Journal of Information Systems & Information Technology (JISIT)
ISSN: 2478-0677; Vol. 8 No.1 2023; [Pages 15 - 24]

User authentication using EEG signal

Barathy Mayurathan
barathym@univ.jfn.ac.lk

Abstract

A security system can be defined as a procedure to give protection for different types of data. A security system must perform a sequential process to achieve good protection. User authentication is a method to verify a user's identity to allow them to access and use the system. A biometric system is one of the best security systems available in this technology-driven era. In today's world, Biometric authentication is widely used to protect personal data against unauthorized access. Electroencephalography (EEG)-based authentication is considered one of the most recent developments. The subject-specific nature of brain responses in EEG, coupled with the uniqueness of everyone's brain wave pattern, poses a significant challenge in artificially recreating it. This paper introduces a user authentication system that utilizes EEG signals. The system selects different groups of electrode channels (8, 16, 32, or 64) from the EEG Motor Movement/Imagery dataset and extracts features from each channel. To condense the feature matrix, principal component analysis (PCA) is employed. The Support Vector Machine (SVM) serves as the classifier for subject authentication. The results demonstrate an accuracy of 92.20%, 92.70%, 95.13%, and 95.41% for the 8, 16, 32, and 64 channels, respectively.

Keywords: *Electroencephalography (EEG) signals, Principal component analysis (PCA), User Authentication, Support vector machine (SVM)*

1. Introduction

Biometric is the process of identifying individuals uniquely based on their physical characteristics or behavioral feature. Traditional biometric systems are failed to provide a robust solution in high-security needed locations. It is very important to have effective and autonomous user authentication systems to prevent intruder attacks and data leaks. Biometric, tokens, and passwords (known to the individual) are the three major parts [2] of Authentication systems. There are so many applications that use biological features [18] for user authentication.

Biometrics can generally be defined as a person's unique behavior or physiological characteristics. These physiological characteristics are very important in person identification [13]. Physiological biometrics means it is the shape of the body such as fingerprint, iris, Face, and DNA, while behavioral biometrics depends on the person's behavior such as typing rhythm, gait, and signature. The electrical activity in the human brain is called Electroencephalogram (EEG). The human nervous system, which includes the brain, is made up of nerve cells called neurons. Neurons communicate with each other by sending electrical signals. These signals create

voltage fluctuations that can be measured using recording electrodes. EEG signals of the brain are found to be unique to every human being [10]. EEG signals are widely used [1],[3],[5],[6] as a biometric tool for person authentication. In the EEG method, Electrical signals are used to record brain activity. Using the internal and external stimuli, neural activities are generated, and then EEG signals are used to reflect this neural activity.

EEG signals, which come from the electrical activity of the human brain, show promise for person authentication because of their unique and personalized features. Unlike conventional physiological biometrics such as fingerprints or iris patterns, EEG signals capture inherent brain activity, offering a distinct advantage in user identification. Existing studies have explored various methodologies for leveraging EEG signals in authentication systems, showcasing the potential of this approach. Techniques such as visual simulation, mental imagery tasks, eye blinking analysis, and convolutional neural networks have been employed to extract features and classify EEG signals accurately. These studies have demonstrated impressive performance in user authentication, with high accuracy rates and robustness against intruder attacks.

However, despite these advancements, there remains a literature gap in the development of EEG-based authentication systems. Existing approaches often lack comprehensive methodologies for feature extraction and classification, leading to suboptimal performance in real-world applications. Moreover, the scalability and efficiency of these systems require further improvement to meet the demands of modern security requirements.

To address these shortcomings, this paper proposes a novel authentication system using EEG signals. By employing empirical mode decomposition and principal component analysis, the system aims to enhance signal processing and feature extraction, thereby improving classification accuracy and efficiency. Additionally, the paper explores the impact of task-related electrode channels on system performance, providing insights into optimal configuration for authentication tasks.

User authentication based on EEG signals is proposed in [9]. In this paper, a novel paradigm based on visual simulation of self-photographs and non-self-photographs is presented. Visual simulation of self-photographs means the reaction of individuals to seeing their photographs and Visual simulation of non-self-photographs means the reaction of individuals to seeing others' photographs. 16 individuals are randomly selected to conduct this experiment. Additionally, this paper extracts EEG features such as fuzzy entropy, sample entropy, approximate entropy, and spectral entropy. The feature selection process employs the fisher-

based distance metric method. Finally, the performance of the system is compared using a support vector machine (SVM) with various kernel functions, resulting in an accuracy of 90.7%. In the study conducted by Ashby et al. [1], a user authentication system based on EEG signals is presented. The system involves four mental imagery tasks: baseline measurement, counting, referential limb movement, and rotation, each performed by individual subjects for 150 seconds. From this data, one-second segments are created. Three sets of features are extracted, including 6th order autoregressive (AR) coefficients, power spectral density, and total power in five frequency bands. Additionally, two more sets of features are obtained from inter-hemispheric power differences and inter-hemispheric linear complexity. These features are combined to construct a feature vector, which is then used by a linear support vector machine (SVM) with cross-validation for classification purposes.

