IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

IMPLEMENTATION OF FIVE CLASSES OF AUTOMATED ECG ARRHYTHMIA CLASSIFICATION USING KNN CLASSIFIER

[1]Dr. Yogesh G S, [2]Dr.Vedavathi G R, [3]Dr. S G Hiremath

[1]Professor, [2]Associate Professor, [3]Professor
[1]Electronics and Communication Engineering Department
[1]East Point College of Engineering and Technology, Bengaluru, India

Abstract: Electrocardiography (ECG) signals are classified into five different classes of signals namely, normal, PVC, LBBB, RBBB, and APC to establish the abnormality of the signal. Initially, to find the peaks in the ECG signal Pan-Tompkins algorithm is used. Then, some of the features such as HOS, min-max features, and temporal, are carried out on MIT/BIH arrhythmia database which consists of the huge number of signals including the label of normal and arrhythmia signal. The performance of the classifier is measured with other classifiers; the results showed that KNN provides higher classification accuracy than the other classifiers.

***Index Terms* - Arrhythmia, MIT/BIH database, KNN classifier, Higher order statistics, Temporal feature.**

I. INTRODUCTION

Electrocardiogram is one of the simple Non-Invasive Diagnostic (NID) scheme for detecting different heart diseases. ECG signal is estimated on the patient's body surface and the outputs from electrical changes are related with the activation of the two small heart chambers first, the atria, and after that 2-bigger heart chambers, the ventricles. The automated ECG acquisition framework plays a crucial part in recording and monitoring the heart signal in practical applications and it helps to identify the abnormality and normality of the heart conditions [1]. ECG recognizes physical cardiovascular events that are obtained by the depolarization and re-polarization of the ventricles and atria of the heart. The heart signals comprise of a few features like T waves, P waves, and QRS complex waves, which plays a crucial part in the detection of arrhythmia disease. The irregular heartbeat conditions are named as arrhythmia and also denoted as dysrhythmia. The normal heartbeat range is between sixty to hundred beats per minute. Arrhythmia and abnormal heartbeat rates are not occurred together. Arrhythmia arises only in normal heartbeat rate. In the present days, more than 8,50,000 people are hospitalized due to arrhythmia diseases each year. The patients who have primarily affected by heart attack that patients are have much chance of heart issue. The risk of cardiovascular disease is higher for the men as well as the women and it remains the world's leading cause of death. The expansive measure of ECG information is needed to categorize the ECG signal and it causes the high system complexity of the classification algorithm. The appropriate classification scheme is needed to identify heart range signal in the troublesome circumstance [2]. Machine learning is a type of artificial intelligence system that delivers automatic data learning with-out human intervention [3-5]. The machine learning methodologies automatically learns to discover the appropriate predictions with respect to the past observations. It learns automatically about how to attain accurate predictions based on past observations. One of the superior classifiers is K-Nearest Neighbour. KNN is applied to improve arrhythmia classification. If there is no earlier information concerned to the data distribution, then KNN classifier is optimal for cardiac arrhythmia classification [6]. To train the large data set KNN is the emergent classification technique with the benefit of minimal usage of resources. In order to find the abnormalities of the human heart, ECG is the widely used by the cardiologists. In this research, new techniques are proposed for obtaining five heart-beat conditions; Normal sinus rhythm, PVC, LBBB, PAC, and RBBB.

A detailed explanation on different types of ECG arrhythmias are represented as follows:

- NSR: Normal heartbeat consists of several peaks like T peak, QRS peak and P peak. Usually, the normal heart beats per minute is 60 to 100 and PR peak ranges between 100ms to 200ms, respectively.
- RBBB: In ECG signal, the QRS complex includes an extra diversion that denotes the quick and slower depolarization of left and right ventricles.
- LBBB: In such type of arrhythmias, activation of left ventricle is delayed, as a result of this left ventricle contracts after the compression of right ventricle. Usually, the QRS complex length exceeds more than 120 ms when such type arrhythmias preset in the EKG signal.
- APC beat: It is depicted at early heart-pulse in the beginning of atria. The pulse of ordinary sinus rhythm is directed by the sinoatrial node.
- PVC beat: It is considered as an extra heart-beat, which begins at two lower pumping ventricular chambers. Here, the QRS complexes are extended due to the issues exists outside the sinoatrial node, which is not associated to the P and T waves.

