International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

ECG Signals Classification using Statistical and Wavelet Features

Anuja S B, Usha Nandini K , Sukanya S.T

Abstract—The technique called signal processing is vitally used as a choice in the case of real-time examination of ECG (electrocardiography) signals. Classification of arrhythmic beat is primarily utilized in the detection of abnormalities normally found in electrocardiogram (ECG) and thus help in identifying problem occurred in heart. This work intentionally carryout, signal preprocessing of electroencephalography, feature extraction using statistical and wavelet and SVM-RFE established classification for arrhythmic is achieved to differentiate regular and irregular constituents of ECG. FFT technique, is used initially in order to remove the noise, to detect R-peaks later then the thresholding technique and windowing technique are consumed. The features of wavelet (such as the information of RR-interval) are calculated and sequenced to features of statistical to organize the final set of feature, which is then consumed to characterize and classify with SVM-RFE for ECG signals. The anticipated classification method arrhythmia is pragmatic to inputting electrocardiography signals gained from the database of MIT-BIH Arrhythmia, along with several international databases of ECG signal. There is no cross validation is required since SVM-RFE works healthy and thus it gives an outstanding performance, it will be beneficial for long-term electrocardiography beat analysis and classification. Results recommend that though the duration of the findings of the recording are of short, with the anticipated model which is able to analyze and classify the idiosyncrasies of heart from an ECG measuring unit with a single lead. The results obtained from the experiment shows an overall accuracy of high performance, sensitivity, and98.97% specificity in contrast with the current approaches mentioned in the literature.

***Index Terms*—: Heart rate variability, Arrhythmia classification, SVM-RFE, ECG, Wavelet**

I. INTRODUCTION

Analysis of computerized electrocardiography says ECG, is a renowned run-through for classifying and analyzing the information obtained from ECG and has been placed substantial research focus by investigators and researcher for years. Enhancement in observing of ECG and periodical analysis of heart irregularities (i.e., arrhythmias) are being noted on the automated corroboration and examination of ECG. These variation in the obtained information are not made-up to be treacherous, but it could exemplify a long period risk, if not treated, it may cause stroke or other heart related issues or sudden myocardial infarction [1], [2]. The analysis of ECG by cardiologist on long-term would be hectic and employs more time. The non-automated classifying and monitoring by human involvement is not sufficient to satisfy the need for the diagnosis of real-time arrhythmias. Relatively, computer-based automated diagnosis and monitoring information obtained from ECG signals is evident, which then assist doctors with enough information to deliver the medical support essential for patients**.** In order to lead a healthy lifestyle those kind of automated monitoring and diagnostic tools is employed by patients as a system of proficient decision support. Hence, computer based automatic classification a monitoring of ECG signals is a crucial thing for proper analysis of cardiovascular related diseases.

In the area of research various methods accompanied with signal processing technique have employed for automated detection and analysis of ECG signals monitoring and classification, although the classification and detection of diseases still remnants a challenge because of the substantial discrepancy in characteristics of ECG signal. On investigation with different techniques in the literature which are having included time-frequency analysis [10]-[11], time domain analysis [3]-[4],frequency-based analysis [9], and statistical approach [5]-[6] intentionally for the abstraction of feature from the ECG signals. The above mentioned extraction method for the detection of the features are associated with the linear discriminants classification method,[3][12]NN [6][11] neuro-fuzzy [17], and SVM [5] [8][15] to place proficient monitoring and analysis of abnormalities related to cardiovascular diseases.

QRS, T waves and Onset and offset of P are various trustful features in the analysis of time-domain, which are employed for the monitoring and analysis of electrocardiography signals. Though, it agonizes from several uncertainties that consists of: 1) Occurrence of inconsistency among the ECG waves on the detection of offset and onset of fiducial points;2) Even if using a permanent fiducial detector, the elimination of noise is not efficient and thus affect the location of fiducial points, and(c) lead to failure. If improbability in defining ECG signal speaks.

To overcome the limitations of the widely used aforementioned a statistical and wavelet features are the predominant things which is then united for the feature set signifying the electrocardiograph signals which are employed to foresee their discrete functionality in classes using RF. MIT-BIH arrhythmia database is the data source and fro where the data of arrhythmias and normal signals are employed for experiments to corroborate the proposed approach mentioned in this work.

Revised Manuscript Received on January 15, 2020
Anuja S B*, Department of Master of Computer Applications, Narayanaguru college of Engineering, Manjalumoodu, India. Email: 111anuja@gmail.com

Usha Nandini K, Department of Master of Computer Applications, Narayanaguru college of Engineering, Manjalumoodu, India. Email: ushanandinik2002@gmail.com

Sukanya S T, Department of Master of Computer Applications, Narayanaguru college of Engineering, Manjalumoodu, India. Email: suganyast93@gmail.com

The benefactions of this work are as follows:

1) R-Peak detection is proposed based on FFT technique, windowing technique and thresholding. Various cardiovascular disorder and the detection of RR interval can be perceived with the practice of the R peaks location.