In the study conducted by Wu et al. [14], a robust user authentication system based on EEG signals, including eye blinking, is proposed for multiple tasks. The system utilizes rapid serial visual presentation (RSVP) of both subject and non-subject faces to develop an EEG-based biometric system. EEG signals and eye blinking signals are used to extract event-related potential (ERP) features and morphological features. Convolutional neural networks and backpropagation neural networks are separately employed to compare the performance of the EEG features and eye blinking features. The performance score is calculated using score fusion technology based on the least square method. Based on their testing results, the proposed multi-task user authentication system demonstrates superior classification performance compared to other methods. To get a better classification rate for user authentication, a single-channel authentication system is introduced in [9]. A dataset that includes five mental activities that are performed by seven subjects. So, a total of 325 samples were selected to evaluate the proposed methodology. Preprocessing, feature extraction, and recognition are the main stages in this proposed authentication system. Entropy features, auto-regressive modeling, discrete wavelet transform, and Fourier transform are extracted for Subject Authentication. Then, Support Vector Machine, Bayesian network, and Neural Network are used as classifiers for this experiment.

In a study by Yu et al. [16], human EEG responses are decoded using a convolutional neural network (CNN) for user authentication purposes. To ensure consistent individualized patterns, the study focuses on the low-frequency components of steady-state visual-evoked potentials (SSVEP). Different parameter configurations are evaluated to optimize the CNN model's discriminating capabilities. The testing results demonstrate that the proposed methodology achieves an impressive accuracy of 97% when tested with 8 subjects. In the research presented by Li et al. [8], a user authentication system is introduced. The system utilizes PCA algorithms

and Convolutional Neural Networks (CNNs) to recognize P300 Electroencephalogram (EEG) signals. The PCA algorithm is employed for dimensionality reduction, enhancing the processing speed of the system. Additionally, a parallel CNN architecture is utilized to improve the traditional CNN framework. This approach increases the network depth and enhances the network's ability to accurately classify P300 EEG signals.

In the study conducted by Yu et al. [16], a framework for user authentication and identification using EEG signals is proposed. To enhance the quality of the signals, noise reduction techniques such as an average filter and low-pass filter are employed. Wavelet packet decomposition is utilized to extract frequency features from the EEG signals, and Artificial Neural Network (ANN) is used for classification. The framework is evaluated across four different scenarios involving 32 subjects. These scenarios include identifying all subjects, identifying one subject from the others, side-by-side identification of all subjects, and identifying a small group of subjects from the rest. Similarly, in the study by Wu et al. [](2018), these four different scenarios are considered for user identification using EEG signals.

In the research conducted by Sun et al. [12], a biometric identification system based on EEG signals is proposed. They introduce a one-dimensional Convolutional LSTM neural network in their work. To evaluate their methodology, they utilize the EEG Motor Movement/Imagery dataset, which consists of data from 109 subjects. The spatial and temporal features of the EEG signals are considered as a set of features. The data is split, with 90% used for training and 10% for testing. Remarkably, they achieved a high level of accuracy of 99.58% using only 16 channels.

The main goal of this paper is to create a new authentication system using EEG signals. The paper utilizes empirical mode decomposition to break down the signals into different components and separate frequency bands based on the data itself. Additionally, signal complexity and frequency band information are calculated. The paper also selects various task-related electrode channels (8, 16, 32, or 64) and constructs matrices of size 18 (features) × 8/16/32/64 (channels) accordingly. Principal component analysis (PCA) is then applied to reduce the size of the feature matrix. Finally, a reduced feature matrix of size 18 × 2 is employed to train the classification model.

