






Article

Towards Real-Time Heartbeat Classification: Evaluation of Nonlinear Morphological Features and Voting Method

Rajesh N V P S Kandala ¹, Ravindra Dhuli ², Paweł Pławiak ^{3,4,*}, Ganesh R. Naik ⁵, Hossein Moeinzadeh ⁵, Gaetano D. Gargiulo ^{5,6,7} and Suryanarayana Gunnam ⁸

¹ Department of ECE, GVPCE (A), Visakhapatnam 530048, India; kandala.rajesh2014@gmail.com

² Department of ECE, VIT University, Andhra Pradesh 522237, India; ravindradhuli@gmail.com

³ Department of Information and Communications Technology, Faculty of Computer Science and Telecommunications, Cracow University of Technology, Warsaw 24 st., F-3, 31-155 Krakow, Poland

⁴ Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, Poland

⁵ The MARCS Institute, Western Sydney University, Milperra, NSW 2214, Australia; ganesh.naik@westernsydney.edu.au (G.R.N.); h.moeinzadeh@westernsydney.edu.au (H.M.); g.gargiulo@westernsydney.edu.au (G.D.G.)

⁶ Department of Electrical Engineering and Information Technology (DIETI), “Federico II” The University of Naples, 80100 Naples, Italy

⁷ School of Engineering at Western Sydney University, Penrith, NSW 2747, Australia

⁸ Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China; surya_gunnam@yahoo.co.in

* Correspondence: plawiak@pk.edu.pl or plawiak.pawel@gmail.com

Received: 20 September 2019; Accepted: 15 November 2019; Published: 21 November 2019

Abstract: Abnormal heart rhythms are one of the significant health concerns worldwide. The current state-of-the-art to recognize and classify abnormal heartbeats is manually performed by visual inspection by an expert practitioner. This is not just a tedious task; it is also error prone and, because it is performed, post-recordings may add unnecessary delay to the care. The real key to the fight to cardiac diseases is real-time detection that triggers prompt action. The biggest hurdle to real-time detection is represented by the rare occurrences of abnormal heartbeats and even more are some rare typologies that are not fully represented in signal datasets; the latter is what makes it difficult for doctors and algorithms to recognize them. This work presents an automated heartbeat classification based on nonlinear morphological features and a voting scheme suitable for rare heartbeat morphologies. Although the algorithm is designed and tested on a computer, it is intended ultimately to run on a portable i.e., field-programmable gate array (FPGA) devices. Our algorithm tested on Massachusetts Institute of Technology- Beth Israel Hospital (MIT-BIH) database as per Association for the Advancement of Medical Instrumentation (AAMI) recommendations. The simulation results show the superiority of the proposed method, especially in predicting minority groups: the fusion and unknown classes with 90.4% and 100%.

Keywords: electrocardiogram signal; nonlinear features; improved complete ensemble empirical mode decomposition; inter-patient scheme; voting; classification; FPGA

1. Introduction

1.1. Aim of the Work

Nowadays, mortality rates are increasing due to noncommunicable diseases (NCDs) over infectious diseases. Annually about 70% of deaths are because of the NCDs worldwide. As per

the World Health Organization (WHO) report, cardiovascular diseases (CVDs) are the primary cause of death among other NCDs [1]. The effect of CVDs is more in low and middle-income countries. The report demonstrates that this impact will continue further. This alarming scenario influences not only the health perspective, but also the socio-economic advancement of the country. Therefore, the need for adequate diagnosis and treatment for NCDs, especially for CVDs, is highly essential. This situation demands advancements in healthcare technology.

Cardiac arrhythmias are one of the significant sources of CVDs. All arrhythmias may not be fatal, but some need immediate treatment for the patient to survive. Arrhythmias may occur owing to erratic electrical impulse conduction or formation in the heart. Electrocardiograms (ECGs) are an essential tool to study the electrical activity of the heart. ECG discloses any variations in the heartbeat pattern. Clinicians have to explore the longer duration ECG records in the diagnosis process. However, this manual examination is tiresome because of low amplitude and subtle variations in the ECG [2]. Hence, computer-aided diagnosis (CAD) helps clinicians remarkably. CAD-based heartbeat classification is a significant task before the arrhythmia recognition.

1.2. State-of-the-Art

If we look at the literature on heartbeat classification, it can be observed that the Massachusetts Institute of Technology-Beth Israel hospital (MIT-BIH) arrhythmia database [3] is the majority choice. The essential literature of heartbeat classification using this database can be categorized into two types based on the assessment process, viz., class-oriented and subject-oriented. The majority among them are class-oriented based works [4–22].

In the class-oriented approach, from the 16 types of beats including the normal ones in the MIT database, a part or an entire collection of beats are preferred for classification. In [4], 17 types of heartbeats including normal and pacemaker are classified using the features based on various power spectrum density methods. Later, a novel genetic algorithm is used to identify the optimum features to enhance the classification process. Finally, these selected features are fed to the various standard machine learning algorithms. In [19], 13 types of heartbeats are classified using the combination of higher-order statistics (HOS) of the ECG and Hermite basis representation features using a support vector machine (SVM) classifier. In [18], six types of heartbeats are classified by using a local fractal dimension based nearest neighbor classifier. In [21], seven types of heartbeats are classified using gray relational analysis. In [20], Ye et al. designed a heartbeat classification algorithm using dynamic and morphological ECG features. For the morphological feature extraction process, the combination of wavelet transform and the dimensionality reduction technique, namely independent component analysis (ICA), is implemented on the heartbeats. R–R intervals are used as dynamic features. These features are then fed to SVM for classifying 16 types of heartbeats. In [8], a novel genetic ensemble of classifiers machine learning method is proposed. A new genetic training coupled with genetic optimization is used to classify 17 types of heartbeats. In [17], statistical and nonlinear features are derived from the modes obtained from the empirical mode decomposition (EMD) algorithm. Later, these features are provided to one-against-one SVM for classifying five types of heartbeats. In [22], ventricular extra systole or ectopic beats are recognized with the help of morphology matching, R–R intervals, and clustering algorithms. In [6], 17 types of EG beats are classified using hexadecimal local patterns calculated from wavelet sub-bands. In [7], five primary types of heartbeats are classified using ensemble empirical mode decomposition (EEMD) based features subjected to sequential minimal optimization-SVM (SMO-SVM). Besides, Neural networks plays a crucial role in biological signal analysis [23]. Recently, deep learning-based class-oriented schemes come into the picture. Deep learning techniques are a part of machine learning techniques implemented based on more hidden neural networks. In [9,10] these works, 17 types of heartbeats are classified using 1D-CNN and a novel 3-layer deep genetic ensemble of classifiers.

In the subject-oriented approach, the entire MIT-BIH database is subdivided into five groups of heartbeats according to the American National Standards Institute/Advancement of Medical

Instrumentation (ANSI/AAMI) EC57:1998 standard. The list of these groups is non-ectopic (N), supraventricular ectopic (S), ventricular ectopic (V), fusion (F), and unknown (Q). Again, two strategies are observed for classifying these distinct groups: intra-patient and inter-patient schemes. The fundamental disagreement between these two strategies is the separation of training and testing datasets. Intra-patient scheme based methods are widely explored in the literature [24–30]. However, these approaches have less impact in real-time scenarios. Because, in real-time applications, an unknown subject that usually undergoes the testing will be foreign to the constructed model. Thus, the model has to be adequate to capture the inter-individual variations among the ECG. While designing the intra-patient based model, there might be a chance of having common subject information in both training and testing. To mitigate such an issue, De Chazal et al. [31] introduced an inter-patient scheme based heartbeat classification. Here, the overall MIT-BIH database is separated into two groups. One group is assigned to training, and the other one is for testing by ensuring that there is no similar subject data in both groups.

The advantage of the aforementioned computer-aided expert systems can be exploited only after developing real-time systems. In literature, in recent years, some of the field-programmable gate array (FPGA) based ECG signal analysis systems are implemented. In [32], an FPGA based heartbeat classification system is developed using the least-squares linear-phase finite impulse response filter and feed-forward neural network. In [33], three types of common arrhythmia beats, namely, premature ventricular contraction, ventricular fibrillation, and heart block beat along with normal beats, are classified using a real-time FPGA implementation. In [34], an intra-patient scheme based on arrhythmia classification is implemented in the FPGA system. However, most of the successful FPGA implemented systems are followed by an intra-patient scheme. Very few methods are developed in real-time systems based on inter-patient schemes [35]. However, still, these systems failed in detecting rare abnormal beats accurately. Hence, there is a need for developing a new expert system that can succeed in identifying rare heartbeats.

1.3. Contribution

In this paper, we presented an efficient inter-patient heartbeat classification algorithm. For any pattern recognition process, identifying an appropriate set of features and classifier is highly significant. From [36], it is noticeable that ECG is a non-stationary, non-Gaussian signal derived from nonlinear systems. Hence, we employed a decomposition method, namely improved complete ensemble empirical mode decomposition (ICEEMD) to obtain features from the ECG beats. This technique is capable of disclosing the implicit information lying in the ECG. Later, different nonlinear measures like entropies and HOS are determined from the modes obtained after ICEEMD. These measures will serve as features for proper discrimination of the heartbeat groups. The fundamental difficulty in processing these groups is the class imbalance. Here, a significant fraction of the heartbeats is non-ectopic. Hence, the results may be biased toward the majority group, which is undesirable. Therefore, to alleviate such an issue, we followed an algorithmic level approach. To achieve this, we employed a majority voting scheme based classification. It is a type of ensemble classification. The advantage of ensemble classification is that it can reduce both variance and bias. In this work, we used different combinations of classifiers, namely, naïve Bayes, linear, and quadratic discriminant functions, J48, and consolidated J48 classifiers for majority voting.

The rest of the paper is ordered as follows: the ECG data set, training, and testing data division of AAMI labeling, experimental details and theoretical background of the methodology are presented in Section 2. Section 3 presents the simulation results of the proposed method. The comparison with existing works, limitations, and future directions are presented in Section 4. The conclusions of the work are presented in Section 5.

2. Methods

The block diagram of the proposed method is illustrated in Figure 1. The methodology consists of three stages including pre-processing, feature extraction on training and testing data, and a classification model for evaluation. In this section, the database used and the theoretical background of the used techniques are discussed.

