Short Communication

Signal Processing Framework for the Detection of Ventricular Ectopic Beat Episodes

Avvaru Srinivasulu[1,2,3], Natarajan Sriraam[4]

[1]Research Scholar, Center for Medical Electronics and Computing, M.S Ramaiah Institute of Technology, [3]Department of Electrical, Electronics and Communication Engineering, GITAM, Bangalore Campus, [4]Center for Medical Electronics and Computing, M.S Ramaiah Institute of Technology, Bengaluru, [2]Department of Medical Electronics Engineering, MSRIT, Affiliated to VTU, Belgaum, Karnataka, India

Abstract

The Holter monitor captures the electrocardiogram (ECG) and detects abnormal episodes, but physicians still use manual cross-checking. It takes a considerable time to annotate a long-term ECG record. As a result, research continues to be conducted to produce an effective automatic cardiac episode detection technique that will reduce the manual burden. The current study presents a signal processing framework to detect ventricular ectopic beat (VEB) episodes in long-term ECG signals of cross-database. The proposed study has experimented with the cross-database of open-source and proprietary databases. The ECG signals were preprocessed and extracted the features such as pre-RR interval, post-RR interval, QRS complex duration, QR slope, and RS slope from each beat. In the proposed work, four models such as support vector machine, k-means nearest neighbor, nearest mean classifier, and nearest RMS (NRMS) classifiers were used to classify the data into normal and VEB episodes. Further, the trained models were used to predict the VEB episodes from the proprietary database. NRMS has reported better performance among four classification models. NRMS has shown the classification accuracy of 98.68% and F1-score of 94.12%, recall rate of 100%, specificity of 98.53%, and precision of 88.89% with an open-source database. In addition, it showed an accuracy of 99.97%, F1-score of 94.54%, recall rate of 98.62%, specificity of 99.98%, and precision of 90.79% to detect the VEB cardiac episodes from the proprietary database. Therefore, it is concluded that the proposed framework can be used in the automatic diagnosis system to detect VEB cardiac episodes.

Keywords: *Cardiac episode detection, cross-database, electrocardiogram, machine learning classification, ventricular ectopic beat*

Submitted: 29-Jan-2022 **Revised:** 27-Apr-2022 **Accepted:** 19-May-2022 **Published:** 12-Jul-2023

Introduction

The first recorded description of perturbations occurring at irregular intervals, which were consistent with ventricular ectopic beats (VEBs), was given by early Chinese physician, the master in pulse palpation (the process of using hands to examine the body while diagnosing the disease), and diagnosis, Pein Ts'Io, around 600BC.[1] He noted that when these irregularities occur occasionally, there is no significant effect on the average life span. However, when these irregularities arise frequently, it causes myocardial infarction, which is more prone to sudden death.

The VEB is an irregular heart rhythm due to premature ventricular contraction (PVC) or escapes ventricular contraction. PVC means the ventricular contraction happens before it could happen the normal depolarization from atrioventricular node, as shown in Figure 1. In other words, the ventricular contract before the atrium optimally fills the ventricular with blood. The PVC happens due to the stimulation of induces such as alcohol, caffeine, cocaine, stress, medication, and amphetamine (a mood-altering drug).[2] As the causes of PVCs are alcohol, tobacco, and caffeine, the likelihood of PVCs can be reduced by eliminating them.

The early studies have shown that VEBs frequently occur in patients with hypertension, which is the present most prominent and dominant heart disease in India in the ratio of 25%–30% in urban and 10%–20% in rural.[3]

Lin and Yang[4] presented a heartbeat (N, S, and V) classification using morphology and interval features. The data (99,827 heartbeats) were taken from the MIT-BIH

How to cite this article: Srinivasulu A, Sriraam N. Signal processing framework for the detection of ventricular ectopic beat episodes. J Med Sign Sens 2023;13:239-51.

Address for correspondence:
Dr. Natarajan Sriraam, Center for Medical Electronics and Computing, M.S Ramaiah Institute of Technology, Bengaluru, Karnataka, India. E-mail: sriraam@msrit.edu

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.jmss_12_22

Quick Response Code:

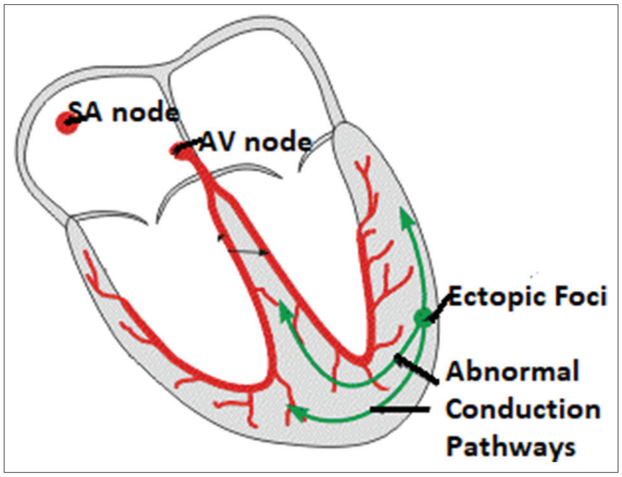

Figure 1: Conduction of ventricular ectopic beat (Source: https://www.cvphysiology.com/Arrhythmias/A017, author: *Richard Klabunde*)

Arrhythmia Database. They achieved a sensitivity of 86.2% and positive predictivity of 73.7% for the VEB classification using the classification model based on Linear Discriminant. Stoyan Tanev[6] described a discrimination method for detecting and classifying VEBs using QRS complex discrimination features. He achieved Se = 99.71%, Sp = 99.66%, SeEB = 92.27%, and SpEB = 94.78% for AHA database and Se = 96.74%, Sp = 97.21%, SeEB = 90.05%, and SpEB = 86.46% for MIT-BIH database. Hammed and Owis[8] proposed a method for classifying VEB and normal beats using template matching. They extracted the features such as RR interval, beat width, and p-wave existence from the MIT-BIH ECG database and achieved an overall accuracy of 97.24% with sensitivities of normal and VEBs 98.93% and 94.54%, respectively. Hu *et al*.[10] presented a real-time cardiac arrhythmia classification system to classify PVC, atrial premature contraction using time-domain features, and layered hidden Markov model. They achieved 99.20% accuracy, 97.75% sensitivity, and 96.63% positive predictivity for PVC classification using the MIT-BIH Arrhythmia Database. Herry *et al*.[11] proposed a heartbeat classification using synchrosqueezing transform. The data were taken from the MIT-BIH Arrhythmia Database and classified the beats into AAMI (N, S, V, F, Q) beat classes using SVM classifier by extracting the features (SST-derived instantaneous phase, the R-peak amplitudes, and R-peak to R-peak interval durations). They achieved an overall accuracy of 93.91%, and VEBs were classified with Se = 77.5% and + P = 79.05%.

To the best of our knowledge, most of the literature failed to discuss the complete details of the datasets used for trained and testing. It was further noticed that annotations were not available for all recordings. As the focus of this research study was to discriminate VEBs from normal, very limited reports were available. Hence, the proposed study made an attempt sincerely to explore the cross-database validation.

The previous work reported a study on VEB classification using the KNN algorithm.[13] The test was conducted on the MIT-BIH Arrhythmia Database. The study achieved an accuracy of 98.67% to classify normal and VEBs.

The present work applied signal processing work on an open-source (MIT-BIH Arrhythmia) database and proprietary database to classify the beats into normal and VEB. The classification was tested with four different machine learning classifiers and different training–testing ratios. VEB cardiac episode detection was done using a cross-database, i.e., training with an open-source database and testing with a proprietary database.

