# IDENTIFYING INDIVIDUALS USING ECG BEATS

*Ramaswamy Palaniappan, Shankar M. Krishnan*

Biomedical Engineering Research Centre, Nanyang Technological University, Singapore

## ABSTRACT

**In this paper, we propose a technique to identify individuals using features extracted from QRS segment of electrocardiogram (ECG) signals. A total of 2000 samples from 10 subjects from the Arrhythmia Laboratory at Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center) database were used. These data are available as MIT-BH Normal Sinus Rhythm database and consist of 18 hour long-term recordings with 2 ECG signals. The commonly used features like R-R interval, R amplitude, QRS interval, QR amplitude and RS amplitude were used. In addition to these features, we propose the use of form factor of the QRS segment. Form factor has been used previously in electroencephalogram analysis and it is a measure of the complexity of the signal. These six features were then used by two neural network classifiers: Multilayer Perceptron - Backpropagation (MLP-BP) and Simplified Fuzzy ARTMAP (SFA). The data were split equally for MLP-BP and SFA training and testing. The results gave classification performance up to 97.6%. This indicates that ECG has the potential to be used as a biometric tool.**

## 1. INTRODUCTION

The most common method of identifying individuals is through the use of fingerprints (thumbprints) [1, 2]. However, in recent years, alternative biometric methods to replace or augment the fingerprint system have been proposed. Some of these methods are like electroencephalogram [3, 4], palmprints [5], hand geometry [6], iris [7] and face [8].

The use of ECG signals for the purpose of identifying individuals is relatively new as compared to the other biometrics. Very few research works have been published using heart signals as a biometric tool to identify individuals. Biel *et al* [9, 10] used ECG signals to classify 20 individuals but the total number of tested patterns were only 50, which is too small to conclude the results. Shen e*t al* [11] used ECG signals for verifying the individuality of 20 subjects, also using a small dataset (20 from each individual). In addition, the latter work was on verification, which is relatively easier as compared to identification.

In a verification system, the user declares his individuality (say subject A) and the system then searches its stored database for subject A's information. If the information matches, then the output is positive, else it is negative. Identification, on the other hand is more difficult because the system has to match the user to the stored databases of multiple subjects. This could mean matching against numerous subject databases instead of one as in the case of verification.

Data from ECG recordings of 10 subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center) is used. This dataset is available on the Physionet[1] website as MIT-BH Normal Sinus Rhythm database. There are two leads in the database but only the first lead from each recording is used here. A total of 200 beats from different times in the 18 hour recording are obtained for each subject.

Noise reduction algorithms are applied to reduce the baseline wander and high frequency noise (including 60 Hz interference). QRS segment is detected through a procedure similar to Pan and Thompkins method [12]. QRS features like R-R interval, R amplitude, QRS interval, QR amplitude and RS amplitude are extracted from these ECG beats. In addition, we propose the use of form factor of the QRS segment. Form factor [13] has been used previously in electroencephalogram (EEG) analysis and it is a measure of the complexity of the signal.

The data set is split randomly into two: one for training and another for testing. Each dataset contained features from 1000 beats, which we shall denote as patterns. These training and testing patterns are used to train and test two different neural network architectures: Multilayer Perceptron - Backpropagation (MLP-BP) [14] and Simplified Fuzzy ARTMAP (SFA) [15] to classify the patterns into the 10 categories representing the individuals. MLP-BP performance is relatively better than SFA [4], but SFA was also studied because of its incremental learning capability. That is SFA need not be retrained should a new category be added to the system (like adding a new subject).

## 2. METHODOLOGY

The database includes 10 long-term ECG recordings of subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center). Subjects included in this database were found to have had no significant arrhythmias; they include 5 men, aged 26 to 45, and 5 women, aged 20 to 50.

ECG signal from the first lead was used. A total of 200 beats were obtained from the 18 hour recording. The beats were obtained at approximately 5 minute intervals. A previous beat was also stored together with the beat being studied. This was to facilitate the computation of R-R intervals.

[1] http://www.physionet.org

### 2.1. ECG Pre-processing

An example of the ECG signals in the database is shown in Figure 1. However, only the first lead ECG signals were used here. Before we could proceed to the feature extraction, it is necessary to reduce noise interferences from the ECG signals. Three commonly encountered noises are high frequency noise, 60 Hz powerline interference and baseline wander. These noises have to removed or reduced in order for accurate extraction of features from ECG signals.

**Fig.1 An example of ECG signals in MIT-BH Normal Sinus Rhythm database**

The signal was low-pass filtered to remove 60 Hz powerline interference and high frequency interferences using a 7th order Butterworth digital filter with 3 dB cut-off at 30 Hz. Order 7 is enough to obtain attenuation of 30 dB at 40 Hz with the 128 Hz sampling frequency of the ECG signals. This cut-off frequency was chosen after some preliminary simulations. Because Butterworth filter is a type of Infinite Impulse Response (IIR) filter, which could cause phase distortion, the output of the filter was run back through the filter gain to offset the phase distortion. This procedure of forward and reverse filtering was done for all times whenever Butterworth filter was used in this paper.

Baseline wander is caused by subject movement or perspiration or excess chest hair. To remove this type of noise, which has a very low spectral range, we first extracted the baseline noise. This was done using a 5th order low-pass Butterworth digital filter with cut-off at 1 Hz. Order 5 is enough to attenuate the signals completely beyond 2 Hz. Once the baseline signal has been detected, the clean signal was obtained by subtracting the baseline signal from the original signal.

Figure 2 shows an example of the original ECG signal and ECG signal after reducing noise interferences.

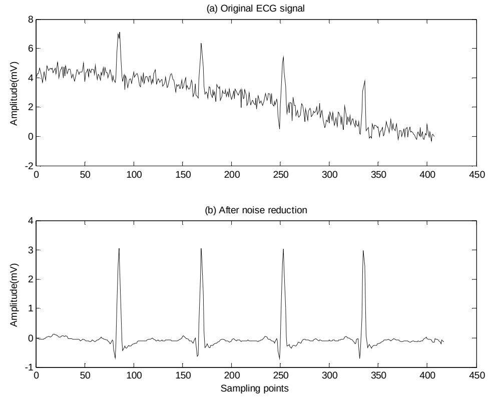


**Fig.2 An example of (a) Original ECG signal, (b) after noise reduction**

### 2.2. ECG feature extraction

A modified version of Pan and Thompkins algorithm [12] was used to detect the QRS segment. The algorithm was modified slightly where the low-pass and high filters used in the algorithm have been replaced by a 9th order bandpass Butterworth digital filter with 3 dB cut-off from 8-14 Hz. The rest of the steps used here like squaring, moving window integration and thresholding were the same as Pan and Thompkins algorithm. Moving window integration length, N was set at 19[2] and thresholding value was set at 50.0. Both these values were chosen after some preliminary simulations. From our manual inspection, it was found that the algorithm had successfully identified approximate QRS segments for R peak identification. Further details of these steps could be found from [16].

With the detection of the QRS segment, R peak was detected at the biggest upward deflected peak in this segment. The Q point was detected as the first downward deflection trough by tracing back 50 ms from R point. Similarly, S point was detected as the first upward deflection peak by tracing forward 100 ms from R point. The features like R amplitude, QR amplitude, RS amplitude and QRS width (interval) were computed. Since each pattern in our signal consisted of 2 beats, we also obtained the R-R interval. Next form factor [13] was computed using the formula

$$FF = [(\mathrm{var}(x'') / \mathrm{var}(x')) / (\mathrm{var}(x') / \mathrm{var}(x))]^{-\frac{1}{2}} \qquad (1)$$

where x is the QRS segment, x' is the first derivative of the QRS segment, while x'' is the second derivative of the QRS segment.

These features were combined to give a total of 6 features.

[2] In [12], the authors used N = 30 for sampling frequency of 200 Hz. Here, we use N = 19 for sampling frequency of 128 Hz.

### 2.3. SFA and MLP-BP neural network

Two NN architectures: MLP-BP [14] and SFA [15], were used to classify the ECG features into the corresponding categories that represent the individuality of the subjects and their performances were compared. NN architectures were considered here because of their ability to give high classification accuracies.

SFA is relatively new as compared to BP. SFA is advantageous as compared to BP due to its high-speed training in fast learning mode and incremental learning abilities. SFA consist of two modules: Fuzzy ART and Inter ART mapping. The Fuzzy ART module receives the input features and outputs the particular category node. The category node is linked to the corresponding class (i.e. the individuality of the subject in this case) through the Inter ART mapping module. Inter ART mapping module works by increasing the vigilance parameter (VP) of Fuzzy ART by a minimal amount to correct a predictive error at the category layer. VP was varied from 0 to 0.9 in increments of 0.1. Figure 3 shows the SFA network architecture as used here.

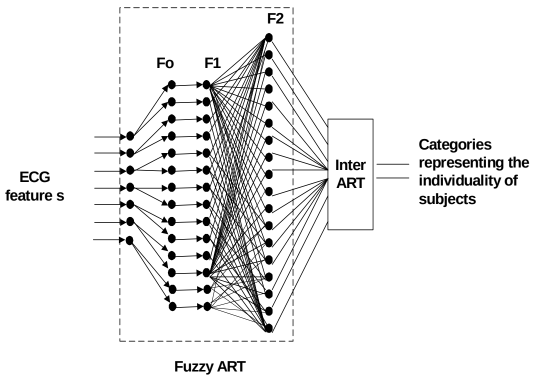


**Fig. 3 SFA network architecture**

BP neural network is a MLP type with a single hidden layer and trained by the BP algorithm. Descriptions on BP NN could be found in common textbooks and will not be discussed here. Figure 4 shows the BP network architecture that was used in the study. The desired target output was set to 1.0 for the particular category representing the individual that was being trained, while for the other 9 categories, it was set to 0.

Half of the data (1000 patterns) were used to train the network while the rest half (1000 patterns) were used in testing. The hidden nodes were varied from 5 to 50 nodes in steps of 5. The input layer consisted of 6 nodes because the number of features was 6, while the output layer was set to 10 nodes for the 10 classes representing the subjects. The target output for the trained pattern was set to 1.0, while for the rest of the classes, it was set to 1.0. Figure 4 shows the architecture of the MLP-BP network architecture as used in this study.

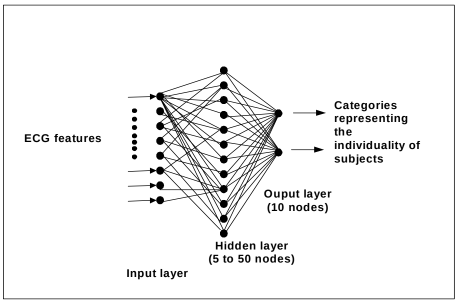


**Fig. 4 MLP-BP network architecture**

## 3. RESULTS

Table 1 shows the MLP-BP and SFA classification results for the varying hidden units and vigilance parameter.

**Table 1: MLP-BP and SFA classification results**

| MLP-BP | | Fuzzy ARTMAP | |
|---|---|---|---|
| **Hidden units** | **(%)** | **VP** | **(%)** |
| 5 | 93.1 | 0 | 83.30 |
| 10 | 94.9 | 0.1 | 84.34 |
| 15 | 95.9 | 0.2 | 84.50 |
| 20 | 96.9 | 0.3 | 84.31 |
| 25 | 96.8 | 0.4 | 83.21 |
| 30 | 97.6 | 0.5 | 84.08 |
| 35 | 96.8 | 0.6 | 83.77 |
| 40 | 96.4 | 0.7 | 83.81 |
| 45 | 96.3 | 0.8 | 82.72 |
| 50 | 97.0 | 0.9 | 82.04 |
| **Average** | **96.17** | **Average** | **83.61** |

## 4. DISCUSSION

As could be seen from Table 1, the performances did not vary much with different hidden nodes or vigilance parameter. SFA performance varies with input pattern ordering, so the SFA performance was averaged over 10 simulations with random input pattern ordering. MLP-BP performed much better than SFA for all the cases. The best performance of 97.6% was achieved for 30 hidden nodes with MLP-BP. However, MLP-BP takes much longer time to train as compared to SFA. But this is not really a problem for classification because the training is done offline, and classification (i.e. testing) times are faster using MLP-BP. SFA has incremental training ability, so adding new subjects will be easier as compared to MLP-BP, which requires retraining of all the subjects.

The whole process of pre-processing, feature extraction and classification takes less than a second when run on P4 machine with MATLAB (Mathworks, Inc.) codes. Therefore, the method could be applied for real-time individual identification.

## 5. CONCLUSION

In this work, we have extracted features from a single ECG beat for individual identification. The features were from QRS segment with a new feature called form factor. These features have been classified by MLP-BP and SFA neural network. The results give up 97.6% classification when tested on 1000 ECG beats from 10 subjects. Further increase in classification accuracy might be possible by including more features from T and P waves of ECG.

This approach of using ECG signals as a biometric could be used to identify people in a small group like operators in factories or in a high security workplace like military bases. However, further work is necessary to determine the changes over a period of time and under different health conditions.

## 6. REFERENCES


[1] S. Pankanti, R. M. Bolle, and A. Jain, "Biometrics: The Future of Identification," *Special Issue of IEEE Computer on Biometrics*, pp.46-49, Feb. 2000.

[2] S. Pankanti, S. Prabhakar, and A. K. Jain, "On the Individuality of Fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1010-1025, August 2002.

[3] R. Palaniappan, and K.V.R. Ravi, "A New Method to Identify Individuals Using Signals from the Brain," *CDROM Proceedings of Fourth International Conference on Information Communications and Signal Processing and Fourth Pacific-Rim Conference on Multimedia (ICICS-PCM 2003)*, Singapore, Dec. 15-18, 2003.

[4] K.V.R. Ravi, and R. Palaniappan, "Neural Network Classification of Late Gamma Band Electroencephalogram Features," *Soft Computing Journal (in press)*.

[5] N. Duta, A. K. Jain, and K. V. Mardia, "Matching of Palmprints," *Pattern Recognition Letters*, vol. 23, no. 4, pp. 477-485, 2002.

[6] A. K. Jain, A. Ross, and S. Pankanti, "A Prototype Hand Geometry-based Verification System," *Proceedings of 2nd International Conference on Audio and Video-Based Biometric Person Identification (AVBPA)*, pp. 166-171, March 22-24, 1999.

[7] J. Daugman, "Recognizing Persons by Their Iris Patterns," *in Biometrics: Personal Identification In Networked Society*, Jain, A.K., Bolle, R., and Pankanti, S., (eds.), Kluwer Academic, 1999.

[8] A. Samal, and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, 1992.

[9] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG Analysis: A New Approach in Human Identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, No. 3, pp. 808-812, June 2001.

[10] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG Analysis: A New Approach in Human Identification," *Proceedings of 16th IEEE Instrumentation and Measurement Technology Conference*, vol. 1, pp. 557-561, 1999.

[11] T. W. Shen, W. J. Tompkins, and Y. H. Hu, "One-lead ECG for Identity Verification," *Proceedinsg of the Second Joint EMBS/BMES Conference*, Houston, pp. 62-63, 2002.

[12] J. Pan, and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230-236, March 1985.

[13] B. Hjorth, "EEG analysis based on time-domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 34, pp.306-310, 1970.

[14] D. E. Rumelhart, and J. L. McClelland, *Parallel Distributed Processing; Explorations in the Microstructure of Cognition*, vol. 1, MIT Press, Cambridge, 1986.

[15] T. Kasuba, "Simplified Fuzzy ARTMAP," *AI Expert*, vol. 8, no. 11, pp. 19-25, 1993.

[16] R. Palaniappan, and S. M. Krishnan, "Detection of Ectopic Heart Beats Using ECG and Blood Pressure Signals," *International Conference on Signal Processing and Communications*, Bangalore, India, 2004.

[17] B. M. Beasley, *Understanding EKGs: A Practical Approach*, Prentice Hall, 2003.