iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VIII **Month of publication:** Aug 2023

DOI: https://doi.org/10.22214/ijraset.2023.55320

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Analysis of ECG Signals using Frequency andTime domain features with SVM

P. Satyanarayana Goud[1], Dr. Panyam Narahari Sastry[2], Dr. P. Chandra Sekhar[3]
[1]*Ph.D. Scholar, Osmania University, Hyderabad, Telangana, INDIA*
[2, 3]*Professor, Department of Electronics and Communication Engineering, CBIT, Hyderabad, Telangana, INDIA*

Abstract: ECG Examination is not limited to diagnosis of cardiovascular disease, but diagnosis of diseases and pathological conditions of a patient. There are numerous uses of ECG data in medical or surgical system namely preoperative / postoperative evaluation, drug efficacy evaluation, detection of side effects and health diagnosis, which can be found using ECG data.Hence, this is an open area of research. This work presents an ECG disease classification modelusing time and frequency domain features with machine learning models. The ECG signal base line wandering (typical noise) problem is solved by filtering. The time domain features such as Heart rate, RR intervals, QRS duration, Shannon entropy, Average Heart rate variability is used along with frequency domain features such as wavelet features, FFT featuresand DWT leader. These features are used for training in the proposed model. A SVM model is used to identify and classify the various heart diseases. In this work, two data sets have beenconsidered to validate the proposed model, which are constructed from MIT-BIH website. Theresults of the proposed model are compared with the results of the existing algorithm. The recognition accuracy for proposed model is 97.95% for "Data set-1" and 94.47% for "Data set-2", which are better than the existing results.

Keywords: Electrocardiogram (ECG), Atrial Fibrillation (AF), Baseline Wandering Correction, Hybrid Features, Support Vector Machine (SVM).

I. INTRODUCTION

One of the leading causes of morbidity and mortality is the cardiovascular disease across the world. The patient's life is reduced due to these diseases and further the cost incurred towardsNational Health Organisations increases. In addition to the prevalence and incidence, the cardiac arrhythmias with the clinical significance is increased that being associated with the population aging. As the Atrial Fibrillation (AF) is the type of sustained arrhythmia [1], it is more common for adults. It includes the significant growing trends specifically in the elder population or obesity disorders. Sometimes, AF is challenging task for diagnosis owing to the possible symptoms absence or the paroxysmal behaviours. The great interest towards the portable devices today as monitoring devices in the clinical settings in the researches. For providing the reliable AG diagnosis, the automatic techniques have been used to obtain the Electrocardiogram (ECG) signals [2] with portable devices. However, a challenge is still existed, especially if they are also considered other normal or pathological rhythms. The ECG shows information on cardiac electrical activity, revealing some of the arrhythmias [3]. Differences from a normal rhythm help to diagnose deviations in driving routes, enlargement of the heart muscle, hormonal imbalances or cellular ion channels, or even the appearance of a myocardial infarction. The most prominent complex on the ECG, called QRS [4], where R indicates the highest peakof the signal. This is why the R-R interval typically serves to indicate the rate at which the heart beats. On the other hand, the P wave indicates atrial activation, while the T wave indicatesthe repolarization of the ventricles. The Figure 1 shows a representation of the most importantelements involved in the generation of the P-QRS-T complex of the ECG, or what is the same,during a heartbeat or cycle.

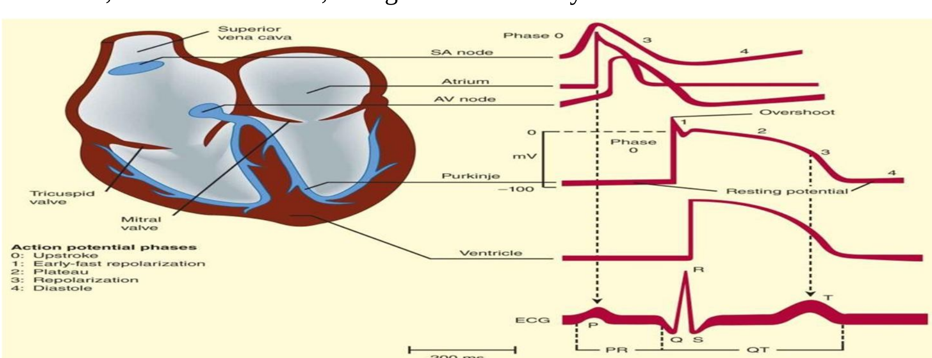


Fig 1: Schematic representation of the heart and Electrocardiogram (ECG)

In adults, the most common sustained cardiac arrhythmia is the Atrial Fibrillation (AF), whichis characterized by the uncoordinated atrial activation predominantly. It is based on the atrial mechanical function with the consequent deterioration [5].

In this regard, Medical Decision Support System (SADAM) provide specific and accurate knowledge for decision making during prognosis, diagnosis, treatments and patient management. The SADMs [6] are related to the concept of evidence-based medicine as they infer knowledge from clinical databases, to later assist in the diagnosis of new patients using the knowledge acquired.

The most complex SADMs are based on the use of Artificial Intelligence (AI) in medicine [7], specifically with the branch of Machine Learning, where algorithms are used to obtain predictive models from the data entrance clinics.

On the other hand, Pattern Recognition (PR) is as follows natural stage of AF, where the resulting predictive models assign a label or class to a set of input values from which it is initially unknown to which group belongs. In classification problems, these labels of classes were known to a priori.

In this research work, the most direct example of Pattern Recognition(PR) would be to determine whether one ECG signal corresponds to that of a person suffering from AF, any other pathological rhythm or conversely, belongs to an individual with a normal heart rhythm. Machine Learning [8] is a subset of techniques within the field of Artificial Intelligence that, through statistical methods, provides the computers of the ability to learn. This learning process consists of the progressive improvement of the performance in a specific task using only data in the form of observations or samples, thus providing models or data structures thatare they progressively adapt to the input data. Current ECG data classification techniques report low accuracy for several classes of arrhythmias due to class conflict and class imbalance problems. The existing machine learningmethods like KNN, Random Forest, Navies Bayes algorithm produce low accuracy. In this paper Multiclass SVM is proposed. This method uses frequency domain and time domain features were used to train the model. So, in test case this model produce better accuracy thanexisting methods.

The organization of this work is as follows: the section-II describes the literature survey, the proposed method is explained in section-III. The results are discussed in section-IV.

II. LITERATURE SURVEY

SK Pandey et al. [9] has introduced three ANN models that categorized into arrhythmia and healthy classes based on UCI repository ECG 12 lead signal extracted data features. Specifically, the ANN models have been used that are tested and trained on the Radial Basis Function (RBF), Recurrent Neural Network (RNN), and back-propagation feedforward neuralnetworks. The testing results are evaluated based on the results of specificity, sensitivity and classification accuracy. The better diagnosis results have showed by the RNN models in termsof 83.1% testing accuracy of classification based on the chosen attributes among the contrast ANN models.