The ECG database collected for classification was reported in section 2. The methods of preprocessing the data, feature extraction, and classifier details were discussed in section 3. The results and discussion of classification performance were presented in section 4. Finally, the study is concluded with an overview for further research work in section 5.

Materials and Methods

The block diagram representation of the automated VEB classification system is shown in Figure 2.

Database

The current study has used the datasets from MIT-BIH Arrhythmia (open source) Database[14,15] and proprietary database (collected by the "Actiwave Cardio" and "CardioS"[16] device from the hospital patients). MIT-BIH Arrhythmia Database (DB-I) consisting of 48 records of each 1-min duration is considered. This study assessed 81 VEB beats and 678 normal beats from all the records of DB-I. Each record of the proprietary database is minimum of 1-h duration and total of 106 h from 30 records: 28 records from Actiwave Cardio (DB-II) and two records from CardioS (DB-III) were used. Data were collected from 27 males and 3 females of age 60.27 ± 18.37 years and BMI of 23.29 ± 3.31 Kg/Cm^2. Table 1 presents the background of the patients of DB-II and DB-III subjects before the hospitalization.

ECG signals were recorded using the Actiwave Cardio and CardioS device with the modified lead-I configuration of electrode placement at a 256- and 200-Hz sampling rate, respectively. Two cardiologists at the RMCH visually marked the VEB and normal events of patient recordings.

Preprocessing

Since DB-II and DB-III have different sampling frequencies from DBI, the datasets of DBII and DBII were upsampled to 360 Hz, which is the sampling frequency of DB-I. Then, ECG recordings of DB-I, DB-II, and DB-III were filtered using BPSD-TQWT[17] to attenuate powerline noise and baseline drift. The power spectral density analysis confirmed the attenuation of 50-Hz powerline noise and baseline drift after BPSD-TQWT implementation.

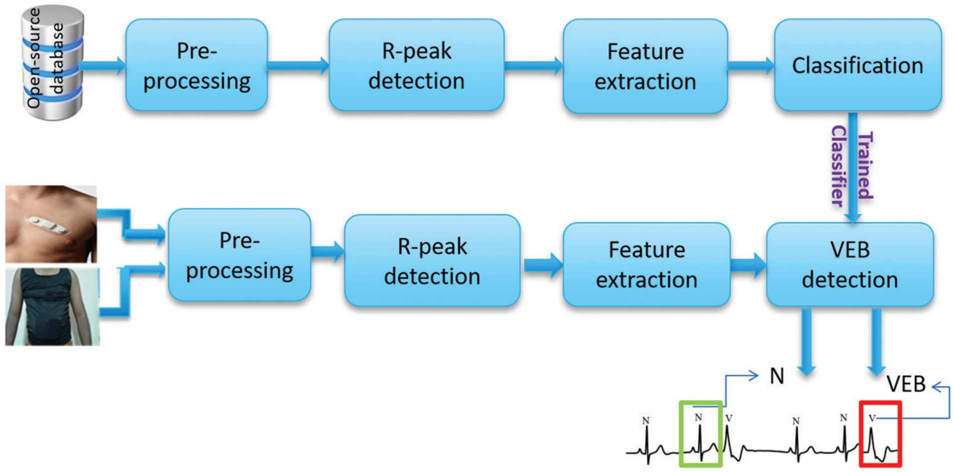

Figure 2: Block diagram representation of automated ventricular beat classification system

Table 1: Background of the patients

Subject	Background
s1-s3	Syncope and palpitation
s5	Syncope and palpitation
s7-s26	Palpitation
s27	Paroxysmal atrial fibrillation
s28	Missing beats 11 years ago due to which RF ablation was done
s29	Trivial MR, trivial TR/PASP - 32 mmHg
s31	Ebstein's anomaly mild MR, mild TR/PASP-36 mmHg partial collapsing dilated right atrium/right ventricle
s32	Chronic small vessel ischemic changes
s33	Renal tubular acidosis

MR – Mitral regurgitation; TR – Tricuspid regurgitation; PASP – Pulmonary arterial systolic pressure; RF – Radiofrequency

R-peak detection

After the noise attenuation from ECG, R-peak detection was done using the traditional Pan–Tompkins's QRS detection algorithm. Pan–Tompkins's algorithm filters the signal to separate the QRS complex from other components. The separated QRS is amplified by squaring and passed through a moving window with an adaptive threshold to identify the peaks known as R-peaks of the ECG signal. It was recognized that the detected peaks did not match the peaks of the preprocessed signal. Therefore, the positions of R-peak are adjusted to the actual positions by tracking (left and right) the nearest peak by avoiding the local peaks. Then, Q and S are detected as extreme minimum points to the left and right of true R-peaks. Figures 3-5 show the raw ECG, preprocessed signal, R-peak detection, and HRV of a sample ECG segment from three databases.

In this article, the features extracted to classify VEBs are selected based on the anatomic features of PVCs,[18] such as

1. QRS width is greater than or equal to 120 ms with an abnormal shape
2. Premature beat, followed by a full compensatory pause, i.e., the next beat occurs at greater than or equal two times the previous RR interval.

The features extracted based on beat intervals that replicate the anatomic features of PVCs are given in Table 2.

The features mentioned in Table 2 are calculated for each beat of the test signal by the Eq. (1)-(5).

$$pre\_RR_i = R_i - R_{i-1} \tag{1}$$

$$Post\_RR_i = R_{i+1} - R_i \tag{2}$$

$$QRS_i = S_i - Q_i \tag{3}$$

$$QR_{slope}(i) = \frac{Amplitude\ difference\ between\ R_i\ and\ Q_i\ points}{Time\ difference\ between\ R_i and\ Q_i\ points} \tag{4}$$

$$RS_{slope}(i) = \frac{Amplide\ difference\ between\ R_i\ and\ S_i\ points}{Time\ difference\ between\ R_i\ and\ S_i\ points} \tag{5}$$

Classification systems

The feature vector consisting of all the features presented in Table 3 was applied on different machine learning classifiers such as support vector machine (SVM), k-means nearest neighbor (KNN), decision tree, naïve Bayes, the nearest mean classifier (NMC), and nearest RMS (NRMS).

Support vector machine classifier

Support vector machine

The SVM is a computational efficient linear and nonlinear classification machine learning approach. The SVM constructs a set of hyperplanes in a high-dimensional space. Therefore, the hyperplane achieves excellent separation with the most significant functional margin, i.e., the distance between the hyperplane to the nearest training data point of any class (A. Kaveh, W. Chung. 2013, Bradley M Whitaker *et al*., 2017). In this article, SVM

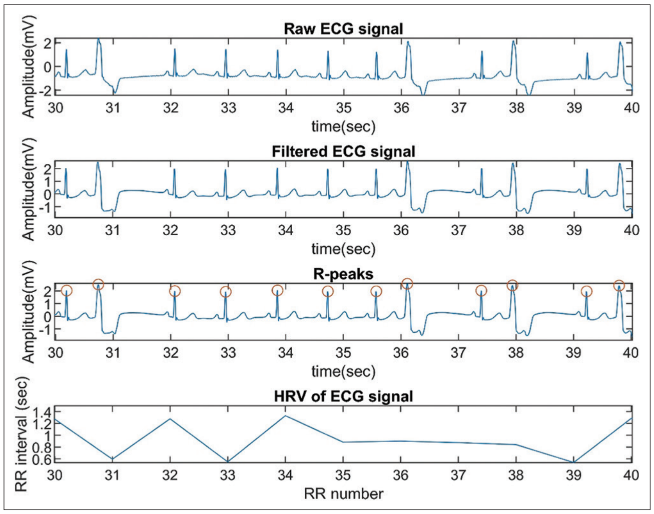

Figure 3: R-peak detection and heart rate variability of an ECG segment taken from DB-I. ECG – Electrocardiogram

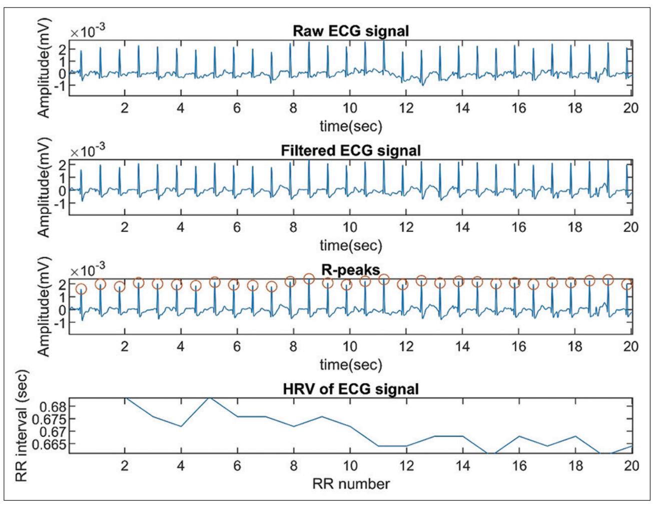

Figure 4: R-peak detection and heart rate variability of an ECG segment taken from DB-II. ECG – Electrocardiogram

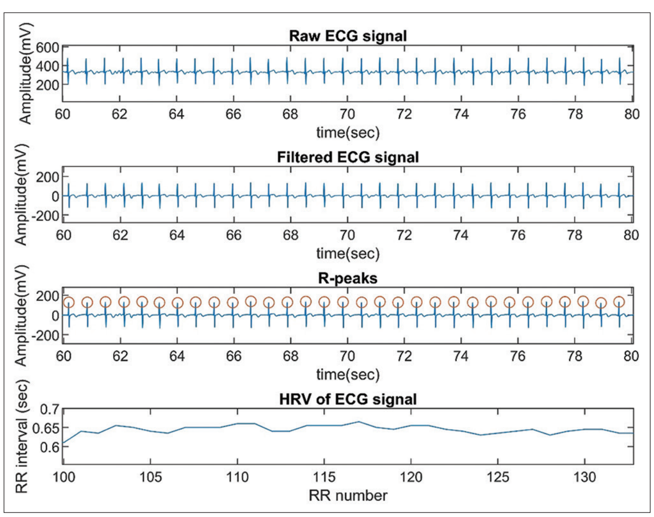

Figure 5: R-peak detection and heart rate variability of an ECG segment taken from DB-III. ECG – Electrocardiogram

Table 2: Description of features extracted

Feature	Description
Pre-RR	Difference between present and past R peak positions
Post-RR	Difference between present and next R peak positions
QRS width	Width of QRS complex
QR slope	Slope between Q and R points
RS slope	Slope between R and S points

Table 3: ANOVA test results on features of normal and ventricular ectopic beat cardiac episodes

Feature	SS	df	MS	*F*	*P*
Pre-RR	2.10E+00	1	2.10E+00	8.73E+01	1.49E-14
Post-RR	1.88E+00	1	1.88E+00	4.13E+01	8.14E-09
QRS	4.72E-07	1	4.72E-07	1.18E+01	9.31E-04
QR slope	4.89E-02	1	4.89E-02	3.01E+01	4.53E-07
RS slope	6.04E-02	1	6.04E-02	2.34E+01	6.04E-06

SS – Sum of square; MS – Mean sum of square

clustering algorithm developed by Vladimir Vapnik (Gradl, Stefan, *et al*., 2012) was used to classify the ECG pattern into normal and abnormal. Here, radial basis function was used as a kernel function with a default scaling factor of $\sigma = 1$, and the separating hyperplanes were found by the least-squares method.

KNN classifier

K-means nearest neighborhood

The k-NN classifier is a nonparametric, most straightforward machine learning algorithm for pattern classification. Using the k-NN classifier, the pattern is classified based on the majority vote of the test sample by its nearest "k" neighbors. This article classified the ECG pattern by finding nine nearest neighbors, measured by the Euclidean distance function (Alan S. Said Ahmad *et al*., 2018) to calculate the distance between the test data and trained data.

Nearest mean classifier

The NMC is developed based on the mean value of the samples, i.e., if a test sample is closest to the mean value of trained pieces of a group, the test sample is classified to the respective group. For example, assume f1 and f2 are the two most significant features of trained data. Let the sample xi = (f1(i), f2(i)), where x1, x2, x3…, xn are the samples of the subject 1 to "n" respectively in Group 1, and let the sample yi = (f1(i), f2(i)), where y1, y2, y3,…., yn are the samples of the subject 1 to 'n' respectively in Group 2. Where f1(i) and f2(i) are the f1 and f2 features of ith subject and n is the number of subjects in each group. Then, the mean values of these features of each group are calculated separately by Eq. 6 and 7 (Veenman, C. J. and Reinders, M. J., 2005).

$$\overline{f_1} = \frac{\sum_{i=1}^{n} f_1(i)}{n} \quad (6)$$

$$\overline{f_2} = \frac{\sum_{i=1}^{n} f_2(i)}{n} \quad (7)$$

where ($\overline{f_1}$) is the mean value of the first most significant feature of all the training samples of a group and ($\overline{f_2}$) is the mean value of the second most significant feature of all the training samples of the same group. The Euclidean distance between the test sample ($f_1(n+1)$, $f_2(n+1)$) and the mean sample ($\overline{f_1}, \overline{f_2}$) of each group is found. If the distance between the test sample and the mean sample of Group 1 is less than that of Group 2, the test sample is classified as Group 1. Otherwise, the test sample is classified as Group 2, as shown in Figure 6. Here, the test sample is compared only with the mean sample. Therefore, the computational time (CT) is reduced significantly. In KNN algorithm, the test sample is compared with k-nearest neighboring samples, which increases the computation time. Therefore, compared with the KNN and SVM classifiers, NMC is simple in structure and provides less CT.

Nearest root mean square classifier

The NRMS classifier is like the NMC classification method. In the NRMS method, the RMS value of each feature is calculated, which avoids the zero mean problem when the feature is floating in the range of (−n, n), where "n" is the maximum of the feature. Here, the features of the test data are compared with the RMS value of the respective feature. The majority nearest neighboring features representing a class separate the normal and abnormal episodes.

Results

Performance on the open-source database

In this experimental study, 678 normal and 81 VEB samples of preprocessed ECG signals from DB-I were used to test the classification performance of the machine learning classifiers. A one-way analysis of variance (ANOVA) test on normal and VEB cardiac episode features tells the statistical significance of $P < 0.001$, as shown in Table 3. The statistical parameters such as mean, median, minimum, and maximum values are shown in the box plot in Figure 7. The performance was observed by selecting randomly 60%, 80%, and 90% data for training and 40%, 20%, and 10% data for testing. Finally, the performance was compared in the recall, specificity, precision, F1-score, and classification accuracy, calculated by Eq. (8-11)

$$Recall\,(Re) = \frac{TP}{TP+FN} \quad (8)$$

$$Specificity\,(Sp) = \frac{TN}{TN+FP} \quad (9)$$

$$Precision\,(\mathrm{Pr}) = \frac{TP}{TP+FP} \quad (10)$$

$$Classification\ Accuracy\,(ACC) = \frac{TP+TN}{TP+FP+TN+FN} \quad (11)$$

where

TN: TRUE NEGATIVE: correctly classified as normal beat (N)

TP: TRUE POSITIVE: correctly classified as VEB (V)

FN: FALSE NEGATIVE: incorrectly classified as N

FP: FALSE POSITIVE: incorrectly classified as V

Figure 8 shows the detected R, Q, and S points of N and V beats. Table 4 reports the comparison of classification performance when the feature vector was applied to machine learning classifiers. The highest abnormal episode detection accuracy was achieved with NRMS classifier with an accuracy of 95.71% and F1-score of 77.97% for 60% training data, NRMS 97.37% of accuracy and 87.5% of F1-score for 80% training data, and 98.68% of accuracy and 94.12% of F1-score for 90% training data. KNN classifier reported equal performance to NRMS for 60% training data, and SVM classifier reported similar performance to NRMS for 80% training data in terms of accuracy and F1-score. NMC classifier also reported above 95% accuracy for the classification of VEB cardiac episodes.

