International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 1479–1485

ECG Signals Classification For Early Detection Of Cardiovascular Diseases (CVDs)

Vrushali Shinde[1] , Rutuja Raut[2], Vrushali Patil[3], Prof. Nandini Babbar[4], Milind Ankleshwar[5]

[1,2,3] Student, Dept. of Computer Engineering, NBN SSOE, Maharashtra, India
[4] Guide-Professor, Dept. of Computer Engineering, NBN SSOE, Maharashtra, India
[5] Director, Mass Technologies, Maharashtra, India

Abstract

CardioVascular Diseases is one of the major cause of human deaths. The increasing threats of CVD can be early detected with various medical tests including electrocardiogram (ECG), and also 2D Echo,Stress Test. With the help of ECG signal, early detection of CVD is possible and proper medication can be provided for human life as and when needed. However to examine all these signals manually can be very much hectic, stressbuster and would require ample lot of time. Discrete wavelet transform (DWT) method combining with nonlinear features for automated characterization of CVDs will be main highlight in this research which will also help overcome manual ECG work. DWT subjects ECG signals upto five levels of normal, dilated cardiomyopathy(DCM), hypertrophic cardiomyopathy (HCM), myocardial infraction(MI) . DWT coefficients extracts fuzzy entropy, sample entropy, fractal dimension, and signal energy etc as relative wavelet. Our proposed methodology is inclusive of multiple CVD devices signal which helps us to increase the accuracy of the data and giving right prediction to save and help human life by taking proper medication.

Key Words:

CVD, ECG, Discrete Wavelet Transform, Hypertrophic cardiomyopathy, Dilated Cardiomyopathy, Myocardial infarction.

1. INTRODUCTION

Cardiovascular Diseases (CVDs) are amongst the key causes which results in increased death rate worldwide consistent with the records provided by the WHO i.e. World Health Organization in 2012, 17.5 million deaths occurred; thanks to CVDs which ultimately increases medical expenditure. The worldwide direct medical expenses thanks to CVDs are approximated to US$863 billion in 2010. The worldwide predicts that these expenses will still rise and can reach US$30 trillion by 2030. The common cardiovascular diseases are infarct (MI) followed by memory stroke, hypertensive heart diseases, cardiomyopathy, rheumatic cardiopathy, congenital cardiopathy and heart arrhythmia. So proposed methodology studies two forms of cardiomyopathy i.e. cardiomyopathy (HCM), dilated cardiomyopathy (DCM) and infarct (MI) which are considered as distinct abnormalities of the myocardium. HCM causes due to enlargement in heart muscles which affects the ventricles and interventricular septum (IVS) leading to thickening of them. While DCM causes thanks to enlargement within the ventricle (LV) due to the stretching and thinning of the myocardium. So this incorporates a dangerous impact on LV systolic function which ends up in congestive failure or arrhythmias . Infarct could be a consequence of arteria coronaria disease (CAD) which ends within the death of myocardium due to prolonged disturbance in oxygenated blood circulation. Further, this condition ends up in the decline of myocardial contractility and damage percentage depends on the dimensions of the affected region. This could vary from a tiny low area of dead myocardial tissue to an autosized area of infarcted myocardial tissue, which is related to acute shock and death. The presence of CVD like left ventricular hypertrophy (LVH) thanks to HCM is reflected on the electrocardiogram (ECG) signals as taller R waves and inverted T waves within the left-sided praecordial and lateral frontal leads, i.e. leads V5-6, and that I and VL, respectively. Tall R waves are found in DCM. ECG with ST segments either elevated or depressed with inverted T waves counting on the position of the lead characterizes the

ISSN: 2233-7857 IJFGCN

MI condition. Clinically, of those variations in ECG parameters are visually assessed and manually interpreted to detect the presence of CVDs. Due to the non-stationary nature of the ECG signal, indicators of CVDs may appear randomly within the timescale. Some vital diagnostic details don't seem to be perceptible with manual examination and may end in errors of interpretation . Therefore, to beat these limitations during the manual assessment, computer-aided techniques could even be more appropriate and useful for accurate diagnosis. Therefore, during this study, we've intended towards on the characterization of three CVDs (HCM, DCM, and MI) by extracting relative wavelet nonlinear features from ECG signals.