II. Literature Review

Automation in arrhythmia classification is a challenging task, as there are large number of variations in the features of different ECG signals under various conditions. Some of the related works are mentioned below:

G. Sannino, and G. De Pietro, [7] "developed a new deep learning system for ECG beat classification. In this literature, Deep Neural Network (DNN) was developed using deep learning library from Google and tensor flow framework. The developed deep learning system comprises of seven hidden layers with 5, 10, 30, 50, 30, 10 and 5 neurons. This experiment was conducted on a publicly available dataset; MIT–BIH arrhythmia and compared the experimental outcomes with the recent scientific literature. The final outcome shows that the developed system was more effective by means of accuracy, sensitivity and specificity. The gradient in DNN was a major issue, which was unstable and tends to either vanish or explode in earlier layers".

M. Mohanty, S.Sahoo, P. Biswal, and S. Sabut, [8] "developed a new sy stem for classifying the ventricular fibrillation arrhythmias and ventricular tachycardia using statistical, spectral, and temporal features. In this literature, MIT-BIH malignant ventricular ectopy and CU ventricular tachyarrhythmia datasets were used for evaluating the performance of the developed system. The extracted features were given as the input for C4.5 classification methodology for classifying the two heartbeat conditions. The experimental section confirmed that the developed system was more effective than the existing methodologies by means of accuracy, sensitivity and specificity. Here, the computational complexity of the developed system was high, while performing with large number of features".

N.K. Dewangan, and S.P. Shukla [9] have made use of discrete wavelet transform to preprocessing and to extracting various features. The different features extracted are, Left Bundle Branch Block, Right Bundle Branch Block, Paced Beat, Atrial Premature Beat and First degree AV Block. The artificial Neural Network is used to classify the signal based on the feature. This method provides the sensitivity of 65% and positive prediction value of 92%. The negative predictive value is also high in this method and provide false results.

F. Akdeniz et al. [10] perform the arrhythmia detection in the ECG signal based on the Wigner-Ville distribution. The database is obtained from the Physio Net database and used it features to identify the arrhythmia. The accuracy, sensitivity and specificity achieved higher than the existing method and also have the less computational time. This method only classifies the signal into arrhythmia or sinus rhythm and not obtain the specific disease.

III. Methodology

To build an automatic arrhythmia classification system, all the three axes should operate efficiently i.e., high accuracy, high scalability and ease of implementation with its utility. The main objective of the work is to build and implement supervised arrhythmia classification system using ECG. An automated system for arrhythmia classifier is majorly divided into five phases; data acquisition, pre-processing, peak detection, feature extraction and classification. In this work, initially, the input EKG signals are collected from MIT-BIH arrhythmia dataset. Then, normalization is carried-out for minimizing the system complexity.

- Pan-Tompkins algorithm is utilized to find the peaks.
- Different features such as HOS, min-max, and temporal features are applied to extract the feature values on the data obtained from PT algorithm.
- Five types of arrhythmia Classification is carried-out by using KNN

