Automatic cardiac Arrhythmia classification using SVM

Prof. Pooja Mishra

Ankita Thorat **Deepak Mustare,**

Nikita Sonar **Siddhesh Chavan**

Dr. D.Y. Patil Institute of Engineering, Management and Research, Akurdi

Computer Engineering Department

Abstract **- Cardiac arrhythmia is a heart condition that affects the electrical system of the heart, leading to abnormal heart rhythms. Arrhythmias are associated with a number of adverse outcomes, including sudden cardiac death, heart failure, and stroke. In this paper, we explore the use of Support Vector Machines (SVM) as a machine learning tool for detecting and classifying cardiac arrhythmias based on electrocardiogram (ECG) data. We evaluate the performance of SVM using a publicly available dataset of ECG recordings and compare the results to those obtained using traditional statistical methods.**

Index Terms - Cardiac Arrhythmia, Machine Learning, deep learning, ECG classification

I. INTRODUCTION

Cardiac arrhythmia is a condition that occurs when the electrical system of the heart fails to function properly, leading to an irregular heartbeat. Arrhythmias are a significant public health concern and can result in sudden cardiac death, heart failure, and stroke. Early detection and treatment of arrhythmias can reduce the risk of adverse outcomes, highlighting the importance of accurate diagnosis.

One approach to diagnosing arrhythmias is through the analysis of ECG recordings, which provide a non-invasive method of measuring the electrical activity of the heart. However, ECG data are complex and difficult to interpret manually, and automated methods of analysis are necessary for accurate diagnosis.

Support Vector Machines (SVM) are a type of machine learning algorithm that has been used successfully in a variety of applications, including pattern recognition and classification. SVM has been applied in the field of biomedical engineering for the diagnosis of various medical conditions, including arrhythmias.

In this paper, we explore the use of SVM for the classification of arrhythmias using ECG data. We compare the performance of SVM with that of traditional statistical methods and evaluate its ability to accurately classify different types of arrhythmias.

II. PEER REVIEW

IN [1], the author presents the analysis of heart diseases that are categorized as arrhythmia based on an Electrocardiogram (ECG). ECG database of different disease conditions was analyzed. The ECG signals are filtered to remove noise which is caused due to power line interface or Electromyogram. This filtered signal is segmented into smaller pieces of ECG so that feature extraction is accurate.

In [2], the proposed method combines both Support Vector Machine (SVM) and Genetic Algorithm approaches. The design of the SVM classifier is optimized by searching for the best value of the parameters that tune its discriminate function and looking for the best subset of features that optimizes the classification fitness function. Experimental results demonstrate that the approach adopted better classifies ECG signals. Four types of arrhythmias were distinguished with 93% accuracy.

In [3], In this paper, Among various existing SVM methods, three well-known and widely used algorithms one-against-one, one-against all, and fuzzy decision function are used here to distinguish between the presence and absence of cardiac arrhythmia and classify them into one of the arrhythmia groups. The results obtained through implementation of all three methods are thus compared as per their accuracy rate in percentages and the performance of the SVM classifier using one-against-all (OAA) method was found to be better than other techniques. ECG arrhythmia data sets are of generally complex nature and SVM based one-against-all method is found to be of vital importance for classification based diagnosing diseases pertaining to abnormal heart beats.

In [4], in this paper support vector machine (SVM) classifier is developed for the classification of two types of arrhythmias i.e. premature ventricular contraction (PVC) and atrial premature contraction (APC). Discrete wavelet transform (DWT) is used for feature extraction of the ECG signal. For the classification purpose MIT-BIH arrhythmia database is used from the physionet.org. The aim of the work is to develop technique which classifies the arrhythmia with higher accuracy. MATLAB 7.8.0(R2009a) is used for the Simulation purpose.

In [5], the current paper, describes a machine learning based approach for computer-assisted detection of five classes of ECG arrhythmia beats using Discrete Wavelet Transform (DWT) features. Further, methodology comprises dimensionality reduction using Independent Component Analysis (ICA), ten-fold cross-validation and classification using Support Vector Machine (SVM) kernel functions. Using ANOVA significant features are selected and reliability of accuracy is measured by Cohen's kappa statistic. Large dataset of 110,093 heartbeats from 48 records of MIT–BIH arrhythmia database recommended by ANSI/AAMI EC57:1998, which are grouped into five classes of arrhythmia beats viz. Non-ectopic (N), Supraventricular ectopic (S), Ventricular ectopic (V), Fusion (F) and Unknown (U) are classified with class specific accuracy of 99.57%, 97.91%, 92.18%, 76.54% and 97.22% respectively and an overall average accuracy of 98.49%, using SVM quadratic kernel. The developed methodology is an efficient tool, which has intensive applications in early diagnosis and mass screening of cardiac health.