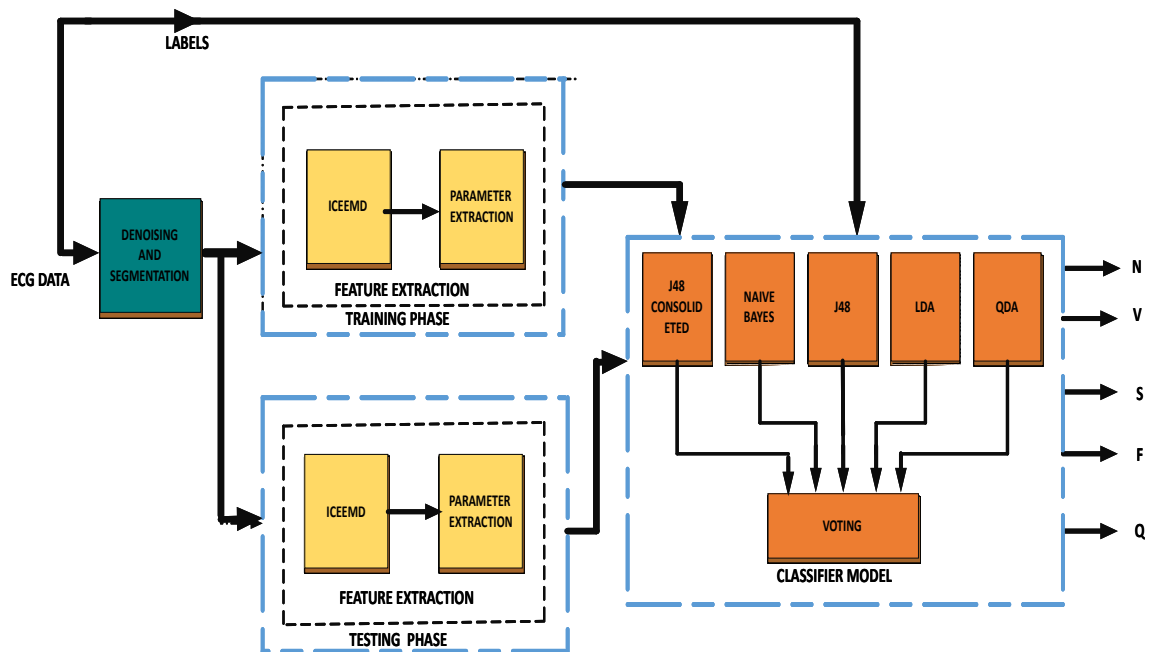


Figure 1. Block diagram of the proposed methodology.

2.1. Database

The proposed method is examined using the MIT-BIH arrhythmia database. MIT-BIH is a standard database widely explored for arrhythmia classification. It comprises of Holter monitoring records from several male and female patients. Each record duration is 30 minutes, sampled at 360 Hz. The records consist of both normal and abnormal beats of 15 types.

The annotation files available in the database are obtained from the chart recordings recognized by the experts. This file describes the 'R' peak locations and the labeling of normal and abnormal beats. Based on the recommendations of AAMI, class-labeling was assigned for discriminating various heartbeat groups.

AAMI Class Labeling Recommendations

According to the ANSI/AAMI EC57:1998 standard, within the annotation files, beat labels are divided into five groups, namely, N, S, V, F, and Q based on the physiological origin of the beats. Here, the mainly N group consists of normal and bundle branch block beats. S and V groups consist of ectopic beats, originated above and below Atrio Ventricular (A–V) junction of the heart, respectively. The F group consists of the combination of ventricular and normal beats. Unclassifiable beats are placed in the Q group. According to [31], the total number of available heartbeats are divided into training for modeling and testing for evaluation. Details of the number of heartbeats utilized for this work are presented in Table 1.

Table 1. Details of the number of beats selected from each group of AAMI classes.

AAMI Classes	MIT-BIH Heartbeats	Total Data	Training (DS1)	Testing (DS2)
N	Normal, left and right bundle branch block atrial and nodal escape beats	83,761	41,746	42,015
S	atrial premature contraction, Aberrated atrial, supra ventricular and junctional premature beats	2614	777	1837
V	premature ventricular contraction, ventricular flutter and escape beats	6893	3787	3106
F	fusion of ventricular and normal beats	526	266	260
Q	paced, unclassifiable, fusion of paced and normal beats	12	6	6

2.2. Pre-Processing

Pre-processing is an initial step in any data processing systems. Raw ECG signals will inherently have some artifacts. These may occur due to instrumental noise (power line interference), a physiological signal disturbance (muscular movements), or the environment where the experiment takes place. These artifacts are undesirable and diminish significant features in the ECG. Therefore, to attenuate the effect of this noise, we perform denoising as one of the pre-processing steps. For this, we used a filtering routine proposed by [37] with minimal modification. This operation comprises the following:

1. Mean separation from the noisy ECG,
2. Moving average filter of order five,
3. High-pass filter with cut-off frequency 1 Hz (for baseline wander suppression),
4. Low pass Butter worth filter with cut-off frequency 45 Hz (To suppress any left out high-frequency noise).

We need individual heartbeats from the long-term ECG recording for heartbeat classification. We perform a segmentation process after denoising. In the segmentation process, the annotation chart records with 'R' peak locations are utilized. From the annotation file, it is observed that there are a lot of variations among R-peak positions time-to-time. The difference between the R-peak positions is dynamic. Hence, we applied a window of length 300 samples on ECG signal to obtain an ECG segment that covers the QRS complex which is an important epoch in the ECG. Our segmentation process retains other important epochs like P and T waves, unlike centered R-peak distribution segmentation methods.

2.3. Feature Extraction

Feature extraction has a critical role in heartbeat classification. A feature provides crucial information about a signal and facilitates better discrimination of classes. From [36], it is evident that ECG is a non-stationary signal stemming from a nonlinear system. Hence, exploration of ECG with nonlinear methods can improve the performance of a model since they extract subtle information lying in ECG. Therefore, in the feature extraction stage, initially, we perform ICEEMD on ECG segments to get intrinsic mode functions (IMFs). Later, entropy and higher-order cumulants are extracted from the selected modes. In this section, the techniques employed and their support in the methodology development are briefly discussed.

2.3.1. ICEEMD

The EMD decomposes a given signal in a full data-dependent approach by exploiting the local characteristics. However, EMD is limited by "mode-mixing" problem while analyzing the real data [38]. Therefore, some noise-assisted data analysis methods can provide a solution. Here, noise is added in a controlled manner for developing new extrema. Thus, the local mean is limited to that of the original version where extrema are generated. A few among these noise assisted methods are EEMD [39] and CEEMD [40]. Among these methods, CEEMD provides a better solution to the mode-mixing problem. However, CEEMD has some limitations:

- (i) Some residual can be present in the modes.
- (ii) During the initial decomposition stages, information may appear "late" with undesired modes, when it is compared to EEMD.

To address these issues, Colominas et al. [41] introduced a new noise aided adaptive data analysis method called ICEEMD. The mathematical details of the ICEEMD are given below [41].

Notation used in algorithm: $E_l(\cdot)$ = l^{th} EMD mode, $M(\cdot)$ = local mean of the signal, $\langle \cdot \rangle =$ averaging operator, $w^{(j)}$ = realization of white Gaussian noise with zero mean and unit variance and x = input signal.

The algorithm steps:

1. Compute the local means of J realizations $x^{(j)} = x + \beta_0 E_1(w^{(j)})$, $j = 1, 2, \dots, J$ using EMD, to obtain first residue $r_1 = \langle M(x^{(j)}) \rangle$.
2. At the first stage ($l = 1$), compute the first IMF:

$$C_1 = x - r_1. \quad (1)$$

3. For $l = 2, \dots, L$, calculate r_l as

$$r_l = \langle M(r_{l-1} + \beta_{l-1} E_l(w^{(j)})) \rangle. \quad (2)$$

4. Calculate the l^{th} mode as

$$C_l = r_{l-1} - r_l. \quad (3)$$

5. Go to step 3 for next l

Here, $\beta_l = \epsilon_0 \sigma_{(r_l)}$ is used to obtain the desired SNR at each stage. We choose $\epsilon_0 = 0.2$. The resultant IMFs provide significant underlying features of the ECG signal. The ICEEMD is a beneficial tool used for analyzing non-stationary signals originating from nonlinear systems such as bio-signals. The main advantage of ICEEMD is: avoiding the spurious modes and reducing the amount of noise in the mode patterns. Thus, the decomposed IMFs capture the morphology of the signal. Later, entropy and statistical measures are calculated from the first six modes of each ECG segment.

2.3.2. Entropy Measures

Entropy measures the uncertainty in a given data. It is often used in signal processing and pattern recognition applications [42]. A high value of entropy maps to higher uncertainty (or) unpredictability. Entropy yields useful information for analyzing non-stationary signals [43]. In this work, we calculated Shannon [44], log energy, and norm entropies [45]. The entropy E must be an additive cost function such that $E(0) = 0$ and

$$E(s) = \sum_i E(s_i),$$

where s is the probability of the given signal and i represents one of the discrete states. Various entropies are defined below:

- **Shannon Entropy:**

$$E_{\text{Shannon}}(s) = \sum_i s_i^2 \log(s_i^2), \quad (4)$$

- **log Energy Entropy:**

$$E_{\text{log energy}}(s) = \sum_i \log(s_i^2), \quad (5)$$

with the convention $\log(0) = 0$.

- **norm Entropy:** The l^p norm entropy with $1 \leq p$ is defined as

$$E_{\text{norm}}(s) = \sum_i |(s_i^p)| = \|s\|_p^p. \quad (6)$$

2.3.3. HOS

HOS provides a meaningful measure for analyzing non-stationary signals originating from nonlinear systems [46,47]. HOS represents the deviation from Gaussianity and can provide useful information from the non-Gaussian nature of ECG signals. In our work, we utilized second, third, and fourth-order cumulants as HOS. The mathematical details of the HOS can be found in [48].

We construct a feature vector of size 36×1 for each heartbeat (6 features \times 6 modes = 36). Later, the training feature set is fed to a classifier for building a model, and that model is evaluated using a testing set.

2.4. Voting Scheme

The final goal of machine learning is to get better-generalized performance. We come across a question “which learning algorithm or classifier is preferred over the other?”. According to a “No free lunch theorem” [49], there is no precise answer to this. One algorithm fits or performs well for a set of training and testing data and may fail for another. The learning algorithm overall performance depends on the prior information, distribution of data, amount of training data, and some cost functions. The performance generalization depends on the bias and variance errors. Always, there will be a trade-off between bias and variance. Ensemble classifiers form a better choice, to improve the performance generalization by reducing bias and variance. Combining several classifiers for the final decision is called an ensemble classification or mixture-of-experts model or modular classification.

The primary motivation behind the classifier ensemble is improving the classification performance using the complementary information offered by various classifiers. Kittler et al. [50] developed a scheme for combining classifiers using voting based on a set of rules: min-rule, max-rule, product-rule, sum-rule, and median-rule. From our experiments, we preferred product rule which outperforms others.

Mathematical Framework: Consider a pattern recognition model where a pattern y is to be assigned with one of the m possible classes ($\omega_1, \omega_2, \dots, \omega_m$). Say there are R number of classifiers used for combining. Let us assume that each classifier possesses a different representation of measurement vector $\mathbf{x}_i, i = 1, 2, \dots, R$.

The density function for each class ω_k in the measurement space is $p(\mathbf{x}_i|\omega_k)$ and the prior probability is $P(\omega_k)$. We assume that the models are mutually exclusive.