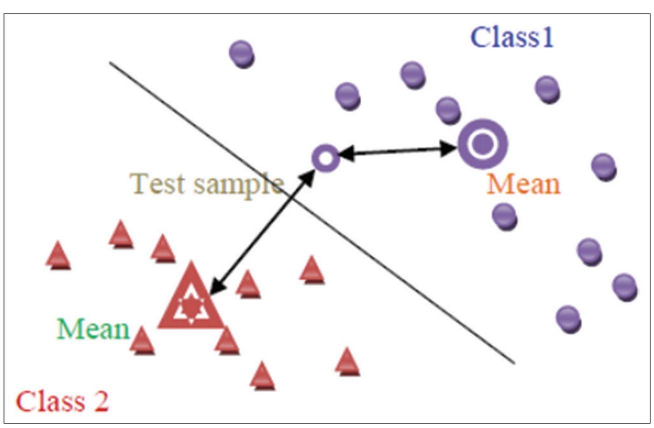

Figure 6: NMC classification method. NMC – Nearest mean classifier

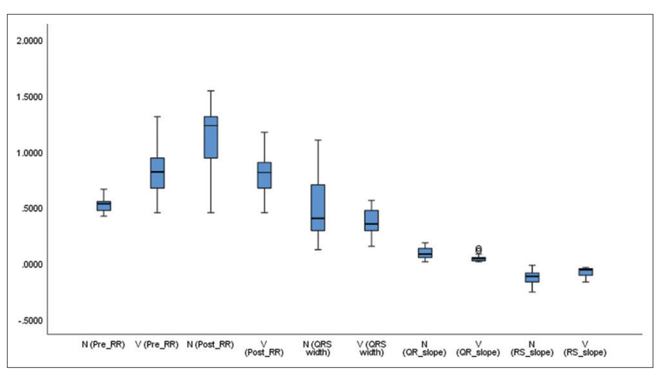

Figure 7: Box plot of normal and VEB cardiac episode features. VEB – Ventricular ectopic beat

The classification results were compared using a radar diagram in Figure 9. The larger hexagon represents the best performance model. The comparison reports that the SVM model performed better than other trained models. Similarly, the comparison of CT is shown in Figure 10. The SVM model reported the highest CT of 0.37 sec, and NMC registered the lowest CT of 0.0031 sec. Therefore, the NMC model has shown the highest relative performance compared to other models due to its lower CT, as shown in Figure 11.

Performance on cross-database

After testing the proposed framework on the open-source database, the trained models were applied to the proprietary database (DB-II and DB-III) to detect the VEB episodes. Training database DB-I and testing database DB-II and DB-III were denoted as cross-database. DB-II and DB-III consist of 514,176 N and 1024 V samples. Since the sampling frequency of DB-II and DB-III was different from trained data (DB-I), the signals of datasets were upsampled to 360 Hz. Next, the upsampled signals were preprocessed, and the features presented in Table 2 were extracted.

The ANOVA statistical analysis of extracted features is shown in Table 5 and tells that the features of normal and VEB cardiac episodes were statistically independent with $P < 0.05$. The statistical mean, median, minimum, and maximum are shown through the box plot in Figure 12. Further, the Area Under the reciever operating characteristic Curve (AUC) for each feature was computed and is shown in Figure 13. The AUC ROC of all features is above 0.5, which indicates the discriminatory capability.

The features were applied to the models trained with 42 N and 42 V randomly selected samples of DB-I. The performance measures on the cross-database are reported in Table 6. In addition, the detected VEB episodes were marked on the ECG signal, as shown in Figure 14. Finally, the performance was compared in the recall, specificity, precision, F1-score, and classification accuracy, calculated by Eq. (8-11).

The detected episodes were compared with the annotations by the cardiologist and calculated the detection accuracy. Table 6 reports that the highest accuracy and F1-score of 99.97% and 94.37%, respectively, were achieved with the NRMS trained model for the VEB episode detection. Other SVM, KNN, and NMC models could classify with above 99% of accuracy and above 90% of F1-score. The results show better precision and F1-score, along with recall and specificity. Due to the unbalance of data, the false positives are still more. Here, the lower precision and F1-score is due

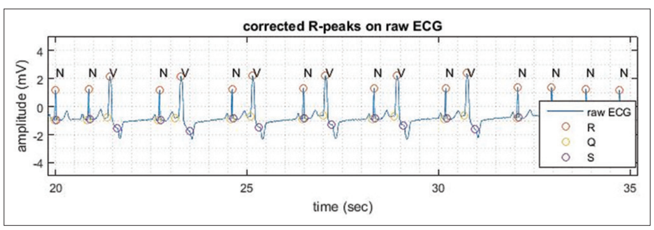

Figure 8: Detected R-peaks, Q and S positioned on raw ECG signal (labeling of N and V beats as per annotations given in database record 119). ECG – Electrocardiogram

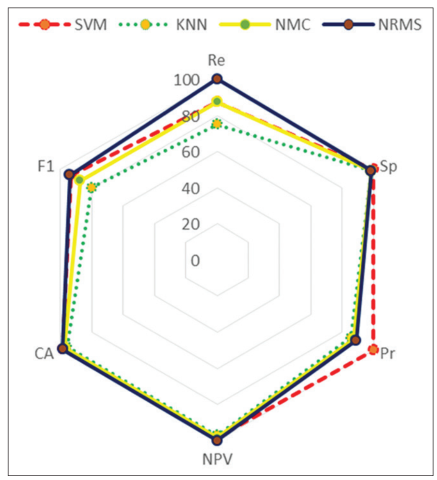

Figure 9: CR comparison of different classification methods for VEB classification. CR – Classification result; VEB – Ventricular ectopic beat

Table 4: Classification performance analysis

Classifier	Recall	Specificity	Precision	F1-score	Accuracy	Training–testing ratio (%)
SVM	71.88	98.15	82.14	76.67	95.38	60-40
	87.50	98.53	87.50	87.50	97.37	80-20
	87.50	100.00	100.00	93.33	98.68	90-10
KNN	71.88	98.52	85.19	77.97	95.71	60-40
	87.50	97.79	82.35	84.85	96.71	80-20
	75.00	98.53	85.71	80.00	96.05	90-10
NMC	71.88	97.79	79.31	75.41	95.05	60-40
	87.50	97.83	82.35	84.85	96.75	80-20
	87.50	98.53	87.50	87.50	97.37	90-10
NRMS	71.88	98.52	85.19	77.97	95.71	60-40
	87.50	98.53	87.50	87.50	97.37	80-20
	100.00	98.53	88.89	94.12	98.68	90-10

SVM – Support vector machine; KNN – K-means nearest neighbor; NMC – Nearest mean classifier; NRMS – Nearest root mean square to fewer V samples available in proprietary datasets DB-II and DB-III. As mentioned earlier, the test data are taken from DB-II and DB-III; each record is applied directly to extract the features and classify each beat with a trained network. Therefore, the test data could not balance the number of N and V beats. Figure 15 shows the comparison of the classification performance of all proposed models. The larger hexagon has resulted from the NRMS classifier representing better performance than other trained models.

