains single ECG signal. Additional information recorded was age, sex, height, weight of the subject [47]. “*Smart Health for assessing the risk of events via ECG (SHAREE)*” included 24-h ECG Holter recordings of 139 hypertensive patients recruited at the Centre of Hypertension of the University Hospital of Naples Federico II, Naples, Italy. Additional information recorded was Gender, Age, Weight, Height, BSA, BMI, Smoking habits, BP, and any vascular event in one-year follow-up [25]. “*Combined measurement of ECG, Breathing and seismocardiogram (CEBS)*” is constructed by measuring ECG signal of 20 healthy subjects. Additional information recorded: age, sex, smoking habits, recent coffee intake and healthy/sedentary lifestyle [48]. “*The MIT-BIH Arrhythmia database*” contains 48 half-hour ECG recording, with many subjects having Cardiac Arrhythmia. All beats are annotated, with Arrhythmia type [37, 44]. By using the data from all the datasets mentioned above, a mixed dataset is created, which contains subjects who are completely healthy, ones with Cardiac Arrhythmia and those who suffered from a vascular event in one year after ECG sample was taken. It also

consists of obese subjects and subjects with abnormally high or low BP levels. Different databases had varied clinical information available for each subject, which was used to validate the developed model. One common screening applied to all records was the minimum availability of 5 minutes of ECG signal. A number of standard HRV measures [21] were calculated: Average heart rate of 5 min duration (Mean RR), standard deviation of RR intervals (SDNN), number of successive intervals differing more than 50 ms taken as percentage of total RR intervals (pNN50), HRV triangular index is the proportion of all accepted RR intervals to their modal measurement at a discrete scale of 1/128s bins (HRVTi). Few frequency domain HRV measures were also calculated: Total power under very low frequency band (0-0.04 Hz) , low frequency band (0.04-0.15 Hz), high frequency band ( 0.15-0.4 Hz ) using PSD of RR interval series estimated for each frequency band and power ratio of total power under low frequency to high frequency (LF/HF). Before any signal processing, ECG signals were carefully selected. Signal intervals having more than 5% ectopic beats from total beats were visually identified and removed. Thus, ECG of 5 minutes duration was identified with only normal beats. Only normal beats are considered for analysis, due to the fact that HRV analysis is done only for normal-to-normal intervals, which are resulting from sinus node depolarization [21]. Including ectopic beats leads to inaccurate HRV analysis. Intervals with less than 5% of ectopic beat duration were accepted to control the error. To maintain uniformity, signals with sampling frequency more than 500 Hz were down-sampled to 500 Hz. Total 140 subjects cleared the initial selection process and were used for further analysis. In order to analyse HRV, the processed ECG signals are further fed in Kubios HRV software (version 2.2). Mainly time domain and frequency domain features are analysed and recorded for each subject. R-peak detection in Kubios HRV software was visually inspected through R-R interval plot, where visible sudden spikes indicated R-peak detection error. Further to classify subjects into risk categories, Self-organizing maps are used to classify subjects into three classes as described in Section III. The relative risk classification explained in Section III can also be validated by analysing the clinical information available from databases. Subjects classified in Class 3 by SOM are the ones who suffered from a vascular event such as myocardial infarction, stroke or syncope (in one year follow up after taking ECG sample). All subjects with tachycardia are also classified in Class 3. Moreover, subjects suffering from serious conditions in Cardiac Arrhythmia, Obese class 2 (BMI>35) also falls under Class 3. Healthy screened patients given in databases falls into Class 1 and Class 2. In 140 subjects, three SOMs are trained, for three Categories A, B & C. Fig 5-7, shows the plots for Total power, SDNN and pNN50% respectively, showing relative parameters values for all three health classes across all three age category subjects.

