Feature Selection of Multichannel EEG for Attention Classification

Atmoko Nugroho
Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia | Informatics Engineering Department, Faculty of Information and Communication, Universitas Semarang, Indonesia
atmoko@usm.ac.id (corresponding author)

Danny Manongga
Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia
danny.manongga@uksw.edu

Hindriyanto Dwi Purnomo
Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia
hindriyanto.purnomo@uksw.edu

Hendry Hendry
Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia
hendry@uksw.edu

Received: 21 July 2025 | Revised: 17 September 2025 and 6 October 2025 | Accepted: 9 October 2025

 DOI: https://doi.org/10.48084/etasr.13062

ABSTRACT

EEG (electroencephalography) is a tool to determine human brain waves and the function of the human brain. A complete EEG device consists of 33 channels and is expensive for independent research. This study aimed to determine which EEG features play an important role in attention and meditation. Data were obtained from 83 respondents with 8 channels and 5 waves in each channel. This study examines three feature selection methods. Using Random Forest, selecting the Fp1_beta, Fp1_gamma, Fp2_beta, Fp2_gamma, Fz_beta, Fz_gamma, C4_beta, C4_gamma, O1_beta, O1_gamma, O2_alpha, O2_beta, O2_gamma features led to an accuracy of 98.2%. Selecting the Fp1_beta, Fp1_gamma, Fp2_beta, Fp2_gamma, Fz_delta, Fz_alpha, Fz_beta, Fz_gamma, C4_beta, O1_beta, O1_gamma, O2_alpha, O2_beta, O2_gamma features with a tree-based model led to an accuracy of 98.0%. Finally, with recursive feature elimination, the Fp1_beta, Fp2_beta, Fz_beta, O2_beta, and O2_gamma features led to an accuracy of 96.7%.

Keywords-EEG channel; feature selection; machine learning

I. INTRODUCTION

Electroencephalography (EEG) is a prevalent neuroimaging method that offers a non-invasive approach to measuring the brain's electrical activity, elucidating the intricate interactions of neuronal processes that govern cognitive, sensory, and motor activities. However, the interpretation and analysis of EEG data, marked by high dimensionality, non-stationarity, and a low signal-to-noise ratio, pose considerable hurdles [1]. A vital step in addressing these issues is the selection of pertinent features from the EEG data, a process termed feature selection, which can enhance the accuracy and efficiency of machine learning models. This study investigates the selection of EEG features to investigate which waves and channels affect attention.

This study aimed to investigate various feature selection methods within EEG-based machine learning contexts, highlighting the distinctive attributes of EEG data. These data typically exhibit a low signal-to-noise ratio, since recorded brain activity is often obscured by various sources of ambient, physiological, and activity-specific noise of comparable or greater magnitude, termed artifacts [2, 3]. In addition, EEG is a non-stationary signal, exhibiting temporal variability, which may result in inadequate generalization of classifiers trained on temporally-restricted data. To address these challenges, researchers have devised various feature selection methods that utilize both domain-specific expertise and data-driven strategies, incorporating techniques from disciplines such as signal processing, machine learning, and neuroscience [4].

The number of channels in EEG scalp devices can differ [5], resulting in variations in features between datasets, thus complicating the feature selection process. A reference is necessary to avoid examining all existing channels [6]. This study aimed to uncover the critical characteristics of various EEG channels related to attention while investigating efficient and effective feature selection techniques.

II. RESEARCH METHODOLOGY

Figure 1 shows the steps taken in this research: collecting the EEG dataset, preprocessing, feature selection, classifier, and accuracy evaluation. The EEG dataset comprised 8-channel EEG data. Preprocessing involved cleaning and removing empty or useless data. Three feature selection methods were used, namely Random Forest (RF), Tree-based, and Recursive Feature Elimination (RFE). The dataset was divided into 70% for training and 30% for testing, and RF was used on all feature selection cases.

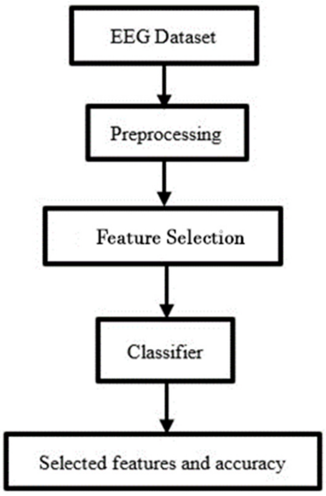

Fig. 1. Research method.

III. EEG DATASET

Data were collected from 83 student participants utilizing an 8-channel EEG, which included the electrodes: Fp1, Fp2, Fz, C3, C4, Pz, O1, and O2. Each channel comprises five waves: delta, theta, alpha, beta, and gamma. To reduce outside distractions, recording sessions were held in a calm space with constant lighting and temperature. During the sessions, participants were told to reduce body movement and were seated comfortably. No personally identifiable information was kept in the dataset, and all subjects were volunteers in good health. During the experiment, participants listened to songs ranging from 2 to 7 minutes in duration until attention data was collected. During the sessions, participants were told to reduce body movement and were seated comfortably. The total number of features in the EEGs studied is 40 (N_{pre}). Figure 2 shows an EEG electrode placement map based on the 10-20 system, which is widely used to measure brain activity through EEG signals [7].

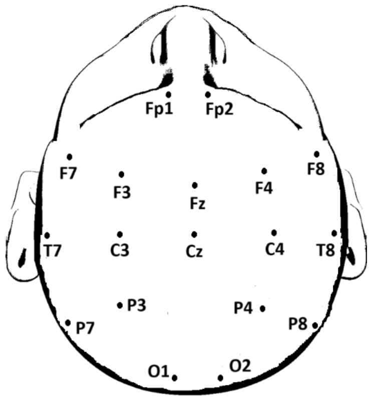

Fig. 2. EEG channels distribution based on the 10-20 system of electrode placement.

The dataset includes pre-extracted EEG features arranged in a tabular format. The last column has the class label that was used for the two-class classification task. The dataset lacks raw EEG waveforms but comprises engineered descriptors, derived from the recorded EEG channels, that capture statistical, temporal, or spectral properties and were used directly as input to the feature selection and classification models. All features are continuous numerical values with no missing entries.

A. Head Shape

Figure 2 shows a top view of the head.

- Right: The right ear (T8).
- Left: The left ear (T7).
- Front: The top of the circle (frontal region).
- Back: The bottom of the circle (occipital region).

B. Electrode Placement

- Fp1, Fp2: Frontal positions near the forehead (prefrontal).
- F3, F4, Fz: Frontal positions closer to the center of the head.
- C3, C4, Cz: Central area, located at the top of the head.
- P3, P4: Parietal area (mid-back region of the head).
- O1, O2: Occipital area, located at the back of the head (near the neck).
- F7, F8: Lateral frontal positions, near the ears.
- T7, T8: Temporal electrodes, located on the left and right sides of the head.
- P7, P8: Lateral parietal electrodes.
- Fz, Cz, Pz: Electrodes along the midline of the head (sagittal line).

C. Electrode Location Function

These locations help record brain electrical activity in specific regions. They are used in studies on cognition, sleep, epilepsy, and the effects of music on brain activity.

D. Number of Electrodes

A total of 18 electrodes were used in EEG measurements. The 10-20 system provides a standard for electrode placement, ensuring consistent EEG signal measurements across different studies and clinical applications.

IV. FEATURE SELECTION

Feature selection is crucial in the analysis of multichannel EEG data, as it enables the identification of the most informative features that can effectively distinguish between different brain states or conditions [8]. Feature selection techniques, such as ensemble learning approaches and embedded methods, have demonstrated superior performance in various applications [9]. It is a reduction process on the number of input variables used in the construction of predictive models [10] that can improve accuracy and performance by focusing on the most relevant and discriminative features [11].

The benefits of feature selection are twofold: it can reduce the computational complexity of the classification task, as a smaller set of features requires fewer computational resources, and it can improve the generalization performance of the model by avoiding overfitting, as the model is less likely to be heavily influenced by irrelevant or redundant features [12].