In [6], This paper focuses on the ECG deflections, cardiac arrhythmia, and its types. The paper further dwells into the development of an automated system to detect and classify arrhythmia. Various Machine Learning algorithms like Support Vector Machine (SVM), Random Forest Classifier (RF) are analyzed that lead to the identification of the optimized machine learning algorithm for classification of cardiac arrhythmia to distinguish the patient with arrhythmia. Kernelized SVM has been identified as the most accurate model.

We used a publicly available dataset of ECG recordings, the data_arrhythmia.csv Database, which contains around 6100 recordings of ECG signals with varying degrees of arrhythmia. The dataset includes recordings of preRR, postRR, pPeak, tPeak, rPeak, sPeak, qPeak, qrs interval, We first preprocessed the ECG signals by filtering out noise and baseline wander using a bandpassam filter and a high-pass filter. After filtration, we extract features from the ECG signals, including amplitude, frequency, and duration measures. These features were used as input to the SVM algorithm.

We implemented a radial basis function (RBF) kernel SVM classifier using the LIBSVM library in MATLAB. We trained the SVM using 80% of the dataset and tested its performance using the remaining 20%. We used 10-fold cross-validation to optimize the SVM hyper parameters.

III. Problem Statement

Using a dataset of electrocardiogram (ECG) signals, the task is to classify the signals into different arrhythmia categories using an SVM algorithm. The input to the SVM algorithm will be a set of features extracted from the ECG signals, such as QRS complex duration, P wave duration, T wave duration, ST segment duration, and heart rate. The output of the SVM algorithm will be a class label indicating the arrhythmia category of the ECG signal.

IV. ALGORITHM

a) Support Vector Machine: SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH).

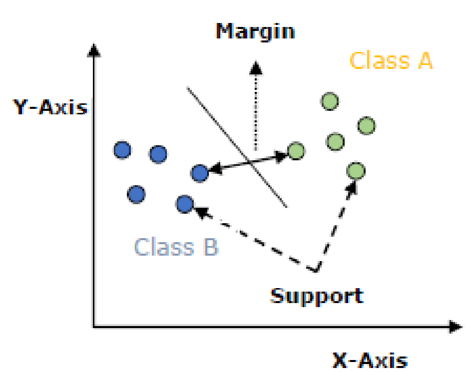

Fig.1 Support Vector Machine representation

The followings are important concepts in SVM:

Support Vectors − Data points that are closest to the hyperplane is called support vectors. Separating line will be defined with the help of these data points.

Hyperplane − As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.

Margin − It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin

SVM KERNEL− the SVM kernel is a function that takes low-dimensional input space and transforms it into higher-dimensional space, i.e. it converts non separable problems to separable problems. It is mostly useful in non-linear separation problems. Simply put the kernel, does some extremely complex data transformations and then finds out the process to separate the data based on the labels or outputs defined. In practice, SVM algorithm is implemented with kernel that transforms an input data space into the required form. SVM uses a technique called the kernel trick in which kernel takes a low dimensional input space and transforms it into a higher dimensional space. In simple words, kernel converts non-separable problems into separable problems by adding more dimensions to it. It makes SVM more powerful, flexible and accurate.

b) Algorithm for arrhythmia classification: Algorithm for arrhythmia classification Once features are detected, SVM classifier is used for classification. SVM generally creates a hyperplane between the classes which can be visualized as data points in a plane. The data set consists of more than hundreds of data which have ECG with tachycardia, bradycardia and normal ECG data out of which 15 data sets 5 of each category are used as training set to train the SVM classifier.

There are two types of comparison beat to beat and beat to bat. The beat to beat variation consists of variation in QRS axis and beat to bat variation depends on P axis. The features used for reference here are heart rate, PR interval and QRS interval which are given to an SVM classifier. Thus, each data point is classified as normal (NL) Tachycardia (TC) and bradycardia (BD). The dimensional spaces use kernel functions in decision making. Using kernel functions the input ECG data are classified into one of the categories.

