A PRAGMATIC AUTHENTICATION SYSTEM using ELECTROENCEPHALOGRAPHY SIGNALS

Ayman Khalafallah[†*], Aly Ibrahim[¶*], Bahieeldeen Shehab[¶*], Hisham Raslan[¶*], Omar Eltobgy[¶*], Shady Elbaroudy[¶*]
[*] Computer and Systems Engineering Department, Alexandria University
[†] Email: ayman.khalafallah@alexu.edu.eg
[¶] Emails: {aliiitarek, engbahy93, eng.hisham.raslan, omareltobgy, shadyelbaroudy}@gmail.com
[*] The authors contributed equally to this work.

***Abstract*—EEG-based authentication is an emerging research field. In this work, a realistic authentication system using Electroencephalography signals, was developed aiming to show that brain signals contain sufficient information to be used in security systems. The dataset used was composed of 29 users on 4 different days via the cheap Neurosky Mindwave headset with a single dry electrode, and 10 users on 3 different days via Emotiv with 14 electrodes. Various techniques, features, and algorithms were examined to achieve the highest security. Experiments indicated that the system proposed can scale with respect to increasing number of users in the datasets. The system successfully handles users authenticating from multiple days not used in training the model with high accuracy. A false acceptance error (FA) of 3% was achieved, with a higher false rejection error (FR) of 48%, yielding an overall accuracy (ACC) of around 80% using the Mindwave dataset, and a FA of 0.3%, a FR of 13.93% resulting in an ACC of 92.88% using Emotiv. These results are promising for an authentication system, because the system is conservative, only allowing correct users to enter -*even at the expense of multiple attempts*- while successfully refusing to grant access to impersonating users.**

***Index Terms*—EEG, Supervised Classifier, Authentication, AR, PSD**

I. INTRODUCTION

Authentication is the process of confirming a person's claimed identity based on the information they provide.

Existing authentication systems use one or more authentication factors:
Knowledge factor (Passwords), Possession factor (ATM card or keys), and / or Biometric factor (Fingerprint, Voice, or Iris).

Recently, a new type of Biometric authentication systems based on Electroencephalography (EEG) signals evolved. Distinctive features, that can differentiate between users, are extracted from the brain's electrical activity during performing various tasks [1].
The advantages of EEG based authentication include:
- Brain damages rarely occur as opposed to hand or eye injuries.
- Disabled users are capable of using the system.
- Inability to force users to authenticate themselves unwillingly, or for users to forget their identity or have it stolen.

II. RELATED WORK

Using EEG signals for authentication has received widespread attention in recent years. Considering passthoughts rather than typing a password dates back to Thorpe et al. [2] in 2005. Ashby el al. [3], using a 14-channel headset (Emotiv) on 5 users, achieved 97.69 - 100%. Chuang et al. [4] using a single channel headset (Neurosky Mindwave) on 15 users with seven different mental tasks (e.g. breathing, listening, visualizing), achieved 99% using cosine similarity on features extracted from the α and β bands. Sohankar el al. [5] using single channel headset on 10 users, achieved 91 - 98%. They used Naive Bayes Classifier on the extracted Fast Fourier Transform features. Marcel and Millan employed a Gaussian mixture model and maximum a-posteriori model for authentication with 9 users [6]. Riera et al. [8] achieved 98.3% ACC while applying a multiday authentication on 51 users with 36 intruders. They used Fisher Discriminant Analysis on features extracted from Autoregression and Fourier transform, mutual information, coherence, and cross-correlation.

III. SYSTEM ARCHITECTURE

A. Signal Acquisition

Experiments were done using both Neurosky Mindwave [10] and Emotiv Epoc+ [11] headsets. Mindwave has one dry electrode with a sampling frequency of 512Hz, while Emotiv has 14 electrodes with a sampling frequency of 128Hz. Mindwave is convenient, as it is cheap and practical, having a single dry electrode in contrast to Emotiv having multiple (wet) electrodes. However, Emotiv captures data from multiple channels giving better indication of user activity.

The Mindwave dataset was gathered from 29 users each doing four sessions, while the Emotiv dataset was gathered from only 10 users each doing three sessions. Each session was 50 seconds long and was collected on different days, spanning 4 months, while resting with eyes closed (REC). Each was composed of five 10-second blocks, with a resting period of 5 seconds between blocks.
The dataset was intentionally gathered in a realistic environment, where users were not constrained to perform the experiments in the same position, location, time of the day, nor mental state, capturing various variabilities in the users.

B. Pre-Processing

The 10-second block of each user was divided into segments; making 1, 2, 5, or 10 seconds the possible seg. sizes.

Each segment was filtered using a 4-order elliptic band-pass filter with a passband in the range of 2 to 100 Hz to reduce the noise and artifacts. To eliminate phase distortion, forward and reverse filtering was performed.

C. Feature Extraction

Features were extracted from each preprocessed segment. The features used in the experiments included:

- Autoregressive Coefficients (AR)
- Power Spectral Density (PSD):
  The power spectral density was obtained by either taking the square of the absolute value of the Fourier transform of the data in each segment, or by Welch's method [7]. PSD was examined at different frequencies:
  α band (7 - 14 Hz), β band (14 - 31 Hz), α - β bands (7 - 31 Hz), δ - β bands (0 - 31 Hz), α β γ bands (7 - 100 Hz), and All frequency bands (0 - 100 Hz).

D. Feature Reduction

Linear Discriminant Analysis (LDA) was tried as a feature reducer for both datasets.

E. Classification

Each user had their own trained classifier; where the training data was labeled +1 for this user's segments and -1 for other users' segments. Several classifiers were examined:
SVM (Support Vector Machine) - Regularized Logistic Regression - LDA (Linear Discriminant Analysis).

F. Testing

The error measure used in this work is Half Total Error Rate (HTER), which is the average between 2 error rates: False Rejection (FR) and False Acceptance (FA)
$HTER = \frac{FA+FR}{2}$, giving $Accuracy = 1 - HTER$
Each user claims to be all the other users including themselves. FR occurs when the user isn't authorized to enter as themselves, while FA occurs when the user is authorized to enter as another user.

Two methods of user authentication were used:

- **Block Test**: The user is authenticated by providing one block (10 seconds) of data, each segment in that block is classified as a +1 or -1 and majority voting between segments is used.
- **Day Test**: The user is authenticated by providing one session (50 seconds) of data, each segment in that session is classified as a +1 or -1 and a single +1 authorizes the user.

K-fold* cross validation is used in testing with two flavors:

- **Shuffled Cross Validation**: Shuffle all the data blocks for each user before partitioning the folds.
- **Day Cross Validation**: Partitioning the folds according to data from different days used for each user, i.e., all the data of each day belongs to only one fold and isn't included in any other fold. This is a more realistic measure as users get authenticated from an unseen day which is similar to actual everyday usage.

* The number of folds (k) equals the number of days in each dataset

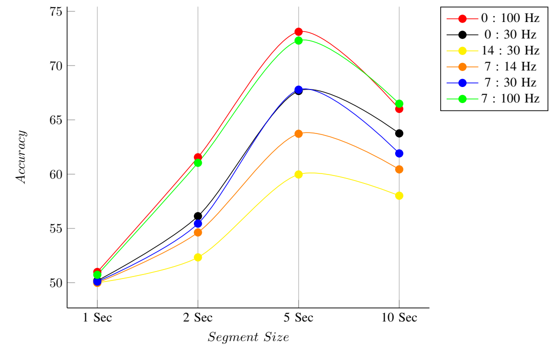

Fig. 1. Comparison between different frequency ranges and segment sizes while using Welch

IV. Analysis of System Parameters

In this study, many different combinations of system parameters are investigated to reach the highest accuracy.
Experiments 1-4 investigated Mindwave dataset.
- Segment Size: {1, 2, 5, or 10}.
- PSD Frequency Range: {7-14, 14-30, 7-30, 0-30, 7-100, or 0-100}.
- PSD Method: Welch with AR.
- Classifier: {Logistic Regression, SVM, or LDA}.
- Number of AR Coefficient: {in range 0 to 100}.

