International Journal of Computer Applications (0975 – 8887)
Volume 38– No.1, January 2012

Analysis and Classification of Cardiac Arrhythmia using ECG Signals

Pooja Bhardwaj
Department of EI&CE,
Govt. Engineering College
Bikaner-334001, India

Rahul R Choudhary
Department of EI&CE,
Govt. Engineering College
Bikaner-334001, India

Ravindra Dayama
Department of EI&CE,
Govt. Engineering College
Bikaner-334001, India

ABSTRACT
ECG is a graphical record of the electrical tension of heart and has established as one the most important bio-signal used by cardiologists for diagnostic purposes and further to adopt an appropriate course of treatment. The difficulties faced in interpretation of ECG signals forced researchers to study about automatic detection of cardiac arrhythmia disorders. The data analysis techniques using specific computer software could easily interpret complex ECG signals, predict presence or absence of cardiac arrhythmia. This provides real time analysis and further facilitates for timely diagnosis. In this paper, Support Vector Machine (SVM) technique, using LibSVM3.1 has been applied to ECG dataset for arrhythmia classification in five categories. Out of these five categories, one is normal and four are arrhythmic beat categories. The dataset used in this study is 3003 arrhythmic beats out of which 2101 beats (70%) are used for training and remaining 902 beats (30%) have been used for testing purpose. Total performance accuracy is found to be around 95.21 % in this case.

Key Words: SVM, arrhythmia, positive prediction

1. INTRODUCTION
In the present scenario, cardiovascular diseases have proved to be one of the major causes of casualties. There is strongly established that a pre-monitoring and a pre diagnostic of these cardiac arrhythmias by using an ECG tool can lead to avoidance of such casualties caused by arrhythmia. Arrhythmia is defined as any sort of disorder that takes place in normal rhythm of heart. Some arrhythmias like ventricular fibrillation are fatal and also can cause death of patient. The classification of arrhythmias is very important as some arrhythmias are severely fatal while others are not. For classification, various methods have been used. The detection of arrhythmia is an important task in clinical reasons which can initiate life saving operations. Quick availability of ECG signal from remote location and providing proper filtering circuit on time can help in analyzing the signal for arrhythmia[1].From early times several detection algorithms have been proposed, such as the sequential hypothesis testing [3], the threshold-crossing intervals [4]**,** algorithms based on neural-networks [6], and wavelets [7]. The classification of detected arrhythmias is also a research field of great interest. Several classification approaches have been presented in the literature viz wavelet analysis combined with radial basis function neural networks [10] and non-linear dynamic modeling [11]. In this paper the dataset has been collected from MIT-BIH. Features have been extracted using AcqKnowledge software (system software of BIOPAC). For classification of arrhythmias SVM is used.

2. METHODOLOGY
2.1 Dataset
Whole dataset has been collected from MIT-BIH arrhythmia database. The database consists of 48, two channel recordings. Annotation is provided along with the signal as shown in figure 1[2].

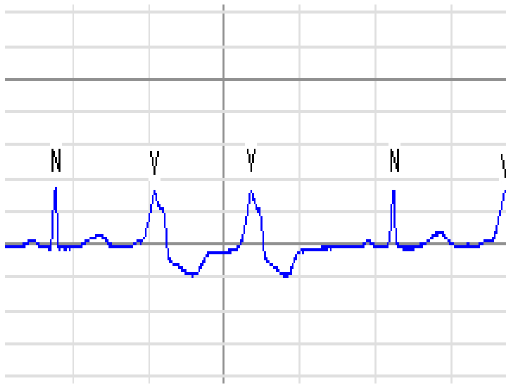

Fig 1: Annotated ECG waveform in MIT-BIH database [2]

These annotations are categorized in five different categories. Start and end of ventricular flutter and fibrillation constitute a specific category named “F”. Atrial, nodal and junctional escape beat are considered in category “E”. Ventricular premature beat and fibrillation beat, both produce irregular contraction of heart muscle and require medical emergency, so these have been kept in category “V”. Atrial, nodal and supraventricular premature beat shows irregular activity of atria and also conduction of impulses from SA node to AV node .Such beats are kept under category “A”. All remaining annotated beats like bundle branch block beats, fusion beats etc., that does not generate any serious situation, are kept under normal category and named “N”. These annotations are shown in table 1.

