An EEG-Based User Authentication System Using Event-Related Potentials and Ensemble Learning

Soudabeh Bolouri
Dept. of Electrical Engineering & Computer Science
University of Wyoming
Laramie, WY, USA
sbolouri@uwyo.edu

Diksha Shukla
Dept. of Electrical Engineering & Computer Science
University of Wyoming
Laramie, WY, USA
dshukla@uwyo.edu

2024 Cyber Awareness and Research Symposium (CARS) | 979-8-3503-8641-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/CARS61786.2024.10778891

***Abstract*—This paper presents an EEG-based user authentication system using Event-Related Potentials (ERPs) to distinguish legitimate users from impostors. Utilizing a publicly available EEG dataset, we implemented a comprehensive data processing pipeline, which included advanced preprocessing and feature extraction techniques. Multiple state-of-the-art machine learning classifiers, such as CatBoost and XGBoost, were evaluated to assess their effectiveness in user authentication. The results showed a very low average Equal Error Rate (EER) of 2.53%. Our study emphasizes the strength of the P300 and N400 responses in biometric authentication and demonstrates the potential of advanced ensemble classifiers in improving system accuracy. This research contributes to the development of EEG-based authentication and lays the groundwork for future studies aiming to create secure and practical biometric systems.**

***Index Terms*—Brainwave Authentication, electroencephalogram (EEG), Biometrics, User Authentication, Brain-Computer Interface (BCI), Event-Related Potentials (ERPs)**

I. Introduction

A user authentication system ensures authorized access to various applications, from personal devices to secure facilities. Traditional user authentication methods such as passwords and PINs are at risk of being forgotten, stolen, or guessed [1]. Biometric authentication identifies users using unique physiological traits (fingerprints, face recognition, iris) or behavioral traits (signature, keystrokes, voice). Compared to traditional methods such as passwords or PINs, it provides higher accuracy and convenience. However, due to the possibility of biometric data imitation and spoofing, this raises privacy concerns [2]. On the other hand, brainwave authentication using electroencephalography (EEG) presents a promising alternative. In contrast to other biometric traits, brainwaves are hidden and difficult to observe or duplicate, so they hold great security against spoofing and fraud [3].

This research explores the potential of ERPs captured through an EEG-based system to effectively differentiate between genuine users and impostors. Researchers have identified several components in ERPs to explain how deviant and unexpected events are processed. The N400 and P300 are two examples of these components [4]. The P300 response, associated with attention and decision-making, occurs approximately 300 milliseconds after a stimulus presentation [5]. The response associated with semantic processing, known as the N400 response, occurs approximately 400 milliseconds after the stimulus presentation [6].

The potential of EEG-based authentication is significant, but there are still several challenges that need to be addressed. These include the influence of artifacts such as eye blinks and muscle movements, as well as the complexity of accurately classifying brainwave data [7]. Previous studies have investigated different preprocessing, feature extraction, and classification techniques to tackle these challenges, with varying levels of success.

In this study, our goal is to advance the field of EEG-based authentication by using a comprehensive data processing pipeline and assessing the performance of various state-of-the-art machine learning classifiers. We utilize a publicly available EEG dataset that is specifically created to evoke ERPs suitable for authentication [8]. Our approach includes preprocessing steps to enhance signal quality, robust feature extraction methods to capture relevant characteristics of the EEG signals, and advanced feature selection techniques to identify the most important attributes. Additionally, we utilize cutting-edge classifiers to achieve high accuracy in differentiating genuine users from impostors.

Our model achieved an Equal Error Rate (EER) of 2.53%, demonstrating a significant improvement of approximately 64.86% compared to previous research [9]. These results reveal the robustness and effectiveness of our approach in the realm of EEG-based user authentication.

The main contributions of this study are as follows:

1) Independent Component Analysis (ICA) [10] was employed to remove artifacts and improve signal quality, and we demonstrated that this method is effective for improving EEG-based authentication.
2) A comprehensive feature set was extracted that included all EEG signals' essential characteristics, and the most relevant ones were selected based on their importance.
3) Evaluation of various classifiers, including ensemble methods, to identify the most effective models for EEG-based authentication [11].

The rest of this paper is structured as follows. Section II reviews relevant literature. The dataset and stimuli utilized in our work for EEG-based user authentication are described in

Section III. Section IV presents the brainwave data cleaning and processing pipeline details. The design of the authentication system is outlined in Section V. Section VI presents performance results and comparisons. In Section VII, we discuss the implications and future directions. Finally, Section VIII summarizes key insights and the importance of our findings.

II. Related Work

EEG-based user authentication is an active area of study that employs the unique and complex nature of brainwave patterns to improve security systems. In this section, we will examine notable investigations that have used EEG data for authentication. We will concentrate on the methodologies and results presented compared to our investigation.

Event-related potentials, specifically P300 and N400, have been broadly utilized in EEG-based authentication studies due to their specific and identifiable patterns [12].

The P300 wave, which happens approximately 300 milliseconds after stimulus presentation, is generally used in oddball tasks where users recognize rare target stimuli among frequent non-targets. For instance, Alzahrani [13] and Koike-Akino et al. [14] utilized the P300 response elicited through an oddball paradigm, and demonstrated the potential of using P300 responses for biometric authentication, highlighting the advantages of using consumer-grade EEG devices for such applications.

The response known as N400, occurring approximately 400 milliseconds after the stimulus, has been extensively studied. Hamm et al. [15] analyzed the N400 effect on pictures that were semantically incongruous to a previously presented object name. They found that the N400 reacted to all semantic mismatches. Moreover, Franklin et al. [16] investigated semantic priming mechanisms and specified an N400 effect for semantic matching, revealing the potential of N400 in indicating between individuals based on their brainwave patterns.

Recent studies have demonstrated that cutting-edge machine-learning algorithms can significantly enhance the performance of EEG-based authentication systems. Safont et al. [17] introduced a biometric authorization and identification method based on EEG signals, utilizing multiple detectors and classifiers to improve performance. The study utilized various classifiers, including Discriminant Analyzers (DA, [18]), Classification Trees (TREE, [19]), and a simple copula-based classifier (COP, [20]). These classifiers were combined in different ways to enrich the system's robustness and accuracy.

Ensemble techniques like Random Forests, XGBoost, and CatBoost are highly effective in handling high-dimensional and complicated data structures. A comparative study by Sezer et al. [21] assessed the performance of Random Forest, XGBoost, and a newly introduced ensemble method called NGBoost in landslide susceptibility mapping. This study illustrated the outstanding performance of ensemble methods in complex classification tasks.

While previous research has laid a solid foundation for EEG-based authentication, many studies have not fully leveraged advanced artifact removal techniques like ICA [22], which can significantly enhance signal quality. In our study, we implemented ICA to eliminate artifacts such as eye blinks and muscle movements, resulting in cleaner EEG signals for feature extraction and a lower signal-to-noise ratio. This improvement makes the data more suitable for further processing and analysis. In addition, most studies do not explore extensive feature extraction methods and the selection of the best features, which is also covered in this study. Although ensemble methods have shown promise, there is a lack of comparative analysis using effective classifiers like CatBoost [23]. Our study includes a detailed comparison of various classifiers, demonstrating that CatBoost outperforms traditional methods. Therefore, ensemble methods can improve the overall performance of the authentication system for high-dimensional data.

III. Dataset Description

In our study, we utilized a publicly available EEG dataset, "Brainwave Authentication Dataset," [8] consisting of 38 participants. in this dataset EEG data was collected using the Emotiv EPOC+ headset, which utilizes 14 channels [24].

The recording procedure started by taking baseline measurements of brain activity, where participants were instructed to open their eyes for 20 seconds and then close them for 25 seconds. Following this, participants engaged in several authentication tasks designed to produce different ERPs. The specific stimuli used are as follows [9]:

- $P300/Selected$: This task employed the oddball paradigm to elicit the P300 response. Each participant was instructed to choose an image from a set and then watch a series of images with the chosen image appearing infrequently. In order to strengthen the P300 response, they were asked to keep track of the occurrences of the target image.
- $P300/Assigned$: This is similar to the previous task, but the target image was assigned to the participants rather than selected by them.
- $N400/Words$: The task involved showing participants a video of cars driving on a highway. After watching the video, participants were presented with words, some related to the video and others random.
- $N400/Sentences$: This task involved reading sentences that ended either congruently or incongruently. For instance, participants might read "Steve sat down to eat his car," where the unanticipated final word "car" evoked a strong N400 response due to its semantic incongruence.
- $N400/Faces$: Participants were presented with sequences of well-known faces (celebrities) followed by unfamiliar faces. The N400 response was triggered when an unfamiliar face appeared after a sequence of familiar faces, as the brain needed to inhibit previously activated semantic representations and activate new ones for the unfamiliar face.

IV. Brainwave Data Processing

The brainwave data processing pipeline is required to extract meaningful features from raw EEG signals, which are then used for classification. The processing pipeline consists of two primary steps: preprocessing and feature extraction [25].

A. Preprocessing

The EEG recordings, captured at a sampling rate of 256 Hz, depict brain activity data over a comprehensive period. To improve the signal quality, we have taken several steps to remove artifacts. First, a bandpass filter (1–50 Hz) was applied to eliminate low-frequency drift and high-frequency noise, ensuring that the signal retained only the relevant frequency components and minimizing the effects of electrical noise and physiological artifacts.

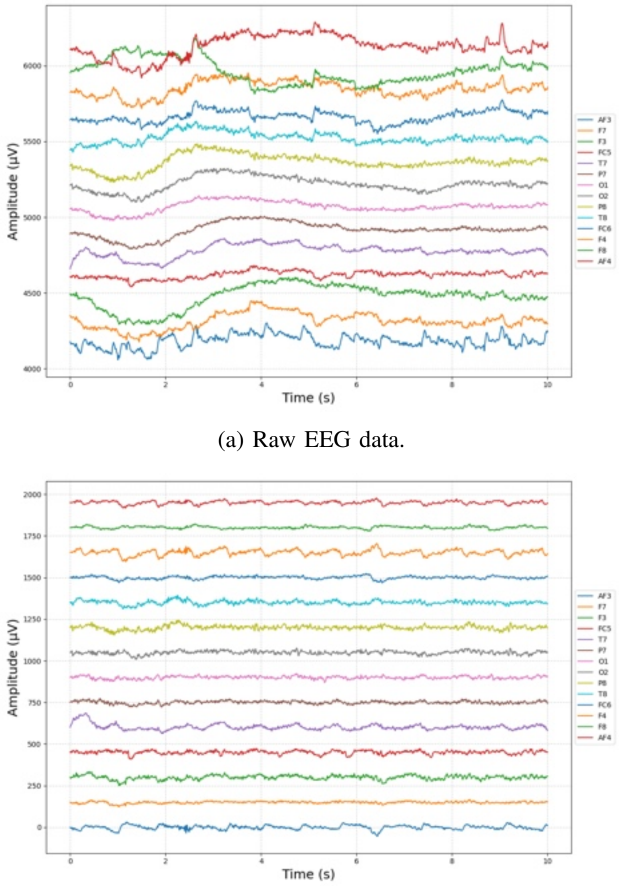
Next, we employed ICA to determine and remove components related to eye blinks, muscle movements, and other non-brain artifacts. ICA separates the mixed signals into statistically independent components, enabling us to isolate and exclude those corresponding to artifacts. The frontal channels, which are most impacted by eye blinks and movements, received particular attention during this technique. By focusing on the channels AF3, AF4, F7, F3, F4, and F8, we sought to identify and remove components associated with eye-related artifacts. The remaining independent components, which were presumed to be free from significant artifact influence, were then recombined to reconstruct the cleaned EEG signals.

