ECG BASED IDENTIFICATION AND AUTHENTICATION

Ahmet Can Ozbek, Ronald Salloum, Jian Li, C.-C. Jay Kuo

Ming Hsieh Department of Electrical Engineering, University of Southern California

ABSTRACT

This paper shows the possibility of using electrocardiogram (ECG) signals as a way of identifying and authenticating individuals. The dataset (CYBHi) we worked on is recorded for 63 subjects by dry contact on the fingertip, as a result preprocessing is required to clean the noise. After preprocessing, we investigate two slightly different feature extraction methods and by using those features we train a support vector machine (SVM) classifier. The first approach for feature extraction involves using the raw data samples along with its Fourier coefficients around the detected R peaks of the heartbeats in the ECG signal. The second approach for feature extraction involves reducing the dimensionality of the raw data samples around the R peak by PCA and extracting analytical and temporal features from the PQRST peaks of the heartbeats.

***Index Terms*—** ECG, Identification, Authentication

1. INTRODUCTION

Physiological characteristic difference among individuals makes it possible to use ECG signal as a biometric. For practical applications of using ECG as a biometric, the recordings of the bio-signals should be with “off-the-person” approach. For this purpose, we decided to use CYBHi dataset [4] in our experiments. In this dataset, there are recordings of 63 subjects and each subject was asked to sit for 2 minutes in resting position while their ECG signals are picked up by dry Ag/AgCl electrodes. After preprocessing and feature extraction stages, we train our model using SVM and we achieve to get 2.9333% and 2.4599% Equal Error Rate with two slightly different feature extraction approaches.

2. METHODOLOGY

As the recorded signal contains some unwanted noises such as power line interference, baseline wander, a filtering step is required in the preprocessing. After the preprocessing stage, we apply two slightly different feature extraction methods as a result we have two different feature vectors extracted from individual heartbeats. At the last stage, we are using SVM with RBF kernel to correctly match individual heartbeats to subjects by classification.

2.1. Preprocessing

Preprocessing involves two steps. In the first step, we are removing noisy segments by a variance-based method [1] for each window. In the second step, we are applying an FIR filter to clean the noise.

2.1.1. Removing noisy segments

In this step, first, the input signal is segmented into windows of one second each. For each window, we calculate a coefficient, called modified coefficient variance as described in [1], let “i” be the index of the window, then for the i^{th} window we have,

$$MCV_i = {}^{\sigma_i}/_{u_i^2}$$

Also we calculate a threshold as the arithmetic average of the modified coefficient of variances times the constant 2.5.

To remove noisy segments, we discard the windows, which has MCV value greater than the threshold.

2.1.2. Filtering

To get rid of undesired noises such as power line interference, baseline wander that are present in the recorded ECG signals, we perform band-pass filtering with an FIR filter. The specifications of the FIR filter we used is as follows;

Filter Order	Low Cutoff Frequency	High Cutoff frequency
75	0.45 Hz	35 Hz

2.2. Feature Extraction

For feature extraction, we are going to investigate two different approaches. Both of the approaches involve detecting the R peaks of the heartbeats and taking the raw data samples around it using Pan-Tompkins algorithm. First approach is very similar to the feature extraction method described in [1], in this method we use raw data samples and frequency components. Second approach utilizes PCA and features extracted from the detected PQRST peaks of individual heartbeats.

2.2.1. Raw Data Samples and Fourier Transform Coefficients

In this approach, R peaks of the individual heartbeats are detected using Pan-Tompkins [3] algorithm. 275 raw data samples before the R peak (R-275) and 265 raw data samples after the R peak (R+265) are taken as a time domain feature vector. For frequency domain feature vector 80 samples before the R peak (R-80) and 80 samples after the R peak (R+80) are taken and 45 point FFT of these samples are considered as frequency domain feature vector. Time domain feature vector and frequency domain feature vector are cascaded to create one feature vector from each heartbeat. The table below summarizes the extracted features and their dimensionality.

	Number of Features
Temporal Component (R-275,R+265)	541
Frequency Component (R-80,R+80) (45 point FFT)	45
Combined Feature Vector	586

2.2.2. Applying PCA to Raw Data Samples and Analytical and Temporal Features from PQRST Peaks

In this approach, similar to the previous approach described, R peaks of the individual heartbeats are detected using Pan-Tompkins algorithm. 275 samples before the R peak (R-275) and 275 samples after the R peak are taken as time domain feature vector and then PCA is applied to reduce the dimensionality of this time domain feature vector to 30. Moreover, PQRST peaks of the individual heartbeats are detected by using the algorithms in the toolbox [2], and then, as a similar feature extraction procedure as described in [5], the absolute time duration difference and amplitude difference between all possible combinations of these 5 peaks (PQRST) is used as another feature vector. The feature vector obtained from performing PCA is cascaded to the feature vector obtained by the PQRST peaks to create one feature vector from each heartbeat. The table below shows the dimensionality of the features used.

	Number of Features
Temporal component (R-275,R+275) (30 Principal Components)	30
Temporal component from PQRST peaks	10
Analytical component from PQRST peaks	10
Combined Feature Vector	50

2.3. Classification

The dataset we are using has 63 subjects' ECG signal recordings. In each recording, 80 heartbeats starting from the second heartbeat of the recording are used as training data. 15 heartbeats starting from the 84^{th} heartbeat are used as testing data.

For the purpose of training our classifier model, we used support vector machine (SVM) with radial basis function (RBF) kernel. The technical implementation is performed in MATLAB environment by using LIBSVM [6]. We optimized our model with respect to the parameters of the RBF kernel using 5-fold cross-validation. In our classification model, we have 63 classes representing each 63 subjects. Each heartbeat is considered as a sample and classified by the model to get results.

3. RESULTS

As explained in section 2.2. Feature Extraction, we have two different approaches for feature extraction as a result, we have two different results with using these two different feature vectors. We are using different metrics to present our results. The first metric we are using is the overall classification accuracy across all heartbeats, the other metric we are using is Equal Error Rate (EER) which is calculated by the true positive rates and false positive rates of individual heartbeats by classification. We also plot the ROC curve of the whole system. ROC curve is calculated by one vs. rest approach among the 63 classes.

3.1. Results with First Feature Extraction Approach (Raw Data Samples and Fourier Transform Coefficients)

With the first feature extraction approach, we get the following results,

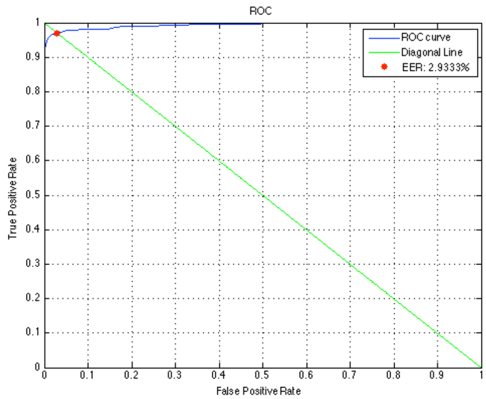
	Overall Accuracy	EER
Training Data	**99.246%**	**-**
Test Data	**91.6402%**	**2.9333%**

As explained in 2.3. Classification section, classification is performed by SVM with RBF kernel. These results are obtained by choosing C value as 1000 and gamma value as 0.02 for the RBF kernel.

Overall accuracy is defined as the ratio of the correctly classified heartbeats to their corresponding subjects over all of the heartbeats.

EER is defined as the point where false positive rate and false negative rates are equal, therefore we calculate EER by intersecting the diagonal line with the ROC curve.

The ROC curve of overall system is calculated in one vs. all fashion across all 63 classes. The macro average of all ROC curves of classes is taken as the ROC curve of the overall system.



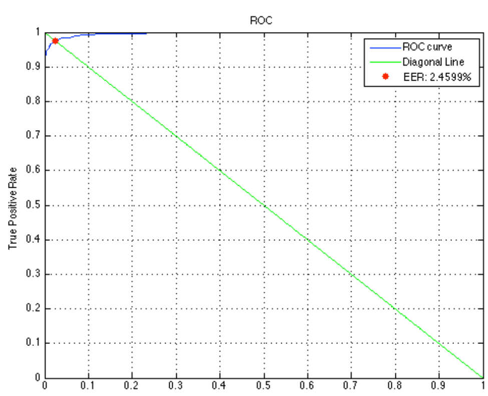
3.2 Results with Second Feature Extraction Approach (Applying PCA to Raw Data Samples and Analytical and Temporal Features from PQRST Peaks)

With the second feature extraction approach, we get the following results,

	Overall Accuracy	EER
Training Data	**99.9008%**	-
Test Data	**91.1111%**	**2.4599%**

Similarly to the procedure in the first feature extraction approach, classification if performed by SVM with RBF kernel. These results are obtained by choosing C value as 50 and gamma value as 2 for the RBF kernel.

From the ROC curve, again we calculate the EER as explained the previous section.



4. CONCLUSION AND FUTURE WORK

In this paper, we showed the possibility of identification and authentication applications from ECG signals. We investigated different feature extraction approaches and presented their corresponding results.

Current results in this paper are obtained from the same recording for each individual. In other words, training and testing samples for each individual are obtained from the same recording session. 80 heartbeats are used as training and 15 heartbeats are used as training data from the same recording of the individual. For future work involves taking training and testing samples from different sessions of recordings.

5. REFERENCES

[1] Karan Singh, Akshit Singhvi and Vinod Pathangay, “Dry Contact Fingertip ECG-based Authentication System using Time, Frequency Domain Features and Support Vector Machine”

[2] An online algorithm for R,S and T wave detection. Hooman Sedghamiz, December 2013. Linkoping university.

[3] Pan, Jiapu, and Willis J. Tompkins. "A real-time QRS detection algorithm." Biomedical Engineering, IEEE Transactions on 3 (1985): 230-236.

[4] Da Silva, Hugo Plácido, et al. "Check Your Biosignals Here: A new dataset for off-the-person ECG biometrics." Computer methods and programs in biomedicine 113.2 (2014): 503-514.

[5] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, “ECG to identify individuals,” Pattern Recognition, vol. 38, no. 1, pp. 133–142, 2005.

[6] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm