Volume 5, Issue 7, July 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

An Impact of Different Feature Extraction Methods on Classification of Electrocardiogram

P. S. Gawande
Department of Electronics, RSCE,
Buldana, India

S. A. Ladhake
Principal, Sipna College of Engineering,
Amravati, India

***Abstract**— This paper presents an impact of different feature extraction methods on classification of electrocardiogram. The classification system consists of three steps data acquisition, feature extraction and classification using MLP NN. ECG features were extracted using two different methods by using Tompkins algorithm and wavelet transform and three feature sets were formed using this. We have used statistical, fiducial and hybrid feature sets as an input to the neural network. The performance of the classification system is analyzed based on percent average classification accuracy and mean squared error. The optimum percent average classification of the ANN was 98.0556 and the mean squared observed to be 0.0064.*

***Keywords**— Feature Extraction, Multilayer Perceptron, Classification*

I. INTRODUCTION

Electrocardiogram is one of the best cardiac investigations available. It provides wealth of information for diagnosis and treatment. It is the result of electrical potential generated by the muscles of the heart. Since last decades ECG is potentially used as a biometric for person authentication.

In security systems biometrics plays important role to increase the security level. Biometrics is identification of an individual based on the physiological and/or behavioral characteristics. It is highly difficult to do falsification in case of biometric systems. Biometrics has a very old history. Many physiological and behavioral biometric features such as fingerprint, palm veins, face recognition, DNA, palm print, hand geometry, iris recognition, retina etc and behavioural parameters such as typing rhythm, gait, signature etc. are in use. The composition mechanism and electrical activity of the human heart inherit uniqueness from the individuality of DNA. Thus along with all these biometric tools, electrocardiogram can also be used as tool to distinguish the individuals. ECG as a biometrics is first time used by Lena Biel. The advantage of using ECG as a biometric is that ECG checks the aliveness of a person.

The ECG signal varies from person to person due to the differences in position, size, and anatomy of the heart, age, sex, relative body weight, chest configuration and various other factors. However, other than the changes in the rhythm, the morphology of the ECG is generally unaltered. Biometric gives less accuracy as compared to traditional recognition system but it can be preferred where it is tolerable to have false rejection than the false acceptance. The benefit of using ECG as a biometric is its security level. Artificial neural network can be used for the classification of the multi class ECG signal.

II. LITERATURE SURVEY

The literature survey is organized based on feature extraction technique whether it is fiducial, statistical or hybrid.

A. *Fiducial Based Approaches*

Biel et al.'s (2001) for the first time used fiducial feature extraction algorithm, which demonstrated the feasibility of using ECG signals for human identification. Total 180 features were extracted and are reduced to 10 with the help of variance using PCA. They have reported the classification accuracy of 100% during experimentation. The disadvantage of this work is that their approach is machine dependent.

Shen et al. reported ECG based recognition method with seven fiducial based features that relate to the QRS complex. The underlying idea was that ECG is less affected by varying heart rates and also very easy to detect, and it is appropriate for biometric recognition. Further in 2005 they have used 17 fiducial features which include different amplitudes, area slope and angle between them.

Masaki Kyoso and Uchiyama (2003), proposed four fiducial based feature parameters *P* wave duration, *PQ* interval, *QRS* complex and *QT* durations. These features were identified on the pulses by applying a threshold to the second order derivative. In all total 9 fiducial features were considered.

Steven Israel et. al.(2004) features extracted were from the P, R, and T complexes. The fiducial points were extracted in the time domain in two stages. Features include start and end of atrial depolarization (P wave) and ventricular repolarization (T wave), this was done by fixing four new fiducial positions at P and T wave. They have also extracted various fifteen normalized distances

Venkatesh and Shrinivasan (2010), after doing preprocessing detected QRS complex using Tompkins method and determined RR interval. Feature vector selected consists of P wave interval, T wave interval, ST interval, PR interval, QRS interval and QT interval.

Yogendra and S. K. Singh (2011), in order to validate ECG as a biometric for individual authentication, they have prepared a feature set from the extracted fiducials of *P*, *Q*, *R*, *S*, and *T* waves from each heartbeat The feature set contains the attributes of different classes: interval features, amplitude features and angle. Thirteen interval features from different time instances of the dominant fiducials of *P*, *Q*, *R*, *S*, and *T* waves were computed. Four amplitude features were computed relative to the amplitude of *R* peak and the angle class of features were related to angular displacement between different peak fiducials of *P*, *Q*, *R*, *S* and *T* waves.