2. Proposed Methodology:

The proposed user authentication system is illustrated in Figure 1. It comprises the following stages. The main stages of the proposed methodology are described below:

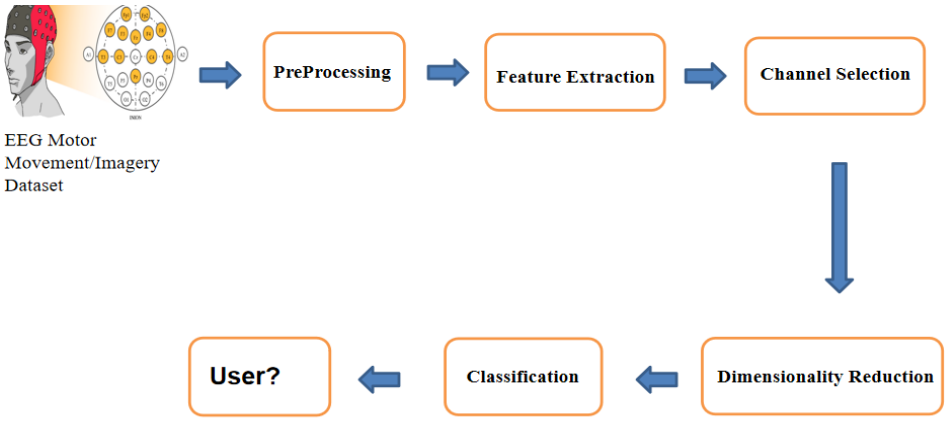

Figure 1: proposed user authentication system

I. Preprocessing

a. Read .edf file: For this experiment, we utilized the EEG Motor Movement/Imagery dataset [4], which is publicly available on the internet. The dataset is in the. edf format, and it contains valuable information related to motor movement and imagery captured through EEG recordings.

b. Filter data and separate Tasks T1 (opening and closing left side) and T2 (opening and closing right side): To process the data, a zero-phase delay filter is applied with cutoff frequencies set at 1Hz and 50Hz. This filtering helps to remove unwanted noise and artifacts from the data. After filtering, the tasks are separated into T1 (left side) and T2 (right side) for further analysis or processing.

c. Cut data: Once T1 and T2 tasks are separated, sliding windows with a length of 2 are applied. This process involves breaking the data into smaller segments. Around 10 trials per subject are obtained for each task, T1 and T2. These trials are then combined and saved as a single .mat file for further analysis or storage.

d. Trial data is collected from multiple different users. Imposters are unauthorized individuals attempting to access the system. 20% of the test data represents a subset of these users for evaluation.

II. Feature extraction

After the preprocessing stage, the data undergoes Empirical Mode Decomposition (EMD), which is a technique introduced by Zeiler et al. [17]. EMD helps to obtain a complete and finite set of components called Intrinsic Mode Functions (IMFs). The stoppage criteria used for EMD include a resolution of 40dB and residual energy of

60dB. From the obtained IMFs, the first four IMFs are considered as they contain the most important information. To analyze the data further, Spectral Density (PSD) is calculated using a multi-taper method from the Chronux toolbox. A total of 18 features are extracted, including Shannon entropy, logarithmic entropy, approximate entropy, and sample entropy. Additionally, the average power of frequencies μ (μ = 7.5 − 12.5Hz) and β (β = 16−31Hz) is calculated using the PSD. In order to analyze the data across multiple channels, the cross-correlation method is used. This method involves calculating the mean and standard deviation of the cross-correlation between all pairs of channels for each feature. This provides insights into the relationships and interactions between different channels.

III. Channel Selection

Channel selection plays a crucial role in improving the performance of the EEG authentication system. By carefully choosing specific channels, we can enhance the system's accuracy and efficiency. In this experiment, we consider four different sets of channel selections and compare their respective performances. This analysis allows us to determine which channel selection yields the best results in terms of classification accuracy and computational complexity. 8 Channels (F4, Fp1, Fp2, C1, C2, Fc1, Fc2, F3), 16 Channels (F4, Fp1, Fp2, C1, C2, Fc1, Fc2, F3, Fz, Fcz, C3, C4, F1, F2, Af3, Af4), 32 Channels (Af4, Af3, F4, Fp1, Fp2, C3, C4, Fc3, Fcz, F3, F7, Fc4, Ft8, T7, Cz, T8, Tp7, Cp3, Cp4, Tp8, P7, P3, Pz, P4, P8, O1, O2, Oz) and all 64 Channels are the four different types of channels used in this work.

IV. Dimensionality Reduction

To handle different channel selections, we construct matrices of varying sizes (18 × 8, 18 × 16, 18 × 32, 18 × 64). However, to reduce the total number of features and streamline the analysis, we employ multichannel analyses using the cross-correlation method. This helps us identify and extract relevant information from the data. Additionally, we utilize Principal Component Analysis (PCA) [7] as a dimensionality reduction technique. By applying PCA, we can condense the feature space and focus on the most significant components for further analysis.