From the Bayesian framework, y is assigned to the class ω_j having a maximum posterior probability out of ω_k classes:

$$\text{assign } y \rightarrow \omega_j \text{ if } P(\omega_j|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R) = \max_k P(\omega_k|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R). \quad (7)$$

Rewriting the posterior probability $P(\omega_k|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R)$ using the Bayes theorem:

$$P(\omega_k|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R) = \frac{p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R|\omega_k)P(\omega_k)}{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R)}. \quad (8)$$

Here, $P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R)$ can be expressed in terms of conditional measurement distribution as

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R) = \sum_{j=1}^m p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R|\omega_j)P(\omega_j). \quad (9)$$

Product Rule: $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R|\omega_j)$ represents the joint probability distribution of the measurements computed by the classifiers. Assuming that these representations are statistically independent, we can rewrite the joint probability distribution as

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R|\omega_k) = \prod_{i=1}^R p(\mathbf{x}_i|\omega_k). \quad (10)$$

Based on $p(\mathbf{x}_i|\omega_k)$, the measurement process model for i^{th} representation is developed. Substituting Equation (10) and Equation (9) into Equation (8)

$$P(\omega_k|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R) = \frac{P(\omega_k) \prod_{i=1}^R p(\mathbf{x}_i|\omega_k)}{\sum_{j=1}^m p(\omega_j) \prod_{i=1}^R p(\mathbf{x}_i|\omega_j)} \quad (11)$$

and using Equation (11) in Equation (7), we obtain the decision rule

$$\text{assign } \mathbf{y} \rightarrow \omega_j \text{ if } P(\omega_j) \prod_{i=1}^R p(\mathbf{x}_i | \omega_j) = \max_{k=1}^n P(\omega_k) \prod_{i=1}^R p(\mathbf{x}_i | \omega_k). \quad (12)$$

Rewriting in terms of the posterior probabilities obtained from the respective learning algorithms,

$$\text{assign } \mathbf{y} \rightarrow \omega_j \text{ if } \frac{1}{P^{(R-1)}(\omega_j)} \prod_{i=1}^R p(\omega_j | \mathbf{x}_i) = \max_{k=1}^n \frac{1}{P^{(R-1)}(\omega_k)} \prod_{i=1}^R p(\omega_k | \mathbf{x}_i). \quad (13)$$

Equation (13) represents the likelihood decision rule obtained after combining the posterior probabilities generated by different classifiers using the product rule.

In this work, we used five different classifiers for ensembling using a voting scheme to enhance the performance of the system: naïve Bayes [51], linear and quadratic discriminant functions [52], J48 [53], and J48 consolidated classifiers [54]. A brief description of these classifiers is given below.

Naïve Bayes Classifier: It is a probability-based learning algorithm developed on the Bayesian framework. According to Bayes theorem, an unknown \mathbf{y} is categorized into the one among the R classes, with high posteriori probability:

$$\mathbf{y} \rightarrow \omega_k \text{ if } \arg \max_{\omega_k \in \omega} P(\omega | \mathbf{y}) P(\mathbf{y}), \quad (14)$$

where $\omega = \{\omega_1, \omega_2, \dots, \omega_R\}$ is a vector of R classes.

Naïve Bayes is a modified version of Bayes classifier, based on the assumption that the features in an unknown example vector are independent. Therefore, posteriori probability can be written as

$$P(\omega | \mathbf{y}) = P(\mathbf{y} | \omega) P(\omega) = P(y_1, y_2, \dots, y_m | \omega) = P(y_1 | \omega) P(y_2 | \omega) \dots P(y_m | \omega) P(\omega). \quad (15)$$

Hence, Equation (14) can be modified as

$$\mathbf{y} \leftarrow \arg \max_{\omega_k} P(\omega = \omega_k) \prod_i P(y_i | \omega = \omega_k). \quad (16)$$

With this final rule, the naïve Bayes classifier operates. The parameters used for the Naïve Bayes Classifier is given in Table 2.

Table 2. Naïve Bayes Classifier parameters used in this work.

Parameters	Naïve Bayes
Use Kernel Estimator	False
Use supervise Discretization	True

In general, the naïve Bayes classifier assumes that the given features follow the normal distribution. In Table 2, use the Kernel Estimator parameter set to false to follow this assumption. Supervised discretization converts a specific range of attribute values to binary values. Here, the term supervised is coined because the class information of the training instances is used for discretization. However, this process is possible only when the class labels are nominal. The advantage of supervised discretization in naïve Bayes classifier is present in [55].

Linear and Quadratic Discriminant Analysis Based Classifiers: The approach of discriminant analysis is to derive a decision boundary or a discriminant function based on the linear combinations of features that best separate the given classes. The assumption made is: examples from different

categories follow Gaussian distribution. For instance, the discrimination function for two-class problems based on Bayes theory can be written as

$$(\mathbf{y} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} (\mathbf{y} - \boldsymbol{\mu}_1) + \ln |\boldsymbol{\Sigma}_1| - (\mathbf{y} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_2^{-1} (\mathbf{y} - \boldsymbol{\mu}_2) > T, \quad (17)$$

where $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ are the mean vectors of class 1 and class 2, $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ are the covariance matrices of class 1 and class 2 and T is the threshold value.

The above function without further assumptions represents the quadratic discriminate function. If the covariance matrices $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, then the discriminant function simplifies to a dot product.

$$\mathbf{x} \cdot \mathbf{y} > \text{constant}, \quad (18)$$

where $\mathbf{x} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, $\text{constant} = \frac{1}{2}(T - \boldsymbol{\mu}_1^T \boldsymbol{\Sigma} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T \boldsymbol{\Sigma} \boldsymbol{\mu}_2)$. This decision rule represents the classification based on linear discriminant.

The parameters used for linear discriminant analysis (LDA) and quadrature discriminant analysis (QDA) classifiers are given below in Table 3.

Table 3. LDA and QDA classifier parameters used in this work.

Parameters	LDA	QDA
Ridge	1.0×10^{-6}	1.0×10^{-6}

Ridge parameters in the discriminant analysis classifiers reduce the overfitting problem by penalizing the large quantity coefficients. In our work, we use the default values as given in Table 3.

J48 Classifier: Recently, decision tree-based algorithms have become popular in machine learning strategies. In practice, J48 is an execution of popular C 4.5 algorithms proposed by Quinlan [53]. According to this algorithm, the decision process involves the construction of a tree based on the feature splitting. The superiority of matching \mathbf{y} to a class label $\omega_k \in \omega$ depends on the choice of feature splitting based on the value of information gain.

Information gain is measured with the help of difference entropy as the difference between the entropy of the central node to the sum of entropies of the leaf nodes. It measures how well a given feature splits the training data under its class label. A feature node having high information gain is preferred.

J48 Consolidated (J48-C) Classifier: It is a consolidated version of C 4.5 classifier. “J48 consolidated” is an implementation of a consolidated tree’s construction algorithm, proposed by Arbelaz et al. [54] in WEKA. The basic idea is building a single tree using several subsamples. In each iteration, we will find a better feature using information gain content similar to J48. After finding the best feature split, all the subsamples are divided using the same features. More details can be found in [54]. The parameters used for J48 and J48-C classifiers are given in below Table 4.

Table 4. J48 and J48-C classifier parameters used in this work.

Parameters	J48	J48-C
Minimum Objects	1000	1000
Use MDL correction	True	True
Number of folds	3	3
Sub-tree raising	True	True

Details of the parameters can be find in WEKA 3.9 version [56].

J48 and J48-C classifiers are decision tree classifiers in which tree splitting criteria play a significant role. The above-mentioned parameters determine the growth and direction of the tree structures that influence the final model accuracy. Sub-tree raising considers raising of a sub-tree when pruning is

enabled. The minimum number of objects determines the number of instances per leaf. Minimum description length (MDL) correction is a statistical measure like information gain to identify the best split tree. The number of folds determines the data used for error reduce pruning; here, one fold is for pruning and the other folds for building the tree.

All the parameters are fixed based on the final results. All the details of the parameters can be found in WEKA 3.9 version [56].

3. Results

In this work, we are classifying the five classes: N, V, S, F, Q. The training set is constructed with the array of records as DS1 = [101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230], and DS2 = [100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234]. Here, the numerical values represent the patient record number. Four records (102, 104, 107 and 217) having paced beats are exempted from both DS1 and DS2 data sets.

We start with scatter plots for justifying the choice of features in discriminating against the heartbeats. Individual performance of five classifiers naïve Bayes, LDA, QDA, J48, and J48 consolidated is presented, and analysis using a voting scheme with various combinations of these classifiers is considered. The performance is illustrated for each combination. We used the WEKA 3.9 version (University of Waikato, New Zealand) [56] for implementing the classification algorithms and scatter plots. Data pre-processing and feature extraction is implemented using MATLAB 2018a (Mathworks, MA, USA). All the experiments are carried out in Windows 8, 8 GB RAM, and 64-bit operating system.

The Performance Measures

An algorithm's efficiency can be validated with appropriate performance measures. In this work, Sensitivity (SEN), False Positive Rate (FPR), Positive Predictive Value (PPV), and Overall Accuracy (OA) are used as performance measures to compare with the state-of-the-art methods, following the AAMI recommendations. The confusion matrix required for calculating these measures is given in Table 5. For V and S classes, the measures are calculated as per [31]. For remaining classes, we followed [57].

Table 5. Confusion matrix.

Actua Labels	Predicted Labels					Sum
	N	V	S	F	Q	
N	N_n	N_v	N_s	N_f	N_q	R_N
V	V_n	V_v	V_s	V_f	V_q	R_V
S	S_n	S_v	S_s	S_f	S_q	R_S
F	F_n	F_v	F_s	F_f	F_q	R_F
Q	Q_n	Q_v	Q_s	Q_f	Q_q	R_Q
Sum	C_N	C_V	C_S	C_F	C_Q	R/C

Performance measure from Table 5 can be calculated as follows:

The sum measures of row-wise and column-wise calculations are:

$$\begin{aligned}
 R_N &= N_n + N_v + N_s + N_f + N_q; & C_N &= N_n + V_n + S_n + F_n + Q_n; \\
 R_V &= V_n + V_v + V_s + V_f + V_q; & C_V &= N_v + V_v + S_v + F_v + Q_v; \\
 R_S &= S_n + S_v + S_s + S_f + S_q; & C_S &= N_s + V_s + S_s + F_s + Q_s; \\
 R_F &= F_n + F_v + F_s + F_f + F_q; & C_F &= N_f + V_f + S_f + F_f + Q_f; \\
 R_Q &= Q_n + Q_v + Q_s + Q_f + Q_q; & C_Q &= N_q + V_q + S_q + F_q + Q_q; \\
 R &= R_N + R_V + R_S + R_F + R_Q; & C &= C_N + C_V + C_S + C_F + C_Q; \\
 R &= C.
 \end{aligned} \tag{19}$$