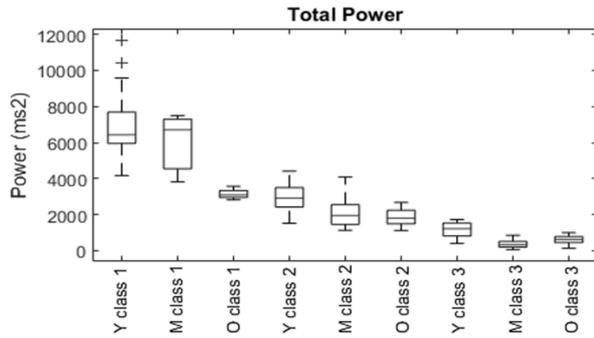


Figure 5. Total power Y- Young (Category A), M- Middle (Category B), O-Old (Category C)

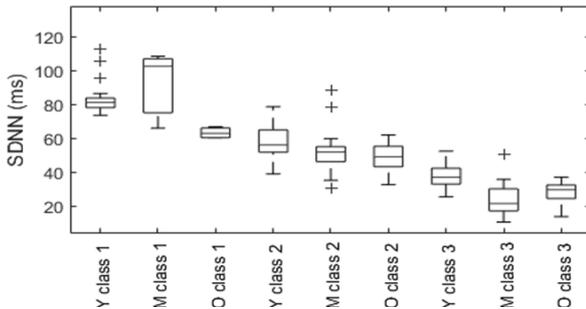


Figure 6. SDNN Y- Young (Category A), M- Middle (Category B), O-Old (Category C)

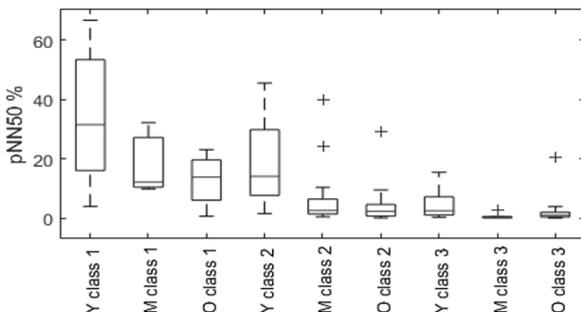


Figure 7. pNN50% Y- Young (Category A), M- Middle (Category B), O-Old (Category C)

As expected, Class 1 has higher HRV value, as compared to Class 2 and Class 3. Also, HRV values decreases, from Category A to Category C. Class 3 subjects have the least HRV values of total power, SDNN, pNN50 and HRVTi. They were hence at larger risk as compared to Class 1 and Class 2 subjects. Nolan et al. (1998) in [24] also shows, that subjects with SDNN less than 50ms (during 24-hour HRV analysis) had an annual mortality rate of 51.4% in 800 days follow up. But, subjects with SDNN between 50-100 ms had 12.7% annual mortality rate, and it decreases to 5.5 % annually for subjects with SDNN greater than 100ms. Similarly, Class 3 subjects in our experiment, also have less than 50ms of SDNN, and are kept at highest risk. Table VIII shows the wise category accuracy of NN SOM. Test cases were chosen from the dataset. For testing purposes, different subjects are chosen, and we classified them based on their risk from given clinical information. Subjects with a vascular event in one year follow up, or within twenty-four hours of ECG sample, with serious Cardiac Arrhythmia are classified as Class 3 subjects. Subjects with abnormally high or low BP as well as obese (BMI> 35) are also classified as class 3 subjects. Normal subjects with a sedentary lifestyle

and no known disease are classified as Class 2. Subjects with a healthy lifestyle, free from any disease are classified in Class 1.

Table VIII. Category wise Accuracy of NN SOM

Category	No of test cases	Accuracy
A	10	90.00%
B	17	94.12 %
C	18	94.44 %

D. Results

The proposed algorithm is tested on 30 subjects chosen from SHAREE database, which was originally developed in order to identify hypertensive subjects at higher risk of having vascular events based on HRV analysis [4]. Each signal file consists of three ECG signals each sampled at 128 Hz with 8-bit precision. This database provides other clinical information such as sex, age, weight, height, BMI, BSA, smoking habits, BP, with one-year follow-up for the occurrence of any vascular event. The ECG of the chosen subjects was used as the input to our algorithm. Before using the ECG of any subject, the ECG selection constraints mentioned in Section IV(C) were considered. After pre-conditioning the ECG, steps mentioned below were followed:

- 1) 15 parameters listed in Table I were extracted from each beat. BP prediction model developed in Section IV(A) was used to extract beat to beat BP based on HR, QRS duration and PT duration. The classification criteria in Section III(B) was used to tag the subject as Class 1, 2 or 3.
- 2) The extracted 15 parameters were fed to the NN trained in Section IV(B). Based on the type(s) and frequency of abnormal beats the subject was classification into Class 1, 2 or 3, as discussed in Section III(C).
- 3) HRV analysis was done to extract different parameters from the ECG, which are fed as input to the SOM to classify the subject into Class 1, 2 or 3 as discussed in Section III(D).

