ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 4, Issue 6, June 2015

Taxonomy of Electrocardiograph Signals

Amandeep Singh Khanna, Er. Priyadarshni

***Abstract*— Electrocardiograph signal also known as ECG Signal is randomly variable signal recorded by cardiac experts to acquire information regarding cardiac state of patient. This information can be permanent or may vary randomly according to the health of person. An approach is used in which cardiac signals are compared to find disorders. This paper develops an approach with the amalgamation of GLCM, SVM and K-NN. The proposed technique extract features from cardiac signals using GLCM and categorize them using machine learning algorithm. SVM has further kernel functions; which make classes of different ECG signals. Categorization is made on the basis of Mean, Skewness, Kurtosis, Energy and Correlation. After categorization; it has been observed that SVM gives 100% accuracy as compared to K-NN. Software used is MATLAB2011a.**

***Index Terms*-: Cardiac, SVM, K-NN, Kernel, GLCM.**

I. INTRODUCTION

ECG signal is studied to analyze the cardiac diseases. It is used to measure regularity of heart rate. They are periodic and oscillatory in nature which is produced by polarization and depolarization of atria and ventricles. Each beat of ECG signal has distinct and characteristic shape. But ECG signal gets altered from different types of noises and artifacts. These noises or artifacts are required to be removed before further processing.

Over the past few decades several techniques have been proposed for categorizing such as SVM, K-NN, Fuzzy logic, genetic algorithm and neural networks are some popular classification methods[11]. Each approach has its own advantages and disadvantages. Automatic categorization of cardiac signals has gained importance over few decades. Automatic ECG signal categorization is essential for long recordings.

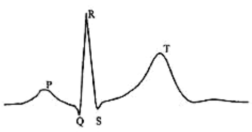

Fig 1: One Cycle of ECG Signal

Manuscript received June, 2015.

***Amandeep Singh Khanna**, Electronics & Communication Engineering,, LCET,Katani Kalan, India.*

***Er. Priya Darshni**, Electronics & Communication Engineering,, LCET. Katani Kalan, India*

II. PROPOSED APPROACH

The proposed method uses the following machine learning algorithms: Grey-Level Co-occurrence matrix (GLCM), K- Nearest neighbor (K-NN) and Support Vector Machine (SVM). It works in two phases: Feature Extraction and its categorization. Researchers have used cardiac signal as database having total 60 signals as training data. Out of these; 17 are normal and 43 are abnormal cardiac signals of different patients. Similarly to test the learning algorithm data set used consists of total 18 signals, out of these; 4 are normal and 14 are abnormal cardiac signals. First phase is to extract valuable features from cardiac signals using GLCM and second phase is to categorize cardiac signals based on those features using K-NN classifier and different kernel functions of SVM classifier into normal and abnormal.

III. FEATURE EXTRACTION

In research; feature is such important part of any data which can easily describe the whole dataset if extracted carefully. It is supposed to be a relevant part which can be used to solve any computational task for any desired application. Feature extraction is a process used to simplify the amount of resources required to describe a large set of data perfectly. Basically it is a process of successful extraction of useful information from signal by discarding redundant information. Features extracted for this research are: Mean, energy, correlation, Skewness and Kurtosis.

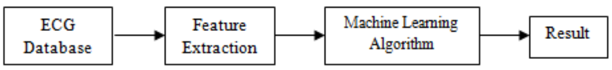

Fig 2: Block Diagram for Proposed methodology

IV. CATEGORIZATION OF FEATURES

Categorization refers the analysis of the properties of an image. Depending upon the analysis; the dataset is further referred into different groups. Input features are categorized as 0 and 1.The process of categorization is divided into two phases: training phase and testing phase. In training phase; known data is given and in testing phase; unknown data is given. Categorization is done by using learning algorithm after training. Algorithms used are K-NN and SVM with different kernel functions.