Sasikala and Wahidabanu (2010) detected QRS complex based on modulus maxima of the discrete wavelet transform. Detection rules were applied to wavelet transform of ECG signal. The fiducials selected as a feature vector matrix are PR, RQ, RS, RT, PS, TS, PQ and TQ amplitudes.

Y Gahi et. al.(2009) in the system proposed they have extracted total 24 features which includes 18 temporal and 6 amplitude features. Later they use a sequential forward selection algorithm to get the optimal subset of features 12 features are finalized.

B. Non Fiducial Based Approaches

Plataniotis et. al. (2006) used non-fiducial approach for the first time to extract features from ECG for biometric authentication. They proposed auto correlation based feature extraction. They proposed that the auto correlation of windowed ECG signal is highly discriminative information in a population. Depending on original sampling frequency of the signal the dimensionality of the signal was considerably high. To reduce dimensionality, discrete cosine transform was used.

Wubbeler et al. (2007), have also reported an ECG based human recognizer by extracting biometric features from a combination of Leads I, II and III to locate and extract position of R peak a thresholding procedure was applied. ECG signals were recorded from 74 subjects and the number of signals recorded for person to person were from 2 to 20.

Chan (2008), used multiplication of backward differences algorithm and temporally aligned using a cross-correlation measurement for each data sequence to detect PQRST complexes. Correlation coefficients were computed between PQRST complexes. Any PQRST complex with a correlation coefficient below one standard deviation of the mean correlation coefficient was discarded to avoid the inclusion of PQRST complexes that had been corrupted by large intermittent noise. The remaining PQRST complexes were used to compute a signal averaged ECG (SAECG), to reduce power-line interference and the effect of low frequency drifts. Classification was performed on the basis of percent residue difference, correlation coefficient and wavelet distance measure. Non-fiducial features considered by them were difference between two signals, correlation and detailed coefficients after decomposing signal by using wavelet transform.

Yongbo and Jianchu (2008), used wavelet transform. With the bior1.1 wavelet decomposition, the signal cycles were converted into wavelet coefficient structures, in the form of vectors with 256 elements as a feature.

Molina et al. (2007) proposed a methodology in which an ECG heartbeat was normalized and compared with its estimate which was constructed from itself and the templates from the claimed identity. The estimated version was produced by a morphological synthesis algorithm involving a modified dynamic time warping procedure.

Mohamed Tawfik and Hany (2010) had taken out features from QRS complex and T wave. They further normalized them in time domain by using Framingham correction formula by assuming constant QT interval. Discrete cosine transform coefficients were used as an input feature vector matrix showing its significance as useful information gets concentrated in few coefficients using DCT.

C. Hybrid Approaches

Yogin Wang et. al. (2008), have used both fiducial and non fiducial features. They have used 15 temporal features and 6 amplitudes fiducial features for recognition. For non fiducial features they have proposed combination of autocorrelation (AC) and discrete cosine transform (DCT) which involved four stages firstly windowing, where the preprocessed ECG trace was segmented into non overlapping windows, secondly estimation of the normalized autocorrelation of each window, thirdly discrete cosine transform over L lags of the auto correlated signal and lastly classification is made based on significant coefficients of DCT.

Tsu Wang Shen et.al. (2011) used correlation coefficients and LDA distance classification was used to prove the match and 17 fiducial features were used for final classification.

Fiducials are specific points of interest on the ECG heart beat. Therefore, fiducial based approaches rely on local features of the heart beats such as the temporal, amplitude difference between consecutive fiducial points or angle between them. The disadvantage of multiple fiducial based approaches is that their identification performance is affected by the accuracy of fiducial point detection. Furthermore, there is no universal standard for defining the boundaries of the ECG wave features. Fiducial points detected by ECG devices are approximate locations and do not satisfy biometric system requirements because even slight variation in the locations will result in misclassification. The P and T waves are sometimes too small to detect and are time variant, preventing exact extraction of fiducial features such as the duration and width of the P and T waves. Also the spectrum of the P wave is considered to be about 10–15 Hz. When the heart rate changes, the magnitude and duration of the P wave also changes slightly, same is the case with S-T segment. When the heart rate changes, the S-T interval is altered significantly. When the heart rate increases, the magnitude of the T wave increases and the duration decreases, with an almost linear correlation to heart rate.