For all three models of BP, Arrhythmia and HRV, Class 1 is at the least risk as compared to Class 2 and 3, and Class 3 at highest risk. Each subject is given a score on the basis of Class which it falls in. Based on the scores, the present and future cardiac health/risk is decided, as already discussed. The present and future cardiac health prediction accuracy of the proposed algorithm is given in Table IX and X respectively.

Table IX. Accuracy of the Algorithm While Predicting Present Cardiac Health

	Actual No. of Subjects	No. of Subjects Predicted Correctly by the Proposed Algorithm	Prediction Accuracy (%)
Low Risk	16	15	93.75
High Risk	14	14	100
Total	30	29	99.67

Table X. Accuracy of the Algorithm While



## Predicting Future Cardiac Health

	Actual No. of Subjects	No. of Subjects Predicted Correctly by the Proposed Algorithm	Prediction Accuracy (%)
Low Risk	10	10	100
High Risk	20	16	80
Total	30	26	86.67

Looking at the results, we can say that a detailed analysis of ECG signals can satisfactorily predict present as well as future cardiac health.

**Table IX. Number of Low Risk And High-Risk Subjects Classified**

	Present Risk	Future Risk
Low risk Subjects	16	10
High-risk Subjects	14	20
Total Subjects	30	30

Out of 16 low present risk subjects, 9 as low future risk subjects and remaining 7 subjects were classified as a high future risk. From 14 high present risk subjects, 13 subjects were classified as high future risk subject and only single subject into low future risk. Therefore, out of 30, 22 subjects had same present and future risk, while 8 subjects had different present and future risk.

A comparison between the predictions obtained using the proposed algorithm with the follow-up provided in the test databases is shown as follows:

- Out of 20 high future cardiac risk subjects, we successfully predicted that 16 subjects will suffer from a cardiovascular event in one-year follow-up, giving a sensitivity of 80%.
- Out of 10 low future cardiac risk, no subject reports any cardiovascular event in one-year follow-up, giving a specificity of 100%.
- Out of 8 obese subjects (BMI > 30), 6 subjects are classified as high present cardiac risk (sensitivity of 75%), and 7 subjects as high future cardiac risk (sensitivity of 87.5%)
- Out of 8 very high Systolic BP (>150 mmHg), all are classified as high present and future cardiac risk (sensitivity of 100%)
- Out of 16 low present cardiac risk subjects, 15 subjects have their BP in the normal range.

## VII. CONCLUSION

This study proposes an algorithm that predicts present and future cardiac health of an individual using his ECG signals. The predictions are done by analyzing BP, Arrhythmias and HRV, extracted from the ECG. BP and Arrhythmias are analyzed conjointly to evaluate the present cardiac health and Arrhythmia and HRV to evaluate the future cardiac health. The proposed algorithm is tested on 30 subjects chosen from SHAREE database. The prediction accuracy of the algorithm suggests that it can be satisfactorily used to predict the presence of any underlying CVD. It can also be used to alarm/alert the user about the CVDs which he may encounter in near future, without any help from trained medical practitioner. In future, the proposed algorithm can be easily incorporated in any physiological signal acquisition system which can record ECG.