Eric Manibardo et al. [10] has reviewed the 4214 AED rhythm analysis Electrocardiogram (ECG) and the rhythm is annotated while using the consensus decision as the ground truth. Atotal of 22 VT, 472 VF, 1009 PEA, 294 PR, and 2418 AS were included. For developing theautomatic rhythm annotator, the intervals of ECG analysis were extracted and used to partitionthe patient wise data into the test (30%) and training (70%). However, the performance results were evaluated based on F-score (F1) and per class sensitivity (Se). Here, the global performance parameters were used as the unweighted mean of sensitivity (UMS) and F-score.By using the random forest classifier and the stationary wavelet transform of ECG, the denoising stage and feature extraction included in the classification technique. As per thythm Se/F1, the best model presented as 81.9/72.2, 94.2/96.1, 85.3/81.3,43.3/52.2,95.8/95.7 for VT,VF, PEA, PR and AS respectively. For the test set, the UMS was obtained as 80.2% while theabove 2-points that of previous solutions. The large datasets of OHCA annotate by this methodretrospectively and the manual annotation of OHCA rhythm workload is ameliorated.

Namrata Singh et al. [11] has presented the model to diagnose the cardiac arrhythmias based on the selection of best features using the filter based feature selection techniques with the integration of three different ML techniques for cardiac arrhythmia dataset. The crucial pre- processing step is the feature selection that determine the responsible factors for patients who suffering from arrhythmia. The health factors of patients could be examined as it is powerful predictor for heart related deaths.

To assess the feature selection methods performance, three types of ML techniques, termed asJRip, random forest, and linear SVM were incorporated. The highest accuracy of 85.58% hasbeen achieved using the random forest classifier based on gain ratio feature selection by analysing the experimental results for a subset of 30 features.

MU Khan et al. [12] has proposed a technique of signal processing for predicting the disease of Coronary artery based on raw ECG signals of 9 to 12 minutes. Firstly, the raw recording ofECG is pre-processed and segmented based on the chosen intrinsic mode function (IMF) 2-5 and the Empirical Mode Decomposition (EMD). The data is classified by features, like Higuchi Fractal Dimension, Quantile, Energy Entropy, Spectral Entropy, Root sum square, Energy, Marginal Factor, Impulse Factor, shape factor, Kurtosis and Skewness. The pre- processed signal feds into the SVM classifier.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

The accuracy of 95.5% is achieved by the system using the self-collected data. The cardiologists will take assistance from the proposed systemto make the effective decisions of treatment.

Saeed Mian Qaisar et al. [13] has contributed the researches on developing the efficient multirate ECG automated detectors of arrhythmia computationally. An integration of wavelet decomposition and multirate denoising is used for realizing the wireless implants of ECG effectively. The mining of decomposed signal sub-band features is performed and the mature classifier of K-Nearest Neighbour (KNN) classifier is used for diagnosing the arrhythmia. Thesystem's processing activity reduces substantially by the multirate nature and a dramatic reduction is allowed in the energy consumption by comparing with the existing methods.

III. PROPOSED METHOD

Flow chart of proposed method is shown in Figure 2 and block diagram is shown in Figure 3.Flow chart shows the overall flow of the work will be done in 3 stages.

1) *Stage 1:* Data of ECG signals is selected from the standard data set MIT-BIH Physionetdatabase.
2) *Stage 2:* Time domain features and Frequency domain features are extracted for both trainingand testing of ECG signals.

Time domain features:

- Mean HR (in bpm)
- Mean RR (in ms): The mean value of the RR interval.
- SDNN (in ms): Standard deviation of RR interval.
- NN 50v1: Total number of beats with an inter-beat dif-ference over 50ms, variant 1
- NN 50v2: Total number of beats with an inter-beat dif-ference over 50ms, variant 2
- pNN 50v1: Ratio of *NN* 50*v*1 to the ECG segment length
- pNN 50v2: Ratio of *NN* 50*v*2 to the ECG segment length
- SDSD (in ms): Differences of standard deviation betweenthe neighboring RR intervalsof each segment or inter- beat differentials.

RMSSD (in ms): Root mean square of inter-beat differentials

Frequency domain features:

- Wavelet Entropy
- Multiscale wavelet variance
- Two fractal coefficients

3) *Stage 3:* Extracted features (both time domain and frequency domain) are used to classify andidentify the diseases using SVM.