2. LITERATURE SURVEY

Adversarial de-noising of electrocardiogram [1] by Jilong Wang , Renfa Li, Rui Li , Keqin Li , Haibo Zeng , Guoqi Xie , Li Liu has designed loss function to both global and native characteristics of signals, utilizing the characteristics to store knowledge on the distribution of ECG noise throughout the game between the generator and therefore the discriminator, and evaluates the standard of denoised signals against SVM algorithm. The extensive experiments show that compared to the state-of-the-art methods, this method achieves up to about 62% improvement on the SNR of denoised signals on the common. **From using electrocardiogram signal Stockwell transform and Hybrid classification scheme based automated detection of cardiovascular diseases** [2] by R.K. Tripathy, Mario R.A. Paternina, Juan G. Arrieta, Alejandro Zamora-Méndez, Ganesh R. Naik 2019 a unique approach for the automated detection of CHF supported on the time-frequency analysis of the ECG signal has been demonstrated. the S-transform coefficients of the ECG signal at different frequency scales, the time-frequency entropy features are computed. A hybrid classifier supported the mix of the residual of SRC and nearest distance for individual classes was used. The entropy measures and various dimension reduction techniques like independent component analysis (ICA) and linear discriminant analysis (LDA) will be want to quantify the diagnostic information from ECG signal for accurate and reliable detection of CHF pathology. **Cardiovascular Diseases detection using random forest classifier** [3] by Zerina Masetic, Abdulhamit Subasi implemented two phases: feature extraction and classification phase. In the phase of feature extraction, AR method is applied for extracting features. In the phase of classification, the classifiers C4.5 decision tree, *k*-nearest neighbor, artificial neural networks, random forest classifier and support vector machine are examined. The ECG signals were acquired from BIDMC Cardiovascular Diseases and PTBDiagnostic ECG databases and classified by applying various experiments. Impressive performance of random forest method proves that it plays significant role in detecting cardiovascular diseases (CHF) and might be valuable in expressing knowledge useful in medicine. **Classification of Infarction with Multi-Lead ECG Signals and Deep CNN**[4 7] by Ulas Baran Baloglu , Muhammed Talo , Ozal Yildirim , Ru San Tan , U Rajendra Acharya 2019 stated that the deep learning model with an toe-to-toe structure on the quality 12-lead ECG signal for the diagnosis of MI is putforth. For this purpose, the foremost commonly used technique, convolutional neural network (CNN) is employed. With yielded impressive accuracy and sensitivity performance easily above 99% for MI diagnosis on all ECG lead signals, the trained CNN model was the proposed architecture. Thus, the proposed model has the potential to provide high performance on MI detection which may be utilized in wearable technologies and medical aid units. **A deep learning approach for ECG-based heartbeat classification for arrhythmia detection** [5 6] by G.Sannino ,G.DePietro a unique deep learning approach for ECG beat classification is studied. Experiments are performed on the well-known MIT–BIH Arrhythmia Database, and compared results with the scientific literature. The last word results show that the model isn't only more efficient than the state of the art in terms of accuracy, but also competitive in terms of sensitivity and specificity[8 9].

3. PROPOSED SYSTEM

Proposed system considers ECG signals for feature extraction but with good thing about fewer features. Firstly, the ready ECG dataset is taken and preprocessing is performed on the identical. Moreover missing,

ISSN: 2233-7857 IJFGCN

duplicate values are excluded and that we get a clean dataset. After the feature extraction to calculate the accuracy, here we apply discrete wavelet transform method therefore the ECG signal is decomposed up to 5 levels and so coefficients are obtained which useful for feature extraction. Considerable features like energy, entropy, fractal dimension, relative wavelet are extracted. Step ahead linear features are extracted from these. Overall the most aim is to think about less number of features for classification purpose.