## REFERENCES

1. Shah, Bela, and Prashant Mathur. "Surveillance of cardiovascular disease risk factors in India: the need & scope." *The Indian journal of medical research* 132, no. 5 (2010): 634.
2. Thaulow E, Erikssen J, Sandvik L, Erikssen G, Jorgensen L, Cohn PF. Initial clinical presentation of cardiac disease in asymptomatic men with silent myocardial ischemia and angiographically documented coronary artery disease (the Oslo Ischemia Study). *Am J Cardiol.* 1993;72:629-633.
3. Cohn PF. Detection and prognosis of the asymptomatic patient with silent myocardial ischemia. *Am J Cardiol.* 1988;61:4B-6B.
4. Goldberger, Ary L., Luis AN Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. "Physiobank, physiotookit, and physionet components of a new research resource for complex physiologic signals." *Circulation* 101, no. 23 (2000): e215-e220.
5. Perkovic, Vlado, Rachel Huxley, Yangfeng Wu, Dorairaj Prabhakaran, and Stephen MacMahon. "The burden of blood pressure-related disease a neglected priority for global health." *Hypertension* 50, no. 6 (2007): 991-997.
6. Ueshima, Hirotugu, Akira Sekikawa, Katsuyuki Miura, Tanvir Chowdhury Turin, Naoyuki Takashima, Yoshikuni Kita, Makoto Watanabe et al. "Cardiovascular disease and risk factors in Asia a selected review." *Circulation* 118, no. 25 (2008): 2702-2709.
7. Fukushima, H.; Kawanaka, H.; Bhuiyan, M.S.; Oguri, K., "Cuffless blood pressure estimation using only photoplethysmography based on cardiovascular parameters," *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, vol., no., pp.2132,2135, 3-7 July 2013.
8. Sujay Deb; Nanda, C.; Goswami, D.; Mukhopadhyay, J.; Chakrabarti, S., "Cuff-Less Estimation of Blood Pressure Using Pulse Transit Time and Pre-ejection Period," *Convergence Information Technology, 2007. International Conference on*, vol., no., pp.941,944, 21-23 Nov. 2007.
9. Forouzanfar, M.; Dajani, H.R.; Groza, V.Z.; Bolic, M.; Rajan, S., "Feature-Based Neural Network Approach for Oscillometric Blood Pressure Estimation," *Instrumentation and Measurement, IEEE Transactions on*, vol.60, no.8, pp.2786,2796, Aug. 2011.
10. Jadooei, A.; Zaderykhin, O.; Shulgin, V.I., "Adaptive algorithm for continuous monitoring of blood pressure using a pulse transit time," *Electronics and Nanotechnology (ELNANO), 2013 IEEE XXXIII International Scientific Conference*, vol., no., pp.297,301, 16-19 April 2013.
11. Samria, R.; Jain, R.; Jha, A.; Saini, S.; Chowdhury, S.R., "Noninvasive cuffless estimation of blood pressure using Photoplethysmography without electrocardiograph measurement," *Region 10 Symposium, 2014 IEEE*, vol., no., pp.254,257, 14-16 April 2014.
12. Suzuki, A.; Ryu, K., "Feature Selection Method for Estimating Systolic Blood Pressure Using the Taguchi Method," *Industrial Informatics, IEEE Transactions on*, vol.10, no.2, pp.1077,1085, May 2014.
13. Soojeong Lee; Joon-Hyuk Chang; Sang Won Nam; Chungsoo Lim; Rajan, S.; Dajani, H.R.; Groza, V.Z., "Oscillometric Blood Pressure Estimation Based on Maximum Amplitude Algorithm Employing Gaussian Mixture Regression," *Instrumentation and Measurement, IEEE Transactions on*, vol.62, no.12, pp.3387,3389, Dec. 2013.
14. Bussink, Barbara E., Anders G. Holst, Lasse Jespersen, Jaap W. Deckers, Gorm B. Jensen, and Eva Prescott. "Right bundle branch block: prevalence, risk factors, and outcome in the general population: results from the Copenhagen City Heart Study." *European heart journal* 34, no. 2 (2013): 138-146.
15. Fernández-Lozano, Ignacio, and Josep Brugada. "Right bundle branch block: are we looking in the right direction?." *European heart journal* (2012): ehs359.
16. Tabrizi, Fariborz, Anders Englund, Mårten Rosenqvist, Lars Wallentin, and Ulf Stenström. "Influence of left bundle branch block on long-term mortality in a population with heart failure." *European heart journal* 28, no. 20 (2007): 2449-2455.
17. Murakoshi, Nobuyuki, DongZhu Xu, Toshimi Sairenchi, Miyako Igarashi, Fujiko Irie, Takuji Tomizawa, Hiroshi Tada et al.



"Prognostic impact of supraventricular premature complexes in community-based health checkups: the Ibaraki Prefectural Health Study." *European heart journal* 36, no. 3 (2015): 170-178.