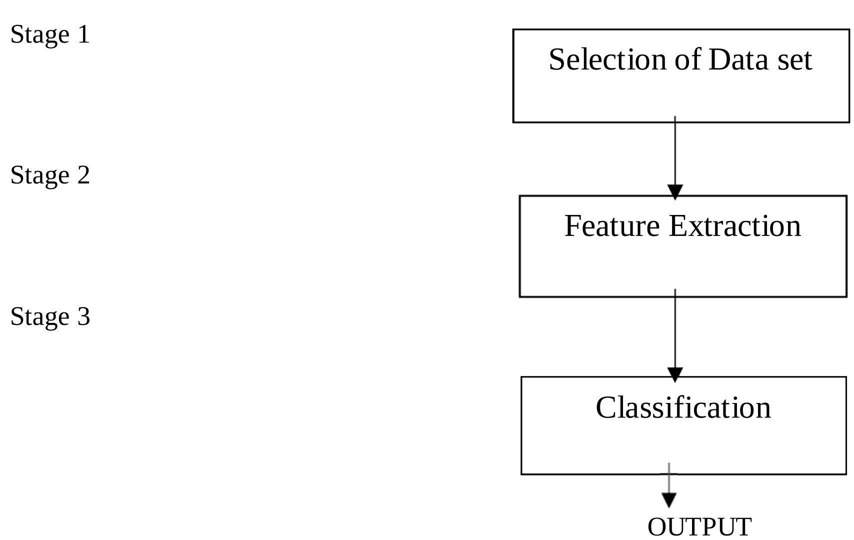


Figure 2: Flowchart of Proposed Model

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

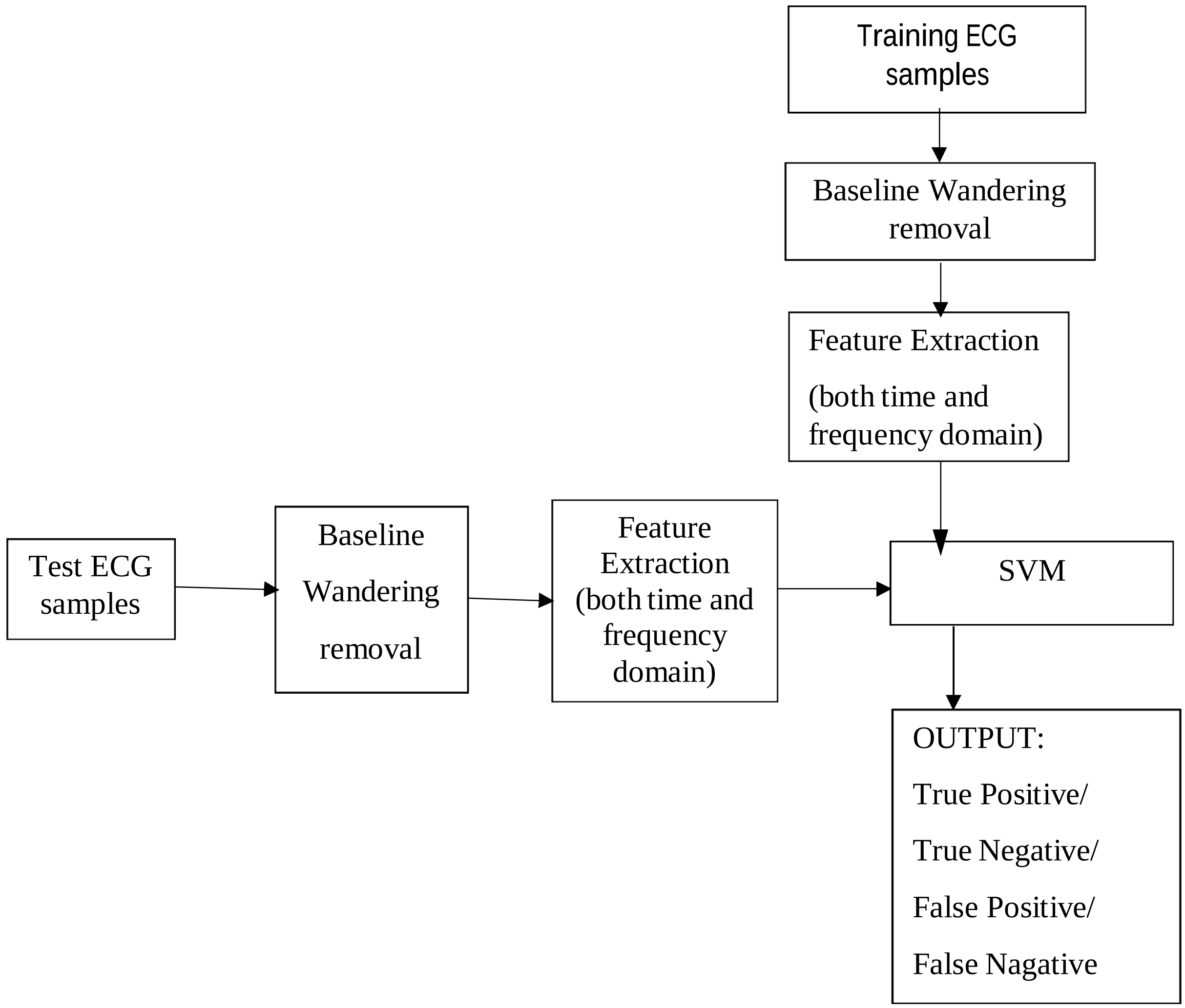


Figure 3: Block Diagram of the Proposed Model

A. Baseline Wandering (BW) Correction

The baseline wandering (BW) can be captures by the coarse approximation $A_{j0}=\sum_k a_{j0,k}$ when the proper wavelet function and resolution level $j0$€$(1,….,J)$ have been chosen. The BW willremove by subtracting the part from the raw signal of ECG.

An open question is included the selection of right wavelet for a particular application, the wavelet is selected for correction of ECG BW that resembles the characteristic and significantwaveform QRS of the ECG signal.

After subtracting the A_{jo} from the raw ECG, the interestingdetail features can be kept and captured the details $\Sigma_{j=j0} D_j$. Without any over-smoothing, theBW can captured by A_{j0} for selection j0. The maximum number of decomposition levels determined based on the signal length.

B. Feature Extraction Techniques

The objective of the feature extraction stage is to transform the segment of the signal to be analysed, in such a way that the relevant clinical information is obtained with reduced number of coefficients. In this way, it is possible to represent the signal in a space whose metric minimizes the distance between patterns of the same class and maximizes the distance between patterns of different classes.

There are two types of features are extracted from the ECG signal. Those are i) Frequency domain features ii) Time domain features.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

1) Frequency Domain Features

The Frequency domain features extracted from wavelet transform and FFT.

a) Wavelet Transformed (WT) Characteristics

The WT decompose the signal into its different spectral components, in such a way that each of these has a resolution according to its scale. The function of the real variable t is knownas the mother wavelet function and must oscillate in time, in addition to being well located in the time domain. The scale parameter "a" is associated with a stretching or shrinking of the parent function. The translation parameter b allows the temporal location of the energy distribution. From the parent function, the wavelet function are generated through joint operations of change of scale and translation, in the form as equation (1)

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \quad \text{........(1)}$$

For the case of the discrete wavelet transform, the dilation parameter "a" and translation parameter "b" take only discrete values. The dilation of the mother wavelet is related to integer powers of a scale of reference to a_0, normally greater than 1, thus $a=a_0^j$. If scales and positions based on powers of 2 (at $a_0=2$, called dyadic positions and scales) are selected, the analysis will be much more efficient and just as accurate as continuous analysis. In this case, the signal f(t)is represented as the approximation series and high frequency details at different resolutions. At each stage, a pair of filters (h, g) are applied as input signal to produce an approximate signaland a detail signal, respectively. The detail signal represents the missing information, from a higher to lower resolution. At all resolutions, the detail coefficients are set by the wavelet representation while at the lowest resolution, the approximation coefficients have been set.

b) Fast Fourier Transform (FFT)

The FFT is used to investigate the distribution of the amplitude of spectrum on ECG and extractthe peak of the spectrum to reflect the different tasks of the signal. From the domain, the mapping of ECG signal is done based on FFT from time to frequency domain. With the decomposition of signal into the corresponding sinusoidal at different frequencies, the signal's frequency spectrum is recognized.

The FFT is given by equation (2), for a discrete signal x[n], where k is each discrete value of the signal:

$$x[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi kn}{N}} \quad \text{........ (2)}$$

Where k is constant, k=0,1,......................... ,N-1

In ECG applications, adopts the Fourier Transform (FT) for the extraction of characteristics in the classification of signals, highlighting that although FT plays an important role is stationarysignal analysis, it has limitations in spectrum analysis of a non-stationary signal.

The determination of FFT can be done quickly or on the real-time basis. The drawbacks of FFT include the used frequencies for breaking down the signal into the sampling frequency of signaland the number of desired frequencies. The frequencies are not chosen without making changes to the two parameters: an energy dispersed spectrum will generate by the simple sinewave, which has the frequency that doesn't fall into one of the transformed frequencies.

C. Shannon Entropy

The average amount of information (entropy) is measure of how much information the sourceis producing. In physics, the word entropy often appears, but its meaning refers to concepts such as "randomness", "irregularity", and "ambiguity". Information theory refers to the exactsame concept, meaning that the more irregular the information, the more information it carrieson average.

Suppose that two alphabets A and B are output randomly. In this way, the information sourcethat the alphabet does not depend on the past and is output independently is called a memorylessinformation source. Each of the probability P_A and P_B, the average amount information givenby equation (3) given below:

$$P_A\,(-\log_2 P_A) + P_B\,(-\log_2 P_B) \text{ bits} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

The entropy decreases when the frequency of the appearance of the alphabet is biased is shownin figure 4.

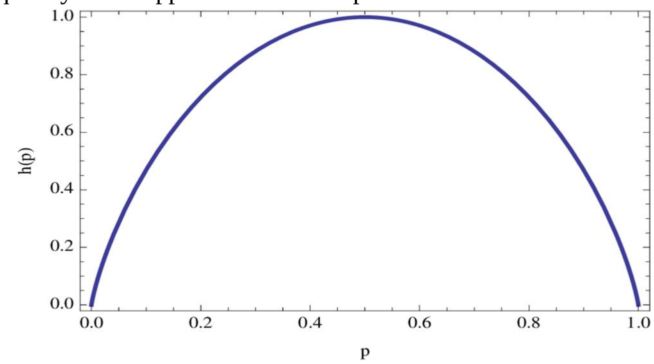


Fig.4: Shannon Entropy

D. AR Coefficients

An autoregressive model (AR model) is a model that regresses a value at a certain time t using data older than the time t. Autoregressive models are useful for modelling time series data with high autocorrelation. The regression equation (4) can be expressed as follows. i is the autoregressive coefficient and p is the order. Also, Et is an error term, which is white noise with an average of 0 and a variance of ^{2} Ψ

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \varepsilon_t \ldots\ldots\ldots\ldots\ldots (4)$$

σ

E. Hybrid Features

Combining the both time domain and frequency domain features into one feature vector is called hybrid features. These hybrid features are used for training and testing.

F. SVM Classifier

In Support Vector Machines, the machine learning process estimates the hyperplane that maximizes the margin between two classes in the training data. The margin is defined as the minimum perpendicular distance between two points in each class to the separator hyperplane.This hyperplane is adjusted during the learning process, selecting from the training data or predictors of those vectors that they best define the hyperplane, which are called support vectors.

An example with the basic idea pursued by SVM in two dimensions is shown in Figure 4. Contextualizing the above to the specific case and assuming that only has two independent variables x_1 and x_2 , the green points could represent people whose ECG shows a normal heartrhythm, while the red points those who suffer from atrial fibrillation or another abnormal heartrhythm.

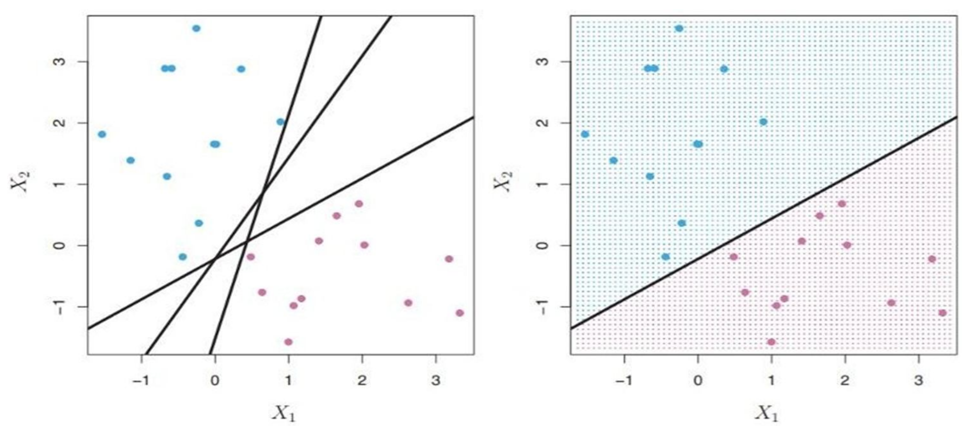


Figure 5: Basic idea of hyperplane in SVM in a two-dimensional space

In an ideal case, a plane could separate two classes with the help of margin of distance to saidplane. For example, 70 independent variables, the space would be made up to 70 dimensions. The hyperplane will be of dimension D-1, where D indicates the number of dimensions or independent variables.

As can be seen in Figure 5, the boundary hyperplanes are the hyperplanes that corresponding to $y\vec{w}.\vec{x}$+b = -1 and $\vec{w}.\vec{x}$+b = -1 while defining the margin. The margin is the distance between two boundary hyperplanes and the 2/|w| is determined the value.

The effect of the classifier complexity reduction has been maximized the separation margin between two classes for given training data. Therefore, the generalization is optimized. The optimal hyperplane is corresponded to minimize the training error and the maximum separationmargin has involved between two classes.