The false-positive and false negative values for each class are defined as below:

$$\begin{aligned}
 FN_N &= R_N - N_n; & FP_N &= C_N - N_n; \\
 FN_V &= R_V - V_v; & FP_V &= C_V - (V_v + F_v + Q_v); \\
 FN_S &= R_S - S_s; & FP_S &= C_S - (S_s + Q_s); \\
 FN_F &= R_F - F_f; & FP_F &= C_N - F_f; \\
 FN_Q &= R_Q - Q_q; & FP_Q &= C_N - Q_q.
 \end{aligned} \tag{20}$$

The other useful measures, true positives, and negatives can be calculated for Classes N,V,S,F,Q:

$$\begin{aligned}
 TP_N &= N_n; & TN_N &= R - (R_N + C_N - N_n); \\
 TP_V &= V_v; & TN_V &= R - (R_V + C_V - V_v); \\
 TP_S &= S_s; & TN_S &= R - (R_S + C_S - S_s); \\
 TP_F &= F_f; & TN_F &= R - (R_F + C_F - F_f); \\
 TP_Q &= Q_q; & TN_Q &= R - (R_Q + C_Q - Q_q).
 \end{aligned} \tag{21}$$

The performance measures are given by

$$SEN = \frac{TP}{TP + FN}; FPR = \frac{FP}{TN + FP}; PPV = \frac{TP}{TP + FP}; OA = \frac{TP_N + TP_V + TP_S + TP_F + TP_Q}{R},$$

where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

We present the scatter plots with marginal histograms on the testing data set DS2, for features in the two-dimensional feature space. These scatter plots reveal how different features spread in feature space, thereby revealing the relationship between different heartbeat classes. Figure 2 shows the two-dimensional scatter plot between cumulant 2 of IMF1 to norm entropy value of IMF2. In this plot, we can observe that N and V beats are dominantly spread across space. In addition, the histogram plots also reveal the good discrimination between N, V, and Q classes out of the five classes. The next plot from Figure 3 gives the relation between the log energy entropy of IMF1 to cumulant 2 of IMF1. In this figure, we can see the spreading of N, V, S, and F classes in the space. In particular, this space provides good discrimination between N, V, and S classes. From Figures 4–6, we can observe that log energy entropy values extracted from different IMFs provide a good perception of N, V, and S classes.

In the same way, Figures 7–8 give better discrimination of Q beats, which are very rare indeed. In these figures, the characteristic feature is the norm entropy. In addition, different combinations of features with norm entropy reveal different class spreads and discrimination capabilities. As a whole, we can say that the combinations of selected features from different IMFs can predict the required hypothesis.

After dividing the training and testing feature sets, we need to learn a model for classification. In this work, we used an ensemble learner for classification. Ensemble classifiers use multiple learning algorithms and combine all the decisions. It can be more accurate than the individual classifiers. The main advantage of the ensemble classifiers is that we can achieve low bias error and low variance error. Ensembles using multiple trained (high variance/ low bias) models can average out of the variance, leaving just the bias. In addition, ensemble classifiers are preferred for imbalanced datasets. Our DS1 and DS2 datasets are highly imbalanced with majority N group class and minor F and Q classes. Therefore, in this work, we used a voting scheme based on product rule to ensemble the classifiers. The individual classifier performance on DS2 (testing data) is presented in Tables 6 and 7. Confusion matrices calculated for LDA, QDA, naïve Bayes, J48 and J48-C classifiers are shown in Table 6. The performance measures for the corresponding matrices based on Table 5 are presented in Table 7.

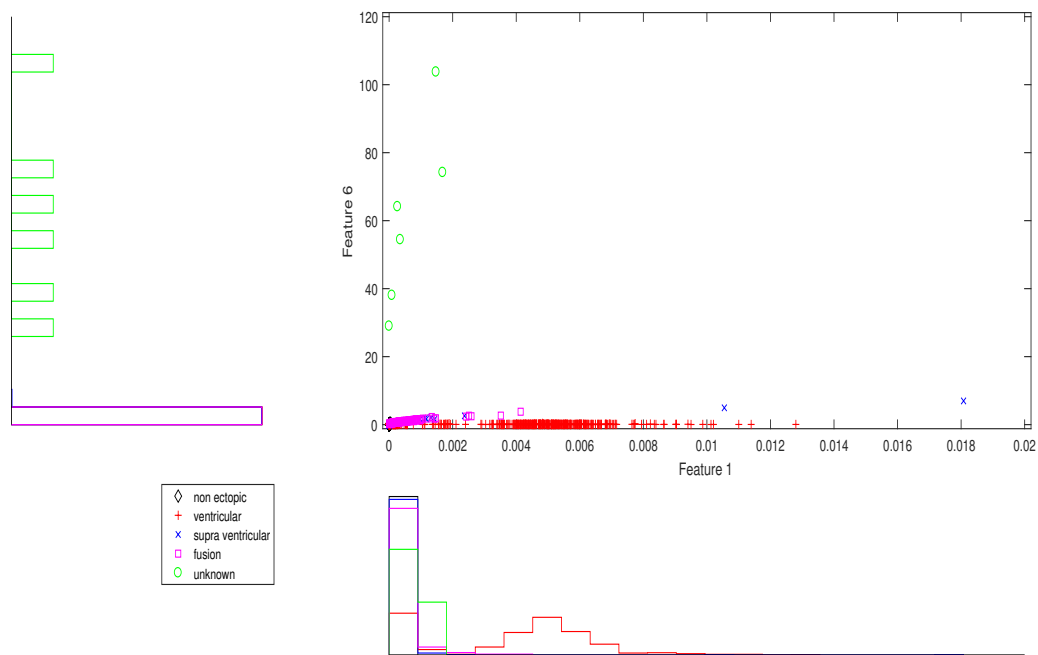


Figure 2. Scatter plot with marginal histogram for CUM2 (IMF1) vs. norm entropy (IMF1).

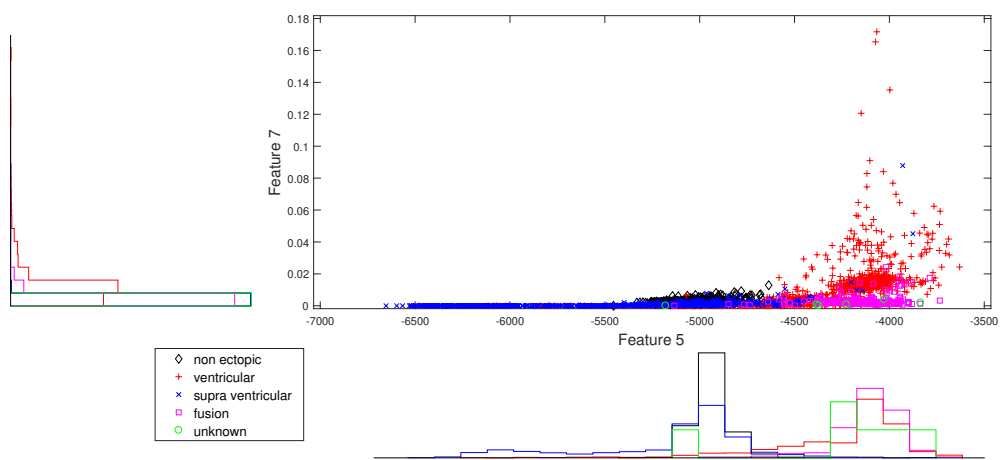


Figure 3. Scatter plot with marginal histogram for log entropy (IMF1) vs. CUM2 (IMF1).

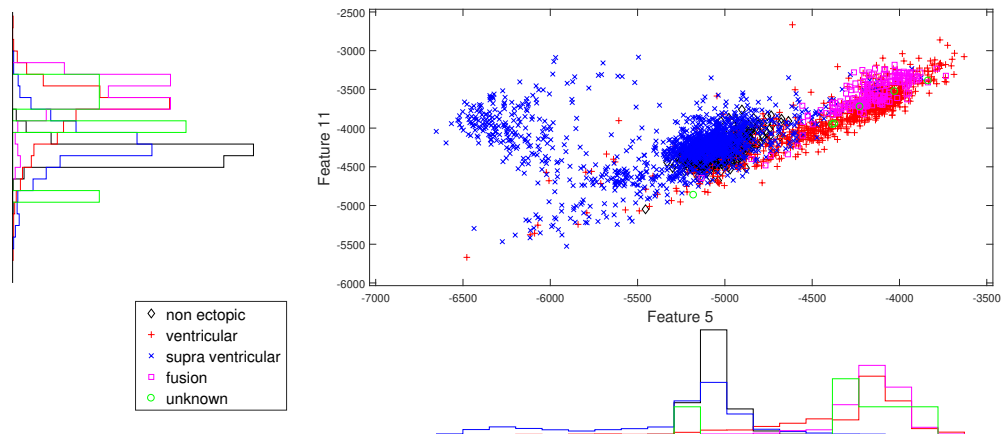


Figure 4. Scatter plot with marginal histogram for log entropy (IMF1) vs. log entropy (IMF2).

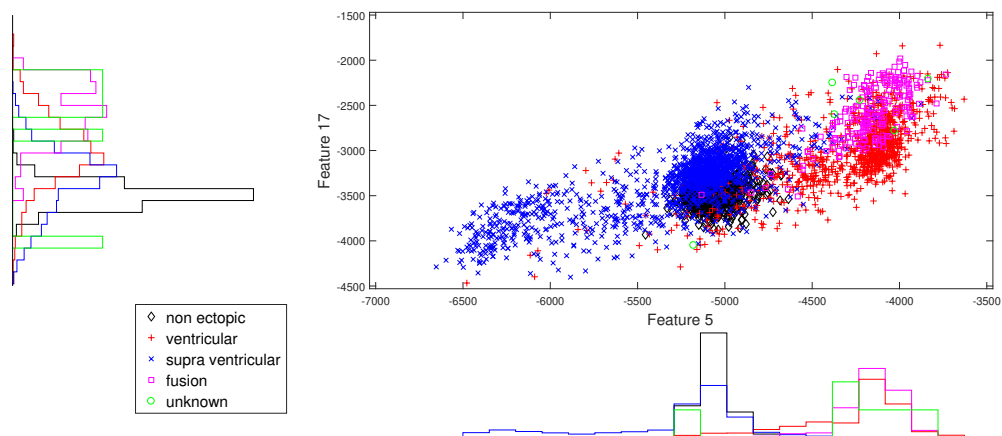


Figure 5. Scatter plot with marginal histogram for log entropy (IMF1) vs. log entropy (IMF3).

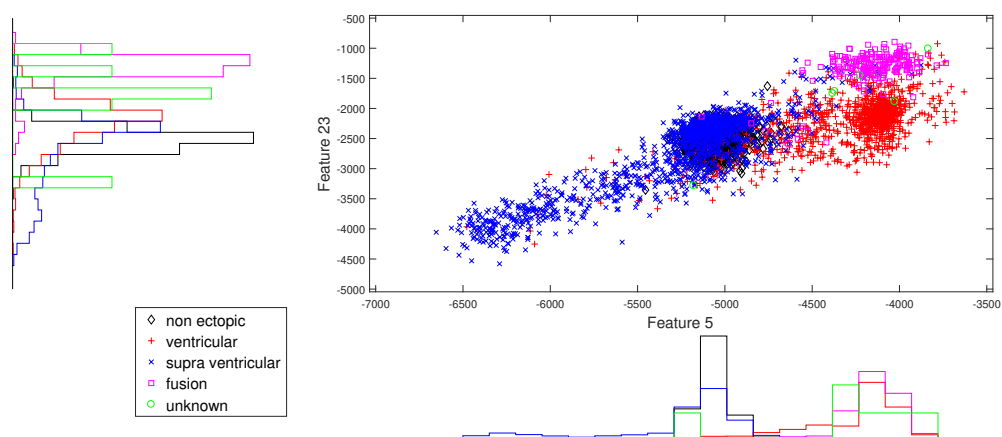


Figure 6. Scatter plot with marginal histogram for log entropy (IMF1) vs. log entropy (IMF4).

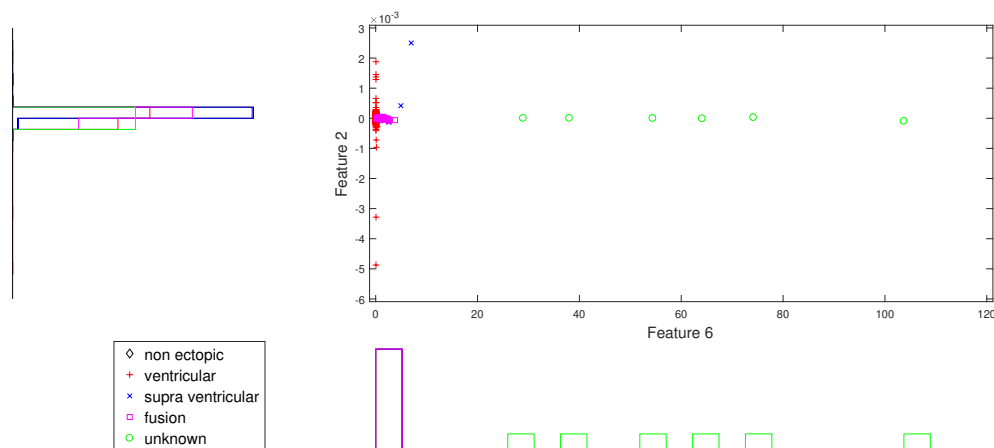


Figure 7. Scatter plot with marginal histogram for norm entropy (IMF1) vs. CUM2 (IMF1).

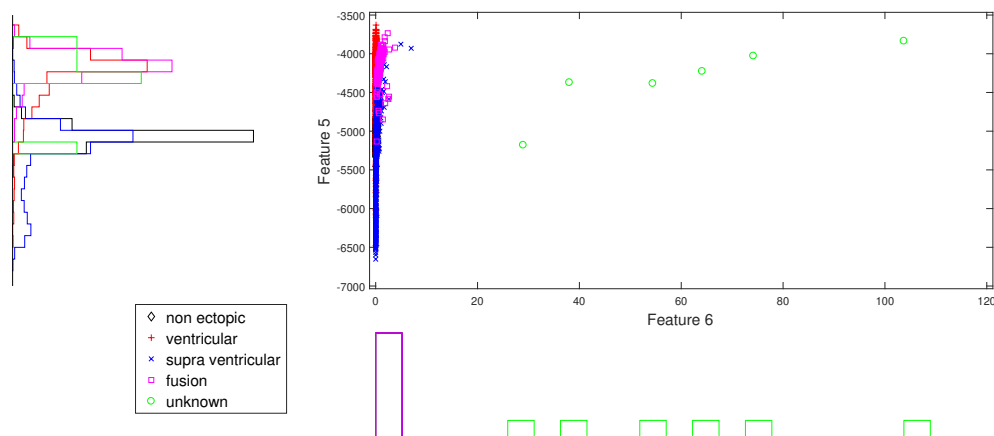


Figure 8. Scatter plot with marginal histogram for norm entropy (IMF1) vs. log entropy (IMF1).

From this table, we can see that each classifier yields different prediction. LDA and J48 give better classification for the N and V classes. It is an understandable phenomenon because of the dominating number of examples in N and V. LDA and J48-C provides better discrimination to Q group. The other classifiers J48-C and naïve Bayes are providing better SEN results for F group. Finally, S class is predicted accurately by J48-C and QDA classifiers. The other important point is, although all classifiers yield better results for the specific group of categories, the OA is dominated by the N class discrimination. Therefore, it is noticeable that OA is no longer a useful performance measure for imbalanced data classification.

In Table 8, the confusion matrix after combining J48, LDA, and naïve Bayes classifiers using the voting scheme is presented. The corresponding performance measures are demonstrated in Table 9. From the results, it is evident that this combination yields better results for N, V, F, and Q classes and average result for S class. The critical point is N, and S classes have more morphological similarities. Therefore, individual classifiers are giving complementary results for N and S. However, this ensemble selection enhances the prediction generalization for both classes.

Table 6. Confusion matrices for LDA, QDA, naïve Bayes, J48, and J48-C classifiers.

LDA					QDA					
	N	V	S	F	Q	N	V	S	F	Q
N	40,842	33	47	1093	0	2777	266	36,394	2578	0
V	319	2787	0	0	0	5	3101	0	0	0
S	1831	1	2	3	0	95	56	1675	11	0
F	171	0	0	89	0	91	64	20	85	0
Q	0	0	0	0	6	0	6	0	0	0

naïve Bayes					J48					
	N	V	S	F	Q	N	V	S	F	Q
N	36,222	745	997	3289	762	41,801	214	0	0	0
V	930	1874	52	231	19	363	2743	0	0	0
S	1321	3	133	27	353	1819	18	0	0	0
F	8	0	2	245	5	259	1	0	0	0
Q	1	0	0	4	1	5	1	0	0	0

J48-C					
	N	V	S	F	Q
N	0	6941	28,966	5290	818
V	0	2205	894	6	1
S	0	24	1785	19	9
F	0	30	23	191	16
Q	0	0	0	0	6

Table 7. Performance measures for LDA, QDA, naïve Bayes, J48 and J48-C classifiers.

LDA				QDA			
OA = 92.60	SEN%	FPR%	PPV%	OA = 16.20	SEN%	FPR%	PPV%
N	97.2	44.6	94.6	6.6	3.7	93.6	
V	89.7	0.1	98.8	99.8	0.9	88.8	
S	0.1	0.1	4.1	91.2	80.2	4.4	
F	34.2	2.3	7.5	32.7	5.5	3.2	
Q	100	0	100	0	0	0	

naïve Bayes				J48			
OA = 81.47	SEN%	FPR%	PPV%	OA = 94.32	SEN%	FPR%	PPV%
N	86.2	43.4	94.1	99.5	47	94.5	
V	60.3	1.7	71.5	88.3	0.5	92.1	
S	7.2	2.3	11.2	0	0	0	
F	94.2	7.6	6.5	0	0	0	
Q	16.7	2.4	0.1	0	0	0	

J48-C				
	OA = 8.86	SEN%	FPR%	PPV%
N	0	0	0	0
V	71.0	15.9	24.0	
S	97.2	65.8	5.6	
F	73.5	11.3	3.5	
Q	100	1.8	0.7	

Table 8. Confusion matrix for combining J48, LDA, and naïve Bayes classifiers using a Voting scheme.

Voting (J48, LDA, naïve Bayes)					
	N	V	S	F	Q
N	39,542	53	489	1931	0
V	395	2708	3	0	0
S	1473	1	353	10	0
F	22	5	1	232	0
Q	0	0	0	0	6

Table 9. Performance measures for combining J48, LDA, and naïve Bayes classifiers using a Voting scheme.

Voting (J48, LDA, naïve Bayes)			
OA = 90.71	SEN%	FPR%	PPV%
N	94.1	36.3	95.4
V	87.2	0.1	97.9
S	19.2	1.1	41.7
F	89.2	4.1	10.7
Q	100	0	100

Similarly, we performed ensemble voting for different combinations and the results are presented in Tables 10–15. Each combination provides various enhanced results in some aspects.

Table 10. Confusion matrix for combining J48, naïve Bayes, and QDA classifiers using a voting scheme.

Voting (J48, naïve Bayes, QDA)					
	N	V	S	F	Q
N	35,629	253	3836	2297	0
V	11	3095	0	0	0
S	1188	11	624	14	0
F	28	61	11	160	0
Q	0	6	0	0	0

Table 11. Performance measures for combining J48, naïve Bayes, and QDA classifiers using a Voting scheme.

Voting (J48, naïve Bayes, QDA)			
OA = 83.6	SEN%	FPR%	PPV%
N	84.8	23.6	96.7
V	99.6	0.8	90.3
S	34	8.5	14
F	61.5	4.9	6.5
Q	0	0	0

Table 12. Confusion matrix for combining J48-C, naïve Bayes, and QDA classifiers using a Voting scheme.

Voting (J48-C, naïve Bayes, QDA)					
	N	V	S	F	Q
N	29,730	255	9089	2491	0
V	8	3098	0	0	0
S	1031	8	779	19	0
F	21	61	12	166	0
Q	0	6	0	0	0

As mentioned earlier, the dataset is dominated by N, V, and S classes, respectively. The F and Q classes are sporadic. Therefore, in some works, only N, V, and S classes are considered for classification. We provide the results for such schemes in Tables 16–25. Here, first results are also presented for individual classifiers; later, the ensemble voting scheme is performed on different combinations of classifiers. Each one gives better classifications than individual classifiers.

Table 13. Performance measures for combining J48-C, naïve Bayes, and QDA classifiers using a Voting scheme.

Voting (J48-C, naïve Bayes, QDA)			
OA = 71.51	SEN%	FPR%	PPV%
N	70.8	20.3	96.6
V	99.7	0.7	90.4
S	42.4	20.1	7.9
F	63.8	6.3	5.3
Q	0	0	0

Table 14. Confusion matrix for combining J48-C, naïve Bayes, and LDA classifiers using a Voting scheme.

Voting (J48-C, naïve Bayes, LDA)					
	N	V	S	F	Q
N	38,329	66	942	2678	0
V	553	2532	20	1	0
S	1405	2	416	14	0
F	20	4	1	235	0
Q	0	0	0	0	6

Table 15. Performance measures for combining J48-C, naïve Bayes, and LDA classifiers using a Voting scheme.

Voting (J48-C, naïve Bayes, LDA)			
OA = 87.91	SEN%	FPR%	PPV%
N	91.2	38	95.1
V	81.5	0.2	97.2
S	22.6	2.1	30.2
F	90.4	5.7	8
Q	100	0	100

Table 16. Confusion matrices for various classifiers (N, V, and S classes).

	LDA			QDA		
	N	V	S	N	V	S
N	41,875	27	113	4569	480	36,966
V	280	2826	0	5	3101	0
S	1835	0	2	105	56	1676
	naïve Bayes			J48		
	N	V	S	N	V	S
N	37,385	2968	1662	41,998	17	0
V	922	2120	64	1095	2011	0
S	1336	18	483	1837	0	0
	J48-C					
		N	V	S		
	N	34,999	7016	0		
	V	651	2455	0		
S	1585	252	0			

Table 17. Performance measures for various classifiers (N,V, S classes).

LDA				QDA			
OA = 95.19	SEN%	FPR%	PPV%	OA = 19.90	SEN%	FPR%	PPV%
N	99.7	42.8	95.2		10.9	2.2	97.6
V	91.0	0.1	99.1		99.8	1.2	85.3
S	0.1	0.3	1.7		91.2	81.9	4.3
naïve Bayes				J48			
OA = 85.15	SEN%	FPR%	PPV%	OA = 93.71	SEN%	FPR%	PPV%
N	89	45.7	94.3		100	59.3	93.5
V	68.3	6.8	41.5		64.7	0	99.2
S	26.3	3.8	21.9		0	0	0
J48-C							
		OA = 79.76	SEN%	FPR%	PPV%		
	N		83.3	45.2	94		
	V		79	16.6	25.2		
	S		0	0	0		

Table 18. Confusion matrix for combining J48, LDA, and naïve Bayes classifiers using a Voting scheme (N,S,V).

Voting (J48, LDA, naïve Bayes)			
	N	V	S
N	40,918	361	736
V	205	2897	4
S	1469	3	365

Table 19. Performance measures for combining J48, LDA, and naïve Bayes classifiers using Voting scheme (N,S,V).

Voting (J48, LDA, naïve Bayes)			
OA = 94.08	SEN%	FPR%	PPV%
N	97.4	33.9	96.1
V	93.3	0.8	88.8
S	19.9	1.6	33

Table 20. Confusion matrix for combining J48, naïve Bayes, and QDA classifiers using Voting scheme (N,S,V).

Voting (J48, naïve Bayes, QDA)			
	N	V	S
N	37,421	574	4020
V	12	3094	0
S	1203	8	626

Table 21. Performance measures for combining J48, naïve Bayes, and QDA classifiers using Voting scheme (N,S,V).

Voting (J48, naïve Bayes, QDA)			
OA = 87.61	SEN%	FPR%	PPV%
N	89.1	24.6	96.9
V	99.6	1.3	84.2
S	34.1	8.9	13.5

Table 22. Confusion matrix for combining J48-C, naïve Bayes, and QDA classifiers using Voting scheme (N,S,V).

Voting (J48-C, naïve Bayes, QDA)			
	N	V	S
N	32011	598	9406
V	7	3099	0
S	1062	8	767

Table 23. Performance measures for combining J48-C, naïve Bayes, and QDA classifiers using Voting scheme (N,S,V).

Voting (J48-C, naïve Bayes, QDA)			
OA = 76.40	SEN%	FPR%	PPV%
N	76.2	21.6	96.8
V	99.8	1.4	83.6
S	41.8	20.8	7.5

Table 24. Confusion matrix for combining J48-C, naïve Bayes, and LDA classifiers using Voting scheme (N,S,V).

Voting (J48-C, naïve Bayes, LDA)			
	N	V	S
N	40,248	688	1079
V	359	2735	12
S	1407	2	428

Table 25. Performance measures for combining J48-C, naïve Bayes, LDA classifiers using Voting scheme (N,S,V).

Voting (J48-C, naïve Bayes, LDA)			
OA = 92.44	SEN%	FPR%	PPV%
N	95.8	35.7	95.8
V	88.1	1.6	79.9
S	23.3	2.4	28.2

4. Discussion

This section contains a discussion on simulating the proposed methodology illustrated in Figure 1. In this work, we employ an adaptive non-stationary and nonlinear decomposition method, namely ICEEMD, to analyze the ECG heartbeats. ICEEMD produces a local and entirely data-driven separation of a signal in the form of fast and slow oscillations called IMFs. The main advantage of ICEEMD is that it successfully avoids the spurious nodes and reduces the amount of noise in the modes.

Later, six nonlinear morphological features: higher-order cumulants, log, Shannon energy, and norm entropies are extracted from the first six IMFs of each heartbeat, to generate a 36×1 feature vector. Then, these feature vectors are divided based on training and testing sets DS1 and DS2 as specified above. Statistics (median and interquartile range) of these features for each class are presented in Table 26. Variation of attributes corresponding to different heartbeats can be observed from this table.

In Table 6, we presented the individual performance of various classifiers on the given problem. Here, all classifier models offer separate results for all the classes. Each model performs well for a specific class or classes. However, it fails in providing the overall better performance. For example, the S class contains 1837 beats, the J48, LDA, and naïve Bayes are predicting 51, 2, 1516 beats, respectively. Whenever we combine these three models using the voting scheme as shown in Table 12, this combined model identified 779 beats correctly. The voting scheme uses the product of probabilities rule. In this scheme, it is assumed that each model representation for a given class is statistically independent. It is because of the different representation capabilities of each model. From this, a final decision rule is formed as described in Section 2.4. This decision rule quantifies the probability of class choice from combined hypothesis models and the same type of results we can observe for other classes. In this work, we implemented four voting schemes with different classifier combinations. Each combination again provides different but better results than individual classifier models. The proposed combinations of classifier details are given in Table 27.

Table 26. Median \pm interquartile range of features extracted on DS1.

Feature Number	Features	N	S	V	F	Q
1	CUM2(IMF1)	$1.0 \times 10^{-6} \pm 4.0 \times 10^{-6}$	$8.0 \times 10^{-6} \pm 4.03 \times 10^{-5}$	$7.50 \times 10^{-5} \pm 0.00045775$	0.000162 ± 0.000326	0.0001405 ± 0.009121
2	CUM3(IMF1)	0 ± 0	0 ± 0	$0 \pm 2.0 \times 10^{-6}$	$0 \pm 5.0 \times 10^{-6}$	0 ± 0.000313
3	CUM4(IMF1)	0 ± 0	0 ± 0	$1.0 \times 10^{-6} \pm 2.58 \times 10^{-5}$	$2.50 \times 10^{-6} \pm 1.1 \times 10^{-5}$	$4.5 \times 10^{-6} \pm 0.006113$
4	Shan(IMF1)	0.004769 ± 0.009686	0.017643 ± 0.06530625	0.112159 ± 0.4199265	0.216149 ± 0.330006	0.153638 ± 1.47632
5	log(IMF1)	-4792.4962 ± 379.828	-4735.1172 ± 286.978	-4532.44 ± 467.507	-4358.371 ± 271.824	-4604.445 ± 1783.20
6	norm(IMF1)	0.1004 ± 0.0726	0.1538 ± 0.1719	0.0815 ± 0.05794	0.5541 ± 0.5434	18.1790 ± 20.55612
7	CUM2(IMF2)	0.001 ± 0.00347	0.00109 ± 0.0035	0.00028 ± 0.00137	0.00107 ± 0.00143	0.00411 ± 0.0134
8	CUM3(IMF2)	$8.0 \times 10^{-6} \pm 5.20 \times 10^{-5}$	$4.0 \times 10^{-6} \pm 5.10 \times 10^{-5}$	$0 \pm 6.0 \times 10^{-6}$	$5.0 \times 10^{-6} \pm 2.10 \times 10^{-5}$	0 ± 0.000387
9	CUM4(IMF2)	$1.50 \times 10^{-5} \pm 0.000185$	$2.10 \times 10^{-5} \pm 0.000128$	$2.0 \times 10^{-6} \pm 4.20 \times 10^{-5}$	$1.60 \times 10^{-5} \pm 5.20 \times 10^{-5}$	0.00046 ± 0.002581
10	Shan(IMF2)	1.3074 ± 2.9180	1.3675 ± 3.4250	0.431 ± 1.429	1.4044 ± 1.44285	2.90440 ± 6.76104
11	log(IMF2)	-4055.72 ± 403.632	-4045.57 ± 565.833	-4064.90 ± 479.11	-3788.21 ± 266.57	-4023.39 ± 1366.58
12	norm(IMF2)	2.41 ± 3.81	2.57 ± 4.95	2.744 ± 1.768	2.69 ± 2.04	18.17 ± 20.556
13	CUM2(IMF3)	0.0069 ± 0.0097	0.00534 ± 0.009013	0.00443 ± 0.00936	0.011997 ± 0.00909	0.004148 ± 0.008015
14	CUM3(IMF3)	$-7.80 \times 10^{-5} \pm 0.000322$	$-8.60 \times 10^{-5} \pm 0.0003605$	$-3.10 \times 10^{-5} \pm 0.000246$	-0.00019 ± 0.00062	$8.95 \times 10^{-5} \pm 0.000266$
15	CUM4(IMF3)	0.000235 ± 0.000863	0.000104 ± 0.0006515	0.000117 ± 0.000747	0.0005255 ± 0.000767	0.0001945 ± 0.000396
16	Shan(IMF3)	6.3075515 ± 6.075866	5.532014 ± 6.254886	4.570122 ± 5.93904725	10.002892 ± 5.49255	3.8205085 ± 6.888019
17	log(IMF3)	-3114.4106 ± 443.70730	-3071.71175 ± 705.504	-3178.659 ± 531.3571	-2731.2660 ± 467.4692	-3341.5431 ± 891.0573
18	norm(IMF3)	9.3564 ± 7.6444	8.69133 ± 8.37546	7.585 ± 4.1465	14.85 ± 7.440	18.179 ± 20.55
19	CUM2(IMF4)	0.013817 ± 0.0210	0.01233 ± 0.01653	0.0215 ± 0.03611	0.0370 ± 0.0297	0.0069 ± 0.0102
20	CUM3(IMF4)	-0.00012 ± 0.00081	$-6.80 \times 10^{-5} \pm 0.000469$	-0.00014 ± 0.001305	-0.001542 ± 0.00207	0 ± 0.00039
21	CUM4(IMF4)	0.000147 ± 0.000734	$4.60 \times 10^{-5} \pm 0.0003225$	0.000404 ± 0.00189	0.00027 ± 0.00124	0.000138 ± 0.00018
22	Shan(IMF4)	13.05585 ± 12.76457	12.2365 ± 11.295	16.104 ± 15.135	25.934 ± 15.110	7.113 ± 10.471
23	log(IMF4)	-2150.494 ± 533.542	-2143.567 ± 611.308	-2206.943 ± 794.75	-1614.17 ± 480.99	-2593.22 ± 1271.81
24	norm(IMF4)	19.367 ± 15.509	18.686 ± 14.240	16.26 ± 9.166	35.59 ± 20.941	18.179 ± 20.556
25	CUM2(IMF5)	0.0084 ± 0.018	0.0066 ± 0.0156	0.033 ± 0.0648	0.056 ± 0.058	0.0075 ± 0.0149
26	CUM3(IMF5)	$1.0 \times 10^{-6} \pm 0.00017$	0 ± 0.000103	$-1.0 \times 10^{-6} \pm 0.00146$	$1.40 \times 10^{-5} \pm 0.0022$	0 ± 0.000405
27	CUM4(IMF5)	$-7.00 \times 10^{-6} \pm 0.00012$	$-6.0 \times 10^{-6} \pm 0.000127$	$-3.0 \times 10^{-5} \pm 0.00193$	-0.00236 ± 0.00621	$-5.50 \times 10^{-6} \pm 0.000287$
28	Shan(IMF5)	10.358 ± 15.6374	8.68112 ± 14.823	25.882 ± 29.195	39.2240 ± 25.737	8.992 ± 16.965
29	log(IMF5)	-1917.009 ± 593.83	-1992.177 ± 872.920	-1567.422 ± 918.993	-1216.140 ± 489.3834	-2179.500 ± 1495.143
30	norm(IMF5)	17.118 ± 19.0735	14.983 ± 19.2726	19.495 ± 14.659	52.96 ± 32.190	18.179 ± 20.556
31	CUM2(IMF6)	0.0040 ± 0.0092	0.0028 ± 0.00723	0.0184 ± 0.04606	0.0571 ± 0.0827	0.0073 ± 0.01072
32	CUM3(IMF6)	$1.0 \times 10^{-6} \pm 0.00010$	$0 \pm 6.80 \times 10^{-5}$	$2.0 \times 10^{-6} \pm 0.0011117$	$8.0 \times 10^{-6} \pm 0.00527$	0 ± 0.0002
33	CUM4(IMF6)	$-1.50 \times 10^{-5} \pm 0.00011$	$-8.0 \times 10^{-5} \pm 6.25 \times 10^{-5}$	-0.00024 ± 0.00205	-0.00389 ± 0.01428	$-4.40 \times 10^{-5} \pm 0.00017$
34	Shan(IMF6)	6.50405 ± 11.11697	5.26119 ± 9.9055	20.0112 ± 31.26347	44.49 ± 39.5131	10.284 ± 13.481
35	log(IMF6)	-1903.169 ± 637.510	-1983.32 ± 713.457	-1470.196 ± 744.176	-1062.5 ± 546.13	-1682.363 ± 1296.591
36	norm(IMF6)	13.0914 ± 15.1495	11.4410 ± 14.268	10.566 ± 12.262	57.357 ± 46.317	18.179 ± 20.556