Table 1- Category based on annotation of Database [1]

Category	Description	Annotation
N	Normal	N, p, f, P, Q, ~, L, R, s, t, +, /
A	Atrial, Nodal and Supraventricular premature beat	A, a, **J, S**
V	Ventricular premature beats	V,F
E	Escape beats	E, n, j, e
F	Ventricular Flutter-fibrillation beats	[, !,]

2.2. Feature extraction

Arrhythmic waves are clinically classified on the basis of heart beat namely bradycardia (<60 bpm) and tachycardia (bpm >150). On the basis of place of origin, these are classified as atrial, ventricular and junctional arrhythmias. Peaks and segments of ECG wave change for different kind of arrhythmias. The arrhythmia classification by SVM requires generation of the input vectors. Filters are used for artifact removal.

Therefore, in this paper, the input vector fed to the classifier was determined using AcqKnowledge software. Nine features, namely: RR interval, P height, R height, heart rate, QT interval, ST interval, QRS width, corrected QT interval and PR interval, are obtained [15].

2.3 Classification

The 18 features extracted work as input to SVM classification. The original idea of SVM classification is to use a linear separating hyper plane to create a classifier. Given training vectors x_i , i = 1 . . . n, of length n, and a vector y_i defined as follows:

$$y_i = \begin{cases} 1 & if\ x_i\ is\ in\ class\ 1 \\ -1 & if\ x_i\ is\ in\ class\ 2 \end{cases} \quad \ldots\ldots\ (1)$$

The support vector technique tries to find the separating hyper plane with the largest margin between two classes, measured along a line perpendicular to the hyper plane.

Assuming all data are at least at distance of 1 from the decision boundary, the following two constraints follow for a training set (x_i, y_i):

$$(w^T x_i + b) \geq 1 \quad \text{if, } y_i=1 \quad \ldots\ldots\ldots(2)$$
$$(w^T x_i + b) \leq -1 \quad \text{if, } y_i=-1 \quad \ldots\ldots\ldots(3)$$

Equation 2 and 3 can be rewritten as

$$y_i\,(w^T x_i + b) \geq 1 \quad \ldots\ldots\ (4)$$

The equation 4 implies the instances to be some distance away from the hyper plane for better generalization. The decision boundary should be as far away from the instances of both classes as possible. Our objective is to maximize the margin, which is the distance between the hyper plane and the instances closest to it.

In order to maximize the margin, $\|w\|$ should be minimized. So the problem can be defined as

$$\text{Minimize } \frac{1}{2}\|w\|^2$$

$$\text{Subject to: } y_i\,(w^T x_i + b) \geq 1$$

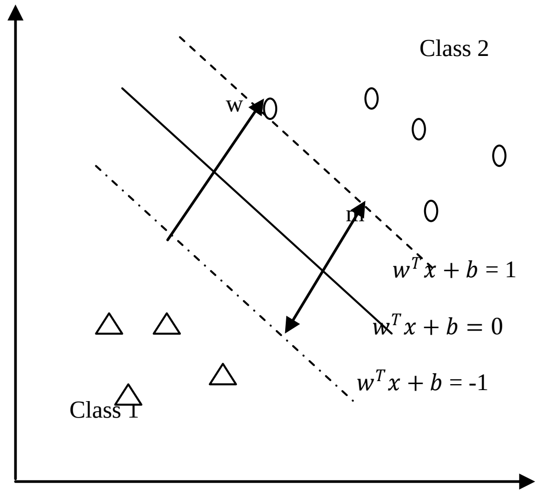

Fig 2: Linear Support Vector Machine [18]

This is a standard quadratic problem, and the solution involves constructing a dual problem where a Lagrange multiplier α_i is associated with every constraint in the primary problem

$$\text{Max } W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{n} \alpha_i \alpha_j\, x_i^T x_j$$

$$\text{Subjected to: } \alpha_i \geq 0, \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

w can be recovered as :

$$w=\sum_{i=1}^{n} \alpha_i y_i x_i$$

The solution suggests that many of the α_i are near to or equal to zero. W is a linear combination of a number of data. x_i with non-zero α_i are called support vectors(SV). Examples closest to the hyperplane are support vectors and the decision boundary is determined only by the support vectors.

During testing, instead of using margin, we can calculate $g(x) = w^T + b$ and choose according to the sign of $g(x)$. Choose C_1 (class 1) if $g(x) > 0$ and C_2 (class 2) otherwise.

The matrix of 3003*18 is obtained from feature extraction which works as an input for support vector machine. For training, 70% data has been used and the remaining dataset is used for testing.

The performance of the SVM is sensitive to its kernel and parameters, so they must be chosen carefully when constructing a predicting model.

