Reduced Features with RBF-SVM for Diagnosing Multiclass Arrhythmia

Gorav Kumar Malik
FET, Gurukul Kangri (Deemed to be) University, Haridwar, Uttarakhand, India
Yatindra Kumar and Manoj Panda
Department of Electrical Engineering
G. B. Pant Engineering College Pauri-Garhwal, Uttarakhand India

Abstract - A Heart disease is the leading cause of death worldwide, accounting for more than half of fatalities. Arrhythmia is a condition in which the heart beats at an abnormally fast or slow pace. Doctors face complex problem in identifying and categorizing different kinds of cardiac arrhythmia. Because only a small percentage of arrhythmias are life-threatening, failing to perform the procedure correctly or promptly puts the patient's life in jeopardy. An excellent method to assist clinicians in the essential identity of a variety of cardiac arrhythmias has been shown in this study; the solution uses GDA-SVM classifying techniques to identify arrhythmia. The model is trained and tested using data from the MIT-BIH database library. Using the solution may increase the accuracy of diagnosis.

Index Terms – Arrhythmia, GDA, SVM, RBF, Data Reduction, MIT-BIH.

INTRODUCTION

The electrocardiogram (ECG) is a vital tool in diagnosing and treating any heart condition [1]. During each heartbeat, electrical impulses may be seen on the ECG machine. The ECG is helpful in determining whether a patient has the cardiac disease. Many cardiac disorders may cause any alterations in a normal ECG waveform.

ECG waves may be used to diagnose arrhythmia, a kind of heart disease [2]. The aberrant electrical activity in the heart is like the heart beats at a rapid or slow pace. It is difficult to obtain relevant information from clinical data because of the manual data processing technique. The medical industry will benefit the most from computer-based automated illness diagnostic systems. Analyzing the ECG machine data, the system will automatically make a judgment. It is one of the latest technologies that reduce human intervention for efficiency and cost-effectiveness of the equipment. A real-time automated ECG analysis system will be the finest assistance for doctors in early diagnosing cardiac arrhythmia. The ECG's characteristics change depending on the patient's physical and temporal circumstances [3]. Manually analyzing the ECG is quite challenging because ECG fluctuations may mislead. Developing an automated ECG classification system is necessary for this. [4] The extraction of meaningful characteristics from data is critical to creating an effective ECG classifier. Four distinct forms of cardiac arrhythmias are classified in the proposed study using the Support vector machine With GDA based Data reduction approach.

Rahul, Jagdeep [5] suggest using RR intervals instead of only the QT interval to classify cardiac arrhythmias better. The DWT and median filters were employed to eliminate high-frequency noise and baseline drift from the raw ECG. After determining the QRS area, the processed ECG was split. This information was used to differentiate between normal, premature ventricular contraction (PVC) and pre-atrioventricular contraction (PAC).

In the existing papers, SVM is used to classify heart disease from an ECG signal [6]. The high processing time of an optimization technique is one such issue. During the training phase attributes, are not exactly discovered, etc. To overcome the above-stated issues, the proposed method gives full attention to the multi-classification model by using innovative techniques.

In 2020 Sanamdikar [7] an efficient neural network and trustworthy algorithms for ECG arrhythmia classification are the primary goals of this article. A 98 % success rate for arrhythmia identification was achieved, which is the maximum degree of accuracy. GSNN's proposed method of classifying and predicting arrhythmia will be more effective than existing methods. Prediction and categorization will be more accurate using the new approach.

In 2016 Raj, S. et al. [8] had presented an analytic methodology for the recognition of cardiac arrhythmias. The diagnosis process includes four steps namely, filtering, peak detection, feature extraction, and classification. Features are extracted by (DOST) approach (Discrete Orthogonal Stockwell Transform), and for classification, PSO optimized Support Vector Machines (SVM) was used.

METHODOLOGY

In order to construct this arrhythmia classification system, the following steps are taken as shown in Figure 1.

