ECG-based Biometric Authentication using Empirical Mode Decomposition and Support Vector Machines

Sumair Aziz
Electronics Engineering Department
University of Engineering and Technology, Taxila
Taxila, Pakistan
sumair.aziz@uettaxila.edu.pk

Muhammad Umar Khan
Electronics Engineering Department
University of Engineering and Technology, Taxila
Taxila, Pakistan
sa.umarkhan@gmail.com

Zainoor Ahmad Choudhry
Biomedical Engineering Technology Department
University of Engineering and Technology, Taxila
Taxila, Pakistan
zainoorah15@gmail.com

Afeefa Aymin
Biomedical Engineering Technology Department
University of Engineering and Technology, Taxila
Taxila, Pakistan
annie162445@gmail.com

Adil Usman
Electronics Engineering Department
University of Engineering and Technology, Taxila
Taxila, Pakistan
adil.usman@uettaxila.edu.pk

***Abstract*—Electrocardiogram (ECG) is an electric signal of cardiac activity posing highly discriminative properties related to human recognition. ECG based authentication has gained much success in recent times however discriminant feature extraction and efficient pattern classification still encounter numerous challenges. This paper proposed a novel methodology for ECG based biometric authentication system. Proposed method first denoise single lead raw ECG signal through empirical mode decomposition (EMD). Region of interest from ECG signals having maximum characteristic information related to subject's recognition is also extracted through EMD. Next, feature extraction is performed by combination of five features from statistical, time and frequency domains. Finally, selected features were categorized with range of different classification methods such as Support Vector Machines (SVM), K-nearest neighbor (KNN) and Decision Tree (DT). 10-fold cross validation based classification evaluation reveals that SVM with cubic kernel achieves best accuracy of 98.7%, sensitivity of 100% and 98.8% specificity for successful classification of 14 subjects.**

Keywords—Biomedical recognition system, Electrocardiogram, Biometric authentication, Empirical mode decomposition, Support vector machines, Feature extraction

I. INTRODUCTION

In today's world of automation and technology, biometrics has become a major tool for the purpose authentication and identification. A major advantage of a biometric system is its full dependency on the individual. At the same time where fingerprints and face recognition are used for this purpose a new type of biometrics i.e. the medical biometrics is gaining pace. Medical biometrics is used to achieve the same motive of identification and authentication by using biological signals like Electrocardiogram (ECG). Biometric features are the characteristics that are unique to a single individual that acts as a basis for the identification of that individual from the rest of the population. Out of all the medical signals, ECG is the ultimate signal used for this purpose because it is unique for every individual and this uniqueness can be exploited by looking at different features of ECG[1]. One of the biggest advantage of using a ECG biometric system is that it also guarantees the presence and aliveness of the person and also it is most difficult to counterfeit as compared to fingerprint of a person which can be forged , voice recording of a person in his absence can be used and iris images can be used in iris based recognition but such tactics cannot be used in ECG biometrics[2].

Electrical currents are generated at the Sinoatrial node (SA) node of heart and travel down to the Atrioventricular (AV) node and spread not only within the heart but also throughout the body[3]. These electrical currents known as the ECG can be measured with the help of surface electrodes. ECG signal consists of P wave that represents atrial depolarization or more commonly known as contraction of atria. The QRS complex in ECG signal shows ventricular depolarization or ventricular contraction and T wave represents ventricular repolarization or ventricular relaxation. Normal ECG signal is shown in Fig. 1.

In order to understand how ECG is suitable for a biometric system, we need to know the requirements of a biometric system. An "ultimate" biometric characteristic as described by [3] should be:

- Universal in terms of each individual possessing it
- Easily measured
- Unique
- Permanent, i.e., it cannot be forged

ECG based authentication was performed through self-developed algorithm using two electrodes ECG [4]. This work utilized Physionet [5] ECG dataset and achieved 84.93% accuracy. Classification of various subjects by extracting

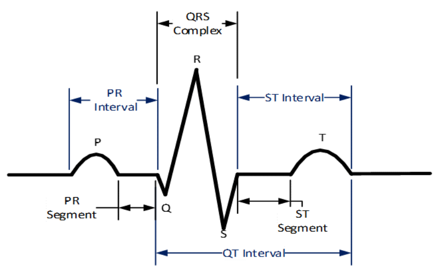

Figure 1: Components of ECG signal

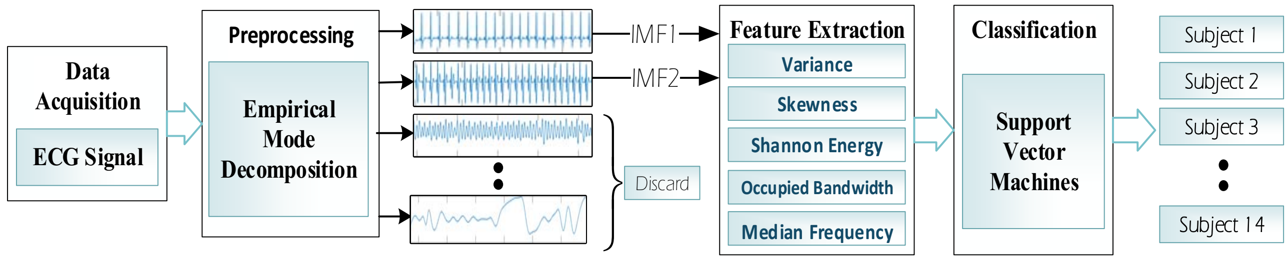

Figure 2: Sketch of the proposed ECG-based Biometric System

multiscale features from ECG signal was performed in [6]. The data used was taken from Physikalisch-Technische Bundesanstalt (PTB) database [7] and reported true positive rate of 91.67%. In [8], non-fiducial point base technique was utilized for defining features. Authors used MIT-BIH Normal Sinus Rhythm database from PhysioNet.

In [9], authors proposed a method for authentication using ECG signals obtained from chest sensors. Feature extraction was performed through dynamic time warping with an error rate of 6% to 13%. In another similar study [10], authors proposed an ECG authentication system. Feature extraction was performed through Discrete Wavelet Transform (DWT). Sum of Squared Difference technique was used for template matching. An accuracy of 91 % was achieved. In [11] ECG signals taken from Physionet database [7] were classified through neural network and SVM. The accuracy rate was 96.6% and 97.6%.

In [12], ECG biometric authentication system including dataset from BioSec Lab [13, 14], based on Gaussian one class and binary SVM classifiers was proposed. Data in this case was preprocessed through DWT. In [15], authors presented authentication technology in which data was collected after extensive exercise. 55 subjects were included in experimentation. Biopac MP150 was used for data acquisition and classification was done by linear discriminant analysis (LDA). A maximum accuracy of 96.22% within 5 minutes recordings was achieved.

Human recognition system using single lead electrocardiogram (ECG) from PTB dataset [7] was proposed in [10]. Finite Impulse Response (FIR) high pass filter was used for preprocessing ECG signal. Haar wavelet transform was applied for detection and an accuracy of 97.12% was achieved. Another similar study used a 12-lead resting ECG database for ECG biometric authentication [16]. Cross correlation method and amplitude measurements were used for feature extraction and LDA as a classifier to achieve maximum accuracy of 96.13%. Authors in [17] proposed strategies for Electrocardiogram (ECG) based identification using Deep Neural Network. An average accuracy of 94.39% was obtained.

In [18], authors proposed a Cascaded PCA Network that was based on principal component analysis network (PCANet) to build an ECG biometric system and an accuracy of 93.3% was achieved. Deep convolutional neural network based on three kernels was utilized in [19] to obtain 4.74% error rate. Reference [20] used a neural network based classification for live ECG biometric system with an accuracy of 97.9%. In [21] four different nonlinear methods were applied for feature extraction and SVM was used as a classifier and an accuracy of 99.06±0.26% was achieved. In [22] ECG data from MIT-BIH database was used on which filters and multi-layer neural networks were used to design an ECG recognition system with a maximum accuracy of 98.65%. Authors in [23] proposed a system based on the diffusion maps algorithm and the Scattering Transform with an accuracy of 97.25%.

In this work, we applied empirical mode decomposition (EMD) for removing artifacts and extraction of region of interest from raw ECG signal. Afterwards, feature extraction is performed by extracting useful and representative features of time, frequency and statistical domain. Selected five features were analyzed through variety of classification methods and SVM based classifier achieved best performance. Rest of the article is structured as follows. Section II explains the data acquisition setup and protocols. Detailed discussion of proposed methodology is also presented here. Section III presents the experimental procedures and results. Section IV concludes this article.

II. Materials and Methods

Fig. 2 illustrates detailed block diagram of the proposed ECG based biometric classification system. Data acquired from ECG electrodes is preprocessed through EMD. EMD decomposes input signal into sub-components called intrinsic mode functions (IMFs). Region of interest is extracted from ECG signal that carry discriminative information about every individual/subject. Redundant information and noise are discarded by removing those signal components from resultant preprocessed signal.

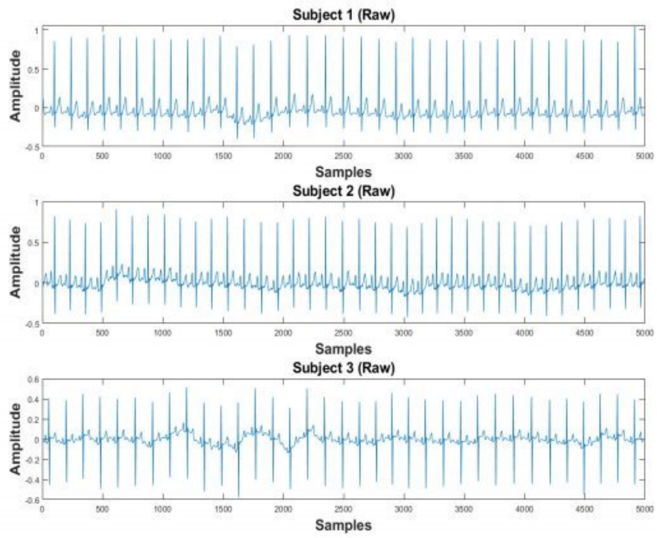

Figure 3: Time domain ECG signal representation of different subjects

IMF1 and IMF2 were added to get the required preprocessed signal, while other IMFs are ignored. Next, feature extraction is performed on preprocessed signal. ECG signal characterization is achieved through combination of Shannon energy, occupied bandwidth, median frequency and statistical features such as variance and skewness. Selected discriminant features are fed to multiclass SVM model for classifying different classes.

A. Data Acquisition

ECG signal data for this study is acquired using BIOPAC systems. SS2L Electrode Lead Set and body surface electrodes were utilized for collecting ECG data. The electrodes were clipped to right forearm, left leg and right leg. The dataset accommodates ECG records of 14 subjects which includes 8 males and 6 females at rest. All ECG records are sampled at sampling frequency of 1000 Hz. Each raw data file is segmented into smaller equal files containing 10,000 samples each. These segmented files are then preprocessed and fed into the classifier. Time domains and frequency domains of the raw signals acquired are shown in Figures 3 and 4.

B. Preprocessing and segmentation

The Raw ECG signals is disrupted due to a number of noises like artifacts, power line interference etc. In order to make it suitable for a biometric system, this signal needs to be preprocessed or denoised. Empirical mode decomposition (EMD) is a recent and adaptive method that expands a signal into a compression of Intrinsic Mode Functions (IMFs) [24-26]. EMD is especially suitable for signals that show non-linearity and are non-stationary like ECG [27].The initial IMFs represent high frequency information while the latter shows low frequency information and in case of ECG represents artifacts. Every IMF should have certain properties;

- Total extrema and zero crossings of the IMFs should either be equal or differ only by one.
- There should be symmetry of IMF with respect to the zero local mean[25].

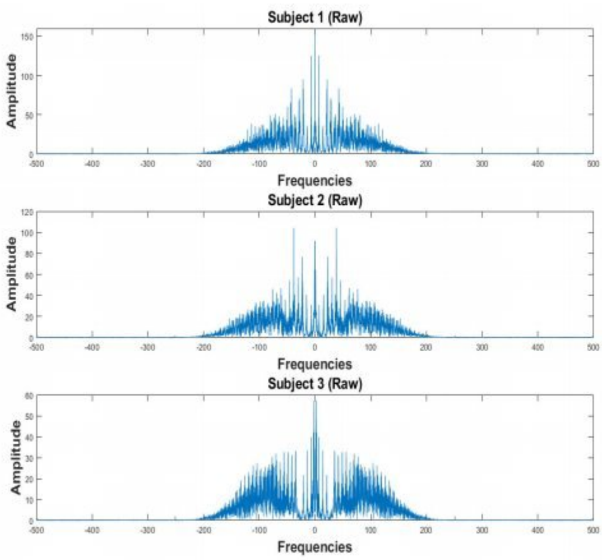

Figure 4: Frequency domain raw ECG signal representation of different subjects

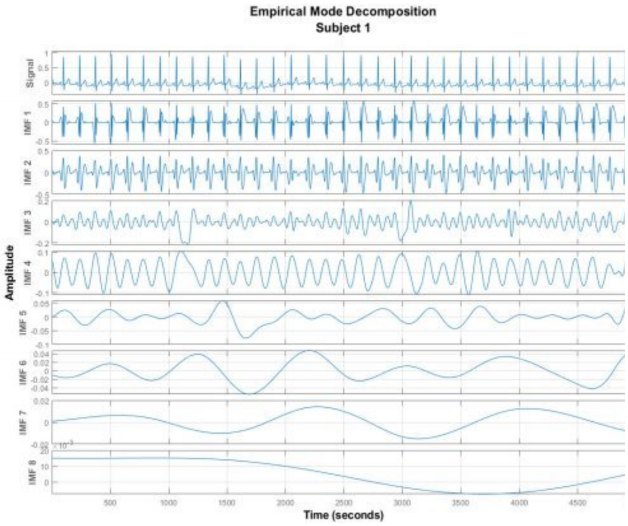

Figure 5: IMFs extracted with EMD for subject 1

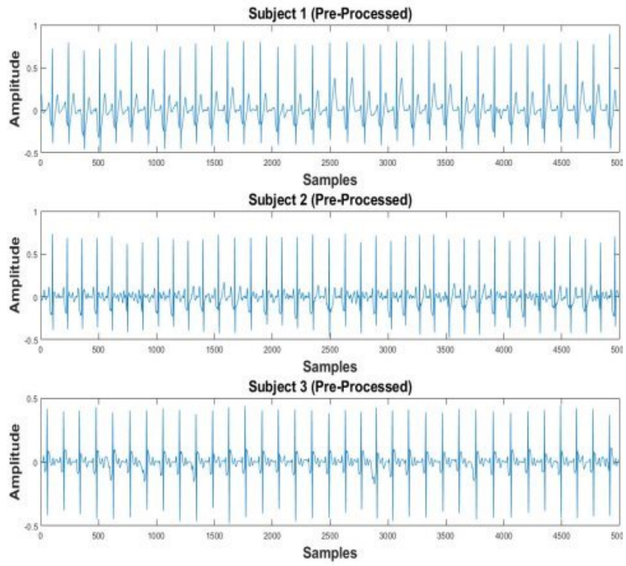

Figure 6: Time domains of preprocessed signals

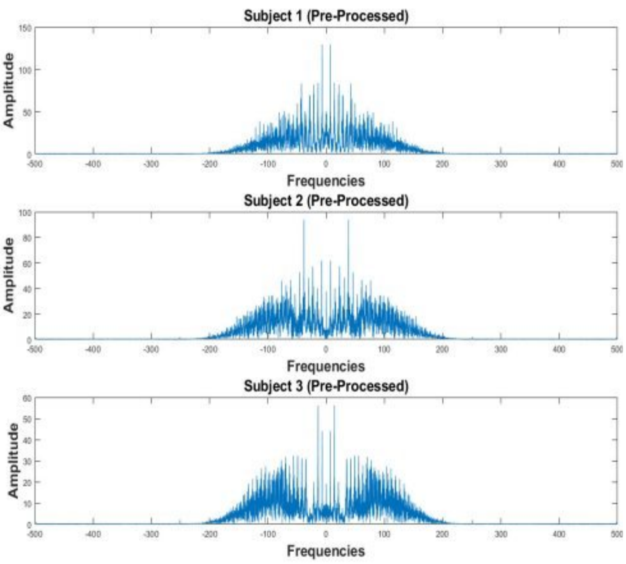

Figure 7: Frequency domains of preprocessed signals

EMD is applied to the raw ECG signal acquired and a number of IMFs and residual are obtained as shown in the Fig. 5.

As it is evident from Fig. 5 that only IMF 1 and IMF 2 represent the denoised ECG while the rest represent noise. So in order to make the signal suitable for the biometric system we make a new signal by combining IMF 1 and IMF 2 .Time domains and frequency domains of the pre-processed signal are shown in Fig. 6 and 7. This selection process also performs region of interest extraction from ECG signal for our target biometric application, as only those IMFs are selected which carry highly decisive information related to different subjects. Information which is redundant and could deteriorate classification performance is removed from resultant preprocessed signal.

C. *Feature Extraction*

An ECG biometric system is based upon the recognition of ECG of different subjects. A good recognition system should depend upon features that are able to distinguish the signal of a particular individual from another individual [28]. After rejecting noisy signal segments and extraction of region of interest, ECG signals were distinguished on the basis of following extracted features;

- Shannon energy
- Skewness
- Variance
- Occupied bandwidth
- Median frequency

i) Shannon Energy

Shannon energy is a parameter that is used to calculate the average spectrum of the energy of the signal. It discounts the high components into low components and hence the input amplitude is of no significance. Shannon energy calculates the spectrum energy of each sample[29]. It is described mathematically in equation (1);

$$SE = -|a[n]|\log(|a[n]|) \tag{1}$$

where, n is after normalization process

ii) Skewness

Skewness is another important feature that is described as the average of the cubed deviation from the mean divided by the cubed standard deviation. Skewness is a measure of the symmetry of the distribution of the samples around the R peak region and its value can be positive, negative or undefined [30]. The mathematical expression for skewness is given in equation (2) below**;**

$$Skewness = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3 / \mathrm{N}}{S^3} \tag{2}$$

Where,

$\bar{Y}$ represents the mean

S is the standard deviation

N shows data points of the normal ECG signal

iii) Variance

Variance is defined as the squared variation of a variable from its mean value. It measures how far a set of random values is spread from its normal mean value. In a given signal like ECG, the larger the variance is, the more big of amplitude variation[31]. Mathematically variance is defined in equation (3);

$$\sigma^2 = \frac{1}{N}\sum_{i=0}^{N-1}[X_i - \bar{X}]^2 \tag{3}$$

Where,

N shows length of the signal

$\bar{X}$ is the mean of the signal

iv) Occupied Bandwidth

Occupied Bandwidth abbreviated as (OBW) is the bandwidth containing 99% of the total integrated power of the transmitted spectrum, centered on the assigned channel frequency. It is usually concerned with the QT complex of the ECG signal [32].

v) Median Frequency

Median frequency is another feature that aids the discrimination of ECG signals in a group of different subjects. Median frequency is defined as the normalized frequency of the median of power spectrum of the R peak region.[33]

D. *Classification – Support Vector Machines*

Support Vector Machines (SVM) is widely applied as a best choice classifier for biomedical signal analysis applications. SVM is a pattern identification method [28] in which a set of training features is segregated by SVM with a maximum margin from hyper plane. In case when linear separation is not possible, non-linear kernel modifications can be applied. Different kernels quadratic, polynomial and radial basis function can be opted [34]. The choice of proper kernel function relies on specific data [35]. The employed features include Shannon energy, Skewness, Variance, Occupied

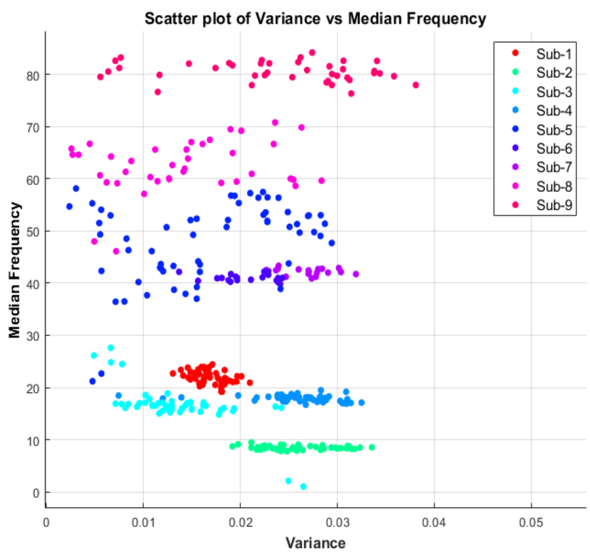

Figure 8: Feature distribution of variance and median frequency

bandwidth and Median frequency. The classifier resulted in accuracy of 99.2%.

III. Results and Discussions

In this work, ECG based biometric authentication framework is proposed. Proposed framework denoise raw ECG signal through EMD and extracts five different features. Comprehensive classification performance with selected features is evaluated over range of different classifiers and their variants. Multiclass SVM with linear kernel (SVM-L), SVM with Quadratic kernel (SVM-Q), SVM with Cubic kernel (SVM-C), SVM with Gaussian kernel (SVM-G), decision tree (DT), weighted KNN (KNN-W), KNN with k =5 (KNN-M) and KNN with cubic distance (KNN-C) were trained and testing through 10-fold cross validation scheme. In 10-fold cross validation, whole dataset is divided into 10 subsets, each subset is used as test set, while remaining 9 sets are combined to form training set. This process is iterated 10 times and results are averaged over all iterations. Standard statistical performance indices such as accuracy, sensitivity, specificity and error rate were computed to analyze the classifier performance. All experiments were performed in MATLAB 2018 software running on computer having Core i5 processor and 8 GB RAM. All results presented in this paper are averaged over 100 iterations.

Table I: Performance evaluation over different classification methods

Classifier	Accuracy	Sensitivity	Specificity	Error
SVM-L	97.03	100.00	99.82	2.97
SVM-Q	98.69	100.00	99.91	1.31
SVM-C	**98.72**	**100.00**	**99.82**	**1.28**
SVM-G	92.67	96.40	100.00	7.33
KNN-M	90.20	99.40	99.60	9.80
KNN-C	89.30	99.20	99.58	10.70
KNN-W	93.29	100.00	99.76	6.71
DT	96.38	98.20	99.67	3.62

Efficient feature extraction leads to better classification performance. Fig. 9 illustrates feature distribution among variance and median frequency, clear interclass separation of features can be observed. Table I presents evaluation results of selected feature with range of classifiers. Except only one classifier i.e. KNN-C, all other classification methods yield

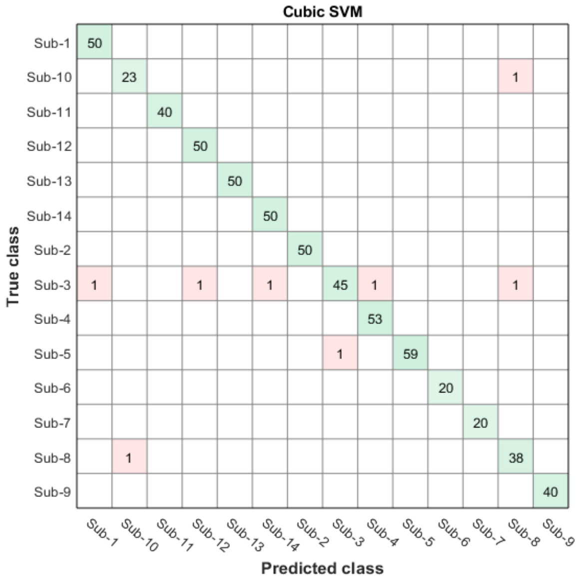

Figure 10: Class-wise performance evaluation in the form of confusion matrix

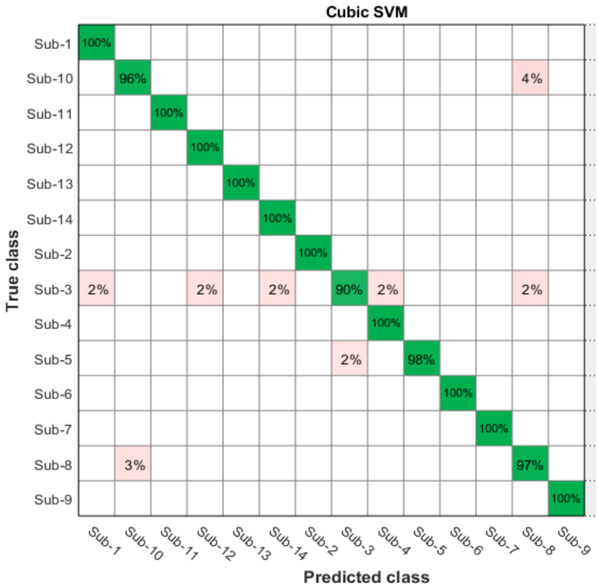

Figure 11: Class-wise accuracy with SVM-C classifier

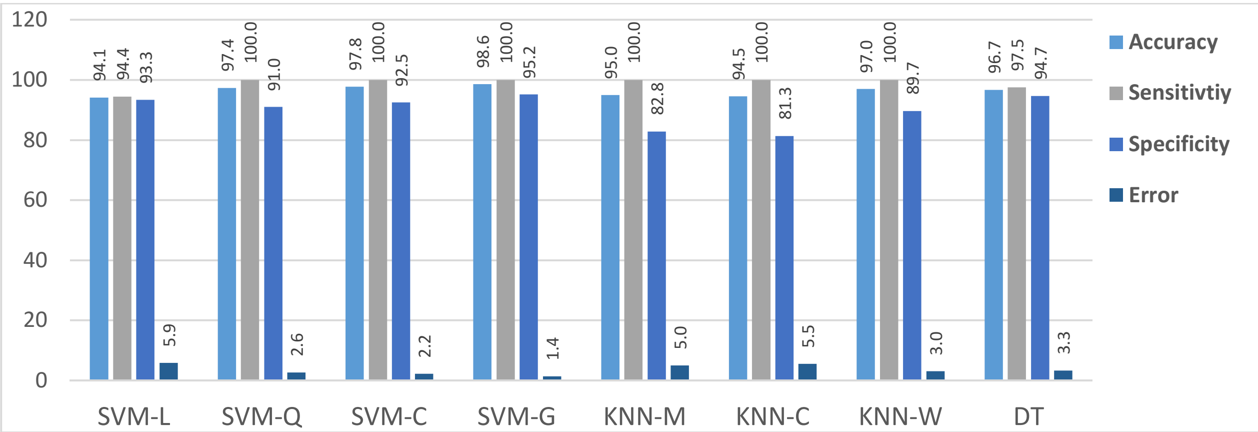

Figure 9: Comparison of different classification methods

accuracy more than 90%. KNN-C achieves 89.30% classification accuracy with highest error rate of 10.70%. Fig. 9 shows graphical comparison of different classifiers in terms of accuracy, sensitivity, specificity and error rate. SVM-C achieved best classification accuracy of 97.8%, followed by SVM-Q which reached 97.4% accuracy.

Fig. 10 presents class-wise performance for 14 subjects with best performing classifier (SVM-C) in the form of confusion matrix. It can be observed that feature data pertaining to subject 5, 8 and 10 got 1 miss-hit each. Whereas subject 3 got 5 miss-hits, i.e. misclassification. Fig. 11 shows the true positive rate for each class on SVM-C classifier. Subject 3 class achieved 90% accuracy due to 5 miss-hits, while all other classes achieved better accuracy.

IV. Conclusions

This study has proposed ECG based biometric authentication system. Proposed method employed EMD for region of interest extraction and signal denoising. Combination of time, frequency and statistical domain features were extracted to distinguish different data classes. Selected features were tested with eight different classification methods. SVM-C achieved highest classification accuracy of 98.72% with 10-fold cross validation strategy. Dataset for this study was collected from 14 different human subjects. Experimental analysis reveals that proposed method is reliable, accurate and computationally less expensive as compared to other studies. In future, we aim to expand the dataset by collecting ECG signals from more subjects in order to design highly reliable solution for authentication. We also aim to design embedded system for real time biometric applications based on proposed methodology. Such embedded systems can be used in various fields like eHealthcare to avoid the breach of medical information and banking etc.

References

[1] P. Huang, L. Guo, M. Li, and Y. Fang, "Practical Privacy-preserving ECG-based Authentication for IoT-based Healthcare," *IEEE Internet of Things Journal,* 2019.

[2] S. S. Abdeldayem and T. Bourlai, "ECG-based Human Authentication using High-level Spectro-temporal Signal Features," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 4984-4993: IEEE.

[3] T. S. J. P. Lugovaya, "Biometric human identification based on ECG," 2005.

[4] J. S. Arteaga-Falconi, H. Al Osman, A. J. I. T. o. I. El Saddik, and Measurement, "ECG authentication for mobile devices," vol. 65, no. 3, pp. 591-600, 2015.

[5] Y. N. Singh and S. K. Singh, "Evaluation of electrocardiogram for biometric authentication," 2011.

[6] M. K. Bashar, Y. Ohta, and H. Yoshida, "ECG-based biometric authentication using mulscale descriptors: ECG-based biometric authentication," in *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, 2015, pp. 1-4: IEEE.

[7] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," vol. 101, no. 23, pp. e215-e220, 2000.

[8] S. Gutta, Q. J. I. j. o. b. Cheng, and h. informatics, "Joint feature extraction and classifier design for ECG-based biometric recognition," vol. 20, no. 2, pp. 460-468, 2015.

[9] S. Šprager, R. Trobec, and M. B. Jurič, "Feasibility of biometric authentication using wearable ECG body sensor based on higher-order statistics," in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2017, pp. 264-269: IEEE.

[10] A. Rehman, N. A. Saqib, S. M. Danial, and S. H. Ahmed, "ECG based authentication for remote patient monitoring in IoT by wavelets and template matching," in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2017, pp. 91-94: IEEE.

[11] Y. Zhang and J. Wu, "Practical human authentication method based on piecewise corrected Electrocardiogram," in *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2016, pp. 300-303: IEEE.

[12] M. Hejazi, S. Al-Haddad, S. J. Hashim, A. F. A. Aziz, and Y. P. Singh, "Non-fiducial based ECG biometric authentication using one-class support vector machine," in *2017 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2017, pp. 190-194: IEEE.

[13] H. P. Da Silva, A. Fred, A. Lourenço, and A. K. Jain, "Finger ECG signal for user authentication: Usability and performance," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1-8: IEEE.

[14] M. Hejazi, S. A. R. Al-Haddad, Y. P. Singh, S. J. Hashim, and A. F. A. J. D. S. P. Aziz, "ECG biometric authentication based on non-fiducial approach using kernel methods," vol. 52, pp. 72-86, 2016.

[15] D. Sung, J. Kim, M. Koh, and K. Park, "ECG authentication in post-exercise situation," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 446-449: IEEE.

[16] V. Krasteva, I. Jekova, and R. J. J. o. e. Abächerli, "Biometric verification by cross-correlation analysis of 12-lead ECG patterns: Ranking of the most reliable peripheral and chest leads," vol. 50, no. 6, pp. 847-854, 2017.

[17] G. Zheng, S. Ji, M. Dai, and Y. Sun, "Ecg based identification by deep learning," in *Chinese Conference on Biometric Recognition*, 2017, pp. 503-510: Springer.

[18] J.-N. Lee, S. B. Pan, and K.-C. Kwak, "Individual identification Based on Cascaded PCANet from ECG Signal," in *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, 2019, pp. 1-4: IEEE.

[19] Y. Chu, H. Shen, and K. Huang, "ECG Authentication Method Based on Parallel Multi-Scale One-Dimensional Residual Network With Center and Margin Loss," *IEEE Access,* vol. 7, pp. 51598-51607, 2019.

[20] L. Zhang, K. Xing, Z. Xu, J. Wang, S. Zhang, and J. Xu, "Human recognizer: An ECG based live biometric fingerprint," in *Proceedings of the 1st ACM Workshop on Privacy-Aware Mobile Computing*, 2016, pp. 21-27: ACM.

[21] F. P. Karegar, A. Fallah, and S. Rashidi, "ECG based human authentication with using Generalized Hurst Exponent," in *2017 Iranian Conference on Electrical Engineering (ICEE)*, 2017, pp. 34-38: IEEE.

[22] S. Yin, C. Bae, S. J. Kim, and J.-s. Seo, "Designing ECG-based physical unclonable function for security of wearable devices," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 3509-3512: IEEE.

[23] J. Sulam, Y. Romano, and R. Talmon, "Dynamical system classification with diffusion embedding for ECG-based person identification," *Signal Processing,* vol. 130, pp. 403-411, 2017.

[24] X. Wang, Z. Chen, J. Luo, J. Meng, and Y. J. E. L. Xu, "ECG compression based on combining of EMD and wavelet transform," vol. 52, no. 19, pp. 1588-1590, 2016.

[25] M. Rakshit, S. J. B. S. P. Das, and Control, "An efficient ECG denoising methodology using empirical mode decomposition and adaptive switching mean filter," vol. 40, pp. 140-148, 2018.

[26] M. U. Khan, S. Aziz, M. Bilal, and M. B. Aamir, "Classification of EMG Signals for Assessment of Neuromuscular Disorder using Empirical Mode Decomposition and Logistic Regression," in *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*, Taxila, Pakistan, 2019, pp. 237-243: IEEE.

[27] M. A. Kabir, C. J. B. S. P. Shahnaz, and Control, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains," vol. 7, no. 5, pp. 481-489, 2012.

[28] Q. Zhao and L. Zhang, "ECG feature extraction and classification using wavelet transform and support vector machines," in *2005 International Conference on Neural Networks and Brain*, 2005, vol. 2, pp. 1089-1092: IEEE.

[29] H. Beyramienanlou, N. J. C. Lotfivand, and m. m. i. medicine, "Shannon's energy based algorithm in ECG signal processing," vol. 2017, 2017.

[30] R. Majhi, S. Chattopadhyay, and A. Ghosh, "Radar assessment of wavelet decomposition based Skewness of ECG signals," 2015.

[31] Y. Chai, J. Ren, R. Zhao, and J. Jia, "Automatic gait recognition using dynamic variance features," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, pp. 475-480: IEEE.

[32] A. K. Bhoi, K. S. Sherpa, and B. J. C. C. Khandelwal, "Arrhythmia and ischemia classification and clustering using QRS-ST-T (QT) analysis of electrocardiogram," vol. 21, no. 1, pp. 1033-1044, 2018.

[33] Q. Zhang, D. Zhou, and X. J. A. M. S. Zeng, "Machine learning-empowered biometric methods for biomedicine applications," vol. 4, no. 3, pp. 274-290, 2017.

[34] U. Desai, R. J. Martis, C. G. Nayak, K. Sarika, and G. Seshikala, "Machine intelligent diagnosis of ECG for arrhythmia classification using DWT, ICA and SVM techniques," in *2015 Annual IEEE India Conference (INDICON)*, 2015, pp. 1-4: IEEE.

[35] E. D. J. D. S. P. Übeyli, "ECG beats classification using multiclass support vector machines with error correcting output codes," vol. 17, no. 3, pp. 675-684, 2007.