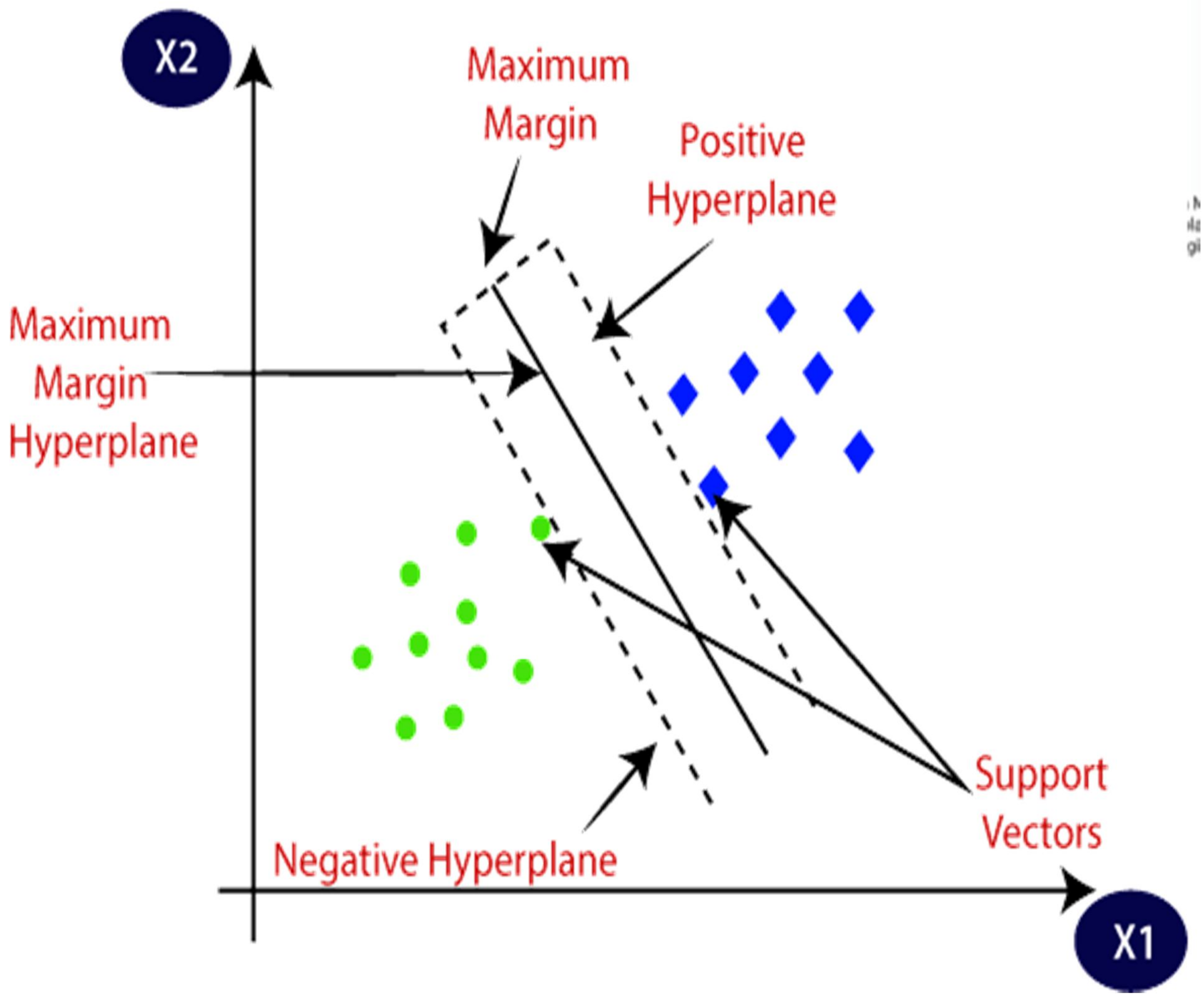


Figure 6: Examples of hyperplane of maximum separation and margin between twoclasses in SVM, in a plane with two dimensions(x1,x2).

The training data is incorporated into the higher dimensionality with another space in the SVMprojects for generalizing the cases, in which the decision limits are not separable linearly. Thedata will always be separable linearly when the new space dimensionality is higher enough.For avoiding to carry out an explicit projection in a larger dimensional space, a kernel(K) function is used. This function K is the one that implicitly transform the data to this larger dimensionalspace to make the linear separation of the classes possible as shown in Figure 6.

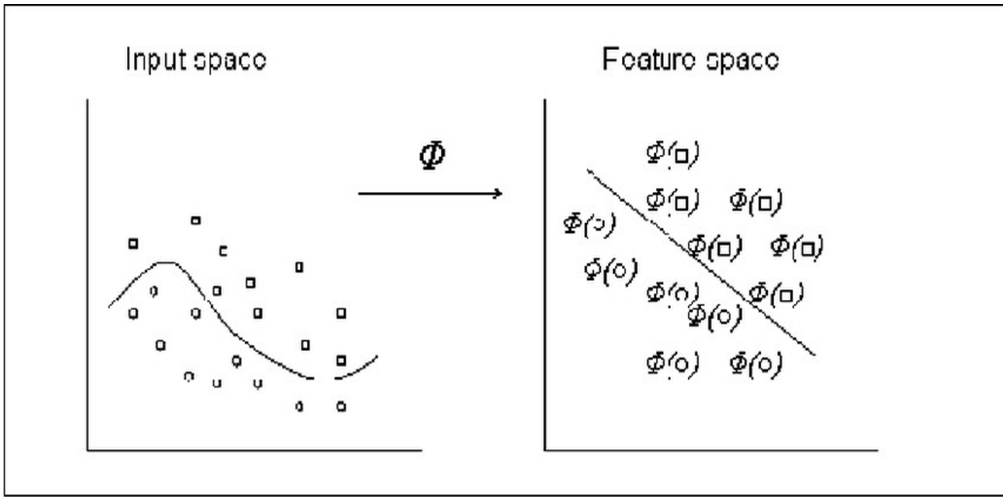


Figure 7: Illustration of how the Kernel (K) functions are used in SVM to transform anon-linearly separable space into a linearly separable one of higher dimensionality.

SVM models were defined as binary classifiers (two classes to be labelled), where there is no selection of any decision threshold in their output, since the classification is carried out by labelling each sample on one side or the other of the created hyperplane is shown in Figure 7. To convert SVM into a classifier of more than two classes, in this work the strategy “one-vs- one” or “one against one” was followed, where it is necessary to train a total of C(C-1)/2 binaryclassifiers in a problem with C classes. With the present case three classes, the trained total ofthree classifiers and implemented the corresponding voting system among the outputs of eachof the models to classify a sample. At each point in the SVM hyper parameter search network,the same values were used to train the six binary classifiers, thus avoiding a combinatorial explosion in unapproachable practice.

On the other hand, a variable C can be handled by SVM models for allowing some flexibility that allows the compensation controls between the rigid margins and training errors. A soft margin is created by allowing some errors in the classification at the same time. Once it penalizes them thus, the hyper parameter C represents the compromise between the size of themargin and the number of misclassification. With everything described above, it can be concluded that the performance of the SVM depends on the Kernel function used, its parameters and the margin penalty parameter C.

IV. RESULTS AND DISCUSSION

A. Data set-1

To evaluate the different methodologies of the participants of the competition, the basis of datawas divided into two data sets: training and testing. The set of training consisted of 113 records. On the other hand, the test set contained 49 recordings of lengths and distributions of three classes similar to training. The dataset1 has 3 different classes, those are Cardiac Arrhythmia(ARR), Normal Sinus Rhythm (NSR) and Congestive Heart Failure (CHF). The process of generating a model for pattern recognition based on Machine learning is dividedinto two main phases: training and recognition. During the training phase a data set is used forbuild the model is called training set. It is in this phase where an adaptive model is adjusted toobtain the best possible generalization, and of that way to resolve new cases during the recognition phase. Once the model is ready, it is possible to incorporate in a computer systemto identify and classify new observations.