Further, proposed study has done an experiment by training the model with two datasets (DB-I and DB-II) tested on the third dataset (DB-III). Total 84 samples (N-44 and V-44) from DB-I and 20 samples (N-10 and V-10) from DB-II were used to train the model, and the remaining samples of DB-II and DB-III were used for testing. The results obtained (refer to Table 7) were close to Table 6. However, the precision and F1-score were increased by ~0.7% and ~0.2%, respectively.

Tables 6 and 7 infer that the VEB episodes from proprietary datasets can be identified by using the models trained with open-source datasets.

Discussion

More populations in developing countries and underdeveloped countries have been suffering from cardiac diseases for the past few decades. Diagnosis of cardiac disease needs an ECG test, a primary requirement for the doctor. When a patient is recommended to undergo long-term ECG monitoring, finding the abnormal segments and diagnosing arrhythmia becomes a time-consuming process, which causes the patient's treatment to be delayed. Therefore, the current research has proposed an automatic diagnostic system to classify and detect VEB cardiac episodes. The proposed framework was tested on the open-source database to classify VEB episodes from the normal. The highest accuracy trained model was used to detect VEB episodes of the proprietary database.

The proposed study makes use of cross-database validation to ensure the robustness of the developed computer-aided automated beat detection algorithm. Demography-driven availability data will be quite helpful. The following aspect shall be considered the combination of the proposed research.

1. Automated classification of normal and VEBs using multi-database and cross-data validation
2. NRMS classifier, a new classification method that outperforms other existing classical classification models like SVM, KNN, and NMC, provides robust pattern classification performance
3. Normal and VEBs were classified with only five features

Table 5: ANOVA test results on features of normal and ventricular ectopic beat cardiac episodes

Feature	SS	Df	MS	*F*	*P*
Pre-RR	0.1126	1	0.11261	58.34	4.459e-10
Post-RR	3.6272	1	3.62725	263.45	7.99901e-56
QRS	0.76629	1	0.76629	346.72	1.30084e-71
QR	0.44868	1	0.44868	272.27	1.6008e-57
RS	0.10942	1	0.10942	39.99	3.12993e-10

SS – Sum of square; MS – Mean sum of square

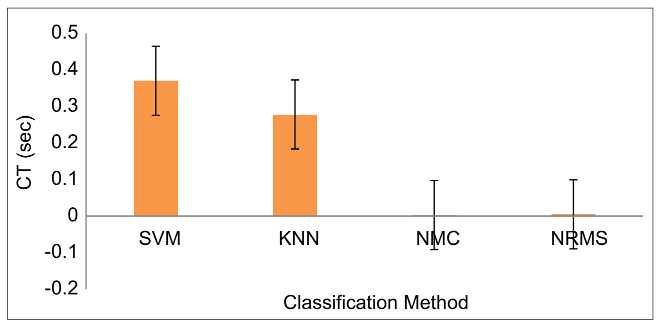

Figure 10: CT comparison of different classification methods for VEB classification. CT – Computational time; VEB – Ventricular ectopic beat

Table 6: Detection of ventricular ectopic beat episodes

	TP	FP	TN	FN	Recall	Specificity	Precision	F1-score	Accuracy
SVM	1011	120	514056	13	98.7305	99.9767	89.3899	93.8283	99.9742
KNN	1003	129	514047	21	97.9492	99.9749	88.6042	93.0427	99.9709
NMC	1014	199	513977	10	99.0234	99.9613	83.5944	90.6571	99.9594
NRMS	1006	102	514074	18	98.2422	99.9802	90.7942	94.3715	99.9767

SVM – Support vector machine; KNN – K-means nearest neighbor; NMC – Nearest mean classifier; NRMS – Nearest root mean square; TP – True positive; FP – False positive; TN – True negative; FN – False negative

Table 7: Detection of ventricular ectopic beat episodes using training samples from DB-I and DB-II

	TP	FP	TN	FN	Recall	Specificity	Precision	F1-score	Accuracy
SVM	1015	174	514002	9	99.1211	99.9662	85.3659	91.7307	99.9645
KNN	1003	115	514061	21	97.9492	99.9776	89.7138	93.6508	99.9736
NMC	1016	205	513971	8	99.2188	99.9601	83.2105	90.5122	99.9587
NRMS	1002	94	514082	22	97.8516	99.9817	91.4234	94.5283	99.9775

SVM – Support vector machine; KNN – K-means nearest neighbor; NMC – Nearest mean classifier; NRMS – Nearest root mean square; TP – True positive; FP – False positive; TN – True negative; FN – False negative

4. Training and testing procedures with multi-database and cross-database processes show the proposed framework's relative efficiency in recognizing VEBs.

Most of the literature was done to classify normal and arrhythmic beats of the MIT-BIH Arrhythmia Database. Therefore, the current work was also done on the same database. As mentioned in the literature, 102, 104, 107, and 217 records had paced beats; these are excluded from the current work. The parts of ECG contaminated with heavy motion artifacts were also excluded from the proposed study. Based on the nature of VEBs, the five most prominent features such as pre- and post-RR interval, QRS duration, and QR and RS slopes were extracted from each beat of the ECG signal.

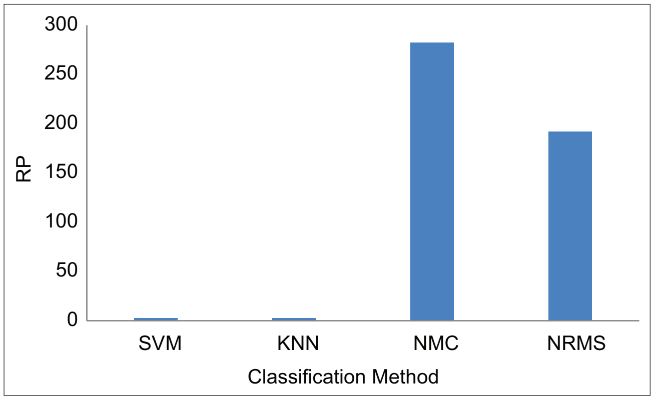

Figure 11: Relative Performance (RP) comparison of different classification methods for VEB classification. RR – Relative Performance; VEB – Ventricular ectopic beat

The feature vector was applied to four machine learning classification models. The experimentation was done with 60%–40%, 80%–20%, and 90%–10% training–testing ratios to verify the performance of the classification model with different data sizes. In three cases, the SVM and NRMS models were able to classify the VEB episodes with the highest classification accuracy. On the other hand, the KNN and NMC models reported comparable performance. The proposed framework intends to verify the inter-database performance, i.e., when the different training and testing databases. Therefore, after training the classification model with the MIT-BIH Arrhythmia Database, the model was applied to the proprietary database. The proprietary database consists of the data collected from the hospital