- DWT is used to remove signal interferences,
- Feature extraction is used to extract unknown parameters of ECG signals from the MIT-BIH database,
- A RBF SVM classifier is used to construct the classification system by assessing a collection of ECG signals with well-known characteristics

- GDA is used to reduce the features dataset in this classification system, which results in the development of optimum multiclass datasets.

Dimension reduction of data sets helps to enhance the performance of the classification system. Finally, the SVM classifier is implemented for ECG signal classification and evaluation of performance utilizing various parameters. This study is concluded in last Section.

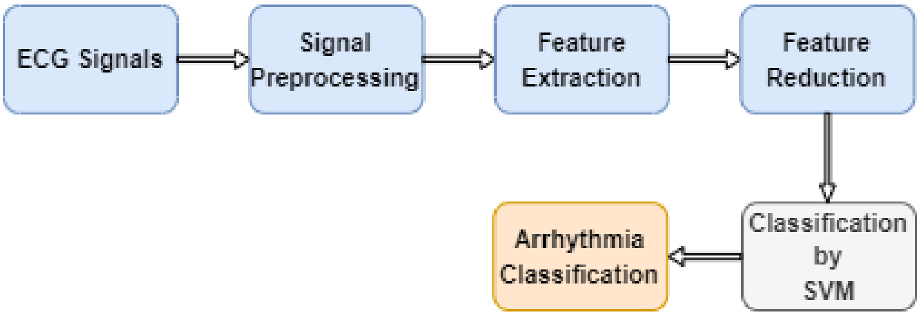

FIGURE1: PROPOSED METHODOLOGY BLOCK DIAGRAM

DATABASE

One of the most significant aspects of signal processing is developing a database. The MIT–BIH Arrhythmia database [9] was utilized in the proposed research. The Beth Israel Hospital Arrhythmia Laboratory acquired the ECGs used in MIT–BIH Arrhythmia. From the set of 45 files, 23 files (numbered from 100 to 124 inclusive, with some numbers missing) were picked at random, and 25 files were chosen to make up the second section of the database (numbered from 200 to 234 inclusive, again with some numbers are absent). Each 45-rpm record clocks in at a little more than 30 minutes. According to AMMI standards, there are a total of 16 subclasses [10]; however, in the proposed research, we take just four main classes [11] and only 29000 beats, which are split by beat segmentation.

PREPROCESSING

The most common issue with noise reduction is that the noise frequency is not always known. The finite impulse response (FIR) is the simplest and most extensively used method for minimizing noise in ECG readings.

These techniques perform well for reducing known frequency bands of noise, such as electrical network noise (50 Hz or 60 Hz), since the band-reject filter can be applied quickly and easily. The signal has a variety of frequency bands. Even while it is possible to employ high-pass and low-pass filters indiscriminately, it distorts the morphology of the ECG, so this article uses the DWT because it is efficient in analyzing non-stationary signals. ECG signals are divided into nine levels using the Daubechies D6 (db6) wavelet basis function [12]. Baseline wander dominates the frequency range 0–0.351 Hz in the ninth level approximation sub-band, which is not utilized to reconstruct the denoised signal. Because the information beyond 45 Hz is not critical for detecting arrhythmias, the denoised and smoothed, ECG signal is obtained by computing the inverse wavelet transform from the third to the ninth level detail sub-bands.

FEATURE EXTRACTION:

The ECG signal may be used to classify cardiac arrhythmias by extracting the characteristics of each heartbeat. Cardiovascular problems may be properly identified using an effective feature extraction technology. Many approaches have been presented for extracting the properties of a single heartbeat. The characteristics of a single cardiac cycle may be time- or frequency-domain specific. In [13], Inan et al. discovered that morphological and temporal information might yield good classification accuracy for huge datasets. According to [14], integrating wavelet domain characteristics with RR-interval features may improve classification accuracy.

Each patient's data's physical and temporal characteristics may provide a high degree of categorization accuracy [15].

