Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Cardiac Arrhythmia Classification using Optimal Feature Selection and K-Nearest Neighbour Classifier

Ritaban Kirtania, Kalyani Mali
Department of Computer Science & Engineering,
University of Kalyani, Kalyani, West Bengal, India

Abstract— In this research paper a new approach for cardiac arrhythmia classification technique is being proposed. This method is based on an optimal feature selection procedure from electrocardiogram (ECG) signal and K-NN classifier. Here, we are solely concentrating on the area of feature selection to get a classification rate more appropriately. Firstly, each and every feature is selected from ECG signal based on time-voltage of P, Q, R, S, and T waves. Then the optimal features of an ECG, based on individual features that have identified every cardiac arrhythmia baet, are being selected. And by using K-NN classifier the arrhythmia classification is done. It is mention worthy that these optimal features are being used to create a pattern. The proposed method obtains a very good classification rate (98.87%) rather than other feature selection procedures and other classifiers like NN classifier.

Keywords— Electrocardiogram (ECG), Cardiac arrhythmia, Daubechies wavelet, Feature reduction, K-Nearest Neighbor classification (K-NN).

I. INTRODUCTION

In Computer Science, cardiac arrhythmia classification is a complex pattern recognition task. However identifying arrhythmia bits by analysing an ECG signal is the most effective and available method in the diagnosis of cardiac arrhythmias. It can provide most accurate classification results by selecting optimal features of an arrhythmia bit. Time and voltage reading of P, Q, R, S, and T waves are the most important aspects of an ECG signal. The morphology and the time durations of these bits are being used to produce optimal features of an ECG bit. Every pattern is produced by extracting these features. After constructing the pattern set, use of K-NN classifier provides a very good classification rate. The dataset, on which feature extraction and selection are applied, is MIT-BIH Arrhythmia. It is one of the most popular and useful datasets.

To classifying cardiac arrhythmia, various methods have been proposed. Among them [1] is provide a method which includes Support Vector Machine and reduced feature using Genetic Algorithm. Another work [2] is proposed to use Mamdani fuzzy inference system on the output of 3 different classifier, Fuzzy K-NN, MLP with Gradient Descent and momentum Back propagation, and MLP with Scaled Conjugate Gradient Back propagation. In [4] a hybrid system is proposed which includes Fuzzy K-NN, MLP, and Fuzzy FIS. Our proposed method is more simple then the above mentioned process, and also produces a better classification rate.

The rest of the paper is structured as follows. Brief description about ECG and arrhythmia is given in section 2, feature extraction is provided in section 3. Proposed method is provided in section 4, Section 5 describes the experimental results. And Conclusion and future work are given in section 6. References are at the end.

II. ECG & ARRHYTHMIA

In this section, a brief description of ECG signal analysis and types of cardiac arrhythmia are being reviewed.

A. ECG

An ECG (electrocardiogram) signal is the representation of electrical activity of the heart muscle in respect of time. It provides important data's that are being produced by the muscle contraction of heart. Every ECG signal contains continuous electrical activity of heart in respect of time, which is called ECG beats. Every ECG beat contains P, Q, R, S, T waves. Every wave has its own time duration, and signal amplitude. Fig. 1 provides the features of a normal ECG beat.

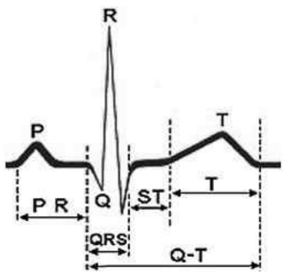

Fig. 1 An ECG beat with P, Q, R, S, and T wave.

B. Arrhythmia

Abnormality or irregularity of the rhythm of heart is known as Cardiac Arrhythmia. In cardiac arrhythmia heart rhythm can be irregular or very slow, or very fast, or it may change the morphology of P, Q, R, S, T waves, and may produce abnormal heart bit. We can classify a heart bit into 4 arrhythmia classes. They are Left bundle branch block (LBBB), Right bundle branch block (RBBB), Atrial premature contraction (APC), Premature ventricular contraction (PVC). There is another class, which is Normal ECG bit.

III. FEATURE EXTRACTION

In this section, a brief description about automatic feature extraction procedure is being provided.

A. Automatic Feature Extraction Procedure

After noise reduction feature extraction process has been executed, we have automatically extracted all features of an ECG beat, based on time-voltage of P, Q, R, S, and T waves. Firstly, we have extracted P, Q, R, S, T peak and the occurring time by the given algorithm:

Input: Say ECG signal = Sig, and ECG time signal = Time.

Output: Extracted amplitude signal of P, Q, R, S, T with time.