2) The statistical and wavelet feature have been used as features to design a classifier based on random forest that classifies EEG signal into normal and abnormal signals.

3) Classification accuracy has been verified with the Arrhythmia database from MIT-BIH.

To the preeminent of the knowledge, a combined statistical and wavelet features for signal modelling approach for ECG analysis and classification would be the first work in the segment. The structure of this papers work is mentioned below. Section II describes the feature extraction methods and learning algorithm. Section III delineate the proposed work and how it analysis electrocardiograph signals, whereas Section IV delineate the assessment and analysis of performance.

II. RELATED WORKS

Computer based automated processing of electrocardiograph signal processing technique is very popular thing nowadays. Using several techniques and employing those sophisticated techniques many complex algorithms have been formulated intentionally for the pre-processing of the signal, extraction of constituents ECG signals and classification of ECG signal set. However, feature extraction methodology utilized is accurate with high efficiency, heavy computational resources and well established architecture such as time-frequency domain conversion, cross-correlation and so on are needed for this, which indeed makes the whole system not usable for arrhythmia detection, real-time monitoring and other diagnosis purposes.

Recording of electrical potential of the heart is generally known as electrocardiogram (ECG).The discrete characteristic variation occurred in the signal of ECG is very worthwhile in the classification of cardiac disorders. There have been many researches involved in the detection of heart disorders in past decades. New steps have been adapted to maximize the possibilities of ECG signal in the specific area of application such as in disease diagnosis and biometric. The no fiducial methods are not supposed to employ to generate the ECG feature set, hence it is essential to have a techniques to overcome the problems in finding of fiducial points in the areas of cardiac irregularities. The classification technique DCT-Discrete Cosine Transform is employed by Platonist is for the auto correlating ECG signal data. 100% recognition rate was achieved on PTB database with the database of 14 relevant subjects. Several methods have been proposed by Agrafioti and Hatzinakos on non-fiducial features. They also found the option of differences occurred in ECG signal data [16].

According to Morteza Zabihi et al. a systematic approach has proposed intentionally to detect and classify AF rhythms occurred in portable handy ECG devices**.** 491 hand crafted features are investigated, and enlisted them on the basis of their prominence. Highest-ranked 150 features are being gotten and feed into this to a classifier called random forest intentionally to detect and analysis AF rhythms with other three ECG types. The technique in this work achieved with an average score of 82.6%.According to this average performance, the anticipated method has the ability for the enhancement, which is what expected in this future work [17].

In order to specifically detect HCM patients a set of amplitude-thresholds for have been proposed by McKenna and Corrado. A small group having the count of 56 HCM patients with these thresholds and healthy control people of 56are been tested under the protocol by Potter et al. It was mentioned that about 90% of the specificity and sensitivity from this study approximately**.** Literally, there were quiter are works have been used in associated with the machine learning algorithm for intentionally determining Hypertrophic Cardiomyopathy patients using ECG signals database**.** The amount of Hypertrophic Cardiomyopathy subjects used in this kind classification examination was about 221, which is likely quite more than other early carried out work held on Hypertrophic Cardiomyopathy detection. In the case of SVM similar performance measures were obtained**.** For valuation, a moderated level of performance related to both SVM and random forests on the application of logistic regression method intentionally for classifying heartbeats. In order to validate system information gain criterion is consumed which is in terms highly explanatory discrete features to distinguish and characterize the rhythm throughout the data set placed as part of training and testing [18].

Difference in features of amplitude such as PQ, QR, and RS, were associated with each rhythm and classified into AAMI classes in the study of Juyoung Park et al. In order to detect and classify heartbeats amplitude difference features relative to random forest classifier is used. To evaluate the performance and throughput of classification, MIT-BIH ECG database were employed with the difference in amplitude. The results of performance showed the detection of arrhythmia. In the forthcoming work, we indeed adapt technique to increase the classification accuracy using an active learning method with the support of the variation in heart beat signal [19].

Diptangshu Pandit, Li Zhang et al. put forward a kind of feature enhancement over our earlier research laterally with performance assessment by means of different classification and assessment methods. The ECG data regarding time period and shape waves were extracted using a different strategy deprived using the findings of definite margin of the ECG waves. The anticipated adapted feature set with extra generated features gives exactly higher accuracy of irregularity detection which is literally reflected in the result of experiment. These mined features would be further tested in forthcoming work in order to prove their efficacy for heart disease recognition from Electro Cardio Graph signals. Novel and Outlier class detection would also be utilized to confirm the effectiveness in dealing with highly noisy and challenging disease diagnosis [20]. To describe the changes in the recordings of the electrocardiograph Ruhi Mahajan et al. filtered and analyzed different domains such as frequency, morphological and time. The PhysioNet 2017 challenge dataset were employed for this, which had 8,524 single lead data of ECG durable over 9 to 60 seconds. Then executed a feature selection method based on GA which waned 37 characteristic features space in order to classify and discriminate ECG rhythms.

We gained AF, total F1-score for NSR and Other rhythms of 0.80withrandom forest on the training and consequently 0.79 on the hidden dataset. These results establish that the relative potential of efficiency of AF detection and classification can be boosted by adaptive learning approach [21].

An unceasing depiction of human kinematics and primarily focus on the automatic controlled devices which have been interpreted by researchers during the last decades. A multi fingered prosthetic hand is said some of the examples of these kinds. Classification methods in the machine learning process were employed in to categorize myoelectric activity among movements of human hand digit then diverse hand postures. EMGs from the forearm were used prosthesis for all control, which classifies three types of grip in real-time Materials [22].

III. PROPOSED METHODOLOGY

First, ECG signals and corresponding HRV signals are modeled. FFT, Window Filtering and Thresholding are methods are utilized to architecture R-Peak detection to rate variability signal. Next, statistical significant features and wavelet features are identified and collected based on the above modeling. Finally design random forest classifier to classify abnormality of the signal.

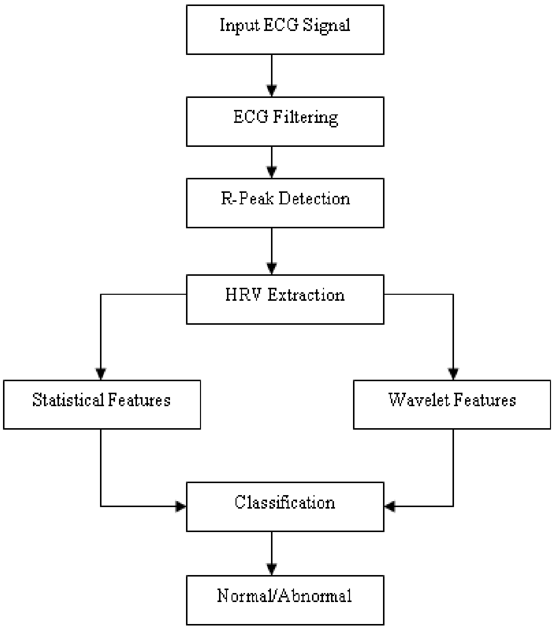

Fig1. Classification algorithm for arrhythmia in proposed work

A) Arrhythmia Database -MIT-BIH

The data set of MIT-BIH is employed in the deep ML process to find the abnormalities in the ECG signals [15].This db contains dual channel electrocardiograph signal recordings of 48 half-hour taken from patients of 47 with heart disorder. The database contains 26 records of arrhythmia that were arbitrarily selected from 4,100 records of outpatients (40%) and inpatients (60%). The rest samples are complex types: ventricular, junctional and supraventricular in 26 records. Among the 48 ECG recordings, 25 belongs to male of 35-87 years old, 23belongs to female subjects of25-85 years old, and 5 recordings with stridden beats. In utmost cases, there is significant improvement in the upper signal limb obtained and thus placing the functional electrodes on places of patient's chest and the minor signal is usually lead V!; though, there are numerous occurrences of V2, V4 and V5. Records for the digitization was at 365 Hz with 11 bits resolution of over 10 mV range. MIT-BIH data set which contains the depiction and classification of rhythm of the measured signal quality and are recognized by expert cardiologists working separately. MIT-BIH dataset is available on Physiobank [12].

B) Detection of R-Peak

The technique concluded in the this work is depicted in the detection structure,Figure-2R-peak

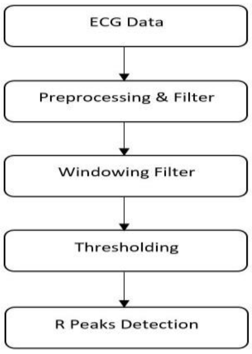

Fig 2. Complete process involves in R -peak detection

The sampled data of electrocardiograph is occupied and the technique called fast Fourier transform (FFT) is applied: The graph represents in Figure-3 is not even hence the first step is the taken to straighten the signal. The low frequency components should be eliminated in order to make straight it all. The values of samplesare in X-axis denotes and the Y-axis which represent voltage. Apply the FFT to the electrocardiograph signal Figure-3 using

$$X(K) = \sum_{n=0}^{N-1} x(n) e^{\frac{-j2\pi nk}{N}}$$

Where K= 0, 1, 2………N-1;

Once more IFFT is applied to the signal to get back the original time domain signal.

$$X(n) = \frac{1}{N} \sum_{K=0}^{N-1} X(K) e^{\frac{-j2\pi nk}{N}}$$

Where n= 0, 1, 2………K-1;

Equation (1) eliminate components of low frequencies by adapting finite impulse response filter and reestablish ECG with the help of inverse FFT

Windowing filter: This is the second step which is intentionally used to calculate local maxima that focus only maximum values in the respective window and reject the others. This technique has its related in the processing of signals in which spectral analysis of non-periodic signals are performed through the windowing operation

Thresholding: Thresholding is the next process which is applicable after windowing filter to eliminate the minor peaks in the signals and reserve the important ones.

Adjusted filter: It is detected that the result is sufficient but it doesn't ensure that it would have all the peaks in overall case. In order to overcome this case, there must be modified in the size of window and methods of filtering are iterated. The final peaks illustrate the overlaid R peaks which is employed for the heart rate calculation

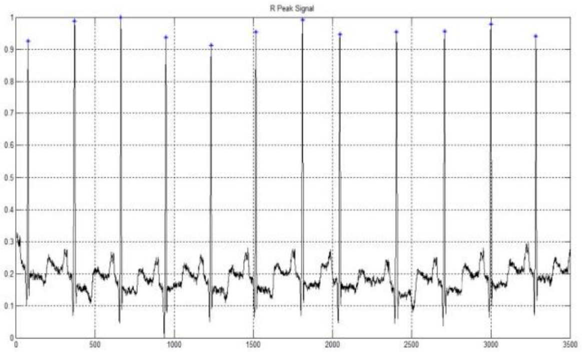

Fig 3. R -peak Signal

C. Heart Rate Variability

Initially, it is quite essential to extract the characterized Heart Rate Variability signals from the ECG signals. The maximum absolute value of signal is represented by R wave in the window of time. In order to quantify the time intervals among the consecutive R waves in signal obtained from the ECG dataset and finally plot these consecutive R wave intervals to time indices. Reduced rate of heart is generally considered as the variability and as an indication of mortality after MCI. Though other evidence have given and idea that the information in heart rate variability is related to acute MCI survival is fully kept in the mean heart rate

D. Feature Extraction

The extraction of feature is said to be the next process which involves in the block diagram. In order to label the ECG literally huge number of possibilities are been proposed. In this paper, proposes a new combination of statistical and wavelet features to represent ECG data to be processed.

Statistical features

Generally, the cardiovascular system indicates both nonlinear behavior and linear. Consequently, linear and nonlinear features are utilized together in this work. The features literally the linear evidences which are attained from frequency and time domains have been used in this method [15].

Linear analysis

Features of time domain: The below is the frequently adapted four types of time domain parameters which are intentionally used for the heart rate variability signal which are being obtained from the intervals of RR are:

Mean HR: The mean HR is an average heart rate value in one minute of time in each segment in each R-R interval (second).

TD HR: TD HR is defined as the SD of immediate heart rate in each segment.

pNN50: The number of consecutive difference of 65 R-R periods which varies more than 52ms, respectively, divided by 65 is said to be pNN50.

HRV triangular index: HRV triangular index is literally defined as the integral of histogram that is total number of RR intervals which is divided by the histogram height.1/128 bin width is designated according to [23].

Frequency domain features: Frequency domain structures: even though the parameters of time domain are literally effective, nevertheless they don't differentiate sympathetic and parasympathetic constituents of heart rate signal [20]. High frequency (HF) heart rate variability signals (0.15-0.4 Hz) make believe the cardiac vagal responses as Respiratory Sinus Arrhythmia (RSA). The early stage of parasympathetic events of the cardiovascular system is considered from HF components. In this proposed work, the PSD and the ratio HF and LF bands are made to be calculated and considered as the feature of frequency domain of signal.

Nonlinear analysis: By using nonlinear approach on the HRV signal monitoring and dynamics leads to evident data for the confirmation of the issues in heart related problems. Hence, in this paper nonlinear components of the heart rate variability signal are employed.

SD1/SD2: The point care plot shows the relation between successive RR intervals of the ECG signals as shown. This is done by marking each RR interval (*RR (n+1)*) on the point care plot as the previous interval (RR (n)) function in time series interval period of RR [23]. The SD time series interval points on lines $y = x$ and $y = x + 2RRm$is considered by quantitatively examining the mark in which *RRm* is the mean values of RR period. These intervals are named as SD1 which means the variability of fast beat to-beat and SD2 shows the variability relatively long-term in the variability signal in ECG [24].

LLE: The advantages about the dependence of method on quite early conditions is defined by the main Lyapunov Exponent and the difficulties in the system model [24] is considered by a positive lyapunov exponent confirms. To calculate LLE a point is selected, in the congregated phase and all the corresponding neighbor points exist in a pre-structured radius ε are been determined. As the improvement of the system, the trajectory of the initial point and the neighbor point's trajectories are taken to compute the distances of mean between them. Mean values corresponding to the slope and the time of the resultant line are being taken as LLE. Lagand implanting dimension are be m = 10 and τ = 1, correlatively. The distance of threshold ε is designated to be *mSD*, in which SD is the standard deviation of RR periods, [25].