1) SVM Training and Validation

The following describes the experimentation performed with support type models vector Machine, which consisted, as in previous models, in the training and cross-validation of classification performance in sets of selected data. In both experiments the same search was used in the different ones hyper parameters which are detailed below.

As described above and SVM is classified by the hyper plane that maximizes the margin between two classes in the data training. The vectors that define this hyper plane, selected from independent predictors or variables are the so called ‘support vectors’.

The machine learning algorithm predicts each element in the verification set, clarifies whetherit is negative or positive, and then classifies all elements into the following four categories based on the prediction and the label of the gold standard: True Negative (TN), True Positives(TP), False Positives (FP) and False Negatives (FN).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

a) Precision: Precision is the proportion of samples that are predicted to be positive and it mentioned below as equation (5).

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$= \frac{True\ Positive}{Total\ Predicted\ Positive} \quad \text{........} \quad (5)$$

b) Recall: In the sample, many positive samples are indicated by the recall rate that have been predicted correctly. It has mentioned as below equation (6)

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$= \frac{True\ Positive}{Total\ Actual\ Positive} \quad \text{........} \quad (6)$$

c) F1_score: The F1_score indicator appears, that is the harmonic average of the recall rate and the precision. It has mentioned as equation (7).

$$F_1\text{–score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \quad \text{........(7)}$$

d) Accuracy: The proportion of the number of prediction pairs to the total number of samplesand it mentioned below as equation (8).

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad \text{........ (8)}$$

2) Training

The confusion matrix for training “Data set-1” shows in figure 8. The data set includes 3 classes, those are Cardiac Arrhythmia (ARR), Normal Sinus Rhythm (NSR) and Congestive Heart Failure (CHF).

BY observing the confusion matrix all the classes are classified correctly and produces almost100% accuracy.

True class	ARR	67	0	0	ARR
	CHF	0	21	0	CHF
	NSR	0	0	25	NSR
		ARR	CHF	NSR	
		Predicted Class			

Figure 8: Confusion matrix for Training

The evaluation parameters for proposed and existing methods represented in table 1. For all KNN classifier, Random forest classifier and SVM classifier are produce better results for training data set. The precision, recall and F1_score for KNN classifier, Random Forest (RF) and SVM classifier are reached to 100%.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Table 1: Evaluation parameters comparison table for training

Parameter	Venkataramanaiah et al.[14] (KNN classifier)			Sudestna Nahak et al.[15] (Random Forest Classifier)			SVM classifier (Proposed method)		
	ARR	CHF	NSR	ARR	CHF	NSR	ARR	CHF	NSR
Precision	100	100	100	100	100	100	100	100	100
Recall	100	100	100	100	100	100	100	100	100
F1_Score	100	100	100	100	100	100	100	100	100

The accuracy comparison between existing methods and proposed method shown in table 2. The accuracy for KNN classifier, Random Forest (RF), SVM classifier are reached to 100%.

Table 2: Accuracy comparison table for training

Methods	Recognition Accuracy(%)
Venkataramanaiah et al.[14](KNN classifier)	100
Sudestna Nahak et al.[15] (Random Forest Classifier)	100
SVM classifier (Proposed method)	100

Testing:
The confusion matrix for testing "Data set-1" represented in figure-9. By observing theconfusion matrix for ARR, CHF classified correctly and for NSR misclassified with 1 value.

True class				
	ARR	29	0	0
	CHF	0	8	1
	NSR	0	0	11
		ARR	CHF	NSR
		Predicted Class		

Figure-9: Confusion matrix for testing

The evaluation parameters for proposed and existing methods with testing "Data set-1" shownin table 3. The precision, recall and F1_score for KNN classifier is 59.184%, 100% and 74.359%. The precision, recall and F1_score for RF classifier is 98.92%, 96.29% and 99.43%.The precision, recall and F1_score for proposed (SVM) classifier is 97.23%, 96.29% and 96.59% respectively.

Table3: Evaluation parameters comparison table for testing

Parameter	Venkataramanaiah et al.[14] (KNN classifier)			Sudestna Nahak et al.[15] (Random Forest Classifier)			SVM classifier (Proposed method)		
	ARR	CHF	NSR	ARR	CHF	NSR	ARR	CHF	NSR
Precision	59.184	NaN	NaN	96.667	100	100	100	100	91.667
Recall	100	0	0	100	88.889	100	100	88.889	100
F1_Score	74.359	NaN	NaN	98.305	100	100	100	94.118	95.652

The accuracy comparison between existing methods and proposed method for testing "Data set-1" shown in table-4. The accuracy for KNN classifier is 59.18%, for Random Forest (RF)is 97.95% and for proposed SVM classifier are reached to 97.95%.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Table 4: Accuracy comparison table for testing data

Methods	Recognition Accuracy (%)
Venkataramanaiah et al.[14](KNN classifier)	59.1837
Sudestna Nahak et al.[15] (Random Forest Classifier)	97.9592
SVM classifier (Proposed method)	97.9592

B. Data Set-2

To evaluate the different methodologies of the participants in the competition, the basis of datais divided into two data sets: training and testing. The set of training consisted of 223 records. On the otherhand the test set consisted of 95 recordings of length and distribution of three classes similar to training. The data set has total 5 different classes and those are Cardiac Arrhythmia (ARR), Atrial Fibrillation (AF), Milignant Ventricular Ectopy (MVE), Normal Sinus Rhythm (NSR) and Supraventricular Arrhythmia (SVA).

1) Training

The confusion matrix for training dataset-2 shown in figure-10. The predicted class is the actualresult of the training data. The training data is classified with the accuracy of 96.41%

True class						
	ARR	62	0	0	2	0
	ATF	1	32	0	0	0
	MVE	0	0	31	0	0
	SVA	1	0	0	65	0
	NSR	1	0	0	3	25
		ARR	ATF	MVE	SVA	NSR
		Predicted Class				

Figure 10: Confusion matrix for training

The evaluation parameters for proposed and existing methods with training "Data set-2" shown in table 5. The precision, recall and F1_score for KNN classifier is 75.04%, 78.1% and 76.04%.The precision, recall and F1_score for RF classifier is 94.62%, 96.58% and 95.9%. The precision, recall and F1_score for proposed (SVM) classifier is 95.94%,97.78% and 96.94%.