A. Support Vector Machine(SVM)

For the classification of linear and non-linear data, Support Vector Machine (SVM) classifier is used. SVM uses anon-linear mapping to transform the original training data into a higher dimension:

1. With the new dimensions, it searches for the linear optimal separating hyperplane i.e. decision boundary.

2. Data between two classes always separated by hyperplane.

3. SVM finds hyperplane using support vectors and margin defined by support vectors.

Classifiers are used to differentiate or distinguish between two or more classes by drawing a line or plane between train data points. There are number of classifiers like Perceptron, KNN, SVM etc. used to classify linear and non-linear data.

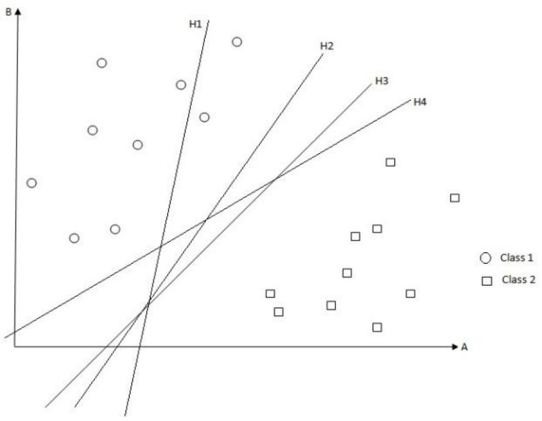

Fig 3: Possible Hyperplanes

Any one hyperplane or decision boundary from H1, H2, H3 or H4 can choose, but question arises here; which would be the best? To find the best hyperplane, find the margin of every possible hyperplane. Best margin is selected based on maximum margin width.

The dotted lines drawn parallel to the separating line or decision boundary; mark the distance between the dividing line and the closest vectors to the line. The distance between the dashed lines is called margin (M). The vectors or points those constraints the width of margin are called Support Vectors.

The classification of unknown instances 'x' is calculated as a sum over all similarities 'x' having support vectors (x_i). This similarity measure is the dot-product between x and x_i. In equation (1), check how the support vector coefficients ($\boldsymbol{\alpha}$) regulate the contribution of a particular support vector. The summed combination of similarity ($x_i . x$) and support vectors strength ($\alpha_i y_i$) will determine the output for a single support vector.

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i (x_i . x) + b \qquad (1)$$

where 'l' being the number of support vectors and y_i the label or class of the corresponding support vector.

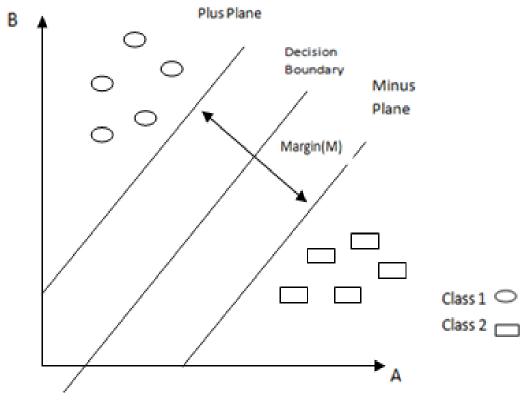

Fig 4: Optimal hyperplane with margin linear classification

NON LINEAR SVM:

But datasets or data points are always not separated by drawing a straight line between two classes.

For example the data points in the following figure can't be separable by using both SVM's discussed cases.

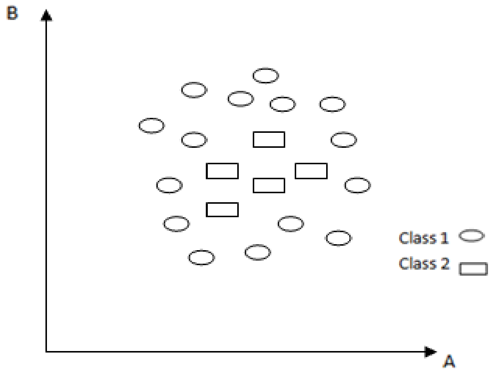

Fig.5: Data points non-linear separable

The Solution for this type of problem is to map the input or gene space data or data points to the feature space and then use classifier in feature space to draw hyperplane between two datasets to differentiate them from each other.

To solve the above given problems, kernel functions have been introduced:

- Solve the computational problem of working with many dimensions.
- Can make it possible to use infinite dimensions-efficiently in time and space.
- Enforcing largest margin permits good generalization.

It can be shown that generalization in SVM is a function of the margin, independent of the dimensionality.

KERNEL FUNCTIONS

Kernel functions are used in many applications as they provide a bridge from linearity to non-linearity for algorithms which can be expressed in terms of dot products.

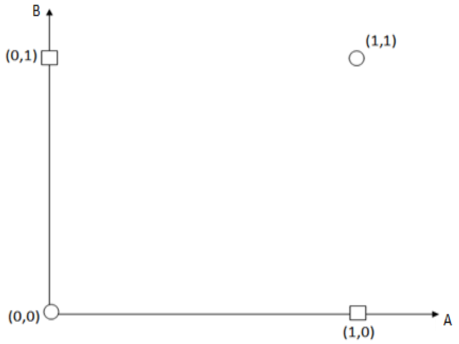

Fig 6: XOR gate data points in 2D

Consider an example of XOR gate whose data points are non-linear and cannot be classified using Simple SVM or by drawing a straight line. The figure 6.14 draw XOR gate for 2 bits, from figure we easily observe data points are not separable. So here is need for kernel function to map these two dimensional data points on three dimensional feature spaces. So after using kernels a cube like three dimensional feature space has been drawn where data points are now linearly separable.

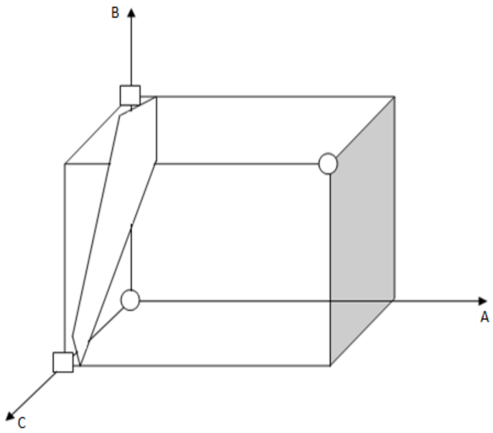

Fig 7: XOR data points in 3D (actual 3D)

By using kernel function, map the two-dimensional input space data points which have only two co-ordinates (A, B) now shows in figure 6.15 the same data points in three dimensional feature space and now having three co-ordinates (A, B, C) and clearly classified.
XOR problem solved by using MATLAB 2011a software and we display result figures below with and without use of kernels. Note: as image is only of two dimensions so here classified image shown only in two dimensions but on background or in actual it is three dimensional.

B. K-Nearest Neighbor (K-NN)

K nearest neighbors is an algorithm which records all the available features and classifies new features based on distance functions. The purpose of the k Nearest Neighbor (K-NN) algorithm is to use a database in which the data points are separated into several separate classes to predict the classification of a new sample point. In this training feature are stored as vectors in multi dimensional space along with class labels where; *k* is a user-defined constant, and an unlabeled vector is classified by assigning the label which is most frequent among the *k* training samples nearest to that query point.

It is a type of instance based and also simplest of all machine learning algorithms. In K-NN; function is locally approximated. An object is classified by majority vote of its neighbors, with the object being assigned to class most common among its K nearest neighbor. Euclidean Distance method is used by researchers to calculate the approximate distance between Query points and the value considered for k=1.

B. RESULT

After categorization; it has been accomplished that for Mean and Kurtosis, K-NN gives accuracy of 95.83 %, on the other hand; SVM gives 100% accuracy for the same. Execution time is more compared with other features. while using SVM with other kernel functions, when signals are categorized with Skewness and Kurtosis, Energy and Correlation; accuracy is less compared to Mean and Kurtosis.
On the other hand; while comparing the performance of K-NN and SVM for same features it has been observed that SVM gives better performance than K-NN.

Below table gives the combined result for extracted features with their performance by learning algorithms in terms of accuracy and execution time.

Table1: Overall result for feature categorization with K-NN and SVM

Features	Classifier	Kernel function	Accuracy	Time(sec)
Mean, Kurtosis	K-NN		95.83	0.8157
Skewness, Kurtosis	SVM	Quadratic	66.67	3.9784
		Linear	83.33	4.2027
		RBF	66.67	4.119
		Polynomial	66.67	4.582
Energy, Correlation	SVM	Quadratic	66.67	3.8374
		Linear	20.83	3.8627
		RBF	66.67	3.8699
		Polynomial	79.16	4.371
Mean, Kurtosis	SVM	Quadratic	100	4.3696
		Linear	100	4.4427
		RBF	100	5.3754
		Polynomial	100	4.0506

Graphical representation of classification results for different features with different kernel functions of SVM classifier.

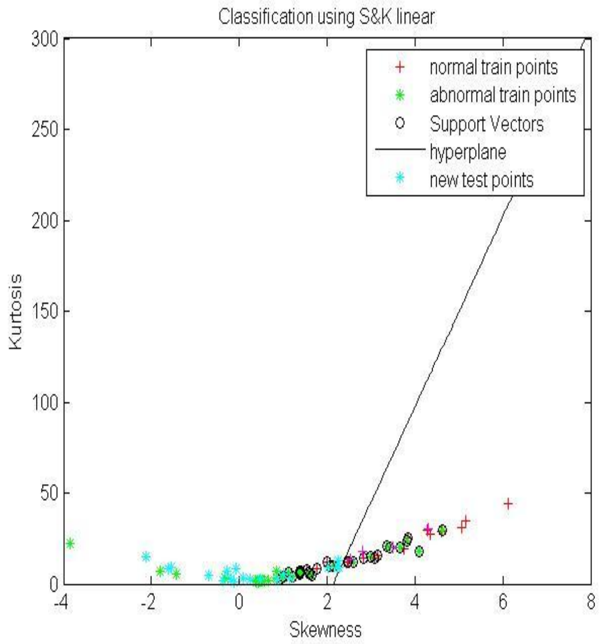

Fig 8: Graph shows the hyperplane constructed between normal and abnormal test points using SVM- linear for Skewness and Kurtosis

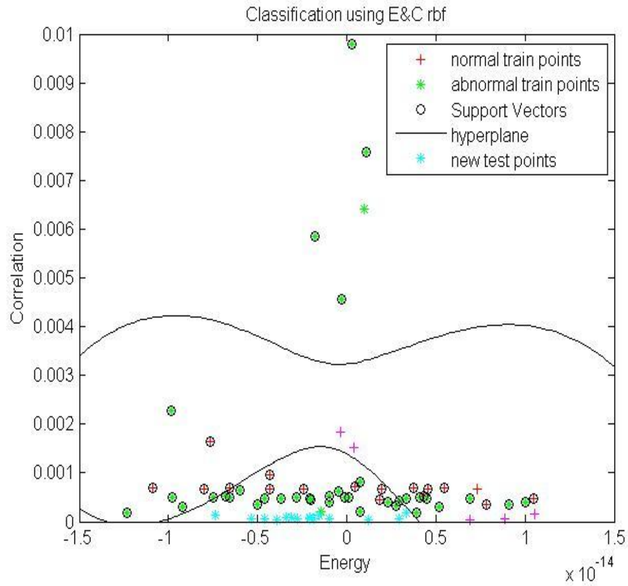

Fig 9: Graph shows the hyperplane constructed between normal and abnormal test points using SVM- RBF for Correlation and Energy