The proposed systems work-flow is denoted in the figure1

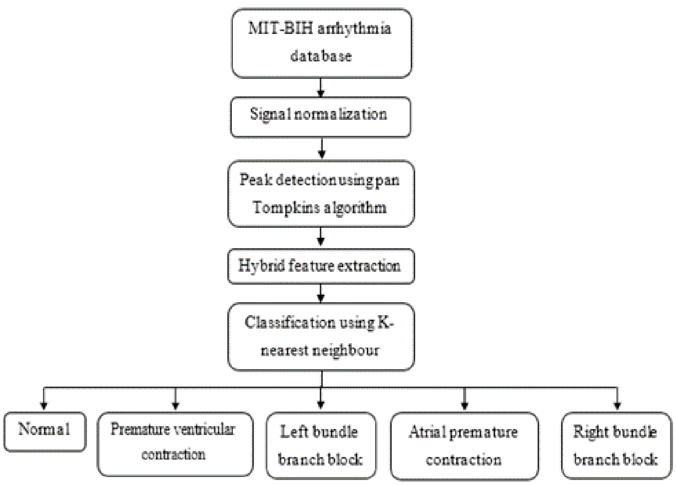

Figure 1 Proposed system work - flow

3.1 Data Collection

The first step of the proposed work is the collection of data from the well-known standard database: MIT-BIH arrhythmia dataset. "It contains forty-eight half-hour excerpts of two-channel ambulatory ECG recordings. In this dataset, the ECG signals are recorded from forty-eight subjects from BIH Arrhythmia Laboratory between 1975 and 1979. Randomly, twenty-three recordings are selected from a set of 4000 records in that 40% are outpatients and 60% are inpatients at Boston's Beth Israel Hospital" [11,12]. Apart from this, the rest of twenty-five recordings are taken from an equivalent dataset during a small random sample. These recordings are digitized at 11-bit resolution over a 10 mV range. Subsequent step after the gathering ECG data, is pre-processing which incorporates signal normalization.

3.2 Signal Normalization

After the collection of ECG signals, normalization is applied for normalizing the signals that ranges between zeros to one. Normalization is the procedure of scaling the ECG signals in identical level. The baseline potential and amplitude of the ECG signal is degraded by several factors that importantly degrades the precision of signal detection. Normalization should be essential in ECG signal to enhance the accuracy of detection. Here normalization work is used to fix the amplitude to one and baseline to zero. The sample normalized ECG signals are graphically denoted in the figures 2 and 3.