A. Experiment 1: Finding the best segment size and PSD frequency range

The fixed parameters were:
- Classifier: Logistic Regression.
- Number of AR Coefficient: 50.
- PSD Method: Welch.
- Testing: 4-Folds *Day Cross Validation* with *Block Test* on 20 users.
The variable parameters were:
- Segment Size: {1, 2, 5, or 10}
- PSD Frequency Range: {7-14, 14-30, 7-30, 0-30, 7-100, or 0-100}
The results of experiment 1 are shown in figure 1. The conclusion drawn from these results is that, a segment size of 5 seconds and frequency range from 0 to 100 Hz gave the best accuracies, hence, should be used.

B. Experiment 2: Finding the best classifier

The fixed parameters were the same as Exp. 1 in addition to:
- Segment Size: 5.
The variable parameters were:
- Classifier: Logistic Regression, SVM, LDA.
- PSD Frequency Range: {7-14, 14-30, 7-30, 0-30, 7-100, or 0-100}.
The results of experiment 2 are shown in figure 2. It is evident that Logistic Regression classifier was superior.

C. Experiment 3: Finding the best number of AR coefficients

The fixed parameters were the same as Exp. 2 in addition to:
- Classifier: Logistic Regression.

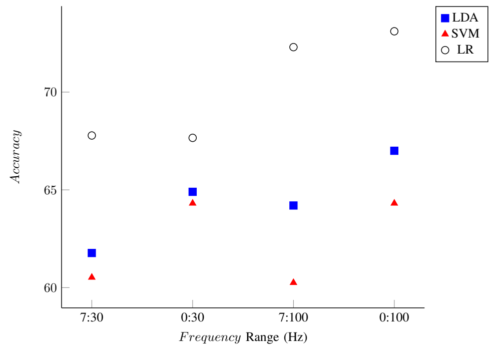

Fig. 2. Comparison between different classifiers

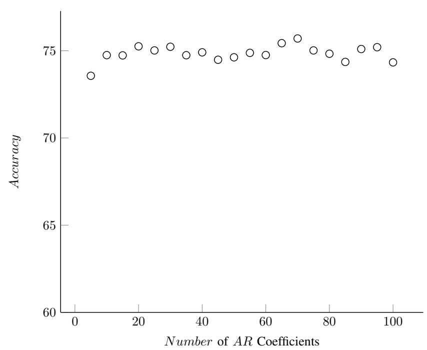

Fig. 3. Comparison between different numbers of AR coefficients

TABLE I
RESULTS OF EXPERIMENT 4

Method	Block Test			Day Test		
	FA	FR	Acc	FA	FR	Acc
None	1.54	48.00	75.23	10.92	28.75	80.16
LDA	0.59	83.45	57.98	4.03	74.14	60.91
PCA	1.29	74.48	62.11	62.07	28.75	64.44

- PSD Frequency Range: 0-100.
The variable parameters were:
- Number of AR Coefficient: {2, 4, 6, ..., 96, 98, 100}.
The results of experiment 3 are shown in figure 3. From these results, it was observed that the apex was in the vicinity of 70 AR coefficients, however, this number of coefficients requires heavy computation and lead to over-fitting. Hence, the second peak was chosen of 30 AR coefficients.

D. Experiment 4: Finding whether feature reduction techniques will improve the accuracy

The fixed parameters were the same as Exp. 3 in addition to:
- Number of AR Coefficient: 30.
The variable parameters were:
- Feature Reduction Technique: {None, Linear Discriminant Analysis (LDA), Principle Component Analysis (PCA)}
The results of experiment 4 are shown in table I. The conclusion drawn from these results is that, using feature reduction techniques decreased the accuracy.

TABLE II
RESULTS OF EXPERIMENT 6

Number of Channels	Best Channel Combination	Accuracy		
		FA	FR	Acc
4	AF3 P8 T7 T8	1.43	5	96.79
3	AF3 P8 T7	2.14	2.5	97.68
2	AF3 P8	1.79	17.5	90.36
1	P8	0.71	35	82.14

E. Experiment 5: Applying experiments 1-4 to Emotiv

The fixed parameters were:
- Features: Welch with AR.
- Testing: 3-Folds *Day Cross Validation* with *Block Test* on 10 users.
- Channels: All 14 channels were used.
The variable parameters were:
- Classifier: Logistic Regression and SVM.
- Segment Size: {1, 2, 5, or 10}.
- Number of AR Coefficient: {in range 0 to 100}.
- PSD Frequency Range: {7-14, 14-30, 7-30, 0-30, 7-100, or 0-100}.
- Feature Reduction: {None or LDA}
The results of experiment 5 indicated the following:
- Classifier: SVM.
- Segment Size: 2.
- Number of AR Coefficient: 28.
- PSD Frequency Range: 7-100.
- Feature Reduction: None.

F. Experiment 6: Finding the best combination of Emotiv channels

The fixed parameters were:
- Classifier: Logistic Regression.
- Number of AR Coefficient: 26.
- Segment Size: 2.
- Features: Welch with AR.
- PSD Frequency Range: 7-100.
- Feature Reduction: LDA
- Testing: 3-Folds *Day Cross Validation* with *Block Test* on 8 users.
The variable parameters were:
- All the possible combinations for various number of channels.
The results of experiment 6 are shown in table II. The conclusion drawn from these results is that, using only 3 electrodes could suffice to achieve acceptable accuracy. Electrodes in positions P8 and AF3 capture variability between users during REC and are recommended to be included in the headset.

V. EXPERIMENT RESULTS

In this section, five experiments were conducted on Mindwave dataset aiming to show the prospect of building an authentication system based on EEG signals.

TABLE III
SHUFFLED CROSS VALIDATION RESULTS

Folds Number	Block Test			Day Test		
	FA	FR	Acc	FA	FR	Acc
2	0.76	25.55	86.84	11.95	0.00	94.03
4	0.38	15.72	91.95	3.93	0.75	97.66
10	0.25	10.82	94.46	0.72	5.85	96.72
20	0.26	10.03	94.86	0.27	21.86	89.03

TABLE IV
RESULTS OF CHANGING NUMBER OF TRAINING AND TESTING DAYS

Training	Testing	Block Test			Day Test		
		FA	FR	Acc	FA	FR	Acc
3 Days	1 Day	1.54	48.00	75.23	10.92	28.75	80.16
2 Days	2 Days	1.93	55.25	71.41	20.26	10.00	84.87
1 Day	3 Days	2.54	67.83	64.82	26.97	18.75	77.14
2 Days	1 Day	2.23	59.67	69.05	11.40	40.00	74.30
1 Day	2 Days	2.57	67.33	65.05	21.40	30.00	74.30
1 Day	1 Day	2.42	72.00	62.79	11.32	55.00	66.84

A. Random Cross Validation

Shuffle Cross Validation was used with *Block Test* and *Day Test* mimicking the methods used in the literature. The results shown in table III indicate that the accuracy is directly proportional to the number of folds. The high accuracies (above 90%) don't reflect a realistic testing measure, as it trains on data from different days, failing to show how the system handles variations in EEG signals.

B. Comparison between Block Test and Day Test

In over 80% of the cases, *Day Test* had higher accuracies than *Block Test*. However, *Day Test* had higher FA, and required longer testing data (50-second session instead of the 10-second block in *Block Test*). *Low FA is critical to authentication systems, while FR can be overcome by retrials.* Hence, Block Test was adopted.

C. Changing the number of training and testing days

Table IV shows the effect of changing the number of training and testing days on the overall accuracy. Several combinations of training and testing days were used. Testing method was 4-Folds *Day Cross Validation* with *Block and Day Test* on 20 users. It is observed that as the number of training days increases, the overall accuracy increases.

D. Changing the number of users

In order to test how the system behaves under varying the number of users, random removal of users from the dataset was done. Testing method was *Day Cross Validation* with *Day and Block Test*. The results for different number of users (23, 20, 13, 8, 5) are shown in figure 4. It is observed that the decrease in accuracy between 5 and 23 users is less than 10%.