The identification performance of single fiducial based approaches is not affected by the accuracy of fiducial point detection, but these methods extract feature coefficients based on a fixed-length heartbeat signal and do not consider the effects of heart rate variability (HRV). In this study, to overcome this we proposed the use of hybrid features. Classification is made on three data sets, first set is of fiducial features, second set consists of statistical features and third set is of hybrid features. Feature selection plays an important role in classifying systems such as neural networks. Features are selected based on best representation of a given class of signals and best distinction between classes.

III. DATA USED

Input to the MLP neural network model is an ECG signals of eight normal persons recorded during thirty six months. Sixty six signals of each person are taken as an input, thus total 528 signals forms the database. The length of each signal is taken in such a way that each of the signal contains atleast five ECG cycles. The ECG signals were recorded using 12 lead ECG recorder "SAMVED". The sampling rate is 400 s/s. ECG recording of one minute is made during each sample. Out of the 12 lead signals received, lead II signal is used for the analysis purpose.

IV. FEATURE EXTRACTION

Three sets of features fiducial, statistical and hybrid were prepared so as to apply as a input to train the neural network.

A. Fiducial Feature Extraction

We have extracted fiducial features related to QRS complex using Tompkin's algorithm. The reason to select features related to only QRS complex is that QRS complex is considered to be fairly constant and doesn't change with the change of heart rate as it reflects the depolarization of ventricular muscle. However the action potential is carried on the ventricle through the high speed Purkinje fibers that travels only a short distance over the ventricle between the endocardium and the epicardium. Feature set 1 (S1) consists of three features average QQ interval, average RR interval and average QR interval, selected from QRS complex.

B. Statistical Feature Extraction

The discrete wavelet coefficients were computed using the MATLAB software tool. The computed discrete wavelet coefficients of the ECG signals of each record were used to calculate statistical features. In this study we selected db5 mother wavelet which are similar in shape to the ECG signal and have scaling function similar to ECG signal and the number of decomposition levels were chosen to be 5. The following seven statistical features (S2) were extracted:

S2= [Standard deviation, Entropy, Covariance, Energy, Maximum, Minimum, Mean]

C. Hybrid Features

Hybrid feature set consists of combination of fiducial and statistical features as stated earlier. Total ten hybrid features forms feature set S3 which were applied as a input to the neural network.

V. METHODOLOGY

The use of MLP NN resides in its simplicity. The MLP NN is referred to as error back-propagation network as the error is propagated in the backward direction and the network weights are adjusted accordingly to move the response closer to the target. Once the network gets trained, it acts in a feed-forward manner. While designing MLP NN we have used six activation functions Tanh, Sigmaoid, LinearTanh, Linear Sigmoid, Softmax and Bias, learning rule used were Momentum, Cinjugate Gradient, Levenberg Marquer, Quickprop and Delta Bar Delta. The number of hidden layer was restricted to one, and the processing elements in the hidden layer were varied from 2 to 50. Each network is trained three a times and number of iterations were 500. Table 1 shows the Percentage Average classification Accuracy (PACA) and mean squared error (MSE) by using single hidden layered MLP classifier, with input Features S1, S2 and S3

Table 1: Percentage Average classification Accuracy and Mean Squared Error using MLP NN for feature set S1, S2 and S3.

Feature Set Used	PACA			MSE		
	Test	CV	Train	Test	CV	Train
Fiducial (S1)	71.80093	73.40241	82.42139	0.035988	0.141079	0.004344
Statistical (S2)	62.69305	66.312	62.4997	0.053773	0.135328	0.006545
Hybrid (S3)	**98.0556**	**100**	**98.7763**	**0.048824**	**0.137592**	**0.002292**

VI. RESULTS

During this experimentation we have analyzed the effect of different feature sets on the classification of ECG signal using multilayer Perceptron Neural Network. The three data sets fiducial, statistical and hybrid were applied as input to the network. The networks were trained as mentioned in the methodology. It can be seen in table 1 that the classification accuracy is less for fiducial and statistical features. From the results it is clear that the percentage average classification accuracy is highest during training, testing as well as cross validation for hybrid features, this can also be observed in figure 1. Further from table 1 and figure 2 we can see that the mean squared error is minimum during training, testing and cross validation for hybrid features.