We also performed baseline removal to correct potential drifts and offsets in the EEG data. To avoid transient effects, we excluded the first and last three seconds of the open-eye and close-eye periods. After this trimming, we calculated the mean signal values across all EEG channels for each period. These mean values were then averaged and subtracted from the related channels in the EEG data. "Fig. 1" demonstrates the effect of the preprocessing steps, showing a clear improvement in signal quality and removing artifacts from the raw EEG data.

We divided the continuous EEG data into 1-second segments to focus on the relevant parts of the signal. This segmentation process captures the brain's response to specific stimuli by breaking the data into smaller, more manageable units. Each segment starts 100 milliseconds before the stimulus presentation and continues for 900 milliseconds after the presented stimulus. The 1-second windows were non-overlapping, meaning each segment represents a distinct, continuous portion of the EEG recording. The segments were labeled based on the type of stimulus presented during the time window they captured. This labeling process allowed us to associate each data segment with its representative stimulus, facilitating subsequent classification and analysis tasks.

B. Feature Extraction

After preprocessing the EEG signals, the next crucial step is feature extraction, which transforms the cleaned data into measurable characteristics for classification.



(a) Raw EEG data.

(b) EEG data after artifact removal.

Fig. 1: EEG signals before and after preprocessing.

We used the *Discrete Wavelet Transform (DWT)* [26] to break down the EEG signal into different frequency components at various resolution levels. We specifically looked at the wavelet coefficients at level *four*, which effectively capture detailed variations in the signal. We calculated the mean and standard deviation for each of the levels (including the approximation at level four and the details at levels 1 to 4). This gave us ten features per channel. With 14 channels, we ended up with a total of 140 features.

We calculated the *Power Spectral Density (PSD)* [27] values for different frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–50 Hz). To get these features, we used the Welch method to assess the PSD of each epoch. Since there are five frequency bands and 14 channels, this resulted in 70 features.

Combining these statistical and frequency domain features, we produced a comprehensive feature set that encapsulates the essential characteristics of the EEG signals. In total, we extracted 210 features for each epoch (140 from the wavelet transform and 70 from the power spectral density analysis). "Table I" provides a summary of the EEG datasets used in this study.

TABLE I: The EEG dataset categorized by stimuli type.

Stimulus	Number of Users	Number of Samples
$P300/Selected$	38	717
$P300/Assigned$	38	720
$N400/Words$	38	1417
$N400/Sentences$	38	222
$N400/Faces$	38	360

V. Authentication System Design

In our EEG-based authentication system, newly acquired EEG signals are compared with stored templates to verify user identities. A classifier is employed to determine whether the presented sample belongs to a genuine user or an impostor. If the matching score exceeds a certain threshold, access is granted. Otherwise, the system classifies the user as an impostor, and access is denied.

A. Classification Method

The authentication system uses a two-class classification approach, in which classifiers are trained to distinguish genuine users from impostors. In this context, a genuine user is someone authorized to access the system, while an impostor is someone who is not. We trained each user separately, considering them as genuine user, and treated all other users as impostors. This means that the data from each genuine user was labeled as class 1, and the data from all other users was labeled as class 0. As a result, the classifier learns to distinguish between the EEG patterns of the authentic user and those of the impostors.

To evaluate the performance of the classifiers, we divided the dataset into training and testing sets, using an 80/20 ratio, with 80% of the data for training and 20% for testing. Stratified sampling was used to maintain the same proportion of genuine and impostor samples in each set.

B. Feature Selection

Feature selection was used to enhance the performance of the classifiers by selecting the most relevant features from the EEG data. The Random Forest (RF) feature selection method was used because it effectively manages high-dimensional data and ranks features according to their importance. During training, the Random Forest algorithm constructs multiple decision trees and outputs the mean prediction from each tree. During feature selection, the importance of each feature is calculated by evaluating the decrease in model accuracy when it is excluded. As a result, higher importance scores are selected for classifier training since they contribute more significantly to the model's prediction [28].

C. Machine Learning Methods

To evaluate our method's effectiveness, we employed different classifiers. These classifiers were trained and tested on both genuine and impostor data. The classifiers used include: *Linear Discriminant Analysis (LDA)* [29], *Neural Networks (NNs)* [30], *AdaBoost* [31], *XGBoost (XGB)* [32], *CatBoost* [23], *Random Forest (RF)* [33], *Logistic Regression (LR)* [34], *Support Vector Machines (SVM)* [35] with linear and RBF kernels, and *Extra Trees (ET)* [36].

The classifiers were trained on a balanced dataset with samples from genuine users and impostors. Before training, we standardized the features. Afterward, classifiers were evaluated regarding how well they detected genuine users and rejected impostors. The process was repeated for each user in the dataset, treating them as genuine while assuming others were impostors. Equal Error Rate (EER) was used as a critical performance metric, representing the point where the rates of false acceptances and rejections are equal. As a final assessment of the system's overall performance, the average EER across all users was reported.

VI. Results

Our results are shown in "Fig. 2", evaluating the performance of various classifiers using the average EER and standard deviation (SD) across different stimulus conditions.

The best EER acquired was 2.53% (SD: 4.58%) with the CatBoost classifier for the $N400/Words$ stimulus, indicating its

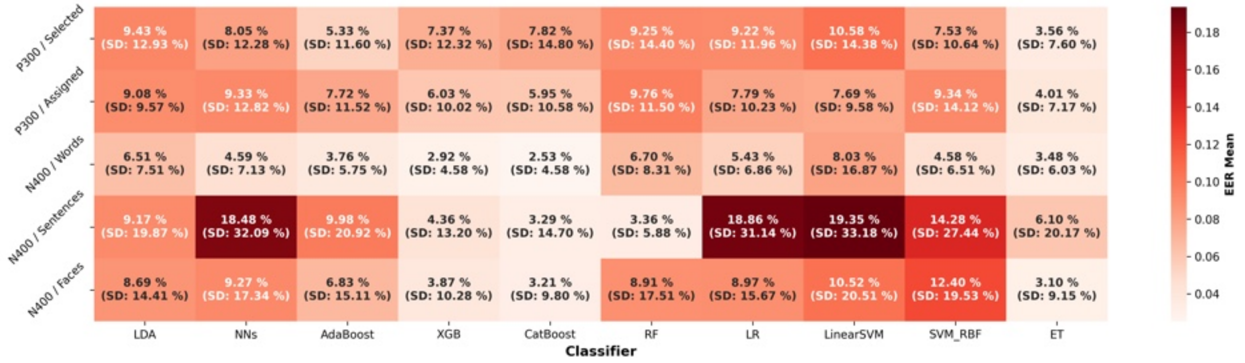


Fig. 2: Average Equal Error Rates (EER) for various classifiers across different EEG stimuli. This heatmap displays the mean EER (in percentage) and the corresponding standard deviations for different classifiers when applied to EEG data recorded in response to various stimuli. The EER values are color-coded, with darker shades indicating higher error rates.

superior ability to handle this specific condition. XGBoost also showed strong performance, achieving an EER of 2.92% (SD: 4.58%) for the same stimulus. Moreover, AdaBoost achieved an EER of 3.76% (SD: 5.75%) for the $N400/Words$ stimulus. These results demonstrate the effectiveness of these classifiers in capturing the complex patterns in EEG signals associated with the semantic processing of words.

The other ensemble method, Random Forest, showed consistent performance with an EER of 3.36% (SD: 5.88%) for the $N400/Sentences$ stimulus.

Extra Trees also showed strong performance, achieving an EER of 3.10% (SD: 9.15%) for the Faces stimulus and 3.48% (SD: 6.03%) for the $N400/Words$ stimulus. These results highlight the robustness of Extra Trees in managing diverse and complex EEG data related to facial and word stimuli.

The results demonstrate that advanced ensemble methods including XGBoost, AdaBoost, CatBoost, Random Forest, and Extra Trees are highly effective for our EEG-based authentication, particularly with $N400$ stimuli. Selecting appropriate tasks and machine learning models is essential to verifying the reliability and robustness of user authentication. Classifiers that can handle high-dimensional and complex data structures perform better in distinguishing genuine users from impostors.

Neural Networks and SVM with RBF kernel have demonstrated their adaptability, achieving an EER of 4.61% (SD: 6.75%) and 4.58% (SD: 6.51%), respectively, for the $N400/Words$ stimulus. Their performance, while mixed, suggests that these classifiers are robust and can adapt to different stimuli. However, their performance heavily depends on the specific characteristics of the data and the complexity of the patterns they need to learn, indicating the need for a nuanced approach in their application.

Comparing the stimuli performance, $N400/Words$ and $N400/Faces$ tasks generally yielded lower EERs across classifiers, suggesting these stimuli are robust and reliable markers for distinguishing between genuine users and impostors. The distinct nature of these responses likely contributes to better classifier performance. $P300$ stimuli, especially $P300/Selected$, also performed well for the Extra Trees classifier but were more variable across classifiers. The $N400/Sentences$ stimulus had higher EERs for many classifiers, indicating that this task may introduce more variability and complexity, making it more difficult for classifiers to generalize effectively. Semantic incongruence in sentence completion tasks may vary particularly among individuals, contributing to increased EERs.

VII. Discussion

Arias-Cabarcos et al. [6] explored ERP-based authentication tasks with consumer-grade EEG equipment, reporting average EERs from 14.5% to 34.14%. In another study, Arias-Cabarcos et al. [9] reported average EERs of 7.2% to 12.3% when examining closed-set and open-set scenarios. A closed-set scenario is described, where classifiers are trained using all user data, including attackers, labeling samples either as authenticated or rejected. In contrast, an open-set scenario involves training only authenticated and known rejected users. The system must identify and reject unseen attackers during testing, simulating a more realistic authentication challenge. A comparison of open-set and closed-set scenarios revealed that the EER increased by 5.1% to 18% in the open-set scenario due to unknown attackers, highlighting the difficulty associated with generalizing to a new, unknown user group.

In this work we used the dataset presented by Arias-Cabarcos et al. [6], [9]. Our research showed significant advancements in EEG-based authentication by achieving a lower EER compared to existing studies. Our system achieved an average EER of 2.53% using the CatBoost classifier with the $N400/Words$ stimulus, which is a notable improvement compared to the other studies.

We further analyze each factor that contributed to the improvement in our study:

- *Preprocessing Methods:* A robust preprocessing pipeline was employed that included ICA for removing eye blinks and muscle movements. As a result, higher-quality EEG signals were obtained for feature extraction.
- *Enhanced Feature Extraction:* Our feature extraction methodology combines time-domain features with frequency-domain features, including Power Spectrum analysis and Discrete Wavelet Transform. As a result of this comprehensive approach, we can extract a wide range of features from EEG signals, which increases the discriminative power of these features.
- *Feature Selection:* We used Random Forests for feature selection, which reduced dimensionality and improved classifier performance. We noted that the model performed better, resulting in a lower EER when we trained it with selected features.
- *Effective Use of Ensemble Classifiers:* In our research, we used ensemble classifiers such as CatBoost, XGBoost, and Extra Trees, which are known for being able to handle high-dimensional and complex data structures. Using these classifiers ensures that the model is robust against overfitting and improves its generalization abilities.

VIII. Conclusion

This study demonstrated significant advancements in EEG-based authentication by showcasing the effectiveness of modern ensemble classifiers, in achieving low EERs. Our system achieved an impressive average EER of 2.53% on the $N400/Words$ stimulus. The results of this study are notable improvements over previous studies, demonstrating the potential of these methodologies to improve the accuracy and reliability of EEG-based biometric systems.

The results highlighted the importance of using sophisticated signal processing techniques to clean and preprocess EEG data, assuring that the extracted features accurately reflect the underlying neural activity. Combining time-domain and frequency-domain features, along with ensemble classifiers' robustness, produced highly effective results for distinguishing genuine users from imposters.

Although the results of the study are promising, it acknowledges some limitations, including the need for comprehensive and diverse datasets to further validate the system's robustness. In addition, future research could explore how these techniques can be applied in real-time and how other advanced feature extraction techniques can be incorporated to further enhance performance.

Overall, this paper showed how modern ensemble classifiers and advanced preprocessing techniques can be utilized to achieve low EERs to implement robust EEG-based authentication. As a result of this work, future developments in secure biometric authentication systems will be able to improve the security measures of a variety of applications, contributing to the broader field of biometrics.

Acknowledgment

This work was supported by the National Science Foundation (NSF) under grant SaTC: 2340997.

References

[1] N. Siddiqui, L. Pryor, and R. Dave, "User authentication schemes using machine learning methods—a review," in *Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2021*, Springer, 2021, pp. 703–723.

[2] D. Bhattacharyya, R. Ranjan, F. Alisherov, M. Choi, *et al.*, "Biometric authentication: A review," *International Journal of u-and e-Service, Science and Technology*, vol. 2, no. 3, pp. 13–28, 2009.

[3] S. J. Alsunaidi, N. A. Saqib, and K. A. Alissa, "A comparison of human brainwaves-based biometric authentication systems," *International Journal of Biometrics*, vol. 12, no. 4, pp. 411–429, 2020.

[4] Y. Arbel, K. M. Spencer, and E. Donchin, "The n400 and the p300 are not all that independent," *Psychophysiology*, vol. 48, no. 6, pp. 861–875, 2011.

[5] M. Yu, N. Kaongoen, and S. Jo, "P300-bci-based authentication system," in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*, IEEE, 2016, pp. 1–4.

[6] P. Arias-Cabarcos, T. Habrich, K. Becker, C. Becker, and T. Strufe, "Inexpensive brainwave authentication: New techniques and insights on user acceptance," in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 55–72.

[7] A. J. Bidgoly, H. J. Bidgoly, and Z. Arezoumand, "A survey on methods and challenges in eeg based authentication," *Computers & Security*, vol. 93, p. 101 788, 2020.

[8] M. Fallahi, *Brainwave authentication dataset and experiment material*, https://github.com/kit-ps/bainwave-authentication, accessed: July, 2024.

[9] P. Arias-Cabarcos, M. Fallahi, T. Habrich, K. Schulze, C. Becker, and T. Strufe, "Performance and usability evaluation of brainwave authentication techniques with consumer devices," *ACM Transactions on Privacy and Security*, vol. 26, no. 3, pp. 1–36, 2023.

[10] J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, and J. Artieda, "Independent component analysis as a tool to eliminate artifacts in eeg: A quantitative study," *Journal of clinical neurophysiology*, vol. 20, no. 4, pp. 249–257, 2003.

[11] P. Jain, S. Tatale, N. Bhirud, R. Pote, V. Kumar, P. Sampat, and N. Jain, "Evaluation of boosting algorithms for p300 detection in eeg signals," in *Disruptive Developments in Biomedical Applications*, CRC Press, 2022, pp. 173–187.

[12] F. Gondesen, M. Marx, and D. Gollmann, "Eeg-based biometrics," *Biometric-based physical and cybersecurity systems*, pp. 287–318, 2019.

[13] S. I. Alzahrani, "Implementation of p300 based bci using a consumer-grade eeg neuroheadset," in *2021 IEEE National Biomedical Engineering Conference (NBEC)*, IEEE, 2021, pp. 59–64.

[14] T. Koike-Akino, R. Mahajan, T. K. Marks, Y. Wang, S. Watanabe, O. Tuzel, and P. Orlik, "High-accuracy user identification using eeg biometrics," in *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, 2016, pp. 854–858.

[15] J. P. Hamm, B. W. Johnson, and I. J. Kirk, "Comparison of the n300 and n400 erps to picture stimuli in congruent and incongruent contexts," *Clinical neurophysiology*, vol. 113, no. 8, pp. 1339–1350, 2002.

[16] M. S. Franklin, J. Dien, J. H. Neely, E. Huber, and L. D. Waterson, "Semantic priming modulates the n400, n300, and n400rp," *Clinical Neurophysiology*, vol. 118, no. 5, pp. 1053–1068, 2007.

[17] G. Safont, A. Salazar, A. Soriano, and L. Vergara, "Combination of multiple detectors for eeg based biometric identification/authentication," in *2012 IEEE international carnahan conference on security technology (ICCST)*, IEEE, 2012, pp. 230–236.

[18] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification, john willey & sons," *Inc., second edition edition*, 2001.

[19] W.-Y. Loh, "Fifty years of classification and regression trees," *International Statistical Review*, vol. 82, no. 3, pp. 329–348, 2014.

[20] R. B. Nelsen, "Archimedean copulas," *An introduction to copulas*, pp. 109–155, 2006.

[21] T. Kavzoglu and A. Teke, "Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (xgboost) and natural gradient boosting (ngboost)," *Arabian Journal for Science and Engineering*, vol. 47, no. 6, pp. 7367–7385, 2022.

[22] C. He and J. Wang, "An independent component analysis (ica) based approach for eeg person authentication," in *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*, IEEE, 2009, pp. 1–4.

[23] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: Unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.

[24] Emotiv, *Emotiv - 14 channel wireless eeg headset*, https://www.emotiv.com/products/epoc, accessed: July, 2024.

[25] F. A. Rosli, S. A. Awang, A. A. Abdullah, and M. S. Salim, "Biometric authentication system using eeg biometric trait–a review," in *AIP Conference Proceedings*, AIP Publishing, vol. 2339, 2021.

[26] B. Tsybenov, M. Svetlakov, and I. Hodashinsky, "Feature selection methods comparison for eeg-based classifier constructed using discrete wavelet transform features," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 2291, 2022, p. 012 003.

[27] Z. Y. Ong, A. Saidatul, and Z. Ibrahim, "Power spectral density analysis for human eeg-based biometric identification," in *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, IEEE, 2018, pp. 1–6.

[28] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, pp. 1–16, 2009.

[29] P. Xanthopoulos, P. M. Pardalos, T. B. Trafalis, P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Linear discriminant analysis," *Robust data mining*, pp. 27–33, 2013.

[30] Z. Zhang and Z. Zhang, "Artificial neural network," *Multivariate time series analysis in climate and environmental research*, pp. 1–35, 2018.

[31] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, "Advance and prospects of adaboost algorithm," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2013.

[32] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, *et al.*, "Xgboost: Extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[33] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," 2013.

[34] M. Maalouf, "Logistic regression in data analysis: An overview," *International Journal of Data Analysis Techniques and Strategies*, vol. 3, no. 3, pp. 281–299, 2011.

[35] V. Jakkula, "Tutorial on support vector machine (svm)," *School of EECS, Washington State University*, vol. 37, no. 2.5, p. 3, 2006.

[36] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, pp. 3–42, 2006.