SpEn: The complication of the HRV segment is shown by the Spectral Entropy [25]. Huge values of SpEn apparently show the high imprudence and reduced values of it demonstrate time series quite ordinary phase. The spectral entropy of the process is evaluated by Shannon's channel entropy as

$$SpEn = -\sum_{f} p_f \log(p_f)$$

*Pf*at frequency *f*is generally meant as the value PDF of the process. Henceforth, the measurement of entropy is uncertain about the frequency f.

D2: The measure of the complication of the time series is defined as dimension of correlation and also it governs the dynamic variables in very less number which can have the system architecture model.

The algorithm stated in[26] is used to evaluate this feature as for the embedding dimension, value of m = 10 is selected.

Wavelet Features

f(x) is the wavelet transform of a signal

$$W_s f(x) = f(x) * \Psi_s(x) = \frac{1}{s}\int_{-\infty}^{+\infty} f(t)\Psi(\frac{x-t}{s})dt$$

In the above s is defined as the scale factor. Ψs(x) = 1 sΨ(xs) is defined as the dilation of a basic wavelet Ψ(x) by the scale factor s. then the wavelet transform is known as dyadic *WT*[7]. WT of signal f (n) is measured with Mallet as follows:

Let s = 2j

$S_{2j}f\ (n) = \sum_{k\in z} h_k S_{2j-1} f\ (n-2^{j-1}k)$

$W_{2j}f\ (n) = \sum_{k\in z} g_k S_{2j-1} f\ (n-2^{j-1}k)$

Smoothing operator is S_{2j}. $S_2j\ f\ (n) = aj$, the stumpy frequency coefficients is a_j which is meant to be original signal's approximation while w_{2j} f(n) = d_j , the high frequency coefficients called d_j is literally pretends original signals. It's identified that for evaluating nonstationary signals WT (wavelet transform) is better, an adaptive time-frequency decomposition along with the discrete wavelet transform (DST) in the mentioned pattern. The multi-resolution depiction is capable to explain the signal structure by only utilizing few coefficients of the wavelet domain. Selection on the amount of decomposition level and proper wavelet in analysis of signals with the WT are very essential. The levels of decomposition is taken generally on the basis of frequency components of the signal. The level of the signal relates to the frequencies essential for the signal classification are maintained in the wavelet coefficient. The level of decomposition was considered to be chosen as 4. Component of electrocardiograph signals are detailed as d1-d4 with a4approximation. Typically, the one which gives extreme efficacy in wavelet is selected for the appropriate application

Hence, 8th order Daubechies wavelet was taken. After the application of wavelet transform on signals, various features are calculated such as:

1) Entropy
2) Standard Deviation
3) Energy
4) Variance.

The Wavelet Transform (WT) produces the representation of a non-redundant image, which offers better spectral and spatial localization of image information, in comparison with other multi scale representations. Based on the result of concatenation of all these characteristic features and thus literally form into a single feature vector. From each single beat 11features are purposely extracted and thus estimating the dimensionality of the feature vector as14. This is taken as the input to the classifier RF.

E. Classification

The following data matrix is engendered and applied to SVM-RFE classifier after the extraction of features. The method backward is employed with SVM-RFE which is principally known as elimination method. Nevertheless, the variables that are top having top score are not certainly the ones of the variable that are independently most appropriate but the peak relevant condition of the subset on the specific rank in the model. The subset variables are optimal if and only if taken together. For instance, On a*p* ranked variable in the model with 1 to *p* ranked variables, The variable of least relevant is called *p* . The following are the major steps:

1. Necessity of optimization of the SVM method and the requirement to tune the parameters.
2. The creation of pseudo-samples matrix is essential with z_*, uniform distanced values from the variable for each variable, during the maintenance of the variables set to median or mean (1). z_q would be the variable meant for an *q*, arbitrary which is the called the number of selected quantiles. The mean considered as 0 as the data is typically normalized. There would be matrices of dimension *q x p*,*p* pseudo-samples.

$$\begin{array}{c c}
\begin{pmatrix} & V_1 & V_2 & V_3 & V_p \\ z_1 & 0 & 0 & \ldots & 0 \\ z_2 & 0 & 0 & \ldots & 0 \\ z_3 & 0 & 0 & \ldots & 0 \\ & & \vdots & & \\ z_q & 0 & 0 & \ldots & 0 \end{pmatrix} &
\begin{array}{l} pseudo-samples_1 \\ pseudo-samples_2 \\ pseudo-samples_3 \\ \vdots \\ pseudo-samples_q \end{array}
\end{array}$$

3. Effort in getting the value of result but not to the class of prediction u sing support vector machine (+tive or –tive value) for every artificial-sample utilizing the support vector machine model tailored in step 1. Mostly, decision value resembles to the intervals of each from the support vector machine margins.
4. By using the univariate robust metric median absolute deviation (MAD) calculate the changeability of each variable's prediction. This measure is articulated for *p* variable as

MADp=median (|Dqp−median (Dp)|)c

the value of decision variable *p* is being D_{qp} for the artificial-sample *q* and being $median(D_p)$ the median of all values for the assessed variable *p*. The value of the constant *c* is 1.4826, which is assimilated in the expression to confirm constancy

E(MAD(D1,…,Dn))=σ

for D_i as $N(\mu, \sigma^2)$ and large *n* .