Note: CUM2-second order cumulat, CUM3-third order cumulant, CUM4-Fourth order cumulant, Shan-Shannon entropy, log-Log entropy, norm- norm entropy.

4.1. Comparative Analysis

To assess the performance of our proposed methodology, we compared our results with the existing methods in the literature. Comparisons are presented in Tables 28 and 29. The features and classification schemes employed by various researchers listed in Tables 28 and 29 are given in Table 27. In Table 28, we compare our four sets of voting schemes with the works which followed AAMI recommendations based on [31] division scheme. In addition, Table 29 shows the performance measure comparison with the literature on only N, V, and S classification. In Table 28, our proposed methods yield almost similar performance compared to the state-of-the-art for N, S, and V classes; however, in case of F and Q classes, our proposed work one and four outperforms the other compared methods. From Table 29, it is evident that our proposed methods one and three are efficiently distinguished the classes N, S, and V. The best results of our method are highlighted in bold. Overall, the measures of our work are appreciable compared with other approaches.

Table 27. Methodology description of recent state-of-the-art compared in our work.

Literature	Feature Extraction	Classification
[58]		
method-1	R-R intervals	Optimum Path Forest (OPF)
method-2	Wavelet based features	OPF
method-3	Mean, standard deviation and average power of wavelet sub-band	OPF
method-4	Auto correlation and energy ratio of wavelet bands	OPF
method-5	Fast-ICA	OPF
method-6	(Wavelet+ICA+RR interval)	OPF
[59]	(ECG+VCG) complex network based features	SVM
[42]	Wavelet packet decomposition based entropy features	Random Forest
[60]		
method-1	Wavelet based features	Hierarchical Classification (tree approach)
method-2	Mean, standard deviation and average power of wavelet sub-band	Hierarchical Classification (tree approach)
method-3	Auto correlation and energy ratio of wavelet bands	Hierarchical Classification (tree approach)
method-4	Fast-ICA	Hierarchical Classification (tree approach)
method-5	(Wavelet+ICA+RR interval)	Hierarchical Classification (tree approach)
[61]	Temporal Vectrcardiogram(TCG) based features	SVM
[62]	A combination of projected features (features derived from the projected matrix and DCT) and RR intervals	SVM
[63]	TCG feature selection by PSO	SVM
proposed work		
method-1	Entropy and statistical features calculated on ICEEMD modes	Voting (J48, LDA, naïve Bayes)
method-2	Entropy and statistical features calculated on ICEEMD modes	Voting (J48, QDA, naïve Bayes)
method-3	Entropy and statistical features calculated on ICEEMD modes	Voting (J48-C, QDA, naïve Bayes)
method-4	Entropy and statistical features calculated on ICEEMD modes	Voting (J48-C, LDA, naïve Bayes)

Table 28. Performance comparison with recent literature (N, S, V, F, and Q classes).

Literature	N	S	V	F	Q
	SEN/FPR/PPV	SEN/FPR/PPV	SEN/FPR/PPV	SEN/FPR/PPV	SEN/FPR/PPV
[58]					
method-1	84.5/-/-	1.0/-/-	77.7/-/-	38.4/-/-	0/-/-
method-2	86.4/-/-	2.3/-/-	40.8/-/-	0.5/-/-	0/-/-
method-3	84.8/-/-	18.3/-/-	77.8/-/-	7.5/-/-	0/-/-
method-4	92.5/-/-	3.0/-/-	61.8/-/-	16.8/-/-	0/-/-
method-5	95.7/-/-	17.7/-/-	74.7/-/-	3.9/-/-	0/-/-
method-6	93.2/-/-	12.1/-/-	85.5/-/-	18.3/-/-	0/-/-
[59]	89.3/25.2/96.6	38.6/6.7/18	81.2/4.9/53.6	0/0/0	0/0/0
[42]	94.67/3.92/99.73	20/3.69/0.16	94.20/0.71/89.78	50/0.78/0.52	0/0/0
[60]					
method-1	92.3/22.2/97.1	28.5/2.6/29.6	83.5/5.51/51.2	19.1/1.07/12.3	0/0/-
method-2	93.6/57.1/93.0	0.49/0.47/3.81	67.9/3.99/54.2	0/1.63/0	0/0/0
method-3	98.2/41.2/95.1	4.72/0.71/20.3	81.7/1.25/82.0	2.58/0.40/4.88	0/0/0
method-4	98.6/39.8/95.3	9.15/0.56/38.6	83.2/1.21/82.7	0.26/0.38/0.53	0/0/-
method-5	94.7/31.2/96.1	37.4/6.19/18.8	43.9/1.48/67.4	0.52/0.72/0.56	0/0/-
proposed work					
method-1	94.1/36.3/95.4	19.2/1.1/41.7	87.2/0.1/97.9	89.2/4.1/10.7	100/0/100
method-2	84.8/23.6/96.7	34/8.5/14	99.6/0.8/90.3	61.5/4.9/6.5	0/0/0
method-3	70.8/20.3/96.6	42.4/20.1/7.9	99.7/0.7/90.4	63.8/6.3/5.3	0/0/0
method-4	91.2/38/95.1	22.6/2.1/30.2	81.5/0.2/97.2	90.4/5.7/8	100/0/100

Table 29. Performance comparison with recent literature (N, S, and V classes).

Literature	N	S	V
	SEN/FPR/PPV	SEN/FPR/PPV	SEN/FPR/PPV
[61]	95/27.9/96.5	29.6/3.1/26.4	85.1/3.01/66.3
[62]	98.4/-/95.4	29.5/-/38.4	70.8/-/85.1
[63] method on VCG	79.1/27.0/96.3	31.2/8.4/13.0	89.5/7.2/46.1
proposed work			
method-1	97.4/33.9/96.1	19.9/1.6/33	93.3/0.8/88.8
method-2	89.1/24.6/96.9	34.1/8.9/13.5	99.6/1.3/84.2
method-3	76.2/21.6/96.8	41.8/20.8/7.5	99.8/1.4/83.6
method-4	95.8/35.7/95.8	23.3/2.4/28.2	88.1/1.6/79.9

4.2. Limitation and Future Scope

Despite the proposed method giving significant results, the performance of the S class is still limited when compared to other classes. Similar behavior is observed in other state-of-the-art methods. Hence, there is a need to explore a new set of attributes and learning algorithms to improve this. In addition, incorporating other physiological signals such as blood pressure, plethysmographic signals along with ECG may improve the description of “heart functioning.”

5. Conclusions

In this work, we implemented a computer-aided inter-patient heartbeat classification algorithm. We employed a nonlinear decomposition method called ICEEMED, to extract some important information lying in ECG. Later, HOS and entropy measures are calculated on the modes obtained after ICEEMD and used as features. Class imbalance is one of the critical challenges in medical diagnosis. We addressed this issue by utilizing the voting scheme as the learning model. The extracted features are then fed to this model for classification. To design this model, naïve Bayes, linear and quadratic discriminating functions, J48 and J48 consolidated classifiers are explored. The proposed method showed promising results compared to state-of-the-art techniques. Our method opens new frontiers to the successful identification of rare heartbeat groups enabling a real-time heart monitoring system.

Author Contributions: Conceptualization, R.N.V.P.S.K.; methodology, K.N.V.P.S.R.; software, R.N.V.P.S.K.; validation, K.N.V.P.S.R.; formal analysis, R.N.V.P.S.K.; investigation, R.N.V.P.S.K.; resources, R.N.V.P.S.K.; data curation, R.N.V.P.S.K.; writing—original draft preparation, R.N.V.P.S.K. and R.D.; writing—review and editing, R.N.V.P.S.K., R.D., P.P., G.R.N., H.M., G.D.G. and S.G.; visualization, R.N.V.P.S.K.; supervision, R.N.V.P.S.K., R.D. and P.P.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Alwan, A. *Global Status Report on Noncommunicable Diseases 2010*; World Health Organization: Geneva, Switzerland, 2011.
- Augustyniak, P.; Tadeusiewicz, R. Assessment of electrocardiogram visual interpretation strategy based on scanpath analysis. *Physiol. Meas.* **2006**, *27*, 597. [[CrossRef](#)] [[PubMed](#)]
- Moody, G.B.; Mark, R.G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [[CrossRef](#)] [[PubMed](#)]
- Pławiak, P. Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Syst. Appl.* **2018**, *92*, 334–349. [[CrossRef](#)]
- Yang, J.; Bai, Y.; Lin, F.; Liu, M.; Hou, Z.; Liu, X. A novel electrocardiogram arrhythmia classification method based on stacked sparse auto-encoders and softmax regression. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 1733–1740. [[CrossRef](#)]