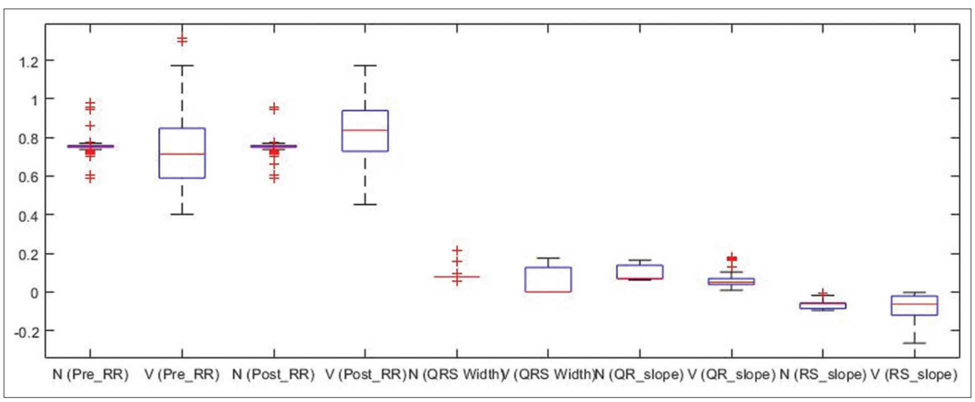

Figure 12: Box plot of normal and VEB cardiac episode features. VEB – Ventricular ectopic beat

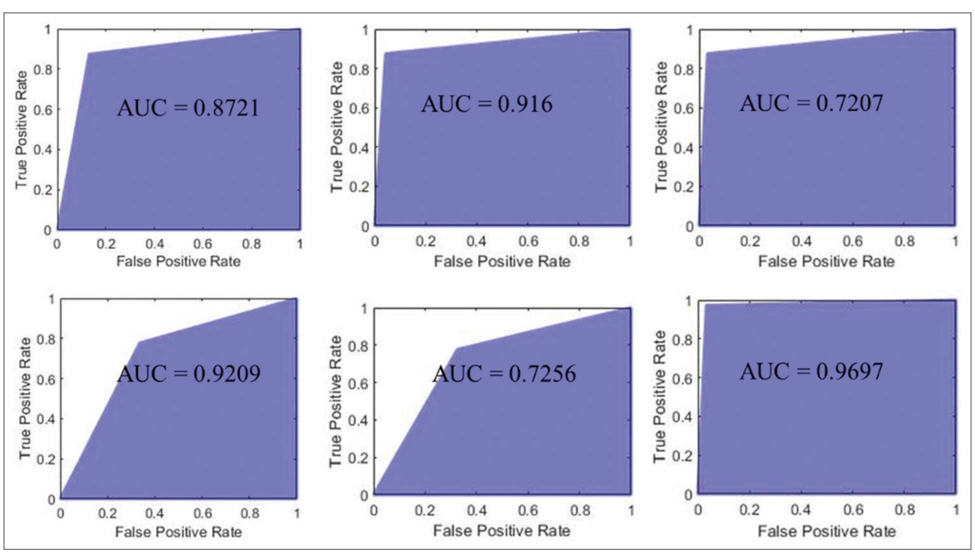

Figure 13: AUC ROC for individual features. AUC ROC – area under the curve of receiver operating characteristics

patients using Actiwave Cardio and CardioS devices. All the training models predicted VEB episodes from the proprietary data with 99% accuracy. As already mentioned, the NRMS model has shown better performance; the proprietary data were predicted with the highest accuracy amongst the other trained models. Further, the proposed study also tried with equal number of N and V beats for VEB detection. Here, the under-sampling technique was adapted for balancing the data. The under-sampling process randomly removes the samples from the majority class until the data distribution gets balanced.[27,28] The samples were selected randomly from the normal class and balanced them with VEB class. The under-sampling technique was used in the current study since there was no loss of information even some samples were removed from the normal class. The imbalanced data for the classification problem may cause overfitting and bias due to the distribution of imbalanced training data but not the testing data. The proposed study experimented with imbalance datasets to classify and predict the VEB episodes from normal. The results presented in Table 8 have reported the comparative results with Table 6. However, a number of false negatives were increased. This infers that the size of the testing data was not significantly affected as the VEB episodes were detected separately for each subject.

Finally, the performance of the proposed framework was compared with the literature in Table 9.

Table 9 reports the comparison of the proposed framework with the literature. Most of the literature focused on ECG classification into different arrhythmic beats as AAMI beats with many features. Due to the most significant and lack of proprietary data availability, the proposed study was limited to VEB episodes. Therefore, the performance of the proposed framework was compared with available minor literature on normal and VEB episode classification. From Table 9, it is observed that the proposed framework has shown better performance with only five features. The proposed work reported less F1-score than literature, resulting in fewer VEB samples than normal samples. However, the proposed framework has shown comparable performance with the literature mentioned in Table 9.

Conclusion

This article presents a novel signal processing framework for detecting VEB (VEB) cardiac episodes using a cross-database, i.e., a trained classification model with an open-source database was used for predicting the VEB episodes from the proprietary database. The open-source data were collected from the MIT-BIH Arrhythmia Database, and proprietary data were collected by Actiwave Cardio and CardioS ECG monitoring devices. In the first step, all the data were normalized to the sampling rate of 360 Hz and preprocessed for removing powerline interference and baseline drift using basis pursuit sparse decomposition of tunable-Q wavelet transform (BPSD-TQWT). The beats in the preprocessed signal were annotated as normal and VEBs by the experts. Five features, namely pre- and post-RR intervals, QRS duration, and QR and RS slopes, were extracted from each ECG beat in the second step. In the third step, all the features were applied to four machine learning classification models to test the classification performance. Here, it is observed that the NRMS model reported the

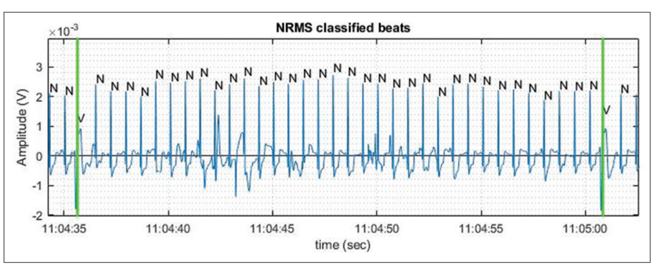

Figure 14: Predicted ventricular ectopic beat episode marking on the ECG signal. ECG – Electrocardiogram

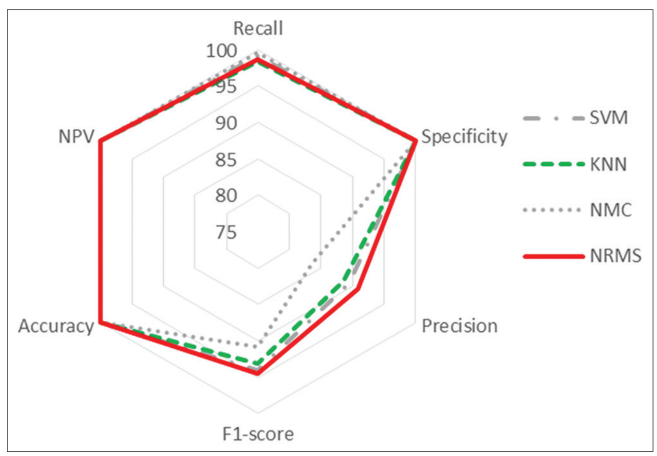

Figure 15: CR comparison of different classification methods. CR – Classification result

Table 8: Detection of ventricular ectopic beat episodes using equal number of N and V beats

	TP	FP	TN	FN	Recall	Specificity	Precision	F1-score	Accuracy
SVM	982	13	1011	42	95.8984	98.7305	98.6935	97.2759	97.3145
KNN	985	16	1008	39	96.1914	98.4375	98.4016	97.2840	97.3145
NMC	992	12	1012	32	96.8750	98.8281	98.8048	97.8304	97.8516
NRMS	989	3	1021	35	96.5820	99.7070	99.6976	98.1151	98.1445

SVM – Support vector machine; KNN – K-means nearest neighbor; NMC – Nearest mean classifier; NRMS – Nearest root mean square; TP – True positive; FP – False positive; TN – True negative; FN – False negative