Filter methods, embedded methods, and wrapper methods have been widely used in EEG data analysis. Filter methods, such as the Student's t-test or Marginal Fisher Analysis (MFA), are often used as a preprocessing step to rank the features based on their statistical significance or discriminative power. Embedded methods, such as Lasso regression or Support Vector Machines (SVM), integrate the feature selection process within the model training, where the model automatically selects the most relevant features during the optimization process [13]. Wrapper methods, such as Sequential Feature Selection (SFS) or Recursive Feature Elimination (RFE), use the performance of the classifier as a guide to select the optimal feature subset. Wrapper methods, although they tend to be computationally more intensive, can often identify feature subsets that are better suited for specific classification tasks compared to filter or embedded methods, as they take into account the complex interplay between the selected features and the classifier's performance.

A. Random Forest (RF)

The first feature selection method used in this study is RF to extract feature importance. This method is a tree-based algorithm that evaluates the importance of features by calculating the reduction in impurity at each decision node across all trees in the forest. Feature importance is calculated based on the total reduction in impurity contributed by each feature across all decision trees in the forest. Feature importance was calculated using [14]:

$$Feature\ Importance = \frac{\sum_{t=1}^{T} Reduction\ in\ Impurity\ for\ Feature\ j}{Total\ Reduction\ in\ Impurity\ for\ All\ Feature}$$

where T is the total number of trees in the forest and j is the feature being evaluated.

RF-based feature selection uses Mean Decrease in Impurity (MDI) to determine which features from the dataset are important. This approach is achieved by calculating the Gini index as the default criterion for impurity using:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2$$

where C is the total number of classes, and p_i is the probability of samples belonging to class i. Entropy quantifies the amount of disorder or uncertainty in the data as:

$$Entropy = -\sum_{i=1}^{C} p_i \log_2(p_i)$$

where C is the total number of classes and p_i is the probability of samples belonging to class i.

The execution of RF is demonstrated in the subsequent Python code snippet.

```
selector =
SelectFromModel(RandomForestClassifier
(n_estimators=100, random_state=42))
selector.fit(X_train, y_train)
```

When selector.fit() is called, the RF classifier model temporarily trains 100 trees (n_estimators=100) on the training data. The SelectFromModel function used the default scikit-learn threshold, which is the mean of the feature importance values computed by the RF estimator. Only features whose importance exceeded this mean value were retained. With this model, this study obtained the following features: Fp1_beta, Fp1_gamma, Fp2_beta, Fp2_gamma, Fz_beta, Fz_gamma, C4_beta, C4_gamma, O1_beta, O1_gamma, O2_alpha, O2_beta, and O2_gamma.

B. Tree-Based Model

This study also utilized a Tree-Based Model with a Random Forest Classifier that uses inherent feature importance scores provided by a tree-based ensemble algorithm [15, 16]. The RF classifier is trained on the dataset, and the SelectFromModel framework uses the learned feature importance scores to find the most useful features. This process is a Tree-Based feature selection method because it utilizes the decision tree structure within the RF to determine the contribution of each feature to the model. In this model, the default threshold of SelectFromModel, corresponding to the average feature importance across all features, was employed. Thus, a feature was selected only if its importance was greater than the mean importance derived from the RF model. Unlike the first model, which compared different feature selection strategies before implementation, this Tree-based model performed feature selection directly based on importance scores from the RF without prior performance comparison. Thus, the result is directly derived from the Random Forest model.

With this tree-based model, 14 features were obtained: Fp1_beta, Fp1_gamma, Fp2_beta, Fp2_gamma, Fz_delta, Fz_alpha, Fz_beta, Fz_gamma, C4_beta, O1_beta, O1_gamma, O2_alpha, O2_beta, and O2_gamma.

C. Recursive Feature Elimination (RFE)

RFE is a widely used feature selection method in machine learning. Its primary goal is to identify the most relevant subset of features by recursively removing less important ones. RFE is particularly effective in reducing dimensionality, improving model performance, and enhancing interpretability. This method is a robust and effective feature selection method, particularly when the number of features is high. Its recursive process ensures that the most important features for the target variable are retained.

This study used RFE with RF to identify the most important EEG features. In this approach, RF is used as an internal estimator to iteratively eliminate the least important features until only the desired number of features remains. The selected features were then used to train a new RF classifier for performance evaluation.

The goal of RFE is to minimize the loss function L by selecting the optimal subset of k features, where $k \leq p$ [17].