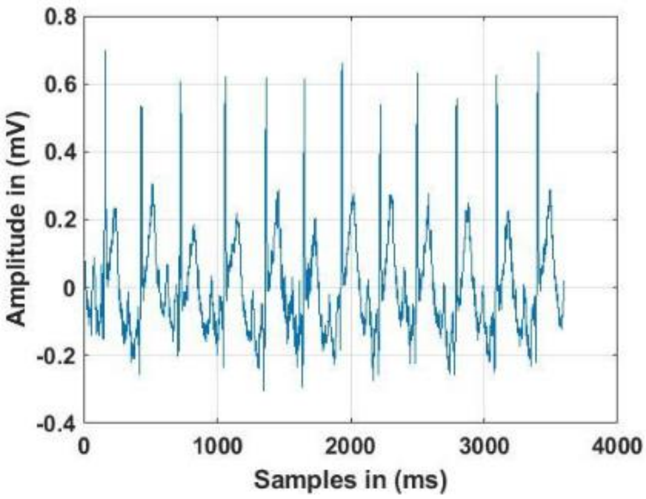

Figure 2 Sample normalized normal ECG signal

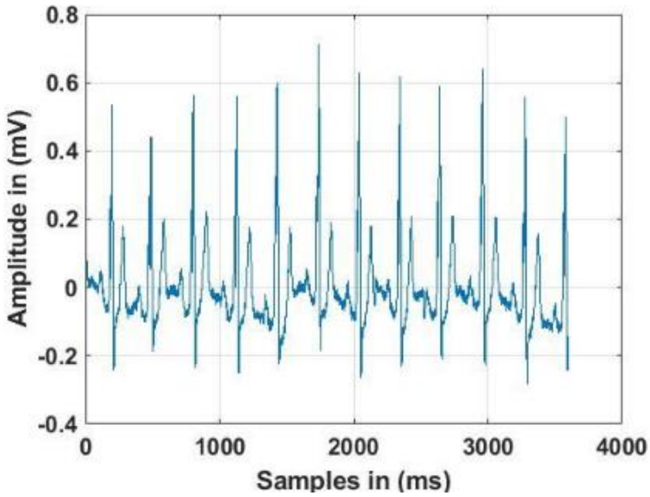

Figure 3 Sample normalized abnormal ECG signal

3.3 Pan-Tompkins Algorithm

The normalized ECG signals are utilized for peak detection by applying PTA. In MIT-BIH, the raw ECG signals includes more artifacts which are needed to be eliminated in the raw signals. The raw ECG signals are contaminated by a few noises like power-line interference, T waves with higher frequency properties similar to QRS complex, baseline wandering, artifacts due to lead movement, and muscle noise. In this work, PT algorithm is adopted to determine the peaks by applying filtering and pre-processing approaches [13-14]. The process of PTA is segregated into four phases such as, R-peak, squaring, derivative, and filtering, which are detailed below.

3.3.1 Filtering

The first phase in PTA is pre-processing the normalized signal by applying band pass filter to reduce T-wave interference, baseline wander influence, 60Hz interference, and noise from the muscles. The desirable range of pass band filter is fixed as 4-15 Hz that effectively reduces the QRS wave. The high and low pass filters are combined to develop a band-pass filter that helps in eliminating the higher frequency artifacts and lower frequency interferences. Hence, the band pass filter permits only a particular frequency in order to investigate the nature of QRS wave, where both low and high frequency signals are getting attenuated.

3.3.2 Derivative

The QRS complex slope (i.e., noise free, ECG signal) can be determined by applying the smoothing technique in a derivative block. The peak width with specific interval of time, wave position appearance of wave position, and peak amplitude is also obtained by smoothing technique.

3.3.3 Squaring Function

The next step is to square the output signal sequentially, so that the derivative output of the ECG signal improved non-linearly, and the QRS complex are highlighted by using the equation (1).

$$y\ (nT) = [x\ (nT)\ ]^2 \qquad (1)$$

Where

x (nT) = Input signal
y (nT) = Input square signal
The higher frequencies obtained, which are mainly due to QRS complex are shown by the squaring function.

3.3.4 Peak Detection

Procedure to be followed in peak detection by applying Pan and Tompkins algorithm are listed below:
1: First, select the value and save it in temp local max.
2: suppose, the new point value is higher than the temp local max, save the new point value in temp local max.
3: A peak is detected and identified, if the new point value is less than the half of temp local max. or else, go to step 2.

Then, the detected peaks are subdivided as noise or QRS complexes. Eliminate all the peaks that are less than 200ms in order to perform multiple detections in QRS complex waves and also to ignore the T waves. The peaks are identified as QRS complexes instead of noise, if the peaks are higher than the adaptive threshold that is mathematically denoted in the equation (2).

$$DT=NPL+TC*(QRSPL-NPL) \qquad (2)$$

Where, DT is signified as threshold value, TC is denoted as coefficient of thresholding, NPL is stated as noise level and QRSPL is represented as QRS wave peak level.

3.4 Feature Extraction

The main objective of this stage is to extract and strengthen the characteristic features of the signal. Therefore, the efficiency of recognition is improved for the arrhythmia classification. According to the estimation of the Higher- Order Statistics (HOS), Temporal and Min-Max, the features are extracted due to the periodic nature of the ECG signals. These feature extractions are briefly described in the below sections.

3.4.1 Higher Order Statistics

In order to estimate the shape parameters such as skewness and kurtosis for measuring the deviation of distribution from the normal value, higher order statistics are used. 3rd order and 4th order statistics are of great importance for representing the nonlinear signals.

3.4.2 Temporal Features

Feature extraction includes an important role in any classification. During this study, the sequence of each statistical and temporal features has been used. The relative unfold of the EKG characteristics is reduced by the cumulant characterization of complexes in QRS features that belongs to identical variety of regular recurrence and this methodology makes the classification comparatively easier.