This kernel type is defined for non-linear classifiers, where simple SVM classifiers cannot be used. The kernel SVM displays data that is non-linear with fewer dimensions as linearly Segre gable data with more dimensions. This way data points belonging to different classes are allocated to different dimensions. But while classification algorithms are being used, the data can either be classified into the correct class or incorrectly. In the case of medical problems, the impact of incorrectly classified data is higher.

V. System Architecture

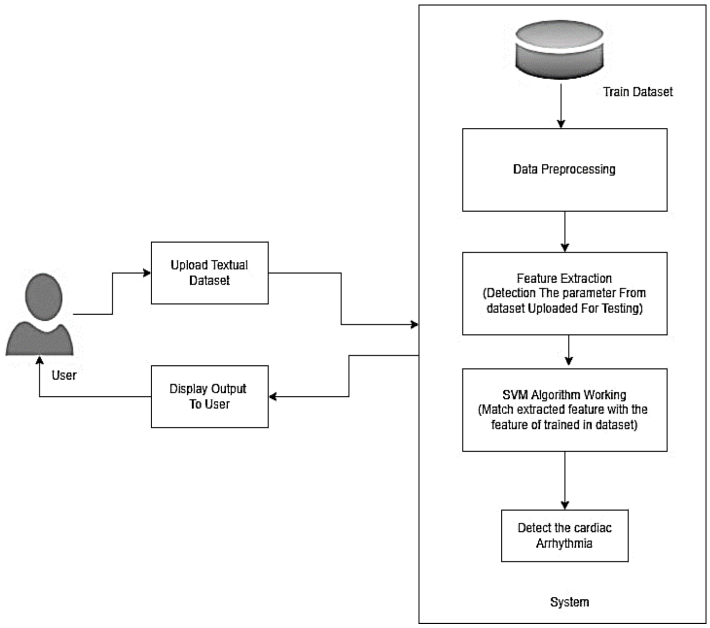

Fig.2 System Architecture

IV. Comparison of performance

The classification results of arrhythmia in ECG database indicate that among various methods, SVM algorithm shows the highest percentage of accuracy rate. The system was trained and optimized by classifying the dataset at various values. The system was trained and optimized by classifying the dataset at various values. It is found that performance wise one-against-one algorithm trails behind one-against-all method in ECG classification but its results could be competitive at times. Naïve Bayes showed poor and constant classification results on ECG dataset. To remove ambiguity over having any potential to improve its performance, Naïve Bayes method was used for classification with value ranging from 1 to 5000 but no improvement in accurate rate was observed. Clearly, SVM method resulted in the highest accuracy rate but in general, the very high accuracy rate is difficult to obtain. This could be due to the presence of a particular class sweeping maximum share of number of instances in the total dataset. This reflects the further potential of SVM method to give even higher results in cases of ECG datasets with more uniform distribution, thereby ensuring better training of the system.

Table.1 Comparison of performances

Parameters	SVM	Naïve Bayes
Avg. Precision	**0.96**	**0.93**
Avg. Recall	**0.96**	**0.94**
Avg. F1-Score	**0.96**	**0.94**
Avg. Support	**412**	**412**
Accuracy	**0.96**	**0.94**
Weighted Average	**0.96**	**0.94**

VI. Performance analysis

For the performance analysis of the classifier, accuracy and F1 score can be calculated

$$\text{Accuracy} = \frac{(TP+TN)}{(P+N)} \quad \text{---------- (1)}$$

In the equation (1), True Positive (TP) which represents data correctly classified and True Negative (TN) which represents data incorrectly classified. P and N are the classes true or false to which classification of data points are done. Accuracy having higher value can be considered that the model has predicted larger members to a class correctly. But the members that are incorrectly classified also come into the picture that is why F1 score is calculated.

$$\text{F1 score} = 2* \frac{precision*recall}{precision+recall} \quad \text{---------- (2)}$$

The harmonic mean of precision and recall gives a better measure for incorrectly classified data. It is based on false negatives and false positives.
F1 score of a perfect model is 1 and that of a failed model is considered as 0. In the equation (2) the value of precision is the fraction of True positive values to the sum of True Positives and False Positive values whereas recall is the ratio of True Positive values and the sum of True Positives and False Negatives.

