International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

# ECG Signal Analysis and Classification using Data Mining and Artificial Neural Networks

[1]K .O. Gupta, [2]Dr. P. N. Chatur

[1]*Student, Govt. College of Engineeering, Amravati*
[2]*Head of Department, Computer Science and Engineering, Govt. College of Engineering, Amravati*
[1]kaps04gupta@gmail.com
[2]chatur.prashant@gcoea.ac.in

***Abstract*— The analysis of the ECG can benefit in diagnosing most of the heart diseases. The electrocardiogram (ECG) provides almost all information about electrical activity of the heart. One cardiac cycle in an ECG consist of the P-QRS-T waves or segments. The ECG signal analysis and classification system gives overall idea about the diseases. In recent years, many research and methods have been proposed and developed for analyzing the ECG signal and extracting features such as amplitude and time intervals for classification of signals. This paper focuses on some of the techniques proposed earlier for the arrhythmia classification and extraction of parameters from the ECG signal which is used for data acquisition and classification system. This paper also gives the brief idea about the proposed work using Artificial Neural Networks (ANN) and data mining techniques using intelligent data miner software.**

***Keywords*— Artificial Neural Networks (ANN), arrhythmia classification, Cardiac Cycle, Data mining, ECG signals.**

## I. INTRODUCTION

ECG [1] is a method to measure and record different electrical potentials of the heart. Willem Einthoven developed the ECG method in the early 1900s. The origin of the electrical activity measured by ECG is in the muscle fibers of different parts of the heart. The ECG may roughly be divided into the phases of depolarization and repolarisation of the muscle fibers making up the heart. The depolarization phases correspond to the P-wave (atrial depolarization) and QRS-wave (ventricles depolarization). The repolarisation phases correspond to the T-wave and U-wave (ventricular repolarisation). The elements in the ECG-complex are shown in Fig. 1.

The human heart contains four chambers: left atrium, right atrium, left ventricle and right ventricle.

Blood enters the heart through two large veins, the inferior and superior vena cava; emptying oxygen-poor blood from body into the right atrium. From the right atrium, the oxygen deficient blood enters the right ventricle.

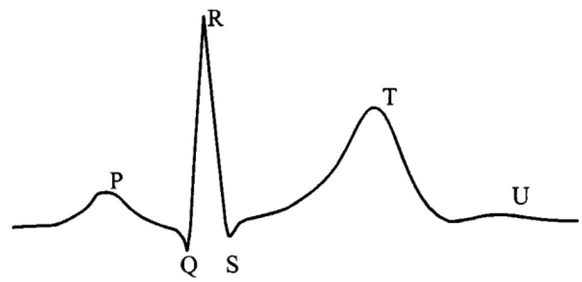


**Fig. 1. Elements of the ECG-complex**

The right ventricle then pushes the blood into the lungs. Inside the lungs a process called 'gas exchange' occurs and the blood replenishes oxygen supply.The oxygen rich blood then enters left atria. From the left atria, blood rushes into the left ventricle. Finally, it is the left ventricle that forces the oxygenated blood to the rest of the body. This mechanical activity of the human heart is powered by electrical stimulations inside the heart. The Depolarization (electrical activity) of a specific portion of the heart (either atria or ventricle) results in mechanical contraction of that specific part. Again, repolarisation results in the mechanical relaxation of the heart chambers. ECG acquisition devices basically pick up these electrical activities via sensors attached to the human skin and draws the electrical activities in milivolt ranges. During the regular activity of the heart, both the atria contract together, followed by ventricular contraction (both the ventricles contract together) [2].

The ECG signal analysis and classification system provides basic features (amplitudes and time intervals) to be used in resultant automatic analysis.

From last decade, a number of methodologies have been proposed to detect features of ECG signal such as amplitude and time interval [3-5]. It is not always adequate to study only time domain method for ECG signal analysis, so there is need of frequency domain representation because small deviation on normal ECG signal means cardiac disorders [6-7].

Changes in ECG reflect the abnormality introduced in the functioning of heart, making diagnosis and treatment of the patient easier. For all types of cardiac disorders and troubles indirectly associated with the heart, doctors advise the patient to have the ECG recorded before complete diagnosis of the patient. ECG analysis has therefore become a preliminary and mandatory requirement for the diagnosis of subjects [8].

In last few decades, several researches have been developed various methods for arrhythmia classification. These methods include digital signal analysis, Fuzzy Logic methods, Genetic Algorithm, Artificial Neural Network, Self-Organizing Map, Support Vector Machines, Bayesian and Wavelet-Domain Hidden Markov Models etc. This paper provides general idea on various techniques and proposed work.

## II. Literature Review

With increase in computational power, sophisticated algorithms have been proposed to improve the prediction accuracy of ECG waveform classification systems. This segment of the paper discusses a variety of techniques proposed earlier in literature for feature extraction, classification and data acquisition of ECG signal.

An approach for effective feature extraction from ECG signal was described in Saxena et al in [9]. Their paper deals with an efficient composite method which has been developed for data compression, signal retrieval and feature extraction of ECG signals. They carried out detailed studies and by training different topologies of error-back-propagation (EBP) artificial neural network (ANN) with respect to variations in number of hidden layers and number of elements in each hidden layer, the best topology with two hidden layers and four elements in each hidden layer has been finalized for ECG data compression using a Military Hospital (MH) data base. After signal retrieval from the compressed data, it has been found that the network not only compresses the data, but also improves the quality of retrieved ECG signal with respect to elimination of high-frequency interference present in the original signal.

The compression ratio (CR) in ANN method increases with increase in number of ECG cycles. The features extracted by amplitude, slope and duration criteria from the retrieved signal match with the features of the original signal. The test results at each stage are consistent and reliable and prove beyond doubt that the composite method can be used for efficient data management and feature extraction of ECG signals in many real-time applications.

Castro et al. in [10] described a wavelet transforms approach for ECG feature extraction. Their paper presented an algorithm, based on the wavelet transform, for feature extraction from an electrocardiograph (ECG) signal and recognition of abnormal heartbeats. A method for choosing an optimal mother wavelet from a set of orthogonal and bi-orthogonal wavelet filter bank by means of the best correlation with the ECG signal was developed. The ECG signal is first denoised by a soft or hard threshold with limitation of 99.99 reconstructs ability and then each PQRST cycle was decomposed into a coefficients vector by the optimal wavelet function. The coefficients, approximations of the last scale level and the details of the all levels, were used for the ECG analyzed. The coefficients of each cycle were divided into three segments, which were related to the P-wave, QRS complex and T-wave, and summed to obtained a features vector of the signal cycles.

Alexakis et al. in [11] used automatic extraction of both time interval and morphological features, from the Electrocardiogram (ECG) to classify ECGs into normal and arrhythmic. Classification was implemented by artificial neural networks (ANN) and Linear Discriminant Analysis (LDA). The ANN gave more accurate results. Average training accuracy of the ANN was 85.07% compared with 70.15% on unseen data.

Ramli et al. in [12] investigate the use of signal analysis technique to extract the important features from the 12 lead system (electrocardiogram) ECG signals. Lead II is chosen for the whole analysis due to it representative characteristics for identifying the common heart diseases. The analysis technique chosen is the cross-correlation analysis. Cross-correlation analysis measures the similarity between the two signals and extracts the information present in the signals. Results show that the parameters signal analysis technique extracted could clearly differentiate between the types of heart diseases analyzed and also for normal heart signal.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

Tayel and Bouridy together in [13] described a technique for ECG image classification by extracting their feature using wavelet transformation and neural networks. Their paper, presents an intelligent diagnosis system for electrocardiogram (ECG) intensity images using artificial neural network (ANN). Features were extracted from many preprocess such as wavelet decomposition (WD), Edge detection (ED), gray level histogram (GLH), Fast Fourier transform (FFT), and Mean-variance (M-V). The ANN supervised feed-forward back propagation using adaptive learning rate with momentum term algorithm used as a classifier. The input data to the classifier is very large so, ECG images data were grouped in batches that introduced to ANN classifier. The objective of their paper was to introduce an expert system for ECG diagnosis, more suitable preprocess for the used 63 ECG intensity images, and simplest ANN architecture classifier, depending on the higher accuracy of the classifier related to the extracted input features.

Tadejko and Rakowski in [14] presented the classification performance of an automatic classifier of the electrocardiogram (ECG) for the detection abnormal beats with new concept of feature extraction stage. Feature sets were based on ECG morphology and RR-intervals. Configuration adopted a Kohonen self-organizing maps (SOM) for analysis of signal features and clustering. In their study, a classifier was developed with SOM and learning vector quantization (LVQ) algorithms using the data from the records recommended by ANSI/AAMI EC57 standard. Their paper compares two strategies for classification of annotated QRS complexes: based on original ECG morphology features and proposed new approach - based on preprocessed ECG morphology features. The mathematical morphology filtering was used for the preprocessing of ECG signal. The performance of the algorithm was evaluated on the MIT-BIH Arrhythmia Database following the AAMI recommendations. Using this method the results of recognition beats either as normal or arrhythmias was improved.

Alan and Nikola in [15] described chaos theory applied to ECG feature extraction. Several chaos methods, including phase space and attractors, correlation dimension, spatial filling index, central tendency measure and approximate entropy are explained in detail. A new feature extraction environment called ECG Chaos Extractor has been created in order to apply these chaos methods. System model and program functions are presented. Some of the obtained results were listed.

Ubeyli et al. in [16] developed automated diagnostic systems employing diverse and composite features for electrocardiogram (ECG) signals were analyzed and their accuracies were determined. Because of the importance of making the right decision, classification procedures classifying the ECG signals with high accuracy were investigated. The classification accuracies of mixture of experts (ME) trained on composite features and modified mixture of experts (MME) trained on diverse features were compared. The inputs of these automated diagnostic systems were composed of diverse or composite features (power levels of the power spectral density estimates obtained by the eigenvector methods) and were chosen according to the network structures. The conclusions of their study demonstrated that the MME trained on diverse features achieved accuracy rates which were higher than that of the ME trained on composite features.

A new wavelet based framework was developed and evaluated for automatic analysis of single lead electrocardiogram (ECG) for application in human recognition by Fatemian et al. [17]. Their proposed system utilizes a robust preprocessing stage that enables it to handle noise and outliers so that it was directly applied on the raw ECG signal. Moreover, it was capable of handling ECGs regardless of the heart rate (HR) which renders making presumptions on the individual's stress level unnecessary. One of the novelties of their paper was the design of personalized heartbeat template so that the gallery set consists of only one heartbeat per subject. This substantial reduction of the gallery size decreases the storage requirements of the system significantly. Furthermore, the classification process was speeded up by eliminating the need for dimensionality reduction techniques such as *PCA* or *LDA*. Experimental results for identification over PTB and MIT healthy ECG databases indicate a robust subject identification rate of 99.61% using only 2 heartbeats in average for each individual.

Pedro et al. [18] was concerned with the classification of ECG pulses by using state of the art Continuous Density Hidden Markov Models (CDHMM's). The ECG signal was simultaneously observed at three different level of focus by means of the Wavelet Transform (WT). The types of beat being selected were normal (N), premature ventricular contraction (V) which was often precursor of ventricular arrhythmia, two of the most common class of supra-ventricular arrhythmia (S), named atrial fibrillation (AF), atrial flutter (AFL), and normal rhythm (N). Both MLII and V1 derivations were used.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

Run time classification errors could be detected at the decoding stage if the classification of each derivation was different. These pulses were selected for a posterior physician analysis. Experimental results were obtained in real data from MIT-BIH Arrhythmia Database and also in data acquired from a developed low-cost Data-Acquisition System.

Jadhav et al. [19] proposed a new approach for cardiac arrhythmia disease classification. The proposed method uses Modular neural network (MNN) model to classify arrhythmia into normal and abnormal classes. They performed experiments on UCI Arrhythmia data set. Missing attribute values of this data set are replaced by closest column value of the concern class. They constructed neural network model by varying number of hidden layers from one to three and are trained by varying training percentage in data set partitions. Their data set is a good environment to test classifiers as it is incomplete and ambiguous bio-signal data collected from total 452 patient cases. The classification performance is evaluated using six measures; sensitivity, specificity, classification accuracy, mean squared error (MSE), receiver operating characteristics (ROC) and area under curve (AUC). The experimental results presented in their paper show that more than 82.22% testing classification accuracy.

## III. Proposed Method

The human heart beats are of different types. The proposed work is on four types of heart beats which is common: Normal, Premature Ventricular Contraction (PVC), Atrial Premature Contraction (APC) and Left Branch Bundle Block Beat (LBBBB) shown in Fig 2. These heart beats show different variations and are nonlinear in nature. Artificial neural networks are very useful in nonlinear problems and gives very capable results.

The methodology for proposed work is shown in the flow of Fig 3. The equal length signal (in terms of time and amplitude) is taken from the raw ECG signal for the processing. This signal is further pre-processed for removing of noises and other DC components. After pre-processing, actual method of feature extraction and classification is done. This extracted feature is given as input to the already trained supervised neural network for classifying ECG signal waves. This ECG signal is also classified using some data mining techniques using intelligent data miner software [20]. Finally, the results of both ANN and data mining algorithms are compared.

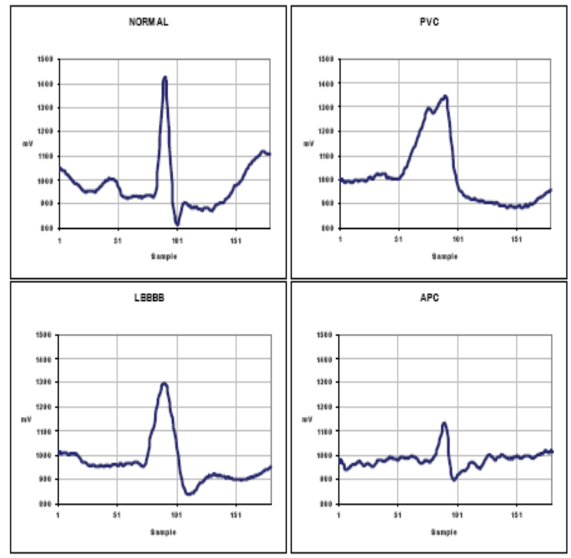


**Fig 2 Four common types of heartbeats. [21]**

## IV. Conclusion

The ECG signal has been studied by various researchers for classification of arrhythmias. The work done in this field is reviewed in literature review. In addition to that proposed method aims to analyze and classify the ECG signal using two approaches namely artificial neural network and data mining. After classifying the signals, both the techniques will compared and accuracy of each method will calculate.

EXPLORING RESEARCH AND INNOVATIONS

**International Journal of Emerging Technology and Advanced Engineering**
**Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)**

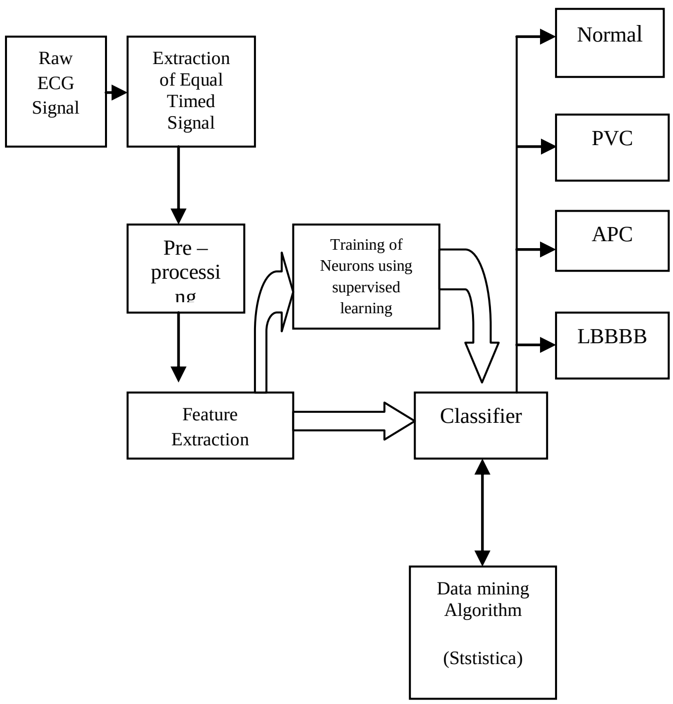


**Fig 3 Methodology for the classification**

## V. FUTURE ENHANCEMENT

In future, more concentration will be given on the classification of all types of heartbeats as proposed paper uses only four types.

## REFERENCES

[1] L. Goldberger and E. Goldberger, "Clinical Electrocardiography A Simplified Approach", 5th ed. St. Louis, MO: Mosby, 1994, vol. 1, p. 341.

[2] Fahim Sufi, Ibrahim Khalil, Jiankun Hu "ECG-Based Authentication". Springer 2010

[3] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, and Pablo Laguna, "A Wavelet-Based ECG Delineator: Evaluation on Standard Databases," IEEE Transactions on Biomedical Engineering Vol. 51, No. 4, pp. 570-581, 2004.

[4] Krishna Prasad and J. S. Sahambi, "Classification of ECG Arrhythmias using Multi-Resolution Analysis and Neural Networks," IEEE Transactions on Biomedical Engineering, vol. 1, pp. 227-231, 2003.

[5] Cuiwei Li, Chongxun Zheng, and Changfeng Tai, "Detection of ECG Characteristic Points using Wavelet Transforms," IEEE Transactions on Biomedical Engineering, Vol. 42, No. 1, pp. 21-28, 1995.

[6] Saritha, V. Sukanya, and Y. Narasimha Murthy, "ECG Signal Analysis Using Wavelet Transforms," Bulgarian Journal of Physics, vol. 35, pp. 68-77, 2008

[7] S.Karpagachelvi, Dr.M.Arthanari, M.Sivakumar, "ECG Feature Extraction Techniques - A survey Approach" International Journal of Computer Science and Information Security Vol. 8, No. 1, pp 76-80, 2010.

[8] Sanjay M. Pati1 (1994) "ECG Analysis - Expert System Approach" ME. – Dissertation, V.J.T.I, Bombay – 19

[9] S. C. Saxena, A. Sharma, and S. C. Chaudhary, "Data compression and feature extraction of ECG signals," International Journal of Systems Science, vol. 28, no. 5, pp. 483-498, 1997.

[10] B. Castro, D. Kogan, and A. B. Geva, "ECG feature extraction using optimal mother wavelet," The 21st IEEE Convention of the Electrical and Electronic Engineers in Israel, pp. 346-350, 2000.

[11] C. Alexakis, H. O. Nyongesa, R. Saatchi, N. D. Harris, C. Davies, C. Emery, R. H. Ireland, and S. R. Heller, "Feature Extraction and Classification of Electrocardiogram (ECG) Signals Related to Hypoglycaemia," Conference on computers in Cardiology, pp. 537-540, IEEE, 2003.

[12] A. B. Ramli, and P. A. Ahmad, "Correlation analysis for abnormal ECG signal features extraction," 4th National Conference on Telecommunication Technology, 2003. NCTT 2003 Proceedings, pp. 232-237, 2003.

[13] Mazhar B. Tayel, and Mohamed E. El Bouridy, "ECG Images Classification using Artificial Neural Network Based on Several Feature Extraction Methods," IEEE, pp113-115, 2008.

[14] P. Tadejko, and W. Rakowski, "Mathematical Morphology Based ECG Feature Extraction for the Purpose of Heartbeat Classification," 6th International Conference on Computer Information Systems and Industrial Management Applications, CISIM '07, pp. 322-327, 2007.

[15] Alan Jovic, and Nikola Bogunovic, "Feature Extraction for ECG Time- Series Mining based on Chaos Theory," Proceedings of 29th International Conference on Information Technology Interfaces, 2007.

[16] Ubeyli, and Elif Derya, "Feature extraction for analysis of ECG signals," Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, pp. 1080-1083, 2008.

[17] S. Z. Fatemian, and D. Hatzinakos, "A new ECG feature extractor for biometric recognition," 16th International Conference on Digital Signal Processing, pp. 1-6, 2009.

[18] Pedro R. Gomes, Filomena O. Soares, J. H. Correia, C. S. Lima "ECG Data-Acquisition and Classification System by Using Wavelet-Domain Hidden Markov Models" 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010.

[19] S. M. Jadhav, Dr. S. L. Nalbalwar, Dr. Ashok A. Ghatol, "modular neural network based arrhythmia classification system using ecg signal data", in International Journal of Information Technology and Knowledge Management January-June 2011, Volume 4, No. 1, pp. 205-209.

[20] www.statsoft.com

[21] Rajesh Ghongade, Dr. Ashok Ghatol "A Robust and Reliable ECG Pattern Classification are using QRS Morphological Features and ANN", in IEEE Region 10 Conference 2008.