Step 1: R = Sig(i), where Sig(i) > Sig(i-1) and Sig(i) > Sig(i+1) and Sig(i) > (Max(Sig)-Avg(Sig))*2/3.

Step 2: P = Max(Sig) between Time(R)-0.15 sec to Time(R)-0.20 sec.

Step 3: Q = Min(Sig) between Time(R)-0.12 sec to Time(R).

Step 4: S = Min(Sig) between Time(R)+0.01 sec to Time(R)+0.10 sec.

Step 5: T = Max(Sig) between Time(R)-0.07 sec to Time(R)-0.20 sec.

B. All Feature

After extracting time-signal of all the waves, we extracted all 37 features that an ECG signal can have. All the features are described in the TABLE I.

TABLE I ALL 37 FEATURES OF AN ECG SIGNAL

Feature Group	Indices	Feature description
Inter-beat Intervals	1-3	Pre-RR, Post-RR, Avg-RR
	4,5	Pre-TP, Post-PT
	6-8	Pre-PP, Post-PP, Avg-PP
Intra-beat Intervals	9-12	PQ, PR, PS, PT
	13-15	QR, QRS, QT
	16,17	RS, RT
	18,19	Avg-PQ, Avg-QRS
Morphological Amplitude	20-24	P, Q, R, S, T
	25,26	Max(P-Q), Max(Q-R)
	27-31	Avg-P, Avg-Q, Avg-R, Avg-S, Avg-T
Morphological Areas	32-34	P, QRS, T
	35-37	Avg-P, Avg-QRS, Avg-T

All 8 inter beat intervals, 11 intra beat intervals, 12 morphological amplitudes and 6 morphological areas are been taken as features. From these 37 features we derived a set of optimal features that are use to create the pattern.

IV. PROPOSED METHOD

The proposed method is to create an optimal subset from the feature set based on those features that can uniquely identify an arrhythmia beat that uses K- NN classifier to classify all other beats. This optimal subset of feature is a pattern which can identify individual beats. In this section we will describe the procedure to get the optimal features from the feature set, and create the pattern. And also we will describe the K-NN classifier to classify arrhythmia beats.

A. K-Nearest Neighbour Classifier

K-Nearest Neighbour (K-NN) classification technique provides a non-parametric classification method which classifies unknown pattern by measuring the distance from a given training Pattern set. First, it calculates the distance of the unknown pattern from all the training pattern set and arranges all the differences by ascending order. Then it selects, at first, K differences with its respective classes. The maximum no. of occurring class, say X, is being taken as the output. It classifies the unknown pattern into class X. K-NN classifier can be used for classification and regression also.

B. Optimal Feature Selection

We have collected all 37 features from an ECG signal. From these 37 features, the optimal feature set has been created. Our experimental studies have identified those features, which are uniquely identify every individual arrhythmia beats. We have used these unique features to construct our optimal feature set.

Broadly we can classify arrhythmia in 4 types. They are Left bundle branch block (LBBB), Right bundle branch block (RBBB), Atrial premature contraction (APC), and Premature ventricular contraction (PVC). The main characteristic features of LBBB are, QRS duration should be greater than 120 ms, S wave should be dominant, and produces multiple R waves denoted as rr. RBBB causes QRS duration more than 120 ms, small r wave between P and Q wave. And it produces a wide, slurred S wave. APC produces abnormal P wave and fluctuation of morphology in P wave. And PVC produces negative R wave, QRS duration more than 120 ms, morphology of T wave should be wide, and next TP duration should be short. Apart from that, a Normal Sinus Rhythm (Normal) produces QRS complex less than 100 ms, normal P wave, and constant PR interval. Optimal feature description is provided in TABLE II.

TABLE II 9 UNIQUE FEATURES OF AN ECG SIGNAL

ECG type name	**Unique features Name**
LBBB	QRS, S, rr
RBBB	QRS, r, S
APC	Morphology of P
PVC	R, QRS, T, TP(Next)
Normal	QRS, P, PR
Total	QRS, S, rr, r, Morphology of P, R, T, TP(Next), PR

From our experimental studies we have created the optimal feature set which consists of 9 features. We have created the training pattern by using these 9 features. Using K-NN classifier and these training patterns, unknown ECG beats can be classified with an accuracy of 98.87%.