$$\text{RECALL} = \frac{TP}{TP+FN} \quad \text{---------- (3)}$$

Recall should ideally be 1 (high) for a good classifier. Recall becomes 1 only when the numerator and denominator are equal i.e. TP = TP +FN, this also means FN is zero. As FN increases the value of denominator becomes greater than the numerator and recall value decreases (which we don't want).

$$\text{PRECISION} = \frac{TP}{TP+FP} \quad \text{---------- (4)}$$

Precision should ideally be 1 (high) for a good classifier. Precision becomes 1 only when the numerator and denominator are equal, TP = TP +FP, this also means FP is zero. As FP increases the value of denominator becomes greater than the numerator and precision value decreases (which we don't want).

Table. 2 Performance Analysis

CLASSIFICATION REPORT	PRECISION	RECALL	F1-SCORE	SUPPORT
F	**0.94**	**0.95**	**0.94**	**435**
N	**1.00**	**1.00**	**1.00**	**356**
VEB	**0.95**	**0.94**	**0.94**	**445**
ACCURACY			**0.96**	**1236**
MACRO AVG	**0.96**	**0.96**	**0.96**	**1236**
WEIGHTED AVG	**0.96**	**0.96**	**0.96**	**1236**

VII. CONCLUSION

We propose an arrhythmic classification framework with SVM Algorithm. We conclude the cardiac Arrhythmia is detected. We could classify multiclass arrhythmia categories, including AFIB, with high accuracy using a model with high classification performance. Because a model trained on data with a large number of classes is fit to a classification type with a large number of classes, the classification performance of a relatively small type is reduced. Finally, our proposed model has the best generalization ability compared to other deep learning models when using few-shot learning and an independent database. Consequently, the proposed model can be useful for long-term ECG monitoring using single-lead wearable devices in clinical settings in the future.

REFERENCES

[1] Machine Learning based Cardiac Arrhythmia detection from ECG signal IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821-1

[2] ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm, 978-0-7695-3886-0/09 $26.00 © 2009 IEEE

[3] SVM based Methods for Arrhythmia Classification in ECG 978-1-4244-9034-/10/$26.00©2010 IEEE

[4] Arrhythmia Discrimination Using Support Vector Machine

[5] Machine Intelligent Diagnosis of ECG for Arrhythmia Classification Using DWT, ICA and SVM Techniques, 978-1-4673-6540-6/15/$31.00 ©2015 IEEE

[6] Classification of Cardiac Arrhythmia using Kernelized SVM, IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

[7] A. B. de Luna, P. Coumel, and J. F. Leclercq, "Ambulatory sudden cardiac death: Mechanisms of production of fatal arrhythmia on the basis of data from

157 cases," Amer. Heart J., vol. 117, pp. 151–159, 1989. CA: Wadsworth, 1993, pp. 123–135.

[8] J. F. Leclercq, P. Coumel, P. Maison-Blanche, B. Cauchemez, M. Zimmermann, F. Chouty, and R. Slama, "Mise en evidence des mecanismes d ´ eterminants ´ de la mort subite," Arch Mal Coeur, vol. 79, pp. 1024–1033, 1986.

[9] 3 H. A. Fozzard, "Electromechanical dissociation and its possible role in sudden cardiac death," J. Amer. Coll. Cardiol., vol. 5, pp. 31B–34B, 1985.

[10] 4 A. K. Dawson, A. S. Leon, and H. L. Taylor, "Effect of pentobarbital anesthesia on vulnerability to ventricular fibrillation," Amer. J. Physiol., vol. 239, pp. H427–H431, 1980..

[11] T. L. Wenger, F. E. Harrell Jr., K. K. Brown, S. Lederman, and H. C. Strauss, "Ventricular fibrillation following canine coronary artery reperfusion: Different outcomes with pentobarbital and alpha-chloralose," Can. J. Physiol. Pharmacol., vol. 62, pp. 224–228, 1984.

[12] 6 F. H. De Jonge, E. A. Bokkers, W. G. Schouten, and F. A. Helmond, "Rearing piglets in a poor environment: Developmental aspects of social stress in pigs," Physiol. Behavior, vol. 60, pp. 389–396, 1996..