6. Tuncer, T.; Dogan, S.; Pławiak, P.; Acharya, U.R. Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. *Knowl. Based Syst.* **2019**, 104923. [[CrossRef](#)]
7. Rajesh, K.N.; Dhuli, R. Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine. *Comput. Biol. Med.* **2017**, *87*, 271–284. [[CrossRef](#)]
8. Pławiak, P. Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals. *Swarm Evol. Comput.* **2018**, *39*, 192–208. [[CrossRef](#)]
9. Yıldırım, Ö.; Pławiak, P.; Tan, R.S.; Acharya, U.R. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput. Biol. Med.* **2018**, *102*, 411–420. [[CrossRef](#)]
10. Pławiak, P.; Acharya, U.R. Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals. *Neural Comput. Appl.* **2019**, pp. 1–25.
11. Pławiak, P.; Abdar, M. Novel Methodology for Cardiac Arrhythmias Classification Based on Long-Duration ECG Signal Fragments Analysis. In *Biomedical Signal Processing*; Springer: Singapore, 2020; pp. 225–272.
12. Khalaf, A.F.; Owis, M.I.; Yassine, I.A. A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines. *Expert Syst. Appl.* **2015**, *42*, 8361–8368. [[CrossRef](#)]
13. Mert, A. ECG feature extraction based on the bandwidth properties of variational mode decomposition. *Physiol. Meas.* **2016**, *37*, 530. [[CrossRef](#)] [[PubMed](#)]
14. Li, H.; Liang, H.; Miao, C.; Cao, L.; Feng, X.; Tang, C.; Li, E. Novel ECG signal classification based on KICA nonlinear feature extraction. *Circ. Syst. Signal Process.* **2016**, *35*, 1187–1197. [[CrossRef](#)]
15. Alickovic, E.; Subasi, A. Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. *J. Med. Syst.* **2016**, *40*, 108. [[CrossRef](#)] [[PubMed](#)]
16. Martis, R.J.; Acharya, U.R.; Lim, C.M.; Mandana, K.; Ray, A.K.; Chakraborty, C. Application of higher order cumulant features for cardiac health diagnosis using ECG signals. *Int. J. Neural Syst.* **2013**, *23*, 1350014. [[CrossRef](#)] [[PubMed](#)]
17. Sharma, P.; Ray, K.C. Efficient methodology for electrocardiogram beat classification. *IET Signal Process.* **2016**, *10*, 825–832. [[CrossRef](#)]
18. Mishra, A.K.; Raghav, S. Local fractal dimension based ECG arrhythmia classification. *Biomed. Signal Process. Control* **2010**, *5*, 114–123. [[CrossRef](#)]
19. Osowski, S.; Hoai, L.T.; Markiewicz, T. Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 582–589. [[CrossRef](#)]
20. Ye, C.; Kumar, B.V.; Coimbra, M.T. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2930–2941.
21. Lin, C.H. Classification enhancible grey relational analysis for cardiac arrhythmias discrimination. *Med. Biol. Eng. Comput.* **2006**, *44*, 311–320. [[CrossRef](#)]
22. Cuesta-Frau, D.; Biagetti, M.O.; Quinteiro, R.A.; Mico-Tormos, P.; Aboy, M. Unsupervised classification of ventricular extrasystoles using bounded clustering algorithms and morphology matching. *Med. Biol. Eng. Comput.* **2007**, *45*, 229–239. [[CrossRef](#)]
23. Tadeusiewicz, R. Neural networks as a tool for modeling of biological systems. *Bio-Algorithms Med.-Syst.* **2015**, *11*, 135–144. [[CrossRef](#)]
24. Kutlu, Y.; Kuntalp, D. A multi-stage automatic arrhythmia recognition and classification system. *Comput. Biol. Med.* **2011**, *41*, 37–45. [[CrossRef](#)] [[PubMed](#)]
25. Martis, R.J.; Acharya, U.R.; Min, L.C. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomed. Signal Process. Control* **2013**, *8*, 437–448. [[CrossRef](#)]
26. Martis, R.J.; Acharya, U.R.; Lim, C.M.; Suri, J.S. Characterization of ECG beats from cardiac arrhythmia using discrete cosine transform in PCA framework. *Knowl.-Based Syst.* **2013**, *45*, 76–82. [[CrossRef](#)]
27. Elhaj, F.A.; Salim, N.; Harris, A.R.; Swee, T.T.; Ahmed, T. Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Comput. Methods Prog. Biomed.* **2016**, *127*, 52–63. [[CrossRef](#)]
28. Li, P.; Wang, Y.; He, J.; Wang, L.; Tian, Y.; Zhou, T.s.; Li, T.; Li, J.s. High-Performance Personalized Heartbeat Classification Model for Long-Term ECG Signal. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 78–86.
29. Desai, U.; Martis, R.J.; Nayak, C.G.; Sarika, K.; Seshikala, G. Machine intelligent diagnosis of ECG for arrhythmia classification using DWT, ICA and SVM techniques. In Proceedings of the India Conference (INDICON), New Delhi, India, 17–20 December 2015; pp. 1–4.

30. Desai, U.; Martis, R.J.; GURUDAS NAYAK, C.; Seshikala, G.; Sarika, K.; SHETTY K, R. Decision support system for arrhythmia beats using ECG signals with DCT, DWT and EMD methods: A comparative study. *J. Mech. Med. Biol.* **2016**, *16*, 1640012. [[CrossRef](#)]
31. De Chazal, P.; O'Dwyer, M.; Reilly, R.B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1196–1206. [[CrossRef](#)]
32. Egila, M.G.; El-Moursy, M.A.; El-Hennawy, A.E.; El-Simary, H.A.; Zaki, A. FPGA-based electrocardiography (ECG) signal analysis system using least-square linear phase finite impulse response (FIR) filter. *J. Elec. Syst. Inf. Technol.* **2016**, *3*, 513–526. [[CrossRef](#)]
33. Raj, S.; Maurya, K.; Ray, K.C. A knowledge-based real time embedded platform for arrhythmia beat classification. *Biomed. Eng. Lett.* **2015**, *5*, 271–280. [[CrossRef](#)]
34. Zairi, H.; Talha, M.K.; Meddah, K.; Slimane, S.O. FPGA-based system for artificial neural network arrhythmia classification. *Neural Comput. Appl.* **2019**, 1–16. [[CrossRef](#)]
35. Jewajinda, Y.; Chongstitvatana, P. FPGA-based online-learning using parallel genetic algorithm and neural network for ECG signal classification. In Proceedings of the ECTI-CON2010: The 2010 ECTI International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Chiang Mai, Thailand, 19–21 May 2010; pp. 1050–1054.
36. Rajesh, K.N.; Dhuli, R. Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. *Biomed. Signal Process. Control* **2018**, *41*, 242–254. [[CrossRef](#)]
37. Amann, A.; Tratnig, R.; Unterkofler, K. Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators. *Biomed. Eng. Online* **2005**, *4*, 60. [[CrossRef](#)] [[PubMed](#)]
38. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
39. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
40. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In Proceedings of the Acoustics, speech and signal processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 4144–4147.
41. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Process. Control* **2014**, *14*, 19–29. [[CrossRef](#)]
42. Li, T.; Zhou, M. ECG classification using wavelet packet entropy and random forests. *Entropy* **2016**, *18*, 285. [[CrossRef](#)]
43. Rosso, O.A.; Blanco, S.; Yordanova, J.; Kolev, V.; Figliola, A.; Schürmann, M.; Başar, E. Wavelet entropy: A new tool for analysis of short duration brain electrical signals. *J. Neurosci. Method.* **2001**, *105*, 65–75. [[CrossRef](#)]
44. Shannon, C.E. A mathematical theory of communication, Part I, Part II. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [[CrossRef](#)]
45. Coifman, R.R.; Wickerhauser, M.V. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theory* **1992**, *38*, 713–718. [[CrossRef](#)]
46. Martis, R.J.; Acharya, U.R.; Ray, A.K.; Chakraborty, C. Application of higher order cumulants to ECG signals for the cardiac health diagnosis. In Proceedings of the 2011 Annual International Conference on Engineering in Medicine and Biology Society (EMBC), Boston, MA, USA, 30 August–3 September 2011; pp. 1697–1700.
47. Nikias, C.L.; Mendel, J.M. Signal processing with higher-order spectra. *IEEE Signal Process. Mag.* **1993**, *10*, 10–37. [[CrossRef](#)]
48. Swami, A.; Mendel, J.M.; Nikias, C.L.M. Higher-order spectral analysis toolbox. *Tech. Support Product Enhanc. Suggest.* **1984**.
49. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evolut. Comput.* **1997**, *1*, 67–82. [[CrossRef](#)]
50. Kittler, J.; Hatef, M.; Duin, R.P.; Matas, J. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 226–239. [[CrossRef](#)]
51. John, G.H.; Langley, P. Estimating continuous distributions in Bayesian classifiers. In Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence, Montréal, QC, Canada, 18–20 August 1995; pp. 338–345.

52. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
53. Quinlan, J.R. *C4. 5: Programs for Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2014.
54. Arbelaz Gallego, O.; Gurrutxaga, I.; Lozano, F.; Muguerza, J.; Pérez, J.M. J48Consolidated: An Implementation of CTC Algorithm for WEKA. 2016. Available online: <https://addi.ehu.es/handle/10810/17314> (accessed on 17 November 2019).
55. Yang, Y.; Webb, G.I. Discretization for naive-Bayes learning: managing discretization bias and variance. *Mach. Learn.* **2009**, *74*, 39–74. [[CrossRef](#)]
56. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2016.
57. Powers, D.M. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. Available online: https://bioinfpublication.org/files/articles/2_1_1_JMLT.pdf (accessed on 19 November 2019).
58. Luz, E.J.D.S.; Nunes, T.M.; De Albuquerque, V.H.C.; Papa, J.P.; Menotti, D. ECG arrhythmia classification based on optimum-path forest. *Expert Syst. Appl.* **2013**, *40*, 3561–3573. [[CrossRef](#)]
59. Queiroz, V.; Luz, E.; Moreira, G.; Guarda, Á.; Menotti, D. Automatic cardiac arrhythmia detection and classification using vectorcardiograms and complex networks. In Proceedings of the 2015 37th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), Ilan, Italy, 25–29 August 2015; pp. 5203–5206.
60. Luz, E.J.D.S.; Merschmann, L.H.; Menotti, D.; Moreira, G.J. Evaluating a hierarchical approach for heartbeat classification from ECG. *Int. J. Bioinf. Res. Appl.* **2017**, *13*, 146–160. [[CrossRef](#)]
61. Garcia, G.; Moreira, G.; Luz, E.; Menotti, D. Improving automatic cardiac arrhythmia classification: Joining temporal-VCG, complex networks and SVM classifier. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 3896–3900.
62. Chen, S.; Hua, W.; Li, Z.; Li, J.; Gao, X. Heartbeat classification using projected and dynamic features of ECG signal. *Biomed. Signal Process. Control* **2017**, *31*, 165–173. [[CrossRef](#)]
63. Garcia, G.; Moreira, G.; Menotti, D.; Luz, E. Inter-Patient ECG Heartbeat Classification with Temporal VCG Optimized by PSO. *Sci. Rep.* **2017**, *7*, 10543. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).