18. Binici, Zeynep, Theodoros Intzilakis, Olav Wendelboe Nielsen, Lars Køber, and Ahmad Sajadieh. "Excessive supraventricular ectopic activity and increased risk of atrial fibrillation and stroke." *Circulation* 121, no. 17 (2010): 1904-1911.
19. Juhl-Møller S, Lilja B, Johansson BW. Ventricular Arrhythmias and left ventricular function: one year of follow-up after myocardial infarction. *Eur Heart J* 1988;9:1181-1187.
20. Dekker, Jacqueline M., Richard S. Crow, Aaron R. Folsom, Peter J. Hannan, Duanping Liao, Cees A. Swenne, and Evert G. Schouten. "Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes The ARIC Study." *Circulation* 102, no. 11 (2000): 1239-1244.
21. Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use, Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, *Circulation*. 1996;93:1043-1065.
22. Binici, Zeynep, Mette Rauhe Mouridsen, Lars Køber, and Ahmad Sajadieh. "Decreased nighttime heart rate variability is associated with increased stroke risk." *Stroke* 42, no. 11 (2011): 3196-3201.
23. Hillebrand, Stefanie, Karin B. Gast, Renée de Mutsert, Cees A. Swenne, J. Wouter Jukema, Saskia Middeldorp, Frits R. Rosendaal, and Olaf M. Dekkers. "Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: meta-analysis and dose-response meta-regression." *Europace* 15, no. 5 (2013): 742-749.
24. Nolan, James, Phillip D. Batin, Richard Andrews, Steven J. Lindsay, Paul Brooksby, Michael Mullen, Wazir Baig et al. "Prospective study of heart rate variability and mortality in chronic heart failure results of the United Kingdom heart failure evaluation and assessment of risk trial (UK-Heart)." *Circulation* 98, no. 15 (1998): 1510-1516.
25. Melillo, Paolo, Raffaele Izzo, Ada Orrico, Paolo Scala, Marcella Attanasio, Marco Mirra, Nicola De Luca, and Leandro Pecchia. "Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis." *PLoS one* 10, no. 3 (2015): e0118504.
26. Krüger, Carsten, T. Lahm, C. Zugck, R. Kell, D. Schellberg, M. W. F. Schweizer, W. Kübler, and M. Haass. "Heart rate variability enhances the prognostic value of established parameters in patients with congestive heart failure." *Zeitschrift für Kardiologie* 91, no. 12 (2002): 1003-1012.
27. Ebrahimzadeh Elias, Mohammad Pooyan, and Ahmad Bijar. "A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals." *PLoS one* 9, no. 2 (2014): e81896.
28. Ramirez-Villegas, Juan F., Eric Lam-Espinosa, David F. Ramirez-Moreno, Paulo C. Calvo-Echeverry, and Wilfredo Agredo-Rodriguez. "Heart rate variability dynamics for the prognosis of cardiovascular risk." *PLoS One* 6, no. 2 (2011): e17060.
29. Sideris, D. A. "High blood pressure and ventricular Arrhythmias." *European heart journal* 14, no. 11 (1993): 1548-1553.
30. Ruiping Wang; Wenyan Jia; Zhi-Hong Mao; Sciabassi, R.J.; Mingui Sun, "Cuff-free blood pressure estimation using pulse transit time and heart rate," *Signal Processing (ICSP)*, 2014 12th International Conference on , vol., no., pp.115,118, 19-23 Oct. 2014.
31. Reyes, I.; Nazeran, H.; Franco, M.; Haltiwanger, E., "Wireless photoplethysmographic device for heart rate variability signal acquisition and analysis," in *Engineering in Medicine and Biology Society (EMBC)*, 2012 Annual International Conference of the IEEE , vol., no., pp.2092-2095, Aug. 28 2012-Sept. 1 2012
32. Mahmoodabadi, S.Z.; Ahmadian, A.; Abolhasani, M.D.; Eslami, M.; Bidgoli, J.H., "ECG Feature Extraction Based on Multiresolution Wavelet Transform," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the , pp.3902-3905, 17-18 Jan. 2006*
33. Rooijackers, M.J.; Rabotti, C.; Bennebroek, M.; Van Meerbergen, J.; Mischi, M., "Low-complexity R-peak detection in ECG signals: A preliminary step towards ambulatory fetal monitoring," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE , vol., no., pp.1761-1764, Aug. 30 2011-Sept. 3 2011*
34. Description of High blood pressure, from <http://www.nhlbi.nih.gov/health/health-topics/topics/hbp>