To measure classification accuracy and compare different classification methods 8 fold cross validation technique is used in the experiments that is the whole dataset is partitioned into 8 subsets. The 7 of the subsets are used as the training set and the eighth is used as test set. This process is repeated 8 times for each subset being the test set. Classification accuracy is the average of these 8 runs. This technique ensures that the training and the test sets are disjoint.

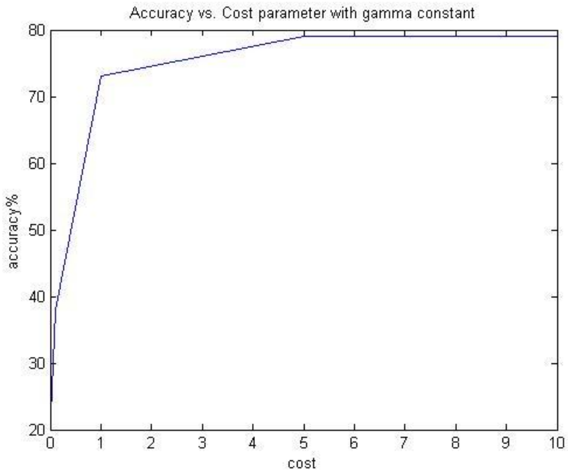

Fig 3: Plot of cost v/s accuracy

While using RBF kernel, the SVM has two parameters, namely C and gamma (γ). It is not known that which C and gamma are best. So, some kind of parameter search must be done. For this we plot graph between accuracy and cost keeping gamma constant as shown in Figure 2 and further accuracy v/s gamma by keeping C (cost) constant as shown in figure 3.

On analyzing the Figure 2 and Figure3, it is clear that the best C and gamma are found to be: **C=5, γ = 0.001**

3. RESULTS

Using 2101 beats for training and remaining beats for testing the results obtained, are shown in table 2.Out of 902 beats used for testing, only 864 beats are correctly classified. 38 beats are misclassified. For escape beat category, total beats classified are 11 out of which 8 are correctly classified and 3 beats, which lie in some other category, are classified as escape beats.

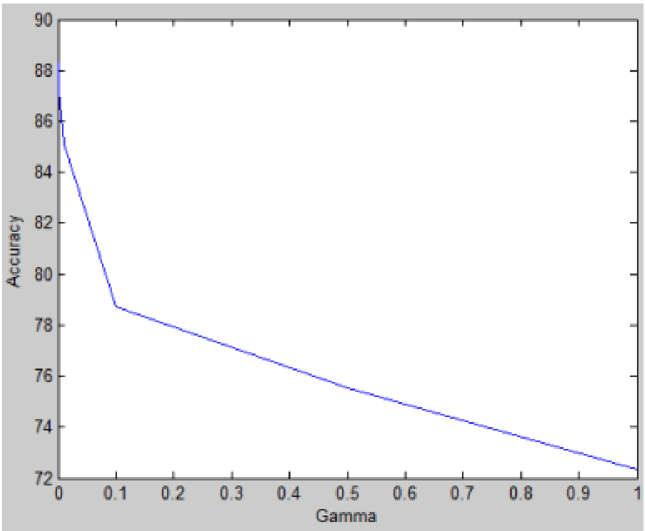

Fig 4: Plot of gamma v/s accuracy

Table 2: Result of Classification of database

Beat Category	N	A	V	E	F	TOTAL
Training beats	1763	85	206	18	29	2101
Testing beats	756	37	89	8	12	902
Correctly classified beats	740	31	74	8	11	864
Misclassified to this beat	15	10	9	3	1	38
Accuracy %	97.8	83.7	83.1	100	91.6	**95.21**

It is necessary to express the data in ways which are clinically relevant for ensuring the maximized usability. In order to communicate the results in a consistent manner, positive prediction and sensitivity are needed to be extracted [16]. These are represented as:

$$Sensitivity = \frac{No.of\ beats\ correctly\ classified\ in\ category}{total\ No.of\ beats\ \ classified\ in\ category}$$

$$Positive\ prediction = \frac{No.of\ beats\ correctly\ classified\ in\ category}{total\ No.\ of\ beats\ \ in\ category}$$

Table 3: Sensitivity and positive prediction of classification

Category	Positive prediction	Sensitivity
N	97.88%	98.01%
A	83.78%	75.64%
V	83.14%	89.15%
E	100%	72.72%
F	91.66%	91.66%

4. DISCUSSION

It is clearly shown in Table 3 that the classification based on SVM is very effective on classifying arrhythmic beats. However, in dataset, the numbers of normal intervals (2519 normal beats i.e. almost 84% of the total beat number) were very large as compared to the other categories (second category 4%, third 9.8%, fourth 0.9% and fifth 1.28%). The computed performance is high (almost 96%, because the achieved sensitivity and positive prediction are high for the first category i.e. 98.01% and 97.88% respectively) as compared with the results for the second and the third category (75.64% sensitivity and 83.78% positive prediction for the second category and 89.157% sensitivity and 83.14% positive prediction for the third).