An ECG signal may be broken down into heartbeats, and each of them reflects a cardiac cycle in terms of physiology. The P, Q, R, S, and T waves denote the smaller functional standard segments of a heartbeat. In addition to the refractory period, the average heartbeat has other inter-wave temporal segments, including the PR interval, QRS complex, QT interval, and ST-segment. Cardiovascular disease may be diagnosed by looking at the heart's arrhythmias during these intervals. Proposed methodology extracts the morphological elements as shown in Figure 2. Temporal characteristics include three heartbeat intervals and nineteen fixed interval morphological features in the feature vector. Thus, for each heartbeat, twenty-six feature vectors are retrieved that may be utilized to classify cardiac arrhythmias using various classifiers. The MIT-BIH arrhythmias database considers all characteristics for a single channel.

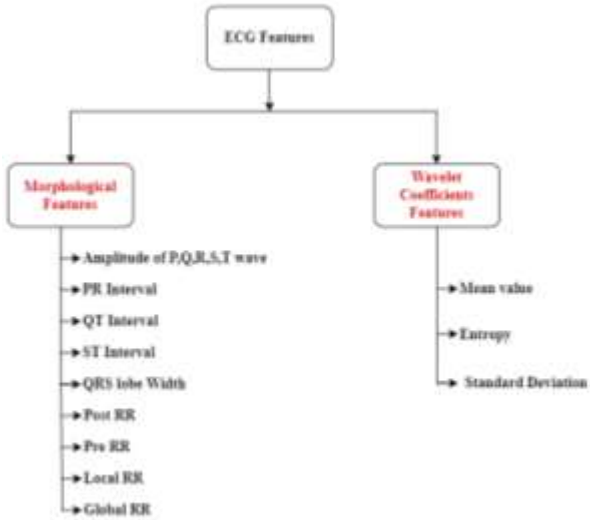

FIGURE 2: ECG FEATURES

DWT FEATURES

Non-stationary signals may be effectively analyzed using DWT [16]. An infinite sequence of sinusoidal waveforms may be generated. As a first harmonic, then as a second harmonic, and so on until the nth harmonic. Even though the harmonics of a sine wave alter in frequency, the wave remains periodic. As a result, signals may be broken down into coefficients using this approach. Low pass approximation and high pass detail coefficients are generated at each method stage. It is also possible to get finer samplings from Daubechies-6 mother wavelet since each resolution is twice the preceding scale. DWT-based feature extraction is a more effective method for classifying. For non-stationary ECG signals, DWT claims to provide acceptable scale values and shifting times [17]. There have been five decomposition layers, resulting in eight feature extraction coefficients.

$$\boldsymbol{DWT(m;n)} = \frac{\mathbf{1}}{\sqrt{\mathbf{2^m}}}\int_{-\infty}^{\infty} \boldsymbol{x(t)\psi(\frac{t-2^n m}{2^m})dt} \qquad (1)$$

In DWT, m is the decomposition level and n is the shifting parameter, data set created using morphological and wavelet characteristics for a total of 28 features.

FEATURE REDUCTION

In the realm of data analysis and feature selection, it has been proposed that not all features are helpful for classification purposes in all cases. On the contrary, some features may behave as noise, reducing the accuracy of classification by distorting the data. It is desired to eliminate the unnecessary features of ECG signals. This study uses feature reduction algorithms: GDA, which are explored in more detail in the following section.

GENERALIZED DISCRIMINANT ANALYSIS (GDA)

The Generalized Discriminant Analysis (GDA) [18] is used to deal with challenges of multi-class classification, i.e., the feature space overlapping problem. A non-linear classification using the kernel function (ϕ) is the general purpose of GDA. Here is a breakdown of the GDA feature reduction procedure:

Step 1: In order to create high-dimensional feature space, assume that the original space S is.

$$T:\phi:S \rightarrow T \qquad (2)$$

Step 2: Using equation (3) and (4) the scatter matrix, There are two ways to extract the non-linear data that is mapped, one within the class, and the other between the class and the scatter matrix.", Within the class data

$$G^{\varphi} = \sum_{k=1}^{k} N_k {n_k}^{\varphi}\left({n_k}^{\varphi}\right)^T \qquad (3)$$

Data between class and scatter matrix $H^{\varphi} = \sum_{k=1}^{k}\sum_{s \in S_k} \varphi(s)\varphi(s)^T$ (4)

Where, N_k represents the number of samples in the s_k in T and n_k^{ϕ} is the mean of s_k and k is a vector of weights.