5. The lowest MAD value are to be removed.
6. Steps 2–5 are required to repeat unless and until there will be left only one variable

In the proposed method the rationale is said that for the variables which are associated to the response, predictions will be affected if changes in the variable is done. On the opposing, for variables which are not associated to the response, predictions and the decision value will be approximately constant and they have not been affected on changes made in the variable value. To calculate the distance to the hyperplane the decision value can be utilized as a score the greater absolute value which belongs to the class defined by the sign.

Statistical and wavelet feature is taken into SVM classifier in order to perform the category based feature extraction and identification. In order to predict the probability of ECG signal as normal or abnormal by employing the methods called RBF-SVM

Radial basis function: The SVM plays a vital role in defining the area of impression and stands for the centre of RBF. The data space of the support vector, it is given in equation:

$$K(x_{i,}x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right)$$

SVM-RFE procedure is employed to optimize the features of the classification and thus provide information relevant to defect detection. A correlation bias reduction strategy is incorporated into SVM-RFE for feature elimination.

The process of Training and Testing: In order train the information using SVM classifier, data features of some kind is used to distinguish the defect class. The classifier will be able to distinguish the candidate region type by the data features fed in to the classifier. The chosen classifiers which have taken for extracting the features are concatenated of feature vectors from statistical and wavelet. The abnormal and normal classes are comparing with the concatenated values of the feature vectors by the SVM classifier. After evaluation, the SVM classifier classifies the signal as normal and abnormal.

IV. ANALYSIS AND PERFORMANCE EVALUATION

The enactment for each instance of class is done on calculating the FN and FP factors of performance evaluation, the correction classification of the regular and irregular ECG signals is denoted by (TP) and (TN) represent. FN mentions the abnormal grouping of regular heart beat as irregular signals while FP mentions the wrong classification of the irregular heartbeats into regular signals. On the basis of these factors, the enactment of signal are calculated. Where the term sensitivity is told as the amount of properly classified instances among the whole instances, while accuracy corresponds to the amount of properly classified instances. Accuracy, specificity and sensitivity can be defined as

Sensitivity (sen) = $\frac{TP}{TP+FN} * 100\%$

Specificity (spec) = $\frac{TN}{TN+FP} * 100\%$

Accuracy (acc) = $\frac{TP+TN}{TP+TN+FP+FN} * 100\%$

All these above mentioned parameters in classification are calculated and analyzed from the MIT-BIH database and simulation with these database references.

The method focused in this feature extraction and classification are done using the software MATLAB 8.1 at the release version ofR-2013a on the Windows 8 platform with the system specification as Intel I5 processor, 3.50 GHz, 8 GB of RAM. These are done on benchmark MIT-BIH arrhythmia database.

In order to abstract the variability of HRV signals initially FFT based detection of R-peak is proposed. The features such as statistical and wavelet parameters are taken for classification to lead to suggest regular and irregular ECG signal using SVM-RFE, ELM and RF. The result shows an accuracy of 98.9% for SVM-RFE, 96.6% for RF and 94.6% for ELM classifier (TABLE-I).

TABLE-I

CLASSIFICATION	ACCURACY	SENSITIVITY	SPECIFICITY
ELM	94.66	93.9	96.33
RF	96.75	96.33	96.93
SVM-RFE	**98.97**	**99.96**	**98.23**

STATISTICAL+WAVELET FEATURESGROUPING&COMPARISON

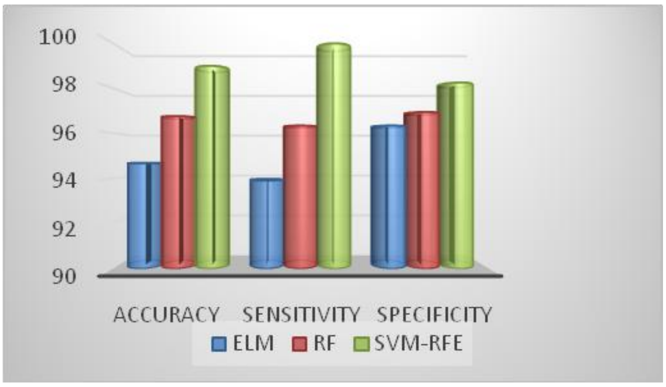

Fig 4. Performance comparison with proposed methodology

Fig 4 shows the enactment of SVM-RFE Classifier on others with the 98.97%accuracy in this work and also noted that the performance of SVM-RFE with the sensitivity and specificity of 99.96% and 98.23% in this work.

To additional efficacy of our work, compare against statistical and wavelet feature extraction techniques using different types of classifier namely, the SVM-RFE, RF and ELM. The assessment graphs of the accuracy, specificity and sensitivity are in Fig. 5, 6.

TABLE-II CLASSIFICATION COMPARISON FOR WAVELET FEATURES

CLASSIFICATION	ACCURACY	SENSITIVITY	SPECIFICITY
ELM	90.5	87.84	**96.33**
RF	88.41	**96.33**	84.81
SVM-RFE	**93.18**	94.1	90.66

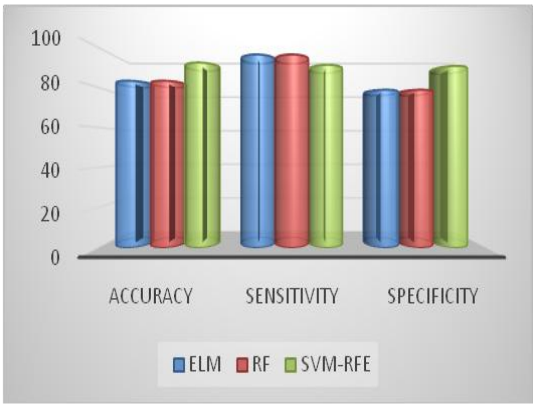

Fig 5. Performance comparison of wavelet feature

By consuming the training set of 25% and 75% for the experiments that are performed. During the training the more the ECG signals samples given the more would be classification accuracy. The feature set representing the discrete characteristics of the abnormalities found in the ECG signal. With the utilization of those feature set will lead to higher accuracy in classification process. Also, these technique is endorsed using the benchmark with MIT-BIH arrhythmia database.

TABLE-III, CLASSIFICATION COMPARISON FOR STATISTICAL FEATURES

CLASSIFICATION	ACCURACY	SENSITIVITY	SPECIFICITY
ELM	83.25	96.33	78.7576
RF	83.25	**96.33**	78.75
SVM-RFE	**92.30**	91.17	90.66

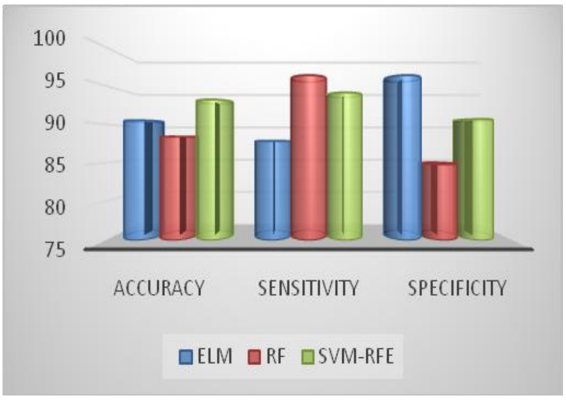

Fig 6. Performance comparison of statistical features

V. CONCLUSION

This research work is especially an automated classification and analysis of ECG signal for long-term monitoring. The major defies in this research work are the choice of appropriate HRV, heart rate variability signal, a joined approach for feature extraction is utilized and thus results better classification accuracy. The wavelet features are computed and categorized to the statistical features to contribute to the processed feature set, which contrast and classify the electroencephalography signal sets. SVM-RFE, RF and ELM classifier are used to classify the irregularities of ECG signal database. The proposed method using SVM-RFE dataset classifier achieved an accuracy and precision of 98.97% on the benchmark MIT-BIH arrhythmia database. The future work of this paper facilitate the use of the additional classes of arrhythmia signals for better analysis. In the future these implementation methods can be adopted in to point of care type devices.

REFERENCES

1. St. Louis, MO, USA: Mosby, 1977. "Research on Clinical Electrocardiography".
2. Da. S, et al, "Athena Scientific", 1991.
3. P. de Chazal, R. B. Reilly & M. O'Dwyer, "Automatic classification of heart-beats using ECG morphology and heartbeat interval features", IEEE Trans. Biomedical Eng., pageno. 1196–1206, 2004.
4. S. Raj, et al., "A knowledge-based real time embedded platform for arrhythmia beat classification", Biomed. Eng. Letter, vol. 5, pageno. 271–280, Dec. 2015
5. F. Melgani & Y. Bazi, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization", vol/no: 12/5, 2008
6. S. Raj, K.C Ray & S. Luthra, "Development of handheld cardiac event monitoring system", IFAC-Pap. Onl., pp. 71–76, vol/no: 48/4, 2015.
7. C. Ye, M.T Coimbra et al, "Heartbeat classification using morphological and dynamic features of ECG signals", IEEE Trans. Biomed. Engg., vol/no: 59/10, pageno. 2930–2941, Oct. 2012.
8. S. Raj, O. Shankar et al, "Cardiac arrhythmia beat classification using DOST and PSO tuned SVM", Comp. Meth. Prog. Biomed., vo.136, pageno. 163–177, Nov. 2016.
9. K. Minami, T. Toyoshima & H. Nakajima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network", IEEE Trans. Biomedic. Engg., vol/no: 46/2, pageno. 179–185, 1999
10. Turker I, Kiranyaz S & Moncef G, "A generic and robust system for automated patient-specific classification of ECG signals", IEEE Trans. Biomedic. Engg., vol/no: 56/5, pageno. 1415–1426, May 2009.
11. Sandeeo R, Kailash C R et al, "Arm-based arrhythmia beat monitoring system", Micro proc. Mic.syst., vol/no: 39/7, pageno. 504–511, 2015.
12. M. Llamedo et al.,"Heartbeat classification using feature selection driven by database generalization criteria", IEEE Transac. Biomedic. Engg., vol/no: 58/3, pageno. 616–625, Mar. 2011
13. T. H. Linh et al., "On-line heart beat recognition using Hermite polynomials and neuro-fuzzy network", IEEE Transac. Instrume. Measu., vol/no: 52/4, pageno. 1224–1231, Aug. 2003
14. G. D. Fraser et al.,"Automated biosignal quality analysis for electromyography using a one class support vector machine", IEEE Trans. Instrume. Measu., vol/no: 63/12, pageno. 2919–2930, Dec. 2014
15. S. Raj et al.,"A comparative study of multivariate approach with neural networks and support vector machines for arrhythmia classification", Inter. Confer. Energ, Power Environ., Tow. Sustain. Growth (ICEPE), Jun. 2015, pageno. 1–6
16. Yue Zhang et al, "A new method for ECG biometric recognition using a hierarchical scheme classifier", 6th IEEE Inter. Conf. on Soft. Engg and Ser. Sci. (ICSESS), 2015
17. Morteza Zabihi et al, "Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier", Comp. in Cardio. (CinC) 2017
18. Quazi Abidur Rahman et al, "Utilizing ECG-based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification", IEEE Transac. on NanoBiosci.., 2015
19. Juyoung Park et al.,"Arrhythmia detection using amplitude difference features based on random forest ", 37th Annual International Conference of the IEEE Engg in Medic. and Biolo. Soci.(EMBC) 2015
20. Diptangshu Pandit et al., "Improved Abnormality Detection from Raw ECG Signals using Feature Enhancement", 12th Inter. Conf. on Nat. Comp., Fuzzy Sys. and Knowl. Disc. (ICNC-FSKD) 2016
21. Ruhi Mahajan et al, Cardiac rhythm classification from a short single lead ECG recording via random forest", Comp. in Cardiol. (CinC) 2017

22. Minas V. et al.,"Task Discrimination from Myoelectric Activity: A Learning Scheme for EMG-Based Interfaces", IEEE 13th Inter. Conf. on Rehab. Robotics (ICORR) 2013
23. "Heart Rate Variability: Stand. of Meas., Physiological Interp. and Clinic. Use", 1996. Available at: http://circ.ahajournals.org/cgi/content/extract/93/5/1043
24. D. Carval al., "Development of a Matlab software for analysis of heart rate variability", In Sig. Process., 6th Inter. Conf., 2002
25. Tul, M.P. et al., "Quantitative beat-to-beat analysis of heart rate dynamics during exercise", 1996. Am. J. Physiolo. Heart. Circu. Physiolo., 271 (1): H244-252.

AUTHORS PROFILE

Anuja S B, received Master Degree in Master of Computer Applications from University of Kerala, India. Presently a full time Assistant Professor in MCA Department of the Narayanaguru College of Engineering, TamilNadu, India. Her Area of Interest includes Image Processing and Networks.

Usha Nandini K, received Master Degree in Master of Computer Applications from Manonmanium Sundaranar University, Tirunelveli, India and Received his Master of Philosophy Degree in Madurai Kamaraj University India. Presently a full time Assistant Professor in MCA Department of the Narayanaguru College of Engineering, TamilNadu, India. Her Area of Interest includes Image Processing and Cloud Computing.

Sukanya S T, received Master Degree in Master of Computer Applications from Anna University Chennai, India and Received his Master of Philosophy Degree in Noorul Islam University, India. Presently a full time Assistant Professor in MCA Department of the Narayanaguru College of Engineering, TamilNadu, India. Her Research Interest includes Data Mining, Image Processing and Neural Networks.