3.4.3 Min Max Features

In Feature extraction phase to extract the features from the peak detected ECG signals are initially decomposed into eight levels. After the decomposition, the features such as standard deviation, mean, median, energy, and entropy of the signal are determined. In this examination, maximum and minimum voltage features are achieved from the EKG signal. The peak voltages are dissimilar in many abnormal and normal signals. So, the min-max features recognize the abnormal and normal signals.

3.5 Classification

The next step after the feature extraction stage is classification of the signal into a sinus and arrhythmia. The research study used KNN classification for classifying the signal into a normal or abnormal signal, because it minimizes the noise between predicted and actual results. Commonly, KNN classifier involves neural layers learning on huge datasets like MIT-BIH arrhythmia dataset [15].

3.5.1 K-Nearest Neighbor Classifier

After extracting the features from normalized signals, the ECG data are utilized for the classification of normal and abnormal ECG signals. In this research study, KNN classifier is used for arrhythmia classification, because it minimizes the noise between predicted and actual results. Commonly, KNN classifier involves neural layers learning on huge datasets like MIT-BIH arrhythmia dataset [16]. Generally, KNN is an automated supervised classifier, which is developed on the basis of mathematics and simple theory [17].

For an arbitrary feature vector x_i, calculation of a defined distance between this feature and the vector y_j is as follows:

$$d(i,j) = f(x_i, y_j) \quad (3)$$

$$\text{Distance function } f(x_i, y_j) = (x_i, y_j)^T \Sigma, (x_i, y_j) \quad (4)$$

$$f(x_i, y_j) = \left(\sum_{k=1}^{p} \left(x_i(k) - y_j(k)\right)^r\right)^{1/r} \quad (5)$$

$$(x_i, y_j) = \frac{1}{p}\sum_{k=1}^{p} abso\left(x_i(k) - y_j(k)\right) \quad (6)$$

f The following equation describes the distance vector D(i),

$$D(i) = \{d(i,j) \lor i = 1,2, \ldots . N_{test}, j = 1,2, \ldots . , N_{train}\} \quad (7)$$

The *D(i)* vector is sorted in an ascending order, and choose the first K elements (which is called K nearest neighbors) as follows

$$D_N(i) = \underset{Ascending}{sort}(D(i)) \quad (8)$$

$$V = \{\delta(D_N(i)(1)), \ldots . , (\delta D_N(i)(K))\} \quad (9)$$

The value of K is suitably chosen for obtaining good accuracy.

IV. Results and Discussion

In the research study, the suggested work was coded by MATLAB software. For assessing the usefulness of suggested work, the working of suggested work was associated with some present methods already exist on MIT-BIH arrhythmia database. The suggested work performance was evaluated by means of sensitivity, specificity, accuracy, and f-measure. Relationship between the output and input variables of the suggested work is can be understand by utilizing the performance metrics like accuracy, error rate, sensitivity, and specificity. The formula to evaluate sensitivity, specificity, accuracy, and f-measure are given in the equations (10), (11), (12), (13).

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (10)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (11)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (12)$$

$$F-measure = \frac{2TP}{(2TP+FP+FN)} \times 100 \quad (13)$$

Where, TN is denoted as true negative, FP is exemplified as false positive, TP is specified as true positive, and FN is stated as false negative.

Table I shows classification of ECG signal using three classifiers, namely NAVY, Radom Forest, and KNN along with the measuring parameters.