Step 3: Projection Matrix Ratio: The primary goal of GDA is to optimize the projection matrix ratio by identifying the optimal projection matrix ratio

$$P_{opt}^{\varphi} = \arg\max \frac{\left|\left(P^{\varphi}\right)^T G^{\varphi} P^{\varphi}\right|}{\left|\left(P^{\varphi}\right)^T H^{\varphi} P^{\varphi}\right|} = \left[p_1^{\phi},.....,p_n^{\phi}\right] \qquad (5)$$

For the generalized Eigen values problem $G^{\varphi} p_i^{\varphi} = \lambda_i H^{\varphi} p_i^{\varphi}$, It is possible to evaluate the vector p^{ϕ} as the solution, and it is included within the time range T, α_{ki} are real weights and s_{ki} is the i^{th} sample of class k.

$$p^{\varphi} = \sum_{k=1}^{K} \sum_{i=1}^{N_k} \alpha_{ki} \varphi(s_{ki}) \qquad (6)$$

$$\alpha = \alpha_{ki}, i = 1,..N_k$$

Step 4: Using the Kernel Matrix and the Diagonal Matrix, we can find the solution to our problem. (11).

$$\lambda = \frac{\alpha^T K_m D_m K_m \alpha}{\alpha^T K_m K_m \alpha} \qquad (7)$$

According to the definition, the kernel and diagonal matrix are expanded as, $K_m = (K_{ab})_{a=1,...K,b=1,...K}$, $D_m = (D_{m(k)})_{k=1,...K}$. The k_{th} on the diagonal has all elements equal to $1/N_k$. and In order to define the projection vectors $p^{\varphi} \in T$, the coefficient vector α must be obtained by solving the Eigen value problem.

A projection of a testing vector s_{test} is computed as

$$(p^{\varphi})^T \varphi(s_{test}) = \sum_{k=1}^{k} \sum_{i=1}^{N_k} \alpha_{ki} W(s_{ki}, s_{test}) \qquad (8)$$

The RBF-SVM classifier is trained using the feature reduction GDA technique, which selects a reduced feature set (i.e., the best subset of features) and applies it to it. Then set of SVM RBF classifier are used to classify the signal as multi-classes like (N, S, V and F) from the ECG signal. Finally, the maximum accuracy is attained for the proposed classification technique.

RBF-SVM CLASSIFICATION

A support vector machine (SVM) [19] is a classifier that uses a separating hyperplane [20] to make a distinction between classes as shown in Figure 3. SVM classification may be separated into linear and non-linear categories based on various kernel functions.

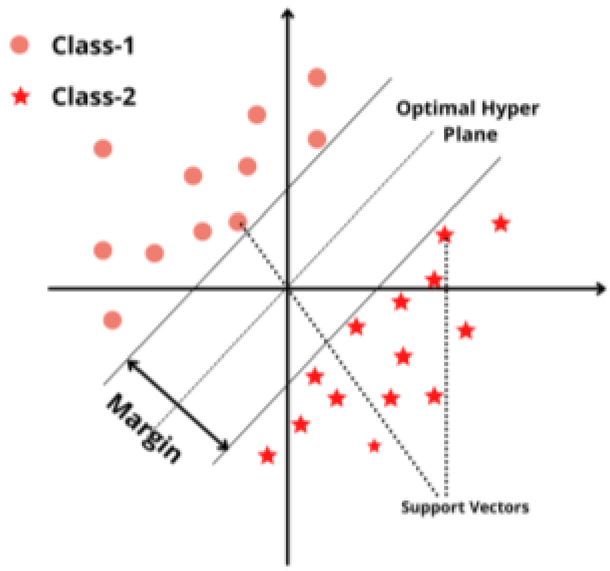

FIGURE 3: STRUCTURE OF SVM

Classifiers such as SVM can successfully address small sample and non-linear classification issues by using linear kernel functions. Input samples are mapped into a high-dimensional feature space using SVM, and then linear regression is used.