35. Blood pressure chart, from <http://www.bloodpressureuk.org/BloodPressureandyou/Thebasics/Bloodpressurechart>
36. Mayo Clinic Family Health Book, 4th Edition
37. Suzuki, Shinya, Koichi Sagara, Takayuki Otsuka, Hiroto Kano, Shunsuke Matsuno, Hideaki Takai, Tokuhisa Uejima et al. "Usefulness of frequent supraventricular extrasystoles and a high CHADS 2 score to predict first-time appearance of atrial fibrillation." *The American journal of cardiology* 111, no. 11 (2013): 1602-1607.
38. Manolis, Antonis S., Hugh Calkins, and Brian C. Downey. "UpToDate", Publisher: Wolters Kluwer, Edition-2014.
39. Voss, Andreas, Rico Schroeder, Andreas Heitmann, Annette Peters, and Siegfried Perz. "Short-Term Heart Rate Variability—Influence of Gender and Age in Healthy Subjects." *PLoS one* 10, no. 3 (2015).
40. Dietrich, Denise Felber, Christian Schindler, Joel Schwartz, Jean-Claude Barthélémy, Jean-Marie Tschopp, Frédéric Roche, Arnold von Eckardstein et al. "Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the SAPALDIA study." *Europace* 8, no. 7 (2006): 521-529.
41. Gari D. Clifford, Daniel J. Scott, and Mauricio Villarroel, "User Guide and Documentation for the MIMIC 11 Database," MIMIC-11 database version 2.6, pp.37, 2012.
42. Liu D, Görges M, Jenkins SA, "University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research", *Anesth Analg*. 2012 Mar; 114(3):584-9.
43. Association for the Advancement of Medical Instrumentation. Manual, Electronic or Automated Sphygmomanometer. Arlington, Virginia, USA: Association for the Advancement of Medical Instrumentation. American National Standard ANSI/AAMI SP10:2002, 2003, pp. 1-70
44. Moody, George B., and Roger G. Mark. "The MIT-BIH Arrhythmia database on CD-ROM and software for use with it." In *Computers in Cardiology 1990, Proceedings*, pp. 185-188. IEEE, 1990.
45. Basu, Jayanta Kumar, Debnath Bhattacharyya, and Tai-hoon Kim. "Use of artificial neural network in pattern recognition." *International journal of software engineering and its applications* 4, no. 2 (2010).
46. Niwas, S., R. Kumari, and V. Sadasivam. "Artificial neural network based automatic cardiac abnormalities classification." In *Computational Intelligence and Multimedia Applications, 2005. Sixth International Conference on*, pp. 41-46. IEEE, 2005.
47. Penzel, T., G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter. "The apnea-ECG database." In *Computers in Cardiology 2000*, pp. 255-258. IEEE, 2000.
48. García-González, M. A., A. Argelagós, Mireya Fernández-Chimeno, and J. Ramos-Castro. "Differences in QRS locations due to ECG lead: relationship with breathing." In *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, pp. 962-964. Springer International Publishing, 2014.

## AUTHORS PROFILE

**Pratishtha Agnihotri** is working as Assistant Professor at PSIT College of Engineering, Kanpur. She has four years of experience in academics and



has been working in the field of VLSI, transistors and biomedical engineering. She has completed her Masters from Institute of Engineering & Technology, Lucknow, India, in 2015, with specialization in Microelectronics. She has completed her Bachelors from Dr. APJ Abdul Kalam Technical University, Lucknow, in 2012, with specialization in Electronics & Communication Engineering.

**Monika Jain** is a PhD student at Queensland University of Technology, Queensland, Australia and Indraprastha Institute of Information



Technology, Delhi, India (IIITD), under a joint PhD program. She has completed her Masters from Institute of Engineering & Technology, Lucknow, India, in 2015, with specialization in Microelectronics. She has completed her Bachelors from Dr. APJ Abdul Kalam Technical University, Lucknow, in 2012, with specialization in Electronics & Communication Engineering. Before PhD, during 2015-16, she has worked as research assistant at IIITD, on a Government of India funded project that focused on biomedical signal capturing and processing for cardiovascular disease diagnosis.