Table 9: Comparison of proposed study with earlier reported work

Study	Database	Total beats	Features	Classifier	Classification	ACC (%)	Se (%)	P+ (%)
Lin and Yang[4]	MIT-BIH Arrhythmia	N=90,042, S=2776, V=7006	Morphology and interval features	Linear discriminants	N, S, V		86.20	73.7
Kanwar and Dewangan[5]	MIT-BIH Arrhythmia	V=111	QRS width and T-amplitude	Decision tree (c4.5)	N, V	96.6	98	96.07
Tanev[6]	MIT-BIH Arrhythmia	47 records	QRS complex discrimination features	Discrimination	Normal and ectopic beats	-	90.05	-
Tanev[6]	AHA database	70 records	QRS complex discrimination features	Discrimination	Normal and ectopic beats	-	92.27	-
Sadrawi *et al.*[7]	AHA database, MITDB, NSTDB	VEB=78+44+12 records	Time domain features	Integrated detection algorithm	VEB		86.52 (AHADB), 87.27 (MITDB), 58.17 (NSTDB)	94.67 (AHADB), 73.26 (MITDB), 50.86 (NSTDB)
Sadrawi *et al.*[7]	MITDB	SVEB=Records	Time domain features	Integrated detection algorithm	SVEB		71.35	36.9
Sadrawi *et al.*[7]	MITDB	AF=44 records	Time domain features	Integrated detection algorithm	AF		70	100
Sadrawi *et al.*[7]	AHA database, MITDB, CUDB	VF=78+44+35 records	Time domain features	Integrated detection algorithm	VF		90 (AHADB), 100 (MITDB), 84 (CUDB)	69 (AHADB), 33 (MITDB), 83 (CUDB)
Hammed and Owis[8]	MIT-BIH Arrhythmia	-	RR interval, beat width and p-wave existence	Template matching	N and V	97.24	94.54	-
Raj *et al.*[9]	MIT-BIH Arrhythmia	-	FFT based DWT coefficients	Feed forward NN	RBBB, LBBB, PVC, APC, VF, N	97.4	94.44	98.19
Hu *et al.*[10]	MIT-BIH Arrhythmia	N=32,655, PVC=2000, APC=144	Time domain features	Layered hidden Markov model	normal, APC, VPC, invalid	99.2	97.75	96.63
Herry *et al.*[11]	MIT-BIH Arrhythmia	-	SST-derived instantaneous phase, the R-peak amplitudes and R-peak to R-peak interval durations	SVM classifier	AAMI beats	93.91	77.5	79.05
Llamedo *et al.*[19]	INCART, MIT-BIH AR, MIT-BIH SUP, Biosigna (private)	N=692,257, S=18,259, V=39,493, F=1044, Q=100	Current RR, next RR, average RR in last 1 min, average RR in last 20 min, zero-cross position of WT ACF in lead-1 and 2, max position of WT ACF in lead 1 and 2	WT-PCA	N, V, S, V, F, Q		N=98, S=86, V=90	N=93, S=91, V=90
Tanis *et al.*[20]	MIT-BIH Arrhythmia	N=89,723, S=2773, V=6986, F=801	Pre-RR, current RR, average RR, average RR of last 10 beats, QRS duration, T-duration, p-wave flag, 10 samples of QRS, 9 samples of T-wave, QRS area, QRS power, QRS max, QRS min, QRS maximum–minimum ratio, peak width at 70% maximum, peak slope, beat area, beat power, beat maximum, beat min, beat minimum–maximum ratio, QRS var, QRS sk,	LDC and MLP based SFFS-feature selection	N, S, V, F	N=90, S=93.3, V=97.3, F=97.3	N=89.6, S=83.2, V=86.8, F=61.1	N=99.1, S=33.5, V=75.9, F=16.6

Contd...

Table 9: Contd...

Study	Database	Total beats	Features	Classifier	Classification	ACC (%)	Se (%)	P+ (%)
			QRS kurtosis, QRS histogram (20 slots) var, beat mean, beat var, beat sk, beat kurtosis, beat histogram (20 slots) var, maximum (WT 3, 4, 5), minimum (WT 3, 4, 5) difference between maximum and minimum (WT 3, 4, 5), distance between maximum and minimum (3, 4), pow (WT 2, 3, 4, 5), pow ratio (WT 3–2, 4–3, 5–4), mean (WT 3, 4), std (WT 3, 4), sk (WT 3, 4)					
Llamedo and Martinez[21]	MIT-BIH AR, MIT-BIH SUP, INCART	AR (training), SUP (validation), INCART (testing)	Past RR, current RR, next RR, RRv (i)=sum|RR (i)-RR (j)|, average RR in last 1 min, 5 min, 10 min, 20 min, QRS width, maximal vector of QRS loop and angle of this vector from 2D VCG loop created by lead II and v1, 7 maximum abs samples from 4^{th} scale of DWT, loc and maximum abs and first zero crossing of ACF and CCF of lead-II and V1, ratio (sum (AsL^*s), sum (AsL))), where AsL=1/D sum(|WsL^*s (ld)|), s=1, 2, 3, 4, 5, 6, ld=Pos of peaks, D=Number of detected peaks, L=Lead	Floating feature selection algorithm, LDC-C (best)	N, V, S	93	N=95, V=81, S=77	N=98, V=87, S=39
Rahman *et al.*[22]	Private database	HCM patients (754) and non-HCM	504 from (10 s) 12 leads by segmenting the beats using open source ecgpuwave tool. Pre-RR, post-RR, average RR, p-duration, QRS interval, T-duration, 10 samples of QRS, maximum and minimum of QRS, 10 samples from P-wave, 10 samples from T-wave, minimum and maximum from p-wave and T-wave samples (total 42×12=504)	Random forest classifier, SVM with 5-fold cross-validation	HCM and non-HCM	-	90	85
De Chazal *et al.*[23]	MIT-BIH Arrhythmia	About 51,020 training and 49,711 testing	Pre-RR, post-RR, average RR, local (10 beats) average RR, QRS duration, T-duration, p-wave flag, normalized and unnormalized 10 samples of QRS and 9 samples from QRS offset to T-wave offset. Normalized and unnormalized 10 samples of FP-50 ms to FP+100 ms and 9 samples from FP+150 ms to FP+500 ms	LDC	N, S, V	-	S=75.9, V=77.7	S=38.5, V=81.9
Li *et al.*[24]	MIT-BIH Arrhythmia	N=18,805, S=1817, V=3427, F=1049, Q=41	Amp between Q and R, amp between R and S, QRS duration, RR interval, QR slope, RS slope, ST slope, QT duration	Parallel general regression neural network	N, S, V	95, V=98.9, S=99.3	V=88, S=85.5	V=92.5, S=92.2
Kiranyaz *et al.*[25]	MIT-BIH Arrhythmia	Total 83,648 beats	Images of FFT of 64 and 128 samples centered on R-peak	Adaptive 1-D CNN (3 CNN+2 MLP	N, S, V	V=98.9, S=96.4	V=95.9, S=68.8	V=96.2, S=79.2

Contd...

Table 9: Contd...