$$S(i) = w^T \mathrm{i} + \mathrm{b} \qquad (9)$$

By applying the kernel method [21] to maximum-margin hyperplanes, from inner-product to non-linear functions. The RBF kernel for a Gaussian distribution is given by:

$$S(i,i') = e^{-\gamma||\mathrm{i}-\mathrm{i}'||^2} \qquad (10)$$

The kernel of the SVM classifier is K. i, and i′ denote the input and support vectors, respectively. The reciprocal number of features is used to calculate parameter γ in this article.

Training and testing are both components of the classification process, including machine learning. Kernel computation is the primary focus of our testing implementations. Consequently, presume that methodology already know all the support vectors created during the training procedure. Take a look at the SVM classifier's RBF kernel, defined in equation (10). The squared Euclidean distance between input and support vectors $||\mathrm{i}-\mathrm{i}'||^2$and the exponential function comprise the stochastic implementation. To simplify things, let's pretend that a dataset has four characteristics. Assume b is the aim of the kernel function.

$$e^{-\frac{||\mathrm{i}-\mathrm{i}'||^2}{4}}$$

An SVM classifier utilizing an RBF kernel has gamma, and C. gamma is an RBF kernel parameter that determines the decision region. When gamma is low, the decision area is wide, and the decision border 'curve' is low. Because the decision boundary 'curve' is high, it produces islands of decision boundaries around data points. The SVC learner parameter C is the penalty for misclassifying a data point. When C is low, the classifier tolerates misclassified data (high bias, low variance). When C is high, the classifier is strongly punished for misclassified data, avoiding it at all costs (low bias, high variance).

RESULT AND DISCUSSION

The MIT-BIH arrhythmia database was used to tests of the automated arrhythmia detection system. Nonlinear and linear approaches were utilized to analyze the data, which had 28 features data set. SVM-RBF received feature vectors built from the altered data points. It was built on the famous Gaussian kernel for the nonlinear SVM (referred to as the SVM-RBF).

Accuracy, sensitivity, specificity, positive predictive value, and negative predictive value (NPV) are the four statistical indices used to evaluate the performance of a classifier [22].

$$\text{Sensitivity (Sen)} = \frac{T_P}{T_P+F_N}$$

$$\text{Specificity(Spe)} = \frac{T_N}{T_N+F_P}$$

$$\text{Accuracy (Acc)} = \frac{T_P+T_N}{T_P+F_N+F_P+T_N}$$

$$\text{Negative Predictive Value (NPV)} = \frac{T_N}{T_N+F_N}$$

$$\text{Positive Predictive Value (PPV)} = \frac{T_P}{T_P+F_P}$$

For this study, 24 records from the MIT/BIH arrhythmia database, which contains a total of 49473 beats, but the proposed study used only 29100 beats, were used to evaluate categorization performance. Complex ventricular, junctional, and supraventricular arrhythmias are included in ECG signals from the Database. For the classification tests, a total of 40% of the data beats (11640) from each kind of N, S, V, and F ECG record are included in the common portion of the training dataset. Hence Support vector machine with decreased features for patient-specific feed-forward. Experimental result shows that the RBF-SVM with reduced features by GDA achieves better results than the standard SVM-RBF classifier as depicted in Figure 4 and Table I and shows a 2.59% improvement in arrhythmia diagnosis. There seems to be a correlation between beats and sensitivity, with class F having the lowest sensitivity but the excellent accuracy due to the lower percentage of training beats, as seen in Table II of the GDA-SVM confusion matrix.

Table I: AVERAGE CLASSIFICATION ACCURACY BY DATA REDUCTION TECHNIQUES.

SVM Approach	Acc	Sen	Spe	PPV	NPV
SVM-GDA	97.39	87.22	97.67	87.64	97.82
SVM	94.8	85.4	96.12	88.69	95.9

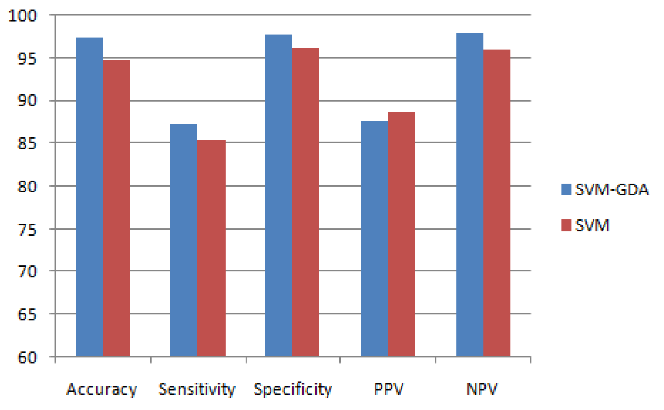

FIGURE 4: ACCURACY VALUES BY FEATURE REDUCTION

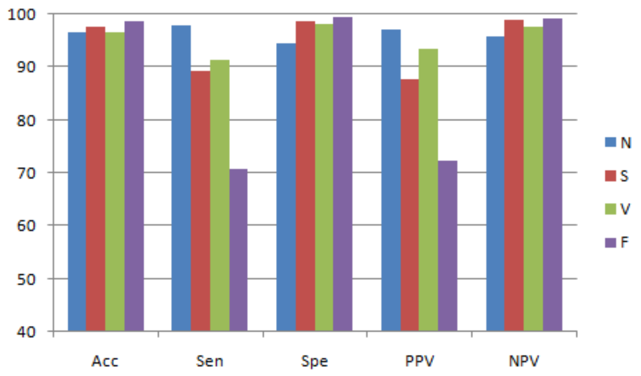

FIGURE 5: ACCURACY PERFORMANCE ANALYSIS

TABLE II: CONFUSION MATRIX FOR RBF-SVM WITH REDUCED FEATURES

Data Classes	N	S	V	F	Total
N	11143	98	118	41	**11400**
S	74	1551	96	19	**1740**
V	196	89	3561	54	**3900**
F	58	31	34	297	**420**
	11143	**1769**	**3809**	**411**	**17460**

TABLE III: PERFORMANCE EVALUATION OF CLASSIFICATION

	Acc	Sen	Spe	PPV	NPV
N	96.64	97.74	94.58	97.14	95.7
S	97.66	89.13	98.61	87.67	98.79
V	96.63	91.3	98.17	93.48	97.51
F	98.64	70.71	99.33	72.26	99.27
Average	97.39	87.22	97.67	87.64	97.82

In the experiments, an accuracy of 97.39% was achieved using the SVM-RBF with decreased features. The RBF-SVM without data reduction has a 94.8 % accuracy rate for the identical tests sets. A high degree of accuracy in signal classification is shown in Table III and Figure 5 for several classes. The Standard SVM RBF approach with GDA achieved an average of 97.39% Accuracy, 87.22% Sensitivity, and 97.67% Specificity across four classes. Consequently, the RBF-SVM with GDA data reduction has a superior classification result than the normal RBF-SVM.

CONCLUSION

The cardiac monitoring system serves an important function as a diagnostic tool in medical professions. In this study, we created a model to detect the various anomalies in cardiac arrhythmias. Only four subclasses were selected from the MIT-BIH arrhythmia dataset for the computational analysis of ECG recordings. Feature extractions to get the morphological and wavelet information while lowering the noise. RBF kernel based SVM and data reductions by GDA categorize the ECG beats. The suggested approach yields a final accuracy of 97.39 %. Using the Generalized Discriminant Analysis (GDA) has shown to be a useful tool for reducing large amounts of data. Medical professionals would benefit greatly from the newly created model since it will make it easier to analyze ECG signals and extract more information regarding cardiac ailments.

REFERENCES

[1] Raj, Sandeep, and Kailash Chandra Ray. "Sparse representation of ECG signals for automated recognition of cardiac arrhythmias." Expert Systems with Applications 105 (2018): 49-64.

[2] De Chazal, Philip, Maria O'Dwyer, and Richard B. Reilly. "Automatic classification of heartbeats using ECG morphology and heartbeat interval features." IEEE transactions on biomedical engineering 51.7 (2004): 1196-1206.

[3] Petrucci, E., et al. "The use of sequential RR distributions to detect atrial fibrillation episodes in very long term ECG monitoring." 2006 Computers in Cardiology. IEEE, 2006.

[4] Acharya, U. Rajendra, et al. "Classification of heart rate data using artificial neural network and fuzzy equivalence relation." Pattern recognition 36.1 (2003): 61-68.

[5] Rahul, Jagdeep, et al. "An improved cardiac arrhythmia classification using an RR interval-based approach." Biocybernetics and Biomedical Engineering 41.2 (2021): 656-666.

[6] Afkhami, R.G., Azarnia, G. and Tinati, M.A., 2016. Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. Journal of Pattern Recognition Letters, Vol. 70, pp.45-51.

[7] Sanamdikar, Sanjay Tanaji, Satish Tukaram Hamde, and Vinayak Ganpat Asutkar. "Analysis and classification of cardiac arrhythmia based on general sparsed neural network of ECG signals." SN Applied Sciences 2.7 (2020): 1-9.

[8] Rai, H.M., andChatterjee, K., 2018. A Novel Adaptive Feature Extraction for Detection of Cardiac Arrhythmias Using Hybrid Technique MRDWT & MPNN Classifier from ECG Big Data. Journal of Big data research.

[9] Moody, George B., and Roger G. Mark. "The impact of the MIT-BIH arrhythmia database." IEEE Engineering in Medicine and Biology Magazine 20.3 (2001): 45-50.

[10] Shimpi, Prajwal, et al. "A machine learning approach for the classification of cardiac arrhythmia." 2017 international conference on computing methodologies and communication (ICCMC). IEEE, 2017

[11] Martis, Roshan Joy, U. Rajendra Acharya, and Lim Choo Min. "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform." Biomedical Signal Processing and Control 8.5 (2013): 437-448.

[12] Mahmoodabadi, S. Z., Alireza Ahmadian, and M. D. Abolhasani. "ECG feature extraction using Daubechies wavelets." Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing. 2005.

[13] Inan, Omer T., Laurent Giovangrandi, and Gregory TA Kovacs. "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features." IEEE transactions on Biomedical Engineering 53.12 (2006): 2507-2515.

[14] Ince, Turker, Serkan Kiranyaz, and Moncef Gabbouj. "A generic and robust system for automated patient-specific classification of ECG signals." IEEE Transactions on Biomedical Engineering 56.5 (2009): 1415-1426.

[15] De Chazal, Philip, and Richard B. Reilly. "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features." IEEE transactions on biomedical engineering 53.12 (2006): 2535-2543.

[16] Zhi, Koh Yi, Oliver Faust, and Wenwei Yu. "Wavelet based machine learning techniques for electrocardiogram signal analysis." Journal of Medical Imaging and Health Informatics 4.5 (2014): 737-742

[17] Lin, H-Y., et al. "Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals." Irbm 35.6 (2014): 351-361.

[18] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach. Neural Comput 2000;12(10):2385—404

[19] C. Cortes and V. Vapnik, “Support-vector networks,” Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[20] M. Ayinala and K. Parhi, “Low-energy architectures for support vector machine computation,” in 2013 Asilomar Conference on Signals, Systems and Computers, pp. 2167–2171, IEEE, 2013.

[21] A. Aizerman, E. M. Braverman, and L. Rozoner, “Theoretical founda tions of the potential function method in pattern recognition learning,” Automation and remote control, vol. 25, pp. 821–837, 1964