Table I Performance of the Proposed Classifier with Existing Classifier

Classifiers	Features	Specificity (%)	Sensitivity (%)	F-Measure (%)	Accuracy (%)
Radom Forest	MIN	72.00	93.00	75.20	75.20
	HOS	52.00	80.00	53.60	53.60
	T Peak	100.00	100.00	65.60	65.60
	QRS Interval	64.00	77.00	60.00	60.00
	Comb	**100.00**	**98.00**	**92.80**	**92.80**
Navy	MIN	8.00	100.00	50.67	53.6
	HOS	100.00	61.00	34.00	53.6
	T Peak	100.00	85.00	100.00	47.2
	QRS Interval	16.00	90.00	42.00	54.4
	Comb	**100.00**	**100.00**	**97.33**	**77.6**
Suggested KNN	MIN	56.00	88.00	45.33	74.40
	HOS	28.00	75.00	20.67	47.20
	T Peak	100.00	100.00	100.00	56.80
	QRS Interval	48.00	94.00	44.00	72.00
	Comb	**100.00**	**100.00**	**100.00**	**98.40**

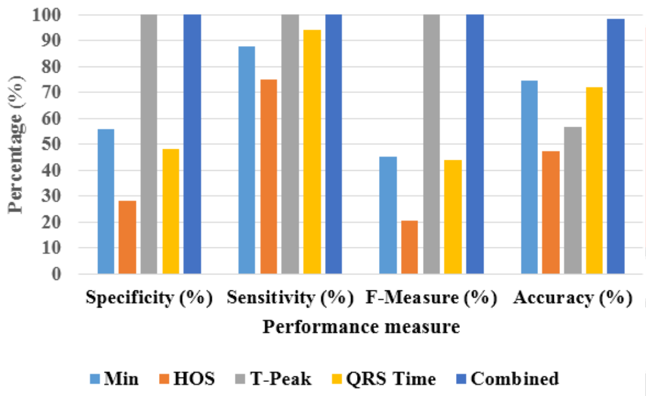

Figure 4 Graphical comparison of KNN classifier for different performance measures.

V. Comparative Study

Table II represents the comparison study of present and suggested work performance. R.J. Martis, *et al*, [18] developed a new system that comprises of high order statistics as a feature extraction approach with other interval features. The classification procedure classifies the extracted features using neural network classifier. This experiment was done on an online available dataset: MIT-BIH arrhythmia. Here, the developed system almost achieved 94.52% of accuracy. Additionally, T. Ince, et al, [19] developed a new arrhythmia prediction algorithm by integrating discrete wavelet transform, PCA, and multi-dimensional particle swarm optimizer. Here, multilayer perceptron's classifier was utilized to classify the ECG signals. The present work was tested on MIT-BIH arrhythmia dataset for validating the developed methodology outcome in light of accuracy. The developed method almost achieved 95.58% of accuracy. In contrast, the proposed work attained 98.40% of classification accuracy, which was superior to the present work already existed.

Table II. comparison study of present and suggested methodology

References	Dataset	Classifiers	Accuracy (%)
R.J. Martis,*et al*, [18]	MIT-BIH database	Neural network	94.52
T. Ince, *et al*, [19]		Multilayer perceptron's	95.58
Suggested work		**KNN**	**98.40**

VI. CONCLUSION

ECG signal is the most used parameter in the medical field for analysis of health of the patient. To classify the signal, the suggested work employs the extraction of four features. The ECG signal is obtained from the MIT-BIH arrhythmia database and it consists of huge number of the ECG signal of both arrhythmia and sinus rhythm. These signals are taken from the 47 subjects of the continuous 48 hours of the data. It also contains the labeled data and it labeled as the abnormal and sinus rhythm, which is utilized in the testing and training process. The features extracted from the signal are HOS, Min/Max and temporal features. KNN used these features to classify the signal into five types, namely PVC, PAC, RBBB, LBBB and sinus rhythm. The other classifiers are also used to classify the signal for analyzing the performance of the classifier and comparison is done with the other classifiers. The classifier accuracy, specificity and sensitivity of the classifiers are calculated by using the proper performance measurement. The experimental result shows that the KNN has the higher classifier accuracy while compared to the other classifier with the hybrid features and accuracy is achieved up to 98.4%. Compared to other present methods in arrhythmia classification, the suggested methodology provided an efficient performance with respect to accuracy and showed 2.8 - 4% of greater classification accuracy.