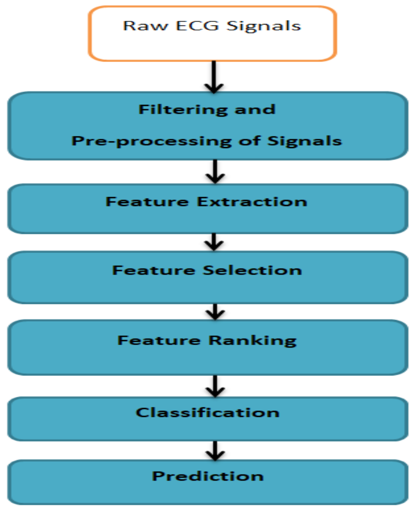

Fig - 1: System Architecture diagram

Feature ranking is completed to settle on the proper subset of feature. So, ANOVA (Analysis Of Variance) and Relief methods are used for the identical. The ANOVA method is where linear relationship between independent and dependent variables is computed and F-value determines the deviation between different classes. The upper F-value more is that the difference between the features. That's the rationale that the highly ranked feature are going to be having higher F-value and is fed to the classifier. For ranking, k nearest is used. The one having closet class selection the price of that feature is taken into account. Also capacity to discriminate among the classes decides the burden of the feature. The selected weights that are crossing the given threshold. Then classification is performed. Any of the classification algorithm may be used for this purpose. Accuracy given the priority by considering those algorithm which has highest accuracy are going to be the aim. The prediction is completed; the ECG segments classified into four classes i.e. normal, HCM, DCM, and MI.

4. METHODOLOGIES

4.1. Methods

As shown in diagram there are six phases ECG signals undergo through. So at last prediction is done that is from which disease patient is suffering from or the signal is normal. For preprocessing median filter is used to remove the noise and then features are extracted using discrete wavelet transform (DWT). By using Relief, ANOVA method feature ranking is done. Naïve Bayes classifier is used for classification and accordingly prediction is done.

4.2 Dataset

The ECG signals were obtained from mitbih_database from kaggle.

ISSN: 2233-7857 IJFGCN

4.3 Pre-processing of ECG signals

The pre-processing is very important and necessary phase to remove the noise which is present in images. If preprocessing is done properly then it will to poor signals and interpreting, evaluating them will be tedious task which will give incorrect result. So in proposed system median filter is applied.

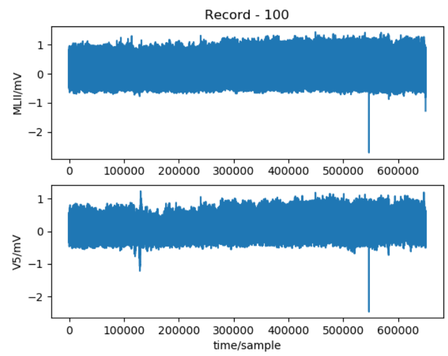

Fig 2 : ECG signals before applying filtering

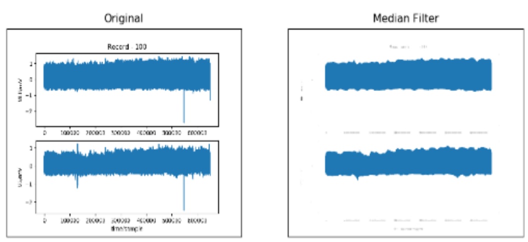

Fig 3 : After applying median filter

4.4 Feature Extraction

Linear and non-linear features are extracted from signals using discrete wavelet transform. In DWT the signals are divided into different frequency sub-bands. These sub-bands are passed through low-pass and high pass filters and approximate (A) and detailed (D) coefficients are evaluated. DWT is performed up to five levels. Energy, entropy, fractal dimension and relative wavelet are examples of few non-linear features. Entropy measures amount of deviation in ECG signals whereas energy gives energy of the signal. Uniformity of the signal is also shown with help of energy. Relative wavelet calculates relative entropy and shows energies of different sub-band signals which are obtained from DWT. Fractal dimension represents the self-similarity of the signal and evaluates complexity of the signal. Fuzzy and sample entropy gives unpredictability and regularity in the ECG signals respectively regardless of the length of the data. Combining linear and non-linear features together to build the model improves the accuracy for the same.

ISSN: 2233-7857 IJFGCN

4.5 Feature Selection

For every non-linear feature 56 related features are calculated. So it becomes important to remove redundant, insignificant features to boost the performance of the system. Otherwise it will be time consuming and complex task to build a system with these number of features. So subset of the feature from the extracted feature is selected to reduce number of features. Sequential Forward Selection (SFS) method is used to perform the feature selection. In this method evaluation function is calculated which reduces mean square error (MSE). Initially the set of selected features is empty and then after each step the features which are selected by SFS method are added to this set. The SFS method chooses the features from the available set of features. As a result set containing less number of distinct features is obtained.