V. Classification

The dataset is split into two sections: 20% for testing and 80% for training. From the dataset, 105 trials from users and 105 trials from non-users (imposters) are randomly chosen. To evaluate the performance of each channel selection, a 5-fold cross-validation approach is employed. This helps ensure robustness and reliability in the classification process. To assess the accuracy of the system, the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are calculated. These metrics

provide insights into the system's ability to correctly identify genuine users (low FAR) and reject imposters (low FRR). By analyzing these rates, we can determine the effectiveness of the different channel selections in achieving accurate user authentication.

3. Results and Discussion:

To compute the performance of the proposed methodology, Motor Movement / Imagery dataset is chosen which is publicly available on the PhysioNet data bank [4]. This dataset has 1500 one and two-minute EEG recordings gathered from a total of 109 subjects. Each subject participated in 14 different experiments: {Baseline, eyes open}, {Baseline, eyes closed} {Task1 (open and closed left or right fist)}, {Task4 (imagine opening and closing both first or both feet)}, {Task1}, {Task2}, {Task3}, {Task4}, {Task1}, {Task2}, {Task3}, {Task4}.

To assess the performance of the proposed methodology, we utilize two important metrics: the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). FAR measures the rate at which the system incorrectly accepts unauthorized users, while FRR measures the rate at which the system incorrectly rejects legitimate users. By analyzing these rates, the accuracy and reliability of the methodology can be evaluated. Additionally, performance indicators such as False Negatives (FN), False Positives (FP), True Positives (TP), True Negatives (TN), and Area Under Curve (AUC) may be considered to provide a comprehensive assessment. Based on testing results presented in Table I, accuracy is highest with 64 channels at 95.41%, with a FAR of 1.36%. For 32 channels, accuracy is 95.13% with a FAR of 1.39%. Experimental results suggest that as the number of channels increases, performance also improves.

Channel	Accuracy	FRR	FAR	AUC
8	0.922	0.133	0.022	0.984
16	0.927	0.123	0.021	0.985
32	0.951	0.084	0.014	0.990
64	0.954	0.781	0.014	0.991

Table I: Performance of the proposed methodology

Performance comparison between proposed methodologies and state-of-the-art approaches:

The classification performances of the proposed methodologies are compared with state-of-the-art methods to calculate the efficiencies of our proposed methodology.

Table II shows the classification performances of the different state of-the-art methods and our method.

Existing Methods	Proposed Methodology	Accuracy (in %)
Mu et al. [10]	SVM	90.7%
Wu et al., [14]	CNN	92.4%
Li et al., [8]	PCA + CNN	95%
Our method	**PCA + SVM**	**95.42%**

TABLE II: Performance Comparison with the State-of-the-art work.

Based on the performances reported in Table II, it can be easily concluded that our proposed methodology gives better classification performances compared to the state-of-the-art methods mentioned in Table II.

4. Conclusion

Dimensionality reduction and channel selection play crucial roles in our proposed methodology. Based on the results of our experiments, Principal Component Analysis (PCA) has proven effective in improving our methodology's performance. It not only reduces computational time but also enhances accuracy.

We tested our methodology with four different channel selections: 8, 16, 32, and 64 channels. The results demonstrate that the 64-channel selection outperforms the others. This indicates that using more channels leads to higher accuracy. Specifically, our PCA-SVM approach achieves the highest accuracy of 95.41% with the 64-channel selection. To further advance our research, we plan to evaluate our approach's performance using different EEG datasets and compare the results with various classification techniques. This will provide a more comprehensive understanding of the effectiveness and generalizability of our proposed methodology.

References

1. C. Ashby, A. Bhatia, F. Tenore and J. Vogelstein, “Low-Cost Electroencephalogram (EEG) Based Authentication”, *In Proceedings of the 5th International IEEE/EMBS Conference on Neural Engineering*, Cancun, Mexico, pp. 442 – 445. 2011. DOI: 10.1109/NER.2011.5910581
2. H. Bojinov and D. Boneh, “Mobile token-based authentication on a budget”, *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, USA, 14–19, 2011. DOI: https://doi.org/10.1145/2184489.2184494
3. J. Chuang, H. Nguyen, C. Wang, and B. Johnson, “I Think, Therefore I Am: Usability and

Security of Authentication Using Brainwaves", *In Proceedings of the Financial Cryptography and Data Security, Lecture Notes in Computer Science*, vol 7862, Springer, Berlin, Heidelberg, 2013. DOI: https://doi.org/10.1007/978-3-642-41320-9_1

4. A. L. Goldberger, L. A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R. G. Mark, J.E. Mietus, G. B. Moody, C.K. Peng, and H.E. Stanley, *PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,* 13;101(23): E215-20, 2000, doi: 10.1161/01.cir.101.23.e215. PMID: 10851218.
5. Q. Gui, Z. Jin, and W. Xu, "Exploring EEG Based Biometrics for User Identification and Authentication", In *Proceedings of 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, USA, pp. 1–6, 2014, DOI:10.1109/SPMB.2014.7002950.
6. I. Jayarathne, M. Cohen and S. Amarakeerthi, "BrainID: Development of an EEG-Based Biometric Authentication System", In *Proceedings of the 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Canada, pp. 1–6, 2016, DOI: 10.1109/IEMCON.2016.7746325.
7. I. Jolliffe, *Principal Component Analysis. In: Lovric, M. (eds) International Encyclopedia of Statistical Science*, Springer, Berlin, Heidelberg, pp. 1094—1096, 2011 DOI: https://doi.org/10.1007/978-3-642-04898-2_455
8. F. Li, X. Li, F. Wang, D. Zhang, Y. Xia, and F. He, "A Novel P300 Classification Algorithm Based on a Principal Component Analysis - Convolutional Neural Network". *Applied Sciences*,10(4), 1546, 2020. DOI: https://doi.org/10.3390/app10041546.
9. Z. Mahsa and S. Hadi, "EEG-based single-channel authentication systems with optimum electrode placement for different mental activities", *Biomedical Journal*, Volume 42(4), pp. 261-267, 2019, DOI: https://doi.org/10.1016/j.bj.2019.03.005.
10. Z. Mu, J. Hu, J. Min, and J. Yin, "Comparison of different entropies as features for person authentication based on EEG signals", *IET Biom.*, 6, pp. 409 – 417, 2017, DOI: https://doi.org/10.1049/iet-bmt.2016.0144.
11. A. Priyanka, W. Bharti, and C. Suresh, *Chapter 2 - Technological Basics of EEG Recording and Operation of Apparatus, Introduction to EEG and Speech-Based Emotion Recognition*, Academic Press, Pages 19-50, 2016, ISBN 9780128044902.
12. Y. Sun, P. W. Frank, and B. Lo, "EEG-based User Identification System Using 1D-Convolutional Long Short-Term Memory Neural Networks", *Expert Systems with Applications*, vol. 125, pp. 259-267, 2019. DOI: https://doi.org/10.1016/j.eswa.2019.01.080.
13. A. Valsaraj, I. Madala, N. Garg, M. Patil, and V. Baths, "Motor Imagery Based Multimodal Biometric User Authentication System Using EEG", In *Proceedings of the 2020 International Conference on Cyberworlds (CW)*, France, pp. 272 – 279, 2020, DOI:10.1109/CW49994.2020.00050
14. Q. Wu, Y. Zeng, C. Zhang, L. Tong, and B. Yan, "An EEG-Based Person Authentication System with Open-Set Capability Combining Eye Blinking Signals", *Sensors (Basel)*, 18(2), 335, 2018. DOI: https://doi.org/10.3390/s18020335.
15. S. Yeom, H. Suk, and S. Lee, "Person Authentication from Neural Activity of Face-Specific Visual Self-Representation" *Pattern Recognition*, 46, Pages 1159–1169, 2013, DOI: https://doi.org/10.1016/j.patcog.2012.10.023.
16. T. Yu, C. S. Wei, K. J. Chiang, M. Nakanishi, and T. P. Jung, "EEG-Based User Authentication Using a Convolutional Neural Network", *9th International IEEE/EMBS Conference on*

Neural Engineering (NER), San Francisco, USA, pp. 1011-1014, 2019, DOI: 10.1109/NER.2019.8716965.

17. A. Zeiler, R. Faltermeier, I. R. Keck, A. M. Tomé, C. G. Puntonet, and E. W. Lang, "Empirical Mode Decomposition - an introduction", *The 2010 International Joint Conference on Neural Networks (IJCNN)*, Barcelona, Spain, pp. 1-8, 2010, DOI: 10.1109/IJCNN.2010.5596829.
18. S. Zhang, L. Sun, Z. Mao, C. Hu, and P. Liu, "Review on EEG-Based Authentication Technology", *Computational Intelligence and Neuroscience*, vol. 2021, 1-20, 2021. DOI: https://doi.org/10.1155/2021/5229576