For category 4 we obtained high positive prediction (100%) but low sensitivity (78.40%), which is clearly due to the high number of false negative (3 beats). For category 5 we obtained the best results, 91.66% for sensitivity and 91.66% for positive prediction. The absolute discrimination achieved among the arrhythmia categories and only 1 false negative is interacted. This is because of category 5 contains start, progress and stop of ventricular flutter-fibrillation beats that are easily detected due to their typical wave pattern.

5. COMPARISION

As compared to other feature extraction techniques like wavelet, AcqKnowledge software provides more number of features. Analysis and preprocessing are also simple with this software. As per the best of our knowledge, it is felt that the first time this software has been used for feature extraction. The accuracy we obtained is around 96% which is much more as compared to that obtained in A.U. Ozkaya et.al.[18]

6. CONCLUSION

We have proposed and developed a method for arrhythmia beat classification. The method is based on the 18 features, extracted from ECGs. Depending upon these features, classification of five beat categories, one for normal and four arrhythmic beats are made. The features were obtained using AcqKnowledge software is used for preprocessing and feature extraction. The proposed method has been evaluated using 3,003 beats. The MIT-BIH Arrhythmia database is used for classification. The accuracy depends mainly upon cost function and gamma function. In this study, we observed that when 70% data taken as training dataset and 30% data taken as testing dataset we achieved 95.21% total performance and high discrimination ability among the five categories has been developed.

7. REFERENCES

[1] Pooja Bhardwaj,Rahul R Choudhary, Ravindra Dayama, "Web based design for collection and filtering of ECG signal", International journal of computer applications (0975-8887) Volume 34 November 2011.

[2] MIT - BIH Arrhythmia database, available: http://www .physionet.org/physiobank/database/mitdb/2010.

[3] Thakor NV, Zhu YS, Pan KY. Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. IEEE Trans Biom Eng 1990; 37:837-843.

[4] Clayton RH, Murray A, Campbell RWF. Comparison of four techniques for recognition of ventricular fibrillation of the surface ECG. Med Bio Eng Comp 1993; 31111-1J7

[5] Clayton RH, Murray A, Campbell RWF. Recognition of ventricular fibrillation using neural networks. Med Bio Eng Comp 1994; 32:217-220.

[6] Yang TF, Device B, Macfarlane PW. Artificial neural networks for the diagnosis of atrial fibrillation. Med Bio Eng Comp 1994

[7] Khadra L, AI-Fahoum AS, AI-Nashash H. Detection of life-threatening cardiac arrhythmias using wavelet transformation. Med Bio Eng Comp 1997;35:626-632.

[8] Minami K, Nakajima H, Toyoshima T. Real-time discrimination of ventricular tachyarrhythmia with Fourier transform neural network. IEEE Trans Biom Eng

[9] Ham FM, Han S. Classification of cardiac arrhythmias using fuzzy ARTMAP. IEEE Trans Bio Engg.

[10] Osowski S, Linh TH. ECG beat recognition using Fuzzy Hybrid Neural Network. 2001.

[11] Docur Z, Olmez T. ECF beat classification by a hybrid neural network. Comp Meth Prog Biomed 2001; 66: 167-181.

[12] Marriot, G.S. Wagner, Practical Electrocardiography, 11th edition (Williams & Wilkins), 2008.

[13] Nello Cristianini and John Shawe-Taylor: "An Introduction to Support Vector Machines" Cambridge University Press 2000

[14] ECG analysis for resting 12 –lead ECG physician's guide, QRS diagnostic, 2006

[15] AcqKnowledge 4 Software Guide, BIOPAC systems, Inc.

[16] Hampton John R., 2003, The ECG Made Easy, sixth edition (chirchill livingstone)

[17] PS 148-Automated ECG Analysis, Available: http://AppNotes/app148proecg/autoecgpro.html, 2006

[18] A.U. Ozkaya. Intelligent arrhythmia classifier using support vector.