- $X = \{x_1, x_2, \ldots, x_p\}$: Set of p features
- $\mathcal{Y}$: Target variable
- $f(X;\theta)$: A model parameterized by θ (e.g., weights or splits)
- $L(f(X;\theta), \mathcal{Y})$: Loss function of the model

This approach selected fewer features than the previous models: Fp1_beta, Fp2_beta, Fz_beta, O2_beta, and O2_gamma.

Efficiency addresses the reduction in model complexity and computational resources resulting from feature selection. It can be measured by the decrease in the number of features and the reduction in computational time:

$$Feature\ Reduction\ Efficiency\ (FRE) = \left(\frac{N_{pre} - N_{post}}{N_{pre}}\right) \times 100\ \%$$

$$Computational\ Time\ Efficiency\ (CTE) = \left(\frac{T_{pre} - T_{post}}{T_{pre}}\right) \times 100\ \%$$

where N_{pre} and N_{post} are the number of features before and after selection, and T_{pre} and T_{post} represent the computational time before and after feature selection.

In this study, the focus was on the number of features obtained; therefore, for time efficiency, it is only based on the time required for each feature selection method (T_{post}). This variable represents the total execution time measured after the initialization of the feature selection and classification process. This variable includes all computational stages, such as data loading, feature selection, model training, and evaluation. The execution time was measured using Python's time.perf_counter() function and reported in seconds.

V. CLASSIFICATION

Classification plays a crucial role within a data analysis pipeline, particularly in machine learning. This study used the Python programming language in the Jupyter Notebook application with an AMD Ryzen 5 processor.

A. Evaluating the Effectiveness of Feature Selection

After the feature selection process, classification is used to assess how effectively the selected features can differentiate between data classes. A similar classification model was used to ensure fair evaluation between these feature selection algorithms. By comparing the performance of the classification model (e.g., accuracy, precision, recall, F1-score) before and after feature selection, it becomes clear whether the feature selection process improves, degrades, or has no effect on the model's performance.

B. Dimensionality Reduction to Enhance Classification Performance

Feature selection typically reduces the number of input features, making classification faster and more efficient, especially for models with high computational complexity. With fewer features, the model also becomes more robust to overfitting, particularly when working with datasets of limited size.

C. Focusing on Relevant Features

Classification after feature selection operates solely on relevant features, producing models that are more interpretable. This is particularly important in fields requiring in-depth interpretation, such as bioinformatics, EEG analysis, or economics.

D. Model Validation and Generalization

Using the selected features, classification helps determine whether the model can generalize well to unseen data (e.g., on validation or test datasets). A model that performs poorly on test data may indicate that the feature selection process failed to capture features that are truly relevant to the classification task.

E. Optimizing the Machine Learning Pipeline

In a machine learning pipeline, feature selection and classification work together as complementary steps. Feature selection filters the data, while classification leverages the filtered features to build predictive models. In this study, RF was used for the classification stage. The program snippet in the Python programming language is as follows:

```
X_train_selected =
selector.transform(X_train)
X_test_selected =
selector.transform(X_test)
model_selected =
RandomForestClassifier(random_state=42)
model_selected.fit(X_train_selected,
y_train)
y_pred_selected =
model_selected.predict(X_test_selected)
```

VI. RESULTS

Table I presents a comparative analysis of the three feature selection methods applied to EEG data: RF, Tree-based, and RFE. Each method selected a different number of features, with Tree-Based Models retaining the highest number (14) and RFE selecting the least (5). The computational time (T_{post}) varies significantly, with RFE requiring considerably more processing time (346.47) compared to RF (40.95) and Tree-Based (37.11).

TABLE I. RESULTS

	Feature selection models		
	RF	**Tree-based**	**RFE**
Selected features	Fp1_beta, Fp1_gamma, Fp2_beta, Fp2_gamma, Fz_beta, Fz_gamma, C4_beta, C4_gamma, O1_beta, O1_gamma, O2_alpha, O2_beta, O2_gamma	Fp1_beta, Fp1_gamma, Fp2_beta, Fp2_gamma, Fz_delta, Fz_alpha, Fz_beta, Fz_gamma, C4_beta, O1_beta, O1_gamma, O2_alpha, O2_beta, O2_gamma	Fp1_beta, Fp2_beta, Fz_beta, O2_beta, O2_gamma
N_{post}	13	14	5
T_{post}	40.954698	37.118637	346.472886
Accuracy	98.2%	98.0%	96.7%
Feature reduction efficiency	67.5%	65%	87.5%

Regarding classification accuracy, all methods yield high performance, with RF achieving the highest accuracy (98.2%), followed by Tree-based (98.0%) and RFE (96.7%). However, RFE demonstrates the highest feature reduction efficiency (87.5%), suggesting its effectiveness in minimizing the number of features while maintaining competitive accuracy.

VII. DISCUSSION

All things considered, the feature selection phase is essential [18]. In [19], a compact CNN-based model was used to classify brain attention states from EEG recordings acquired with a limited number of channels. This CNN achieved satisfactory performance with minimal input, despite substantial architectural complexity and model processing requirements. However, its accuracy was limited by the reduced spatial information from the limited number of electrodes. In contrast, modern techniques using a basic RF classifier with multi-channel EEG data have achieved good classification accuracy. This improvement arises from the additional channels in multi-channel EEG and the inherent resilience of RF to overfitting, compensating for the lack of deep hierarchical feature learning. This comparison suggests that in the classification of EEG data, the quality and completeness of the input data may be more important than the advantages of a complex model. This makes RF-based multi-channel systems a superior and more pragmatic choice for obtaining reliable results.

VIII. CONCLUSION

The comparative analysis of feature selection methods for EEG classification highlights a trade-off between accuracy, computational efficiency, and feature reduction. RF achieved the highest accuracy (98.2%) while maintaining a relatively low processing time, making it a practical choice for EEG-based machine learning applications. Tree-based feature selection performed similarly in terms of accuracy but retained the highest number of features, which may impact model interpretability. On the other hand, RFE demonstrates the highest feature reduction efficiency (87.5%), significantly reducing model complexity, but at the cost of increased computational time and a slight decrease in accuracy. These findings suggest that while feature selection is crucial for improving model performance and efficiency, the choice of method should be tailored to the specific requirements of the application, balancing accuracy, feature interpretability, and computational feasibility.

Future research should focus on optimizing feature selection methods for EEG analysis by addressing computational efficiency, robustness, and applicability in real-world scenarios. Enhancing RFE through heuristic techniques, parallel computing, or reinforcement learning can reduce its processing time while maintaining high feature reduction efficiency. Additionally, hybrid approaches that combine RF classification and RFE could leverage their complementary strengths to improve both accuracy and feature reduction. Exploring deep learning-based feature selection methods, such as autoencoders or attention mechanisms, may offer a more adaptive approach to handling high-dimensional EEG data. In addition, evaluating model robustness across diverse EEG datasets, including pathological conditions such as epilepsy and neurocognitive disorders, is essential to ensure generalizability. Finally, integrating optimized feature selection techniques into embedded systems, such as wearable EEG devices, would enable real-time applications in brain-computer interfaces (BCI) [20], neurofeedback, and cognitive state monitoring, enhancing their practicality in clinical and consumer-grade neurotechnology.

DATASET AVAILABILITY

The corresponding author will provide the dataset generated and analyzed in this study upon reasonable request.

REFERENCES

[1] X. Wei *et al.*, "2021 BEETL Competition: Advancing Transfer Learning for Subject Independence & Heterogenous EEG Data Sets." arXiv, 2022, https://doi.org/10.48550/ARXIV.2202.12950.

[2] C. Zhang, Y. K. Kim, and A. Eskandarian, "EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification," *Journal of Neural Engineering*, vol. 18, no. 4, Aug. 2021, Art. no. 046014, https://doi.org/10.1088/1741-2552/abed81.

[3] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review." arXiv, 2019, https://doi.org/10.48550/ARXIV.1901.05498.

[4] F. A. Alturki, M. Aljalal, A. M. Abdurraqeeb, K. Alsharabi, and A. A. Al-Shamma'a, "Common Spatial Pattern Technique With EEG Signals for Diagnosis of Autism and Epilepsy Disorders," *IEEE Access*, vol. 9, pp. 24334–24349, 2021, https://doi.org/10.1109/ACCESS.2021.3056619.

[5] A. Al-Nafjan and M. Aldayel, "Predict Students' Attention in Online Learning Using EEG Data," *Sustainability*, vol. 14, no. 11, May 2022, Art. no. 6553, https://doi.org/10.3390/su14116553.

[6] Z. Gao, X. Cui, W. Wan, Z. Qin, and Z. Gu, "Signal Quality Investigation of a New Wearable Frontal Lobe EEG Device," *Sensors*,

vol. 22, no. 5, Feb. 2022, Art. no. 1898, https://doi.org/10.3390/s22051898.

[7] R. M. Mehmood and H. J. Lee, "Exploration of Prominent Frequency Wave in EEG Signals from Brain Sensors Network," *International Journal of Distributed Sensor Networks*, vol. 11, no. 11, Nov. 2015, Art. no. 386057, https://doi.org/10.1155/2015/386057.

[8] R. Hussein, H. Palangi, R. Ward, and Z. J. Wang, "Epileptic Seizure Detection: A Deep Learning Approach." arXiv, 2018, https://doi.org/10.48550/ARXIV.1803.09848.

[9] H. H. Htun, M. Biehl, and N. Petkov, "Survey of feature selection and extraction techniques for stock market prediction," *Financial Innovation*, vol. 9, no. 1, Jan. 2023, Art. no. 26, https://doi.org/10.1186/s40854-022-00441-7.

[10] C. Y. Liao, S. K. Tai, R. C. Chen, and H. Hendry, "Using EEG and Deep Learning to Predict Motion Sickness Under Wearing a Virtual Reality Device," *IEEE Access*, vol. 8, pp. 126784–126796, 2020, https://doi.org/10.1109/ACCESS.2020.3008165.

[11] S. L. Oh *et al.*, "A novel automated autism spectrum disorder detection system," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2399–2413, Oct. 2021, https://doi.org/10.1007/s40747-021-00408-8.

[12] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, no. 1, Dec. 2020, Art. no. 52, https://doi.org/10.1186/s40537-020-00327-4.

[13] N. V. Kimmatkar and B. V. Babu, "Human Emotion Detection with Electroencephalography Signals and Accuracy Analysis Using Feature Fusion Techniques and a Multimodal Approach for Multiclass Classification," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 9012–9017, Aug. 2022, https://doi.org/10.48084/etasr.5073.

[14] E. Izquierdo-Verdiguier and R. Zurita-Milla, "An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, June 2020, Art. no. 102051, https://doi.org/10.1016/j.jag.2020.102051.

[15] W. Wan, X. Cui, Z. Gao, and Z. Gu, "Frontal EEG-Based Multi-Level Attention States Recognition Using Dynamical Complexity and Extreme Gradient Boosting," *Frontiers in Human Neuroscience*, vol. 15, Jun. 2021, Art. no. 673955, https://doi.org/10.3389/fnhum.2021.673955.

[16] S. M. K. Alam, P. Li, M. Rahman, M. Fida, and V. Elumalai, "Key factors affecting groundwater nitrate levels in the Yinchuan Region, Northwest China: Research using the eXtreme Gradient Boosting (XGBoost) model with the SHapley Additive exPlanations (SHAP) method," *Environmental Pollution*, vol. 364, Jan. 2025, Art. no. 125336, https://doi.org/10.1016/j.envpol.2024.125336.

[17] M. Y. Shams, Z. Tarek, and A. M. Elshewey, "A novel RFE-GRU model for diabetes classification using PIMA Indian dataset," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, Art. no. 982, https://doi.org/10.1038/s41598-024-82420-9.

[18] X. Xu, X. Nie, J. Zhang, and T. Xu, "Multi-Level Attention Recognition of EEG Based on Feature Selection," *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, Feb. 2023, Art. no. 3487, https://doi.org/10.3390/ijerph20043487.

[19] C. H. Lee, Y. X. Zhang, C. M. Huang, and C. M. Wu, "Classification of Brain Attention based on EEGNet with Fewer Channels," in *Proceedings of the 2024 12th International Conference on Computer and Communications Management*, Kagoshima-shi, Japan, Jul. 2024, pp. 55–61, https://doi.org/10.1145/3688268.3688277.

[20] X. Du, J. Wu, X. Tang, X. Lv, L. Jia, and C. Xue, "Predicting User Attention States from Multimodal Eye–Hand Data in VR Selection Tasks," *Electronics*, vol. 14, no. 10, May 2025, Art. no. 2052, https://doi.org/10.3390/electronics14102052.