REFERENCES

[1] M. Hendel, et al, “Automatic heartbeats classification based on discrete wavelet transform and on a fusion of probabilistic neural networks”, Journal of Applied Sciences (Faisalabad), Vol.10, No.15, pp.1554-1562, 2010.

[2] R. Kumar, et al, “Electrocardiogram signal compression based on singular value decomposition (SVD) and adaptive scanning wavelet difference reduction (ASWDR) technique”, AEU-International Journal of Electronics and Communications, Vol.69, No.12, pp.1810-1822, 2015.

[3] Acharya, U.R., Fujita, H., Lih, O.S., Hagiwara, Y., Tan, J.H. and Adam, M., 2017. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Information sciences, 405, pp.81-90.

[4] Yıldırım, Ö., Pławiak, P., Tan, R.S. and Acharya, U.R., 2018. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in biology and medicine, 102, pp.411-420.

[5] Sannino, G. and De Pietro, G., 2018. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems, 86, pp.446-455.

[6] Bhagyalakshmi, V., Pujeri, R.V. and Devanagavi, G.D., 2018. GB-SVNN: Genetic BAT assisted support vector neural network for arrhythmia classification using ECG signals. Journal of King Saud University-Computer and Information Sciences.

[7] G. Sannino, and G. De Pietro, “A deep learning approach for ECG-based heartbeat classification for arrhythmia detection”, Future Generation Computer Systems, 86, 446-455 2018.

[8] M. Mohanty, S. Sahoo, P. Biswal, and S. Sabut, “Efficient classification of ventricular arrhythmias using feature selection and C4. 5 classifier”, Biomedical Signal Processing and Control, vol.44, pp.200-208, 2018.

[9] N.K. Dewangan, and S.P. Shukla, “ECG arrhythmia classification using discrete wavelet transform and artificial neural network”, In Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE International Conference on pp. 1892-1896, IEEE, 2016.

[10] F. Akdeniz, İ. Kayikçioğlu, İ. Kaya, and T. Kayikçioğlu, “Using Wigner-Ville distribution in ECG arrhythmia detection for telemedicine applications”, In Telecommunications and Signal Processing (TSP), 2016 39th International Conference on pp. 409-412, IEEE, 2016.

[11] U.R. Acharya, et al, “Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network”, Information sciences, Vol.405, pp.81-90, 2017.

[12] M. Carrara, et al, “Classification of cardiac rhythm using heart rate dynamical measures: validation in MIT–BIH databases”, Journal of electrocardiology, Vol.48, No.6, pp.943-946, 2015.

[13] M.A. Hashim, et al, “Efficient QRS complex detection algorithm implementation on soc-based embedded system”, Jurnal Teknologi, Vol.78, No.7-5, 2016. J. Park, et al, “Cascade classification with adaptive feature extraction for arrhythmia detection”, Journal of medical systems, Vol.41, No.1, pp.11, 2017.

[14] J. Park, et al, “Arrhythmia detection from heartbeat using k-nearest neighbor classifier”, IEEE International Conference on Bioinformatics and Biomedicine, pp.15-22, 2013.

[15] J. Park, et al, “Arrhythmia detection from heartbeat using k-nearest neighbor classifier”, IEEE International Conference on Bioinformatics and Biomedicine, pp.15-22, 2013.

[16] R. Saini, et al, “Classification of heart diseases from ECG signals using wavelet transform and kNN classifier”, In International Conference on Computing, Communication & Automation, IEEE, pp.1208-1215, 2015.

[17] R.J. Martis, et al, “Application of higher order cumulant features for cardiac health diagnosis using ECG signals”, International journal of neural systems, Vol.23, No.04, pp.1350014, 2013.

[18] T. Ince, et al, “A generic and robust system for automated patient-specific classification of ECG signals”, IEEE Transactions on Biomedical Engineering, Vol.56, No.5, pp.1415-1426, 2009.