A Survey of Algorithms for Feature Extraction and Feature Classification Methods

Prachi Lamba
Electronics and Communication
Lovely Professional University
Phagwara, India
prachilamba659@gmail.com

Kirti Rawal
Electronics and Communication
Lovely Professional University
Phagwara, India
kirti.20248@lpu.co.in

***Abstract*— Recording of Electrocardiogram (ECG) signal is a non-invasive method for examining the electrical actions of the heart. The characterization of ECG beat is an important step in finding the location of various heart ailments. The abnormal beats are the symptoms of several diseases such as myocardial infarction and ischemic heart diseases. The classification of beats in healthy and diseased subjects aids the researchers to detect various abnormalities in case of any arrhythmia. The arrhythmias of heart like structural, circulatory or electrical can be diagnosed by classifying the abnormal beats. Thus, the classification of ECG beats is necessary in detecting and diagnosing the diseases. In this paper, the computational methodologies adopted for ECG beat classification and the issues related to it are presented. Further, the feature extraction and feature classification methods of ECG signal have been reviewed.**

Keywords—Beat classification, ECG, Feature Extraction, Feature classification.

I. INTRODUCTION

In order to measure the electrical activity of heart, ECG signal is recorded by placing the electrodes on the skin. Using these electrodes, the heart produces tiny electrical changes that appears from the heart muscles during each heartbeat. Heart muscles follow the pattern of polarization and depolarization. Therefore, it is a medical test which identifies cardiac abnormalities that occurred due to the contraction of the heart [1]. The machine which is used to record the patient's ECG is known as electrocardiograph. The ECG signal is used to analyze the variations in heart rate, the size and position of the heart chambers, heart's muscle cells etc. Further, ECG is also used determine the effects of cardiovascular medications, and the capacity of embedded pacemakers. Therefore, ECG analysis is preferred for most of the cases in detecting and diagnosing the heart related diseases.

The features and parameters of ECG signal changes due to any illness or irregularities in the heart rhythm. A specialist or radiologist usually records the ECG of heart patients or the subjects who might be at the danger of the coronary illness. Thus, ECG signal analysis is not limited to diagnose the heart diseases, but it also includes the subjects having hypertension, elevated cholesterol, diabetes, overweight, smoking propensities and coronary illness [1].

In order to prevent the life of numerous heart patients, early diagnosis is necessary to cure the diseases. For early and timely diagnosis, the ECG signal is classified by using several machine learning techniques. These computers-based machine learning techniques for classification the ECG beats are not only used to reduce the doctor's load but also improve the accuracy of diagnosis. The proper treatment of patient can be done by recognizing the abnormalities of heart.

ECG beat classification is well known for classifying variety of diseases particularly in the diagnosis of heart diseases. The different classes of ECG beats are Normal (N), Paced (P), Ventricular ectopic beat (V), Supra Ventricular ectopic beat (S), Fusion (F), Unknown (Q), Left bundle branch (LBBB or L) and Right bundle branch (RBBB or R) [2] as shown in TABLE 1.

The primary step of ECG beat classification is the collection of Raw ECG data from the standard database followed by Pre- processing of the ECG signal. In pre-processing, the noise in the signal which occurs due to power line interference, baseline wandering, muscle noise etc is removed. After pre-processing, the next step is to extract the features from the signal known as Feature extraction. It determines the amplitude as well as interval of the signal using various techniques [3] .Some of the techniques which are used for extracting ECG features are DWT (Discrete wavelet transform), PCA (Principal component analysis), ICA (Independent component analysis), LDA (Linear discriminate analysis) etc. After extraction, the last step is Feature classification. It is used to classify the normal and the abnormal beats by using various methods. Some of the methods used for classifying the ECG beats are SVM (Support vector machine), KNN (K-nearest neighborhood), ANN (Artificial neural network), RBF (Radial basis function) etc.