Based on the number of times a user's session succeeded to get authenticated correctly as themselves, in *Day cross validation* with *Day Test*, a ranking of the users was established. In the 29-user dataset, 6 users weren't correctly authenticated

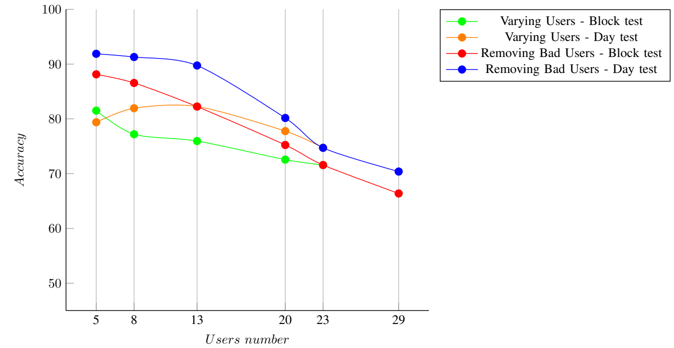

Fig. 4. The effect of reducing number of users or filtering out bad users

with any session. These were the users with the worst data (bad users). Upon removing them from the dataset, the accuracy increased by 5% as indicated in fig. 4. The accuracy increased 20%, reaching above 90% by removing these 'bad users'. The reason why these users' data was considered bad and how to accommodate for that will be investigated in future work.

E. Identification

Identification is the process of identifying a person based on the information they provide (with no claimed identity). Identification is considered harder than authentication [9]. Classifier's input isn't just +1 or -1 for each user, instead there is a single classifier for all the users, each user's data is labeled with a different class label.

The accuracy of using *Day Cross Validation* on 20 users and *Block Test* was **13.52%** which is better than the 1/20 = 5% probability of randomly guessing the correct user. While the accuracy of using *Day Cross Validation* on 20 users and *Day Test* was **67.5%**. This is much higher than Block test and it is an interesting starting point for future exploration of identifying users using EEG signals.

VI. CONCLUSION AND FUTURE WORK

The accuracy in *Shuffled Cross Validation* was high (above 90%), which is similar to the literature, despite having a larger dataset in number of users (29 users) and days (4 days). Using a more realistic testing method, i.e. *Day Cross Validation*, the accuracy was lower (between 75% and 80%), but with a low FA (below 3%) for Mindwave, and 92% for Emotiv with a FA of 0.3%. This is acceptable for an authentication system, eliminating FA, at the cost of multiple attempts from users for entry.

Increasing the number of training days from 1 to 3 days resulted in improving the overall accuracy by 11%, indicating that the model should contain data from various days to account for variations in EEG signals. Further study of the minimum interval between various sessions is needed.

Doubling the number of users from 5 to 10 yielded a 4.3% decrease in accuracy at a slope of -0.862, however, doubling the users from 10 to 23 decreased the accuracy by 5.6% with a slope of -0.432, suggesting that the system is apt to scale with increasing the number of users.

REFERENCES

[1] R. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles. The electroencephalogram as a biometric. In Electrical and Computer Engineering, 2001. Canadian Conference on, volume 2, pages 1363–1366. IEEE, 2001.

[2] J. Thorpe, P. C. van Oorschot, and A. Somayaji. Pass-thoughts: authenticating with our minds. In Proceedings of workshop on New security paradigms, pages 45–56. ACM, 2005.

[3] C. Ashby, A. Bhatia, F. Tenore and J. Vogelstein, "Low-cost Electroencephalogram (EEG) based Authentication", 5 International IEEE/EMBS Conference on Neural Engineering, pp. 442-445, 2011.

[4] J. Chuang, H. Nguyen, C. Wang, and B. Johnson. I think, therefore i am: Usability and security of authentication using brainwaves. In Financial Cryptography and Data Security, pages 1–16. Springer, 2013.

[5] Javad Sohankar , Koosha Sadeghi , Ayan Banerjee , Sandeep K.S. Gupta, E-BIAS: A Pervasive EEG-Based Identification and Authentication System, Proceedings of the 11th ACM Symposium on QoS and Security for Wireless and Mobile Networks, November 02-06, 2015, Cancun, Mexico.

[6] S. Marcel and J. del R. Millan. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4), April 2007.

[7] P.D. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms", IEEE Trans. Audio and Electroacoustics, vol. 15, pp. 70-73, 1967.

[8] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini. Unobtrusive biometric system based on electroencephalogram analysis. EURASIP Journal on Advances in Signal Processing, 2008:18, 2008.

[9] J. Sohankar, K. Sadeghi, A. Banerjee, and S. K. S. Gupta, "E-bias: A pervasive eeg-based identification and authentication system"

[10] Neurosky MindSet. http://www.neurosky.com.

[11] Emotiv Epoc+. https://www.emotiv.com/epoc/