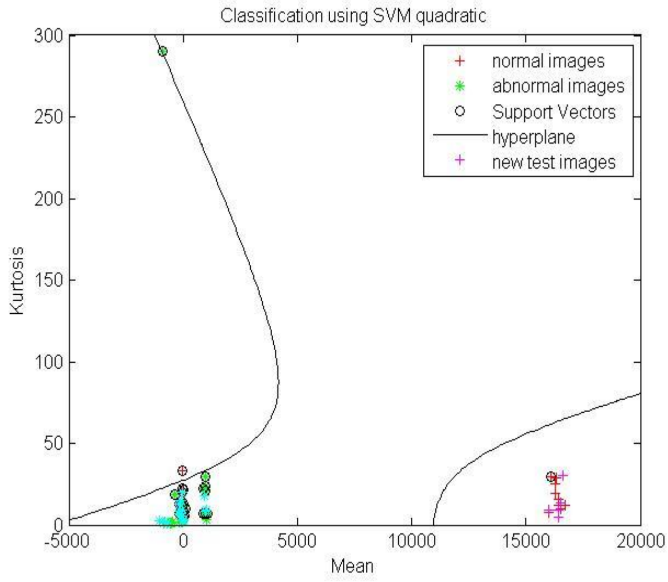

Fig 10: Graph shows the hyperplane constructed between normal and abnormal test points using SVM- Quadratic for Mean and Kurtosis

Conclusion

In this paper; an automatic classification approach has been developed to categorize between normal and abnormal Cardiac signals. It helps the medical practioners to discriminate normal and abnormal cardiac state of any patient for further treatment by experts.

REFERENCES

[1] A Fattah, A.S.M. Mahdee Jameel, R. Goswami, S.K. Saha, N. Syed, S. Akter and C. Shahnaz, "An approach for human identification based on time and frequency domain features extracted from ECG signals", IEEE Region 10 conference, 2159-3442, Nov 2011 pp 259 – 263.

[2] Agrafioti and E. S. Rogers, "An approach for human identification based on time and frequency domain features extracted from ECG signals" IEEEconference, 2011

[3] E.Tatara and A. Cinar, "Interpreting ECG data by integrating statistical and artificial intelligence tools," IEEE Eng. Med. Biol., pp. 36-41, 2002.

[4] I. Khalil and F. Sufi, "Legendre polynomials based biometric authentication using QRS complex of ECG", in proc. International Conference of Intelligent Sensors, Sensor Networks, Info. process 2008, pp. 297-302.

[5] J. M. Irvine, S. A. Israel, W. Todd Scruggs, and W. J. Worek, "Eigen- Pulse: robust human identification from cardiovascular function," Pattern Recognition, vol. 41, no. 11, pp. 3427-3435, 2008.

[6] K .O. Gupta, Dr. P. N. Chatur, " ECG Signal Analysis and Classification using Data Mining and Artificial Neural

Networks" International Journal of Emerging Technology and Advanced Engineering , (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

[7] L. Biel, O. Petterson, L. Philipson, and P. Wide, "ECG analysis: a new approach in human identification," IEEE Trans. Instrum. Meas., vol. 50, no. 3, pp. 808-812, 2001. (Pubitemid 32650858)

[8] Mrs. M.D. Ingole, Mr. S.V. Alaspure, Dr. D.T. Ingole, "Electrocardiogram (ECG) Signals Feature Extraction and Classification using Various Signal Analysis Techniques" INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY ISSN: 2277-9655 pp 39-44

[9] Mansi Varshney, Chinmay Chandrakar, Dr.(Mrs)Monisha Sharma, "A Survey on Feature Extraction and Classification of ECG Signal" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 1, January 2014, pp 6572-6576

[10] Muhammad Fahad Shinwari, Naveed Ahmed, Hassan Humayun, Ihsan ul Haq, Sajjad Haider and Atiq ul Anam, " Classification Algorithm for Feature Extraction using Linear Discriminant Analysis and Cross-correlation on ECG Signals" International Journal of Advanced Science and Technology Vol. 48, November, 2012 pp 149-161.

[11] Sambhu D., Umesh A. C., "Automatic Classification of ECG Signals with Features Extracted Using Wavelet Transform and Support Vector Machines" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Special Issue 1, December 2013 pp 235-241.