4.6 Feature Ranking

The priority of the selected features should be decided and that comes under feature ranking. The feature which will give highest accuracy will be given higher priority. ANOVA and ReliefF are the methods used for feature ranking. ANOVA that is Analysis Of Variance is checks the means of two or more groups that are significantly different from each other. It assumes two hypothesis that is H_0 and H_1. H_0 assumes that means of the all of the groups is same whereas H_1 assumes that at least one of mean of the groups is different. So the relation between dependent and independent variables is computed using ANOVA. In ReliefF method the feature is selected by using near-hits and near-miss. Method makes use of Manhattan distance instead of Euclidean distance and updates the weight factor accordingly. Also ability to discriminate between the classes is considered in this method.

4.7 Classification

Several classification algorithms are applied on the dataset and the accuracy is found. From the below Table it is observed that Naïve Bayes classifier gave the highest accuracy. This classifier is based on bayes theorem and its assumption is the effect of one feature on the class is independent of others. All the features are considered independently. Bayes classifier is fast, accurate and reliable.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Fig 4 : Formula For Naïve Bayes Classifier

Where ,

P(h): the probability of hypothesis h being true (regardless of the data). This is known as the prior probability of h.

P(D): the probability of the data (regardless of the hypothesis). This is known as the prior probability.

P(h|D): the probability of hypothesis h given the data D. This is known as posterior probability.

P(D|h): the probability of data d given that the hypothesis h was true. This is known as posterior probability.

ISSN: 2233-7857 IJFGCN

Sr. No.	**Name Of The Classifier**	**Co-Relation Coefficient**
1.	Naïve Bayes	0.8932
2.	Random Forest	0.8456
3.	Linear Regression	0.2219
4.	kNN	0.8625
5.	Random Tree	0.7646

Table 1: Classification Results.

5. CONCLUSION

ECG signal dataset can detect Cardio Vascular diseases . Being the one that reveal important information about the heart, ECG signal inspects accurate diagnosis. But the problem here is that ECG signals are quite non-stationary making it tedious task to visualize every sub-segment to interpret. That is where our system will be helpful. Moreover the reduced feature makes it time feasible giving faster results. The median filter used in this system helped to remove noise which made the signals clear helping to interpret things significantly. The information reveals that following classifier algorithm is studied with the help of Weka Tool to find their co-relation coefficient and temporarily decide with what to move for further study. The results speaking for itself show that Naïve Bayes is making itself dominant so is decided to go for it. Also this proposed system is cost efficient and beneficial for human health.

6. REFERENCES

1) (Jilong Wang , Renfa Li, Rui Li , Keqin Li , Haibo Zeng , Guoqi Xie , Li Liu): Adversarial de-noising of electrocardiogram.

2) (R.K. Tripathy, Mario R.A. Paternina, Juan G. Arrieta, Alejandro Zamora-Méndez, Ganesh R. Naik 2019): . Automated detection of congestive heart failure from electrocardiogram signal using Stockwell transform and hybrid classification scheme.

3) (Zerina Masetic, Abdulhamit Subasi): Congestive heart failure detection using random forest classifier.

4) (Ulas Baran Baloglu , Muhammed Talo , Ozal Yildirim , Ru San Tan , U Rajendra Acharya 2019): Classification of Myocardial Infarction with Multi-Lead ECG Signals and Deep CNN.

5) (G.Sannino ,G.DePietro): . A deep learning approach for ECG-based heartbeat classification for arrhythmia detection

ISSN: 2233-7857 IJFGCN

6)Shwetambari Kharabe, C. Nalini,” Robust ROI Localization Based Finger Vein Authentication Using Adaptive Thresholding Extraction with Deep Learning Technique”, Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 07-Special Issue, 2018.

7)Shwetambari Kharabe, C. Nalini,” Using Adaptive Thresholding Extraction - Robust ROI Localization Based Finger Vein Authentication”, Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 13-Special Issue, 2018.

8)Shwetambari Kharabe, C. Nalini,” Evaluation of Finger vein Identification Process”, International Journal of Engineering and Advanced Technology (IJEAT), International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6S, August 2019.

9)Udayan Birajdar, Sanket Gadhave, Shreyas Chikodikar, Shubham Dadhich, Shwetambari Chiwhane, “Detection and Classification of Diabetic Retinopathy Using AlexNet Architecture of Convolutional Neural Networks”, Proceeding of International Conference on Computational Science and Application, online 05 January 2020, pp 245-253, Springer paper

ISSN: 2233-7857 IJFGCN