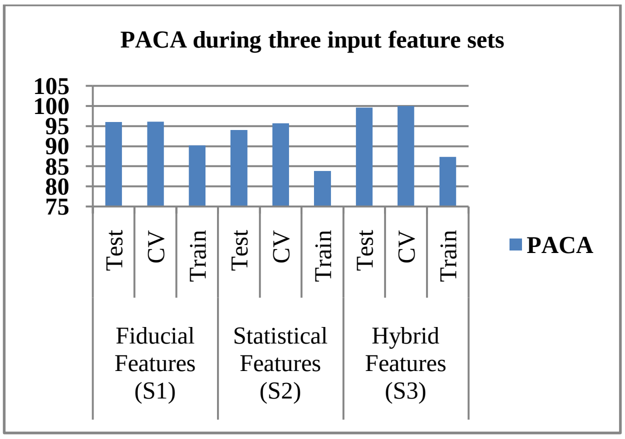

Figure 1: PACA of MLP NN for fiducial, statistical and hybrid features

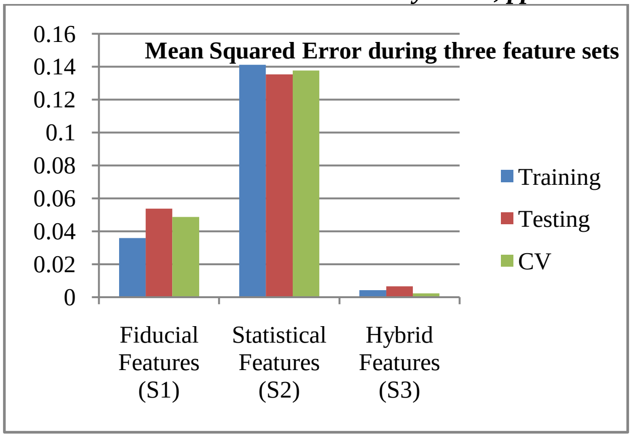

Figure 1: MSE of MLP NN for fiducial, statistical and hybrid features

VII. CONCLUSION

Feature extraction plays very important role in classification, it helps to increase classification accuracy. From the results it is clear that the hybrid features have showed beat percentage average classification accuracy and minimum mean squared error.

REFERENCES

[1] A.K. Jain, A. Ross, S. Prabhakar, "An introduction to biometric recognition," IEEE Trans. Circuits. Syst. Video 14(0), 4–20 (2004).

[2] Mykola Pechenizkiy, "The Impact of Feature Extraction on the Performance of a Classifier: kNN, Naïve Bayes and C4.5", Springer-Verlag Berlin Heidelberg 2005, pp. 268 – 279.

[3] K. Akilandeswari, R. Sathya, "Feature extraction of ECG signals for early detection of heart arrhythmia", *International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 12, December 2014.*

[4] L. Biel, O. Petersson, L.P. PhilipsonWide,"ECG Analysis: a new approach in human identification," IEEE Trans. on Instrumentation and Measurement,Vol. 50(3), pp.808–812 (2001).

[5] T.W. Shen,W.J. Tompkins, Y.H. Hu, "One-lead ECG for identity verification," In Second Joint EMBS/BMES Conference Housten, TX USA (2002) pp. 62–63.

[6] T. W. Shen, W. J. Tompkins , "Biometric Statistical Study of One-Lead ECG Features and Body Mass Index (BMI)", Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005, pp. 1162-1165.

[7] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognit.*, vol. 38, pp. 133–142, 2005.

[8] Y.Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 19, pp. 19:1–19:11, 2008.

[9] J. Yao and Y. Wan, "A wavelet method for biometric identification using wearable ECG sensors," in *Proc. 5th Int. Summer School and Symp. Medical Devices and Biosensors, 2008 (ISSS-MDBS 2008)*, pp. 297–300.

[10] H. Silva, H. Gamboa, and A. Fred, "One lead ECG based personal identification with feature subspace ensembles," in *Proc. 5th Int. Conf. Machine Learning and Data Mining in Pattern Recognition*, Leipzig, Germany, 2007, pp. 770–783.

[11] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 2, pp. 248–253, Feb. 2008.

[12] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, and J. W. Rohrbaugh, "ECG biometrics: A robust short-time frequency analysis," in *Proc. 2010 IEEE Int. Workshop Information Forensics and Security (WIFS)*, Dec. 2010, pp. 1–6.

[13] F. Sufi, L. Khalil, and J. Hu, "ECG-based authentication," in *Handbook of Information and Communication Security*, P. Stavroulakis and M. Stamp, Eds. New York: Springer, 2010, pp. 309–331.

[14] M. M. Tawfik and H. S. T. Kamal, "Human identification using QT signal and QRS complex of the ECG," *Online J. Electron. Elect. Eng.*, vol. 3, no. 1, pp. 383–387, 2011.