Table 5: Evaluation parameters comparison table for training

Parameters	Venkataramanaiah et al.[14] (KNN classifier)					Sudestna Nahak et al.[15] (Random Forest Classifier)					SVM classifier (Proposed method)				
	ARR	AF	MVE	SVA	NSR	ARR	AF	MVE	SVA	NSR	ARR	A F	MVE	SVA	NSR
Precision	89.5	79.8	69.3	87.4	49.2	97.2	91.3	100	98.0	86.6	97.2	97.3	100	98.5	86.7
Recall	74.4	84.8	81.2	77.7	72.4	95.7	100	97.2	93.3	96.7	95.3	100	100	93.6	100
F1_Score	81.3	82.4	75.5	82.3	58.7	96.8	96.4	98.5	96.3	91.5	96.6	98.2	100	96.8	93.1

The accuracy comparison between existing methods and proposed method for training "Data set-2" shown in table 6. The accuracy for KNN classifier is 77.13%, for Random Forest (RF)is 95.52% and proposed SVM classifier are reached to 96.41%.

Table 6: Accuracy comparison for training

Methods	Recognition Accuracy (%)
Venkataramanaiah et al.[14] (KNN classifier)	77.13
Sudestna Nahak et al.[15] (Random Forest Classifier)	95.52
SVM classifier (Proposed method)	96.41

2) Testing

The confusion matrix for testing "Data set-2" shown in figure 11. The predicted class is theactual result of the testing data. The testing data is classified with the accuracy of 94.74%.

True class						
	ARR	26	1	0	1	0
	ATF	1	13	0	0	0
	MVE	0	0	12	0	0
	SVA	1	0	1	28	0
	NSR	0	0	0	1	11
		ARR	ATF	MVE	SVA	NSR
		Predicted Class				

Figure 11: Confusion matrix for testing

The evaluation parameters for proposed and existing methods with testing "dataset-2" shown in table 7. The precision, recall and F1_score for KNN classifier is 69.6%, 69.64% and 70.06%.The precision, recall and F1_score for Random Forest (RF) classifier is 93.48%, 92.74% and 93.06%. The precision, recall and F1_score for proposed (SVM) classifier is 95.74%, 95.16%and 95.52%.

Table 7: Evaluation parameters comparison table for testing

Parameters	Venkataramanaiah et al.[14] (KNN classifier)					Sudestna Nahak et al.[15] (Random Forest Classifier)					SVM classifier (Proposed method)				
	ARR	ATF	MVE	SVA	NSR	ARR	ATF	MVE	SVA	NSR	ARR	ATF	MVE	SVA	NSR
Precision	78.3	79.3	60.1	82.8	47.5	89.2	100	100	94.7	83.5	93.2	100	100	93.5	92.3
Recall	67.2	79.4	69.8	77.1	54.7	89.5	93.3	92.2	97.1	91.6	96.0	93.6	92.7	93.5	100
F1_Score	72.2	79.4	64.6	79.6	54.5	89.3	96.4	96.5	95.7	87.4	95.2	96.3	96.0	93.3	96.8

The accuracy comparison between existing methods and proposed method for "testdata-2" shown in table 8. The accuracy for KNN classifier is 71.58%, for Random Forest (RF) is 92.63% and proposed SVM classifier is reached to 94.74%.

Table 8: Accuracy comparison for testing

Methods	Recognition Accuracy (%)
Venkataramanaiah et al.[14] (KNN classifier)	71.58
Sudestna Nahak et al.[15] (Random Forest Classifier)	92.63
SVM classifier (Proposed method)	94.74

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

V. CONCLUSION AND FUTURE SCOPE

In this work the diseases, Cardiac Arrhythmia (ARR), Normal Sinus Rhythm (NSR), and Congestive Heart Failure (CHF) were considered for "Data set-1" and Cardiac Arrhythmia (ARR), Atrial Fibrillation (ATF), Malignant Ventricular Entropy (MVE), Normal Sinus Rhythm (NSR), and Supra Ventricular Arrhythmia (SVA) were considered for "Data set-2" from MIT-BIH data base. The confusion matrix for both training and testing are found using SVM classifier. The recognition accuracy obtained is found to be 96.41% in the proposed method for "Data set-1" which is in line with existing random forest classifier. However, thisresult is far better than the existing KNN classifier.

The recognition accuracy obtained is 94.47% for "Data set-2" which is better than random forest classifier and KNN classifier. The proposed method gave the superior recognition accuracy compared to existing KNN classifier for both "Data set-1" and "Data set-2" . The precision, recall and F1-score values of the proposed algorithm are 97.23%, 96.29% and 96.59% respectively for "Data set-1" and 96%, 95.36% and 95.58% for "Data set-2" respectively. The future scope of this work can be a hybrid classifier consisting of two differentclassification techniques.

BIBLIOGRAPHY

[1] Linz. Dominik. Adrian D. Elliott. Mathias Hohl. Varun Malik, Ulrich Schotten, Dobromir Dobrev, Stanley Nattel et al. " Role of automatic nervous system in atrial fibrillation.". International Journal of Cardiology 287(2019):181-188

[2] Turker Tuncer, Sengul Dogan, Pawel Plawiak, and U.Rajendra Acharya, "Automated Arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals", Knowledge based systems 186(2019):104923

[3] S.Sahoo, M.Dash, S.Bahera and S.Sabut, " Machine learning approach to detect cardiacarrhythmia in ECG signals:A survey", IRBM41, no4(2020):185-194.

[4] Jagdeep Rahul, Marpe Sora and Lakhan Dev sharma, "Exploratory data analysis basedefficient QRS complex detection technique with minimal computational load", Physicaland Engineering science in Medicine 43, no.3(2020):1049-1067.

[5] Vignesh Kalidas and Lakshman S.Tamil, "Detection of atrial fibrillation using discrete –state Markov models and Random Forests", Computers in biology and medicine 113(2019):103386.

[6] Samia Sbissi, Mariem Mathfoudh and Said Gattoufi, "A medical decision support system for cardiovascular disease based on ontology learning", in 2020 International multi-conference on "Orgaization of knowledge and Advanced Technologies"(OCTA),pp.1-9, IEEE-2020.

[7] Zhanquan Sun, Chaoli Wang, Yangyang Zhao and Chao Yan, "Multi label ECG signalclassification based on ensemble classifier", IEEE access8(2020):117986-117996.

[8] Zhaoyang Ge, Zhihua Zhu, Panpan Feng, Shuo Zhang, Jing Wang and Bing Zhou. "ECG signal classification using SVM with multi feature", in 2019, 8th International Symposium on Next Generation Electronics(ISNE). Pp.1-3, IEEE,-2019.

[9] Saroj Kumar Pandey, and Rekh Ram Janghel, "ECG arrhythmia classification using artificial neural networks", In proceedings of 2nd International conference on communication, computing and networking, pp.645-652, Springer, Singapore-2019.

[10] Eric Manibardo, Unai Irusta, Javier Del Ser, Elisabete Aramendi, Iraia, Mikel Olabarria, Carlos Corcuera, Jose Veintemillas, and Andima Larrea, "ECG based random forest classifier for cardiac arrest rhythms", in 2019, 41st International conference of the IEEE Engineering in Medicine and Biology society(EMBC), pp.1504-1508, IEEE-2019.

[11] Namrata singh and Pradeep singh, " Cardiac arrhythmia classification using machine learning techniques", in Engineering vibration, Communication and Information Processing, pp.469-480, Springer, Singapore,2019.

[12] Muhammad Umar Khan, Sumair Aziz, Syed Zohaib Hassan Naqvi and Abdul Rehman,"Classification of Coronary Artery diseases using Electrocardiogram signals". In 2020,International conference on Engineering trends in smart technologies(ICETST), pp.1- 5, IEEE.2020.

[13] Saeed Mian Qaisar, Moez Krichen and Fatma Jallouli, "Multirate ECG processing andk-nearest neighbor classifier based efficient arrhythmia diagosis', in International conference on smart homes and health telematics, pp.329-337, Springer, Cham, 2020.

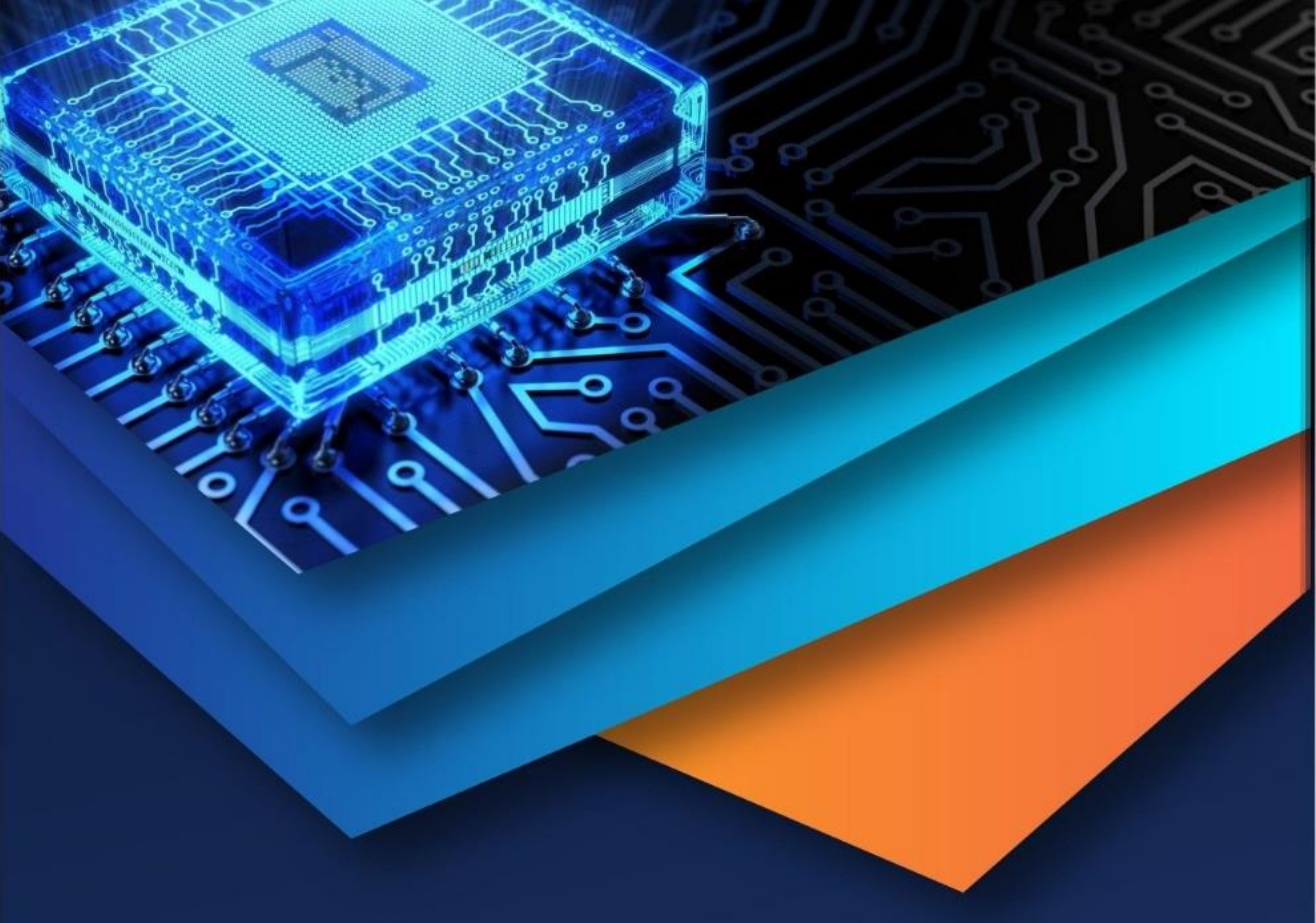
[14] B.Venkataramanaiah and J.Kamala, "ECG signal processing and KNN classifier based abnormality detection by VH-doctor for remote cardiac health care monitoring", Soft computing, 24, no.22(2020):17457-17466.

[15] Sudestna Nahak and Gouthm Saha, "A fusion based classification of normal, arrhythmia and congestive heart failure in ECG", in 2020 National conference on Communications(NCC),pp.1-6, IEEE,2020

AUTHORS PROFILE

Mr. P.Satyanarayana Goud pursued his B.Tech from JNTU, Hyderabadin 2005. He pursued his M.Tech from JNTU, Hyderabad in 2010. He has two International Journals to his credit. His area of interest in research is Signal Processing. He has 16 years of teaching experience. Currently he is working in G.Narayanamma Institute of Technology and Science (for Women), Hyderabad.

Dr. P. Narahari Sastry is working in CBIT since 2003, and is involved in Teaching and Research. He has recognized as Supervisor for JNTU Hyderabad, Osmania University Hyderabad and K.L. University Vijayawada. He has guided three (03) PhD students and also guiding eight (8) Research scholars from Osmania University Hyderabad presently. He has published 6 SCI Journals in the last 5 years and 2 of them are published in the reputed journals which are "Electrical and Computer Engineering" and IEEE Communication letters. Also he has applied and published his work in Indian Journal of patents, Chennai, India.He was the PI for an AICTE project titled "Design and Development of Palm Leaf CharacterRecognition System" granted under RPS scheme, which was successfully completed during the period 2013-16.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY