# Individual Identification Using EEG Features

Mona F. M. Mursi Ahmed
email: monmursi@yahoo.com

May A. Salama
email: msalama@megacom-int.com

Ahmed Abdullah Hussein Sleman
email: mindhunter74@gmail.com

Electrical Engineering Dept.
Faculty of Engineeringat Shoubra, Benha Univ.
Cairo, Egypt

***Abstract*— Electroencephalography (EEG) is a method of monitoring electrical activity along the scalp by measuring voltage variations resulting from neural activity of the brain. A number of published research papers have indicated that there is enough individuality in the EEG recording, rendering it suitable as a tool for person authentication. In recent years there has been a growing need for greater security for person authentication and one of the potential solutions is to employ the innovative biometric authentication techniques. In this research paper, we investigate the possibility of person identification based on features extracted from person's measured brain signals electrical activity (EEG) with different classification techniques; Radial Basis Functions (RBF), Support Vector Machines (SVM) and Backpropagation (BP) neural networks. The highest identification accuracy was achieved using modular backpropagation neural network for classification.**

***Keywords*—*EEG; identification; biometrics; brain-waves;***

## I. INTRODUCTION

The brain is one of the largest and most complex organs in the human body. It is involved in every thought and movement produced by the body, which allows humans to interact with their environment, communicating with other humans and objects. It consists of several parts as indicated in Figure 1 [1] and every part is responsible for certain functions and activities.

There are several different methods used for measuring the activity of the brain such as positron emission tomography (PET), functional magnetic resonance imaging (FMRI), Magnetoencephalography (MEG), and EEG.

EEG is the recording of electrical activity along the scalp. EEG measures voltage fluctuations resulting from ionic current flows within the neurons of the brain [2]. In recent years there has been a growing need for greater security for person authentication. Using EEG as a biometric has some advantages over other biometrics like fingerprint and iris image. Unlike other biometrics, we find that brain-waves are almost impossible to be mimicked; even similar activities produce different brain-waves per person, can't be easily stolen – requires special equipment touching the scalp and can't be produced by forcing the person to do so being sensitive to the person's mental state.

EEG data could be collected with single or multi-electrodes device. This depends on the EEG device and the number of signals needed to be processed. All electrode names mentioned hereafter are based on the 10-20 system for EEG electrodes locations [3]. An overview of this system is shown in Figure 2.

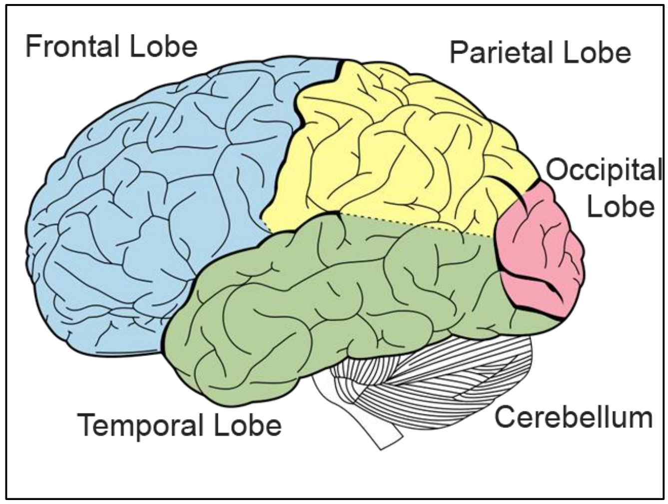


Figure 1. Brain Structure

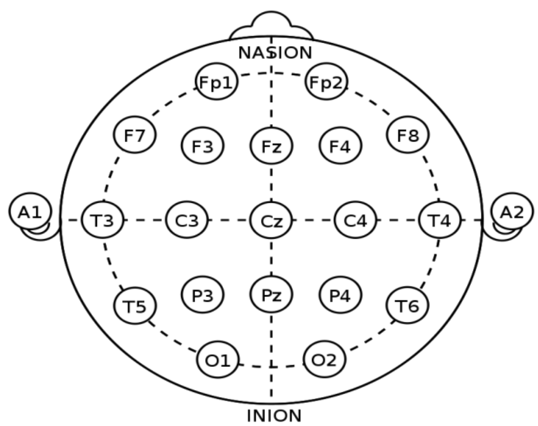


Figure 2. 10-20 Standard System for EEG electrodes locations

One of the potential solutions to identify individuals is to employ the innovative biometric authentication techniques. In this paper, we present a biometric authentication system based on EEG and using offline dataset. After presenting an overview of the previous work (Section II) in this area of research, we first describe the used dataset (Section III) and what the feature vector is composed of. Then we elaborate on using 3 different classification techniques: Radial Basis Function, Support Vector Machines and modular backpropagation neural networks (Section IV). Finally, a conclusion of our work and future work are discussed (Section V). We use MATLAB in all our experiments.

## II. PREVIOUS WORK

Different methods have been applied for EEG based person identification. Based on our survey, different methods differ in data collection and Brain Computer Interface (BCI), Preprocessing and feature extraction, and/or classification techniques.

Both Autoregressive (AR) and Power Spectral Density (PSD) were used in [4] and [5] to produce the input feature vector of collected EEG data. AR model of order 19 was selected after testing the orders 10 – 50 as being the optimal order. PSD of frequency range 4 Hz – 32 Hz has been applied and added to the feature vector to produce a final vector of 127 features. A maximum identification accuracy of 97.5% was reported in [4] using K-nearest neighbor (KNN) and Fisher's Discriminant Analysis (FDA) as classifiers while [5] reported a 95.4% accuracy for a consistent person state and 84.5% for persons on diet using same classification techniques. Arguing that autoregressive model coefficients may not have a remarkable effect on the system performance as a feature extraction method, as mentioned in [6], relying only on PSD for the frequency range (5 Hz to 32 Hz) enabled them to an obtain identification accuracy of 90% and 93.7% using dual space Linear Discriminant Analysis (LDA) based on simple regularization and KNN for classification. Independent Component Analysis (ICA) was used in [7] by separating multi-channels EEG data into independent sources. After testing different ICA algorithms using ICALAB Signal Processing Toolbox [8], JADEop ICA algorithm was found to give the highest percentage of identification accuracy (100%) with 5, 10, and 20 subjects using backpropagation neural networks for classification and in order to find the minimum number of relevant channels for person identification, all possible combinations of 4, 3, and 2 channels were tested to find that the best combination of channels to use is {ch1, ch11, ch14} i.e., {FP1, T5, C4}.

Instead of determining a set of features for classification, [9] uses convolutional neural networks to select the most distinctive features that can be used for classification leading to an identification accuracy of 80% with a dataset of 10 subjects that are in a resting state with their eyes open.

## III. DATASET AND FEATURE VECTOR

The dataset used in our work is the large version of the KDD Dataset [10], which contains EEG recording for 10 alcoholic subjects and 10 control subjects. A subject's sample is a 1-second recording of EEG. The dataset contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz. Statistics about this KDD dataset are shown in **Error! Reference source not found.**.

TABLE I. KDD Dataset Statistics

| | |
|---|---|
| Subjects | 20 subjects |
| Sample length | 1 second |
| Samples per subject | 60 samples |
| Dataset size | 20 x 60 = 1200 samples |

Although this dataset examines EEG correlation of genetic predisposition to alcoholism, we used the EEG data for person identification regardless of the state of the person. First, we derived the feature vector, which had four types:

- AR Coefficients (order 6)
- Spectral Power
- Power Spectral Entropy
- Approximate Entropy

We started by finding out the best combination of features to use by attempting every different valid combination of the suggested features while choosing backpropagation neural networks for classification being it used in many previous of the researches and giving good results. The results indicated in Figure 3 show that using all 4 types of features together gives the best classification accuracy (87%).

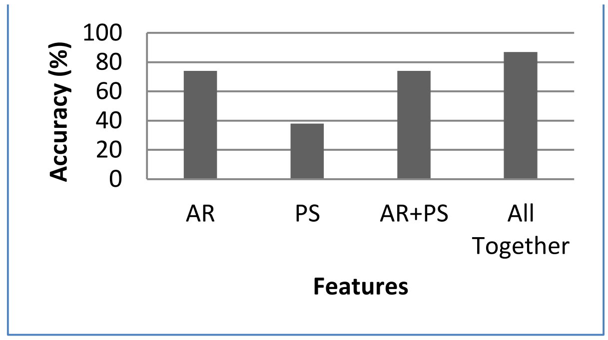


Figure 3. Results of using different combination of proposed features

Principal Component Analysis (PCA) was then applied to reduce the dimensionality of the obtained feature vector to a length of 36 to speed up the classification process.

## IV. CLASSIFICATION

Various classification techniques have been experimented. The results of classification using RBF, SVM, and modular backpropagation neural networks are discussed below. In all classification techniques, we use 2/3 of the mentioned dataset for training the test its accuracy against the remaining 1/3 of it.

### *A. RBF*

Different dataset sizes (number of subjects and samples per subject) and different numbers of centers were tested for classification. The results are shown in Table II.

TABLE II. RBF CLASSIFIER RESULTS

| # | Subjects | Samples | Max Training Accuracy | | Max Testing Accuracy | |
|---|---|---|---|---|---|---|
| | | | *%* | *Centers* | *%* | *Centers* |
| 1 | 10 | 10 | 82 | 6 | 50 | 2 |
| 2 | 10 | 20 | 83 | 6 | 60 | 2 |
| 3 | 10 | 30 | 82 | 13 | 66 | 5 |
| 4 | 20 | 10 | 70 | 4 | 38 | 9 |
| 5 | 20 | 20 | 69 | 9 | 40 | 19 |
| 6 | 20 | 30 | 68 | 16 | 44 | 18 |

The best classification accuracy obtained was 44% with the whole dataset and using 18 centers.

### *B. SVM*

Different SVM model types and kernel functions - mentioned in **Error! Not a valid bookmark self-reference.** - were tested.

TABLE III. DIFFERENT SVM MODEL TYPES THAT WILL BE TESTED

| | |
|---|---|
| **Model Types** | Weston and Watkins (WW) |
| | Crammer and Singer (CS) |
| | Lee, Lin, and Wahba (LLW) |
| | Guermeur and Monfrini (MSVM2) |
| **Kernel Functions** | Linear kernel |
| | Gaussian RBF kernel |
| | homogeneous polynomial kernel |
| | non-homo. polynomial kernel |

First, the results of testing different model types with half of the dataset (2/3 of the half for training and 1/3 of the same half for testing) shows that CS model type is the best one to use as indicated in Figure 4.

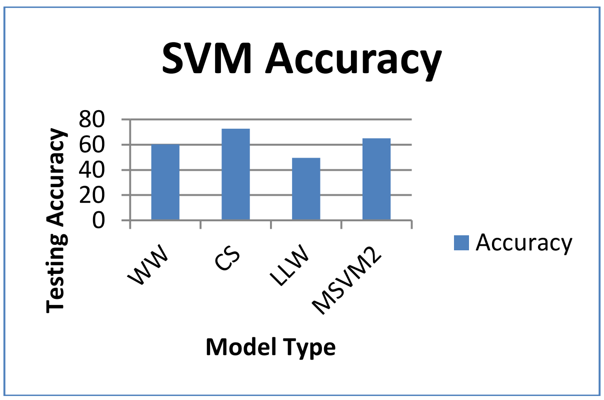


Figure 4. Results of using different SVM model types for classification

Second, testing different kernel functions with the CS model but now with the whole dataset shows that the non-homogenous polynomial kernel function gives the best classification accuracy (63%) as shown in Figure 5.

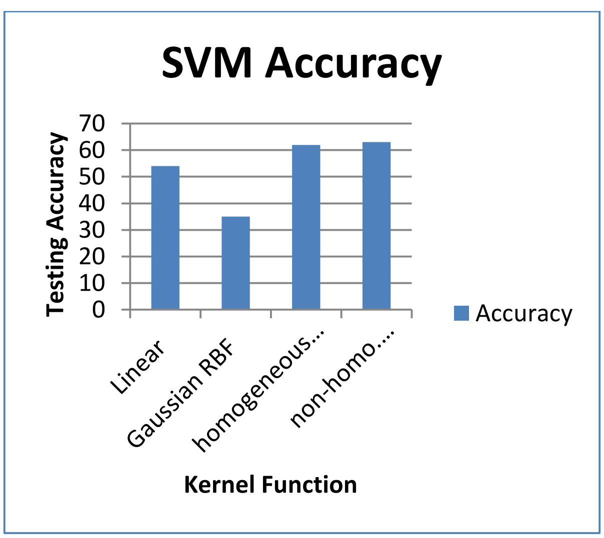


Figure 5. Results of using SVM CS model type with different kernel functions

### *C. Modular Neural Network*

In attempt to achieve better accuracy for identification taking into consideration that being an individual alcoholic affects his EEG measurement, a modular backpropagation neural network is used for classification as follows. A separate BP network, BP2, is used to classify control subjects while BP1 is used to classify alcoholic subjects.

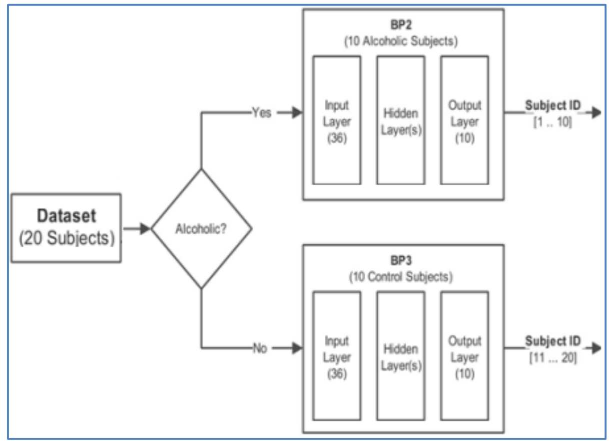


Figure 6. Modular Neural Network design for classification

The property of a subject being alcoholic or not is fed into BP1 to decide onto which network to use to identify that person. The result of the design shown in Figure 6 was 93.5% for the whole dataset.

## V. DISCUSSION

After testing different combinations of the four proposed feature types, using them all together was shown to give best accuracies. Moreover, the lower the number of channels used to extract the features, the less the identification accuracy we get, which was the reason we have chosen to use all the 64 channels used in the dataset to extract the proposed features. Finally, after attempting different classification techniques to identify the 20 subjects in the dataset, the best obtained result (93.5%) was using a modular backpropagation neural network at which there is a separate network for identifying alcoholic subjects and another for identifying control subject where the property of being alcoholic or not was a pre-given property to the whole network design. Although being an alcoholic subject has a noticeable effect on its EEG, we found that separating alcoholic and control subjects yielded better identification results – having a single classifier for all subjects yielded accuracies of 44%, 63%, and 87% using RBF, SVM and backpropagation neural network while the modular design yielded 93.5% identification accuracy. The best accuracy we have got (93.5%) is lower than that obtained in [7] while it used a different dataset and used ICA instead of PCA that was used in our system. In comparison to [9] that used Convolutional Neural Networks for classification to get an identification accuracy of 80% with a dataset of 10 subjects, our system outperformed that yielding better 93.5% identification accuracy with a dataset of 20 subjects. Keeping in mind the particularity of the dataset used (EEG of alcoholic/control subjects), we could better improve the accuracy of the final proposed classification network by having alcoholic and control group of subjects each identified by a separate network. The latter piece of information might not be generally available in practice and we would have to use a single network for classification regardless of the subject state.

## VI. CONCLUSION

In this paper, it was shown that EEG can be used effectively for individual identification. A combination of 4 feature types were used to construct the feature vector; Autoregressive model of order 6, Spectral Power, Power Spectral Entropy, and Approximate Entropy, which was found to give best accuracy results. Different approaches were proposed that yielded identification accuracies of 44%, 63%, and 93.5% using RBF, SVM and modular backpropagation neural network respectively. In future work, we would consider measuring EEG from volunteering individuals to construct the EEG dataset. Also, the measurements would be performed in different mental states so that it would be more efficient to identify individuals when they are doing certain activities.

## REFERENCES


[1] "Structure of the Brain," [Online]. Available: http://controlmind.info/human-brain/structure-of-the-brain. [retrieved: 10, 2016].

[2] E. Niedermeyer and F. H. L. d. Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields., Lippincott Williams & Wilkins, 2005.

[3] "10-20 System," 05 2015. [Online]. Available: http://www.wikiwand.com/en/10-20_system_(EEG) [retrieved: 10, 2016].

[4] F. Su, L. Xia, A. Cai, Y. Wu, and J. Ma, "EEG-based Personal Identification, from Proof-of-Concept to A Practical System," in *International Conference on Pattern Recognition*, 2010.

[5] F. Su, L. Xia, and A. Cai, "Evaluation of Recording Factors in EEG," in *Systems Man and Cybernetics (SMC), IEEE International Conference*, 2010.

[6] F. Su, H. Zhou, Z. Feng, and J. Ma, "A Biometric-based Covert Warning System Using EEG," in *Biometrics (ICB), 5th IAPR International Conference*, 2012.

[7] P. Tangkraingkij, C. Lursinsap, S. Sanguansintukul, and T. Desudchit, "Selecting Relevant EEG Signal Locations for Personal Identification Problem Using ICA and Neural Network," in *Eigth IEEE/ACIS International Conference on Computer and Information Science*, 2009.

[8] "ICALAB for Signal Processing," 05 2015. [Online]. Available: http://www.bsp.brain.riken.jp/ICALAB/ICALABSignalProc/ [retrieved: 10, 2016].

[9] L. Ma, J. W. Minett, and T. Blu, "Resting State EEG-Based Biometrics for Individual Identification Using Convolutional Neural Networks," in *37th Annual International Conference of the IEEE EMBC*, 2015.

[10] KDD Dataset [Online]. Available: https://kdd.ics.uci.edu/databases/eeg/eeg.html [retrieved: 10, 2016].


 ISBN: 978-1-61208-518-0 