Study	Database	Total beats	Features	Classifier	Classification	ACC (%)	Se (%)	P+ (%)
Li *et al.*[26]	MIT-BIH Arrhythmia	N=3, L=3, R=3, P=3, V=2, A=3 records	Statistical features of wavelet packet decomposition (db6) 4 levels derived 16 coefficient	Genetic algorithm-BPNN reduced the feature size	N, L, R, P, V, A	97.78	97.86	97.81
Proposed	MIT-BIH Arrhythmia	N=678, V=81	Pre-RR, post-RR, QRS width, QR slope and RS slope	SVM, KNN, NMC, NRMS	N, V	98.68	100	88.89
Proposed	Cross-database	N=514,176, V=1024	Pre-RR, post-RR, QRS width, QR slope and RS slope	SVM, KNN, NMC, NRMS	N, V	99.97	98.62	90.79

N – Normal beat; S – Supraventricular beat; V – Ventricular ectopic beat; ACC – Accuracy; Se – Sensitivity; LBBB – Left bundle branch block; RBBB – Right bundle branch block; PVC – Premature ventricular contraction; APC – Atrial premature contraction; VF – Ventricular fibrillation; AAMI – Association for the Advancement of Medical Instrumentation; FFT – Fast Fourier transform; DWT – Discrete wavelet transform; AHADB – American Heart Association Database; MITDB – MIT-BIH Arrhythmia Database; CUDB – Creighton University Ventricular Tachyarrhythmia Database; AF – Atrial fibrillation; NSTDB – MIT-BIH Noise Stress Test Database; AHA – American Heart Association; VEB – Ventricular ectopic beat; HCM – Hypertrophic cardiomyopathy; SVEB – Supraventricular ectopic beat; SVM – Support vector machine; KNN – K-means nearest neighbor; NMC – Nearest mean classifier; NRMS – Nearest root mean square; CNN – Convolutional neural networks; MLP – Multilayer perceptron; WT – Wavelet transform; WT-PCA – WT-principal component analysis; NN – Nearest neighbor; SFFS – Sequential forward floating search; VCG – Vectorcardiography; AR – Arrhythmia; INCART – St. Petersburg Institute of Cardiological Technics; SUP – Supraventricular; SST – synchrosqueezing transform; WT – Wavelet Transform; BPNN – Back Propogation Neural Network; LDL – (it is LDC - Linear discriminant Classifier); LDC-C – Linear discriminant Classifier- C type; ACF – Autocorelation Function; CCF – Crosscorelation Function; MIT-BIH – Massachusetts Institute of Technology-Beth Israel Hospital.

highest accuracy of 98.68% and F1-score of 94.12%, recall rate of 100%, specificity of 98.53%, and precision of 88.89%, which has shown better performance to other trained models. In the fourth step, the trained models were used to predict VEB episodes of the proprietary database. The NRMS model achieved a better prediction rate among four trained models with a reported accuracy of 99.97%, F1-score of 94.54%, recall rate of 98.62%, specificity of 99.98%, and precision of 90.79%. Finally, the performance results were compared with the literature and found that the proposed framework can be used in the automatic diagnosis system to detect VEB cardiac episodes. Further, the research work can be extended to other arrhythmia episodes, which lead to a more accurate and fast diagnosis of cardiac diseases.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The experimental design and performance results were validated and approved by the ethical committee (MSRMC/EC/2016/13.01.2016).

Acknowledgment

We would like to thank Dr. V S Prakash and Dr. Sarthak Sahoo, Department of Cardiology, M. S. Ramaiah Medical College and Hospitals, Bangalore, India, for providing the annotations of the ECG recordings and validation of the results.

Consent to participant

Informed consent forms were obtained from the volunteers before recording their vital signal.

Consent for publication

Not applicable.

Financial support and sponsorship

The authors would acknowledge the Department of Bio-Technology (DBT) funding for carrying out this research work (BT/PR14751/MED/32/422/2015).

Conflicts of interest

There are no conflicts of interest.

References

1. Ng GA. Treating patients with ventricular ectopic beats. Heart 2006;92:1707-12.
2. Von Rotz M, Aeschbacher S, Bossard M, Schoen T, Blum S, Schneider S, *et al*. Risk factors for premature ventricular contractions in young and healthy adults. Heart 2017;103:702-7.
3. Gupta R, Gupta S. Hypertension in India: Trends in prevalence, awareness, treatment and control. RUHS J Health Sci 2017;2:40-6.
4. Lin CC, Yang CM. Heartbeat classification using normalized

RR intervals and morphological features. Math Probl Eng 2014;2014:1-11.
5. Kanwar G, Dewangan NK. Decision tree based premature ventricular contractions beat classifier of ECG. Int J Adv Res Electr Electron Instrum Eng 2016;5:9021-7.
6. Tanev S. Ventricular beat detection and classification in long term ECG recordings. Int J Bioautomation 2012;16:273.
7. Sadrawi M, Lin CH, Lin YT, Hsieh Y, Kuo CC, Chien JC, *et al.* Arrhythmia evaluation in wearable ECG devices. Sensors (Basel) 2017;17:2445.
8. Hammed NS, Owis MI. Patient Adaptable Ventricular Arrhythmia Classifier Using Template Matching. In: 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE; 22 October, 2015. p. 1-4.
9. Raj S, Chand GP, Ray KC. Arm-based arrhythmia beat monitoring system. Microprocessors Microsyst 2015;39:504-11.
10. Hu S, Wei H, Chen Y, Tan J. A real-time cardiac arrhythmia classification system with wearable sensor networks. Sensors (Basel) 2012;12:12844-69.
11. Herry CL, Frasch M, Seely AJ, Wu HT. Heart beat classification from single-lead ECG using the synchrosqueezing transform. Physiol Meas 2017;38:171-87.
12. Baali H, Mesbah M. Ventricular ectopic beats classification using Sparse Representation and Gini Index. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:5821-4.
13. Srinivasulu A, Sriraam N. Ventricular Ectopic Beat Classification Using KNN Multi-Feature Classifier. In: 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT), IEEE; 15 February, 2018. p. 69-72.
14. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 2001;20:45-50.
15. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101:E215-20.
16. Srinivasulu A, Sriraam Natarajan, Prakash VS. Wireless Cardioc Framework for Cardiac Episode Detection and HRV Analysis. Communicated to Expert Systems With Applications, Elsevier; 2022.
17. Srinivasulu A, Sriraam N. Basis pursuit sparse decomposition using tunable-Q wavelet transform (BPSD-TQWT) for denoising of electrocardiograms. Phys Eng Sci Med (2022). https://doi.org/10.1007/s13246-022-01148-w.
18. James E Keany, Aseem D Desai. Premature ventricular contraction. Medscape 2017;2017:761148.
19. Llamedo M, Khawaja A, Martínez JP. Cross-database evaluation of a multilead heartbeat classifier. IEEE Trans Inf Technol Biomed 2012;16:658-64.
20. Mar T, Zaunseder S, Martínez JP, Llamedo M, Poll R. Optimization of ECG classification by means of feature selection. IEEE Trans Biomed Eng 2011;58:2168-77.
21. Llamedo M, Martinez JP. Heartbeat classification using feature selection driven by database generalization criteria. IEEE Trans Biomed Eng 2011;58:616-25.
22. Rahman QA, Tereshchenko LG, Kongkatong M, Abraham T, Abraham MR, Shatkay H. Utilizing ECG-based heartbeat classification for hypertrophic cardiomyopathy identification. IEEE Trans Nanobioscience 2015;14:505-12.
23. de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng 2004;51:1196-206.
24. Li P, Wang Y, He J, Wang L, Tian Y, Zhou TS, *et al.* High-performance personalized heartbeat classification model for long-term ECG signal. IEEE Trans Biomed Eng 2017;64:78-86.
25. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng 2016;63:664-75.
26. Li H, Yuan D, Ma X, Cui D, Cao L. Genetic algorithm for the optimization of features and neural networks in ECG signals classification. Sci Rep 2017;7:41011.
27. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 2018;106:249-59.
28. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: Review of methods and applications. Expert Syst Appl 2017;73:220-39.