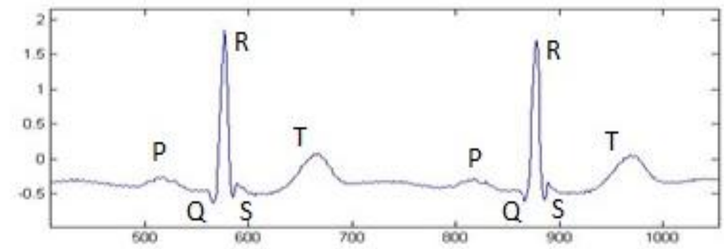

Fig. 4 Normal Beat

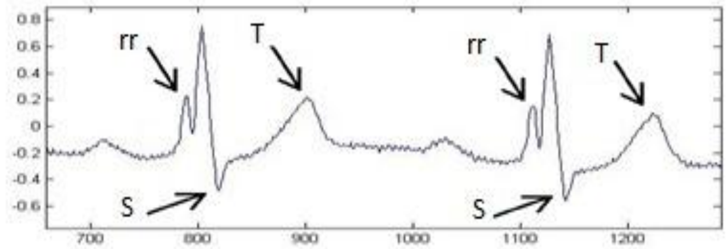

Fig. 5 LBBB

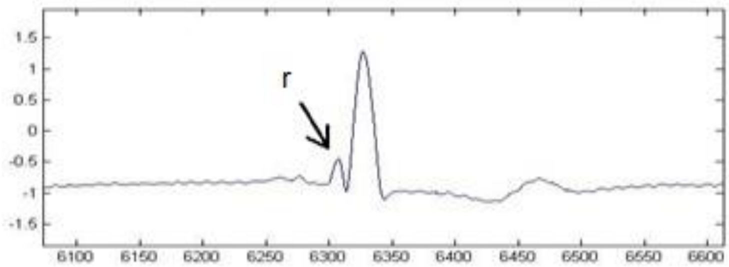

Fig. 6 RBBB

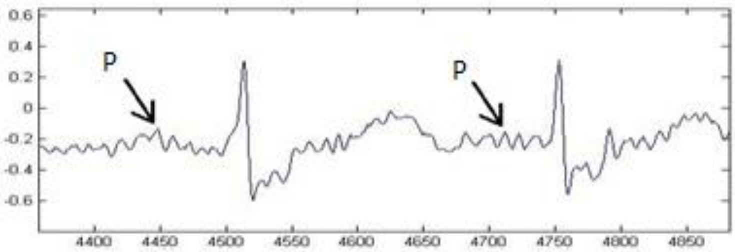

Fig. 7 APC

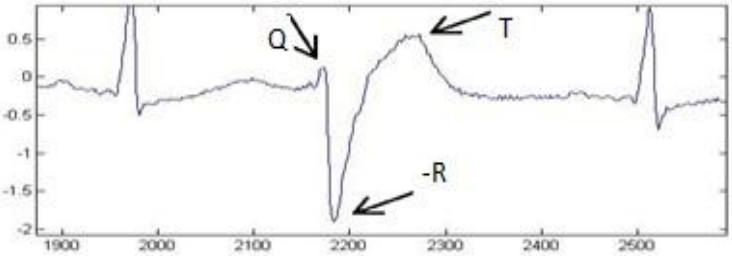

Fig. 8 PVC

V. EXPERIMENTAL RESULTS

A. Database Description

Dataset, on which our experiment has taken place, is one of the most important aspects of our research. MIT-BIH Arrhythmia database [6] is being used for this research paper. It is one of the most useful and important databases, currently available. Most of the research papers used this particular arrhythmia database for the evaluation of their proposed work.

MIT-BIH arrhythmia database has a total of 48 records, obtained from 47 subjects. Each record is an ECG signal with duration of slightly over 30 minutes. Each was recorded by using two channels, Modified Limb Lead II (mill) and Modified lead VI (VI).

We have used total of 5 records from MIT-BIH arrhythmia database for our research work. These are 103, 111, 124, 200, and 232. Beat selection from these records are in TABLE III.

TABLE IIi MLIST OF SAMPLE DATA TAKEN FROM MIT-BIH DATABASE

Types of beat	Normal	LBBB	RBBB	APC	PVC
ECG Record no.	103	111	124	232	200
No. Of beat taken	228	190	197	185	143

We have used MATLAB 13.b as platform to evaluate our experimental work. It provides better environment, with respect to time and effort.

B. Noise Reduction

Before feature extraction, Daubechies D4 ([7]) wavelet is being applied on every ECG signal for reducing the noise. Due to power line interference and instrumentation error, noise may effect on an ECG signal. So that noise reduction is a mandatory task. Other high frequency noise removal techniques are also being applied, but the most suitable is Daubechies D4 wavelet transformation, based on PSNR value. The Peak Signal to Noise Ratio (PSNR) is the ratio between possible highest power of a signal and the strength of corrupting noise. The expression of PSNR is given bellow.

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (1)$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (2)$$

Eq. 2 Calculation of PSNR value

The table of PSNR values for every noise removal technique is given in TABLE IV using Eq. 1 and Eq. 2. And Fig. 2 and Fig. 3 show the noise reduction result.

TABLE IV LIST OF PSNR VALUES FOR NOISE REMOVAL ALGORITHMS

Noise reduction algorithm	**PSNR**	**Noise reduction algorithm**	**PSNR**
D4	53.48	Haar	47.23
D6	50.89	Avg-3	43.80
D8	49.80	Avg-5	36.13
D10	49.74	Avg-7	39.39
D12	50.34	Avg-9	28.19
D14	51.10	Med-3	48.80
D16	51.52	Med-5	42.62
D18	52.56	Med-7	36.43
D20	51.78	Med-9	32.01

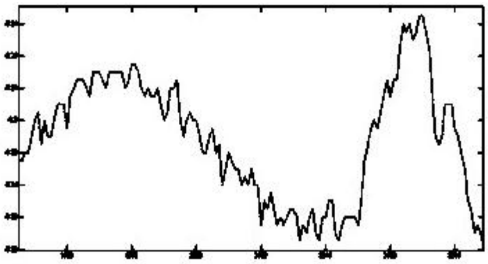

Fig. 2 Before Noise reduction.

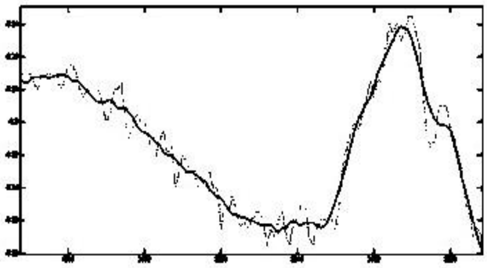

Fig. 3 After Noise reduction.

C. Restlts

MIT-BIH arrhythmia database is used for the evaluation of our proposed method. Both training set and unknown data set are selected from this database. Using proposed method, the optimal features are being provided to the Nearest Neighbour classifier and K-Nearest Neighbour classifier. Both produce a very good classification rate of 97.74% and 98.87% respectively.

TABLE V CLASSIFICATION RESULTS

Classifier	**Normal**	**LBBB**	**RBBB**	**APC**	**PVC**	**Over all**
NN	100%	94.44%	94.84%	100%	100%	97.74%
K-NN	100%	96.67%	97.84%	100%	100%	98.87%

With respect to [1], our proposed method achieved more than 5% accuracy. Reference [1] achieved a classification rate of 93.46% while our proposed method achieved 98.87%.

VI. CONCLUSIONS

In this research paper a new method for classifying cardiac arrhythmia has been proposed, by using K-NN classifier and optimal feature section from electrocardiogram signal. This method works efficiently with respect to other classifier like Nearest Neighbour classifier. It produces more accurate results than other classification method. Unnecessary features are reduced to produce better classification rate. For future works, more accurate classifier as well as feature will be used to get more accurate result.

ACKNOWLEDGMENT

This research work has been fully supported by Dept. of CSE, University of Kalyani where all the guidance and resources required for this work are used and DST, PURSE scheme. This guidance and support is gratefully acknowledged.

REFERENCES

[1] Jalal A. Nasiri, Mahmoud Naghibzadeh, H. Sadoghi Yazdi, Bahram Naghibzadeh, "ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm", Computer Modeling and Simulation, 2009. EMS '09. Third UKSim European Symposium, 25-27 Nov. 2009, IEEE, DOI: 10.1109/EMS.2009.39.

[2] Oscar Castillo, Patricia Melin, Eduardo Ramirez, Jose Soria, "Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system", Volume 39, Issue 3, 15 February 2012, doi:10.1016/j.eswa.2011.08.156.

[3] M.G. Tsipouras, D.I. Fotiadis, D. Sideris, "An arrhythmia classification system based on the RR-interval signal", Artificial Intelligence in Medicine (2005).

[4] Ramirez, E. ; Castillo, Oscar ; Soria, J. "Hybrid system for cardiac arrhythmia classification with fuzzy k-nearest neighbors and Multi Layer Perceptrons combined by a fuzzy inference system", Neural Networks (IJCNN), The 2010 International Joint Conference on 18-23 July 2010, Barcelona, IEEE, DOI: 10.1109/IJCNN.2010.5597548.

[5] Zhancheng Zhang, Jun Dong, Xiaoqing Luo, Kup-Sze Choi, Xiaojun Wu, "Heartbeat classification using disease-specific feature selection.", Computers in Biology and Medicine 46 (2014).

[6] MIT-BIH Arrhythmia Database.

[7] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis, IEEE Trans. Inf. Theory, 36 (5), pp.961-1005, 1990.