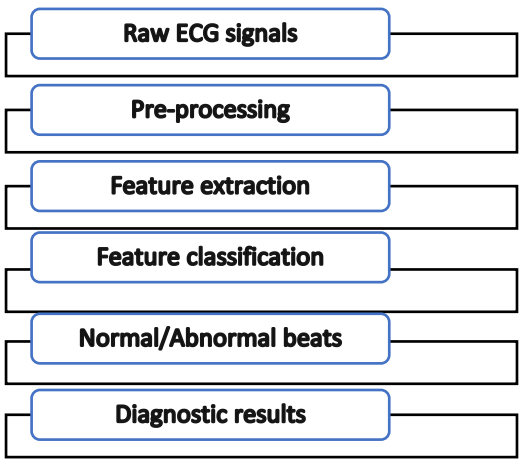

Fig. 1. Flowchart of ECG beat classification system

Once the classification is done, then on the basis of classification of normal/abnormal beats, the diagnosis can be done [4]. The steps used for classifying the ECG beats are shown in Fig. 1. In spite of the huge developments in this field there are still lots of improvement needs to be done in order to make the extraction and classification system more robust, reliable and accurate. So, in this paper we tried to present the summary of the work done so far in order to facilitate the researchers to have an idea of existing work. The comprehensive analysis and the basics related to ECG beat classification can help the beginners to understand the research area.

This paper incorporates the fundamental concepts of ECG and its characterization (segment 1), foundation information (segment 2) and a detailed study of ECG include extraction and classification of ECG beats. In this area, diverse strategies are utilized for highlighting the extraction of ECG features, classification of ECG beats (segment 3) and the various databases used for ECG analysis (segment 4). At last, the paper presents conclusion (segment 5).

The different types of beats which are used to classify the normal and abnormal subjects are shown in TABLE I.

TABLE I CLASSIFICATION OF ECG BEATS

Name of the Beat	*Explanation*
N	A typical resting pulse ranges from 60 to 100 BPM (Beats per Minute).
L	A postponement or blockage of electrical driving forces to left side of the heart.
R	A postponement or blockage of electrical driving forces to right side of the heart.
P	When heart rhythm is interrupted, irregular or too slow.
S	A beat that is pre-mature, narrow in width and is different from normal patient's beat which may even lead to heart attack or heart stroke. Also, called as PAC (Premature atrial contraction).
V	A beat which shows irregular heart rhythm due to premature heartbeat. It occurs due to skipped beat or stopped beat, when heart is in a fluttering condition. Also, called as PVC (Premature ventricular contraction).
F	When ventricular and supra ventricular electrical impulses coincide at the same time to produce a hybrid complex beat.

II. BACKGROUND KNOWLEDGE

An ECG tracing consists of several components that indicate the electrical events during one heartbeat. ECG signal comprises of various ECG beats which in turn contains some ECG features such as peaks, intervals and segments. Peaks namely P, Q, R, S, T and U while that of intervals namely PR, RR, QRS, ST, QT and segments namely PR and ST. ECG beats includes P wave, QRS complex and T wave in one ECG signal.

P wave and QRS complex lasts for 0.08 seconds and their duration is not more than 0.10 seconds. P wave occurs due to depolarization from SA node throughout atria. But in case of QRS complex, Q and S both are immediately deflecting downwards. Q arises before ventricular contraction while S arises after ventricular contraction and R occur due to polarization of ventricles. R beat deflected in positive direction, but Q and S beat deflected in negative direction. For such beat, ventricular systole & atrial diastole takes place. T wave lasts for 0.16 seconds and it occur due to repolarization of ventricles or when ventricular diastole takes place. Lastly, U wave is a kind of short wave which occurs after the occurrence of T wave, but it is not mostly observed.

III. LITERATURE SURVEY OF ECG BEAT CLASSIFICATION METHODS

Numerous methodologies are adopted by the investigators for the ECG beat classification. It includes 3 major stages namely, pre-processing stage, feature extraction and feature classification. MIT-BIH Arrhythmia database (The Massachusetts institute of Technology Beth Israel hospital arrhythmia database) [5] is most commonly used in most of the papers. Based on various classifiers, some of the parameters are calculated which gives the information about the performance of classification of beats.

Ref. [6, 7 and 8] extracted some ECG features with the help of XWT (Cross wavelet transform), DWT and ICA techniques respectively and classified the beats by SVM classifier having accuracy of 94.8%, 98.94% respectively and classification rate for [8] is 99.75%. Ref. [9 and 11] used CWT (Continuous wavelet transform) for feature extraction and classification of arrhythmia beats is done by RBF and SVM respectively having accuracy of 98.92% and 95.03% respectively. Ref. [10] determined the characteristics of cardiac beats by EMD (Empirical mode decomposition), DWPD (Discrete Wavelet Packet Decomposition) and KNN which yielded an average accuracy, sensitivity, selectivity of 99.2% and average specificity of 99.8%. Ref. [12] extracted ECG parameters by using GA (Genetic algorithm) which was broadly classified into temporal, spectral, complexity and wavelet features. They further used ANN and SVM, among them ANN gave better results when provided with 100% training. Ref. [13] diagnosed different arrhythmias by DT (Decision tree) and KNN resulted in accuracy of 96.3%, 93.3% and sensitivity of 99.3%, 97.5% respectively. Analysis of different arrhythmia beats was done with the help of DWT and SVM methods [14], and achieved specificity and NPV (Negative predictive value) of 92%. Ref. [15] examined the performance of normal and abnormal heartbeats with DSNT (Discrete sinc transform), DCT and SVM got the accuracies of 95.45%, 90.91% for recognition, with specificity 100%, 91.67%, and sensitivity 90% respectively.

Ref. [16, 17, and 18] applied DWT and SVM for the detection of ECG beats and obtained accuracy 99.59%, 96% and 88.24% respectively. For extraction of ECG beats, various authors [19, 20, 21, 22 and 23] utilized DWT and PCA technique and further used different classifiers for beat

classification and gave better results in [23]. An Accuracy of 99.08% was achieved in [24] who classified 5 different heart beat types with the help of DWT and RF classifier. Ref. [25] classified 5 types of beats with ST (Stockwell transform), BFO (Bacteria foraging optimization) algorithm and LMS (Least mean square) based SVM resulted in average accuracy and sensitivity of 98.2%, 74.7% (S-detection) and 98.6%, 91.7% (V-detection).

IV. Database and Methods

This section includes ECG databases, various feature extraction methods, feature classification methods and performance measures which are discussed below:

A. ECG Database

MIT-BIH arrhythmia database have been used by the researchers for carrying out their study based on ECG or other cardiac disease analysis or diagnosis. According to the AAMI (Association for the advancement of Medical Instrumentation) recommendation, this database was taken from PhysioBank ATM that is available at physioNet website [5] contains various data sets of biomedical signals. Most of the researchers [6-11, 12, 13-18, 20-23, 24 and 25] used MIT-BIH database and European database for analyzing the ECG signals [19].

B. Feature extraction methods

Feature extraction is the method of gathering discriminative information from a set of samples. Features are required to be computed for extraction of clinically valuable information from the textures. The features can be considered as supplement to the visual skills of the researchers that are relevant to the diagnostic problem but may not be visually perceptible. The distinguishing and potent features are extracted by using different feature extraction methods. Some of the feature extraction methods are explained below:

1) DWT:

DWT deals with the decomposition of signal into a set of mutually orthogonal wavelet basis function which have smaller bandwidths and slower sample rates. It consists mainly of two frequency ranges i.e low frequency leads to Approximation component (cA) and high frequency leads to Detail components (cD) of the signal.

2) PCA:

PCA is a statistical technique and a dominant tool which is used to analyze and compress a large amount of data that captures the essence of original data. This method indicates various similarities and differences present in various samples of data. Hence, it focuses on calculating the correlation between the samples.

3) LDA:

LDA is a straightforward arrangement procedure which depends on the idea of looking for direct mix of factors that best isolates among two classes. It mainly focuses on conversion from high dimensional data into low dimensional data. Maximization of between-class distance and minimization of within-class distance is the key concept of LDA which concludes in the dimensionality reduced space.

C. Feature classification methods

Feature classification is the process of grouping of features based on some criteria which categorizes data based on different classes and it includes various techniques such as:

1) SVM:

SVM classifier is a sort of administered learning model which is utilized to break down information for arrangement and relapse. It is characterized utilizing an isolating hyper plane which groups all the preparation vectors into two classes. The best decision will be the hyper plane that departs the most extreme edge from the two classes. The edge is the separation between the hyper plane and the nearest component from hyper plane. Hence, the margin is maximized between the two classes which mean seperability is also maximized.

2) RBF:

RBF is a classification approach which is used for functional approximation and classification that uses nonlinear activation functions like Sigmoidal and Gaussian Kernel. Therefore, the response of Gaussian function is positive for all the values of x and it decreases to zero as $|x \rightarrow 0|$. RBF as the name suggests, is found to be radically symmetric as it produces identical output for all the value of the input from the centre of the kernel.

3) KNN:

KNN is a simplest algorithm which is used for classification as well as regression of k nearest neighbour data sets given. The KNN model splits the given data into a number of classes to predict the classification of a new dataset. It acts as a "*clustering model*". As a classifier, it predicts the membership of a class based upon the majority vote of its k nearest neighbours. In case of regression, the mean (average) of its k nearest neighbours depicts its class. It is determined using Euclidean distance.

V. Conclusion

In this paper, the computational methodologies adopted for ECG beat classification and the issues related to it are presented. Additionally, the review of feature extraction and feature classification methods of ECG signal has been presented in this paper. Classification of beats in normal and abnormal class plays an important role in early diagnosis of heart related diseases. Abnormal beats can be the cause of the cardiac ailments like Ischemic heart disease or related Arrhythmias or can be the vital sign provider for life threatening. This diagnosis helps in detecting the cardiac diseases at an early stage so as to reduce the mortality rate caused by cardiac arrhythmias, cardiac arrests, and overall cardiac deaths. A considerable amount of work has been already done by different researchers to extract the features from the ECG signal and then classify it. Investigation of the ECG signal relies upon precise recognition of different features of ECG signal. DWT, PCA, ICA, LDA and Pan-Tompkins algorithm are some of the techniques which have been applied to extract the ECG features. Similarly, SVM, KNN, RBF and ANN are some of the classifiers which have been used for ECG classification. However, from the survey it is examined that DWT with the combination of SVM has been verified with several measures which provide promising

classification results in case of arrhythmia. Therefore, it is concluded that there is a lot of scope of research in extracting the features of ECG signal and then classification of the following ECG signal. The classification of the ECG beats is the open issue of research. ECG beat classification is important to diagnose the possible illness of human being along with ECG signal analysis.

REFERENCES

[1] Jambukia, Shweta H., Vipul K. Dabhi, and Harshadkumar B. Prajapati, "Classification of ECG signals using machine learning techniques: A survey," In Computer Engineering and Applications (ICACEA), pp. 714-72, March 2015.

[2] Ahmed, Waqas, and Shehzad Khalid, "ECG signal processing for recognition of cardiovascular diseases: A survey," In Innovative Computing Technology (INTECH), pp. 677-682, August 2016.

[3] Luz, E.J.D.S., Schwartz, W.R., Cámara-Chávez, G. and Menotti, D., "ECG-based heartbeat classification for arrhythmia detection: A survey," Computer methods and programs in biomedicine, pp.144-164, April 2016.

[4] Houssein, E.H., Kilany, M. and Hassanien, A.E., "ECG signals classification: a review," International Journal of Intelligent Engineering Informatics, pp.376-396, 2017.

[5] https://physionet.org/ (Internet Source).

[6] Jacob, Neenu, and Liza Annie Joseph, "Classification of ECG beats using cross wavelet transform and support vector machines," In Intelligent Computational Systems (RAICS), pp. 191-194, December 2015.

[7] Desai, Usha, Roshan Joy Martis, C. Gurudas Nayak, K. Sarika, and G. Seshikala, "Machine intelligent diagnosis of ECG for arrhythmia classification using DWT, ICA and SVM techniques," In India Conference (INDICON), pp. 1-4, December 2015.

[8] Barhatte, Alka S., Rajesh Ghongade, and Abhishek S. Thakare, "QRS complex detection and arrhythmia classification using SVM," In Communication, Control and Intelligent Systems (CCIS), pp. 239-243, November 2015.

[9] Harkat, A., R. Benzid, and L. Saidi, "Features extraction and classification of ECG beats using CWT combined to RBF neural network optimized by cuckoo search via levy flight," In Electrical Engineering (ICEE), pp. 1-4, December 2015.

[10] Shahnaz, C., T. B. Anowar, R. H. M. Rafi, I. Ahmmed, and S. A. Fattah, "Cardiac beat classification based on wavelet analysis of empirical mode decomposed ECG signals," In TENCON 2015-2015 IEEE Region 10 Conference, pp. 1-6, November 2015.

[11] Shahbudin, S., S. N. Shamsudin, and H. Mohamad, "Discriminating ECG signals using Support Vector Machines," In Computer Applications & Industrial Electronics (ISCAIE), pp. 175-180. IEEE, April 2015.

[12] Karthika, J. S., Jan Mary Thomas, and Jubilant J. Kizhakkethottam, "Detection of life-threatening arrhythmias using temporal, spectral and wavelet features," In Computational Intelligence and Computing Research (ICCIC), pp. 1-4, December 2015.

[13] Acharya, U. Rajendra, Hamido Fujita, Muhammad Adam, Oh Shu Lih, Tan Jen Hong, Vidya K. Sudarshan, and Joel EW Koh, "Automated characterization of arrhythmias using nonlinear features from tachycardia ECG beats," In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 000533-000538, October 2016.

[14] Dewangan, Naveen Kumar, and S. P. Shukla, "ECG arrhythmia classification using discrete wavelet transform and artificial neural network," In Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 1892-1896, May 2016.

[15] Gad, Mohamed Moustafa Azmy, "Feature extraction of electrocardiogram signals using discrete sinc transform," In Biomedical Engineering International Conference (BMEiCON), pp. 1-4, December 2016.

[16] Ranaware, Preeti N., and Rohini A. Deshpande, "Detection of arrhythmia based on discrete wavelet transform using artificial neural network and support vector machine," In Communication and Signal Processing (ICCSP), pp. 1767-1770, April 2016.

[17] Venkatesan, C., P. Karthigaikumar, Anand Paul, S. Satheeskumaran, and R. Kumar, "ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications," IEEE Access 6, pp. 9767-9773, 2018.

[18] Sabeeha, S., and C. Shiny, "ECG-based heartbeat classification for disease diagnosis," In Computing Methodologies and Communication (ICCMC), pp. 1113-1117, July 2017.

[19] Desai, Usha, "Automated detection of cardiac health condition using linear techniques," In Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 890-894, May 2017.

[20] El-Saadawy, Hadeer, Manal Tantawi, Howida A. Shedeed, and Mohamed F. Tolba, "Electrocardiogram (ECG) heart disease diagnosis using PNN, SVM and Softmax regression classifiers," In Intelligent Computing and Information Systems (ICICIS), pp. 106-110, December 2017.

[21] Elhaj, Fatin A., Naomie Salim, Taqwa Ahmed, Arief R. Harris, and Tan Tian Swee, "Hybrid classification of Bayesian and extreme learning machine for heartbeat classification of arrhythmia detection," In Student Project Conference (ICT-ISPC), pp. 1-4, May 2017.

[22] Akin, Zahide Elif, and Süleyman Bilgin, "Classification of normal beat, atrial premature contraction and ventricular premature contraction based on discrete wavelet transform and artificial neural networks," In Medical Technologies National Congress (TIPTEKNO), pp. 1-4, October 2017.

[23] Martis, Roshan Joy, U. Rajendra Acharya, and Lim Choo Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," Biomedical Signal Processing and Control 8, pp. 437-448, September 2013.

[24] Emanet, Nahit, "ECG beat classification by using discrete wavelet transform and Random Forest algorithm," In Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, ICSCCW, pp. 1-4, September 2009.

[25] Das, Manab K., and Samit Ari, "Patient-specific ECG beat classification technique," Healthcare technology letters 1, no. 3, pp. 98-103, September 2014.