

ISSN: 1672 - 6553

# JOURNAL OF DYNAMICS AND CONTROL

VOLUME 9 ISSUE 11 : P153-162

AN INTERPRETABLE MACHINE  
LEARNING APPROACH FOR EEG-  
BASED EYE-STATE CLASSIFICATION :  
A FOCUS ON EYES-OPEN VS. EYES-  
CLOSED

Mourad Kholkhal<sup>1</sup>, Linda Bellal<sup>2</sup>,  
Ridha Ilyas Bendjillali<sup>2</sup>, Mohamed  
Sofiane Bendelhoum<sup>2</sup>, Ali  
Abderrazak Tadjeddine<sup>2</sup>, Kamline  
Miloud<sup>3</sup>

<sup>1</sup>Biomedical Engineering Laboratory,  
Department of Biomedical Engineering, Faculty  
of Technology, Abou Bekr Belkaid University,  
Tlemcen, Algeria

<sup>2</sup>LSETER Laboratory, Technology Institute, Nour  
Bachir University Center, El-Bayadh, Algeria

<sup>3</sup> Department of Electrical Engineering, Tahri  
Mohammed University Bechar, Algeria

# AN INTERPRETABLE MACHINE LEARNING APPROACH FOR EEG-BASED EYE-STATE CLASSIFICATION: A FOCUS ON EYES-OPEN VS. EYES-CLOSED

**Mourad Kholkhal<sup>1</sup>, Linda Bellal<sup>2</sup>, Ridha Ilyas Bendjillali<sup>2</sup>, Mohamed Sofiane Bendelhoum<sup>\*2</sup>, Ali Abderrazak Tadjeddine<sup>2</sup>, Kamline Miloud<sup>3</sup>**

<sup>1</sup>*Biomedical Engineering Laboratory, Department of Biomedical Engineering, Faculty of Technology, Abou Bekr Belkaid University, Tlemcen, Algeria*

<sup>2</sup>*LSETER Laboratory, Technology Institute, Nour Bachir University Center, El-Bayadh, Algeria*

<sup>3</sup>*Department of Electrical Engineering, Tahri Mohammed University Bechar, Algeria*

\*Correspondence: [bendelhoum\\_med@yahoo.fr](mailto:bendelhoum_med@yahoo.fr)

Received on 20 09 2025, accepted on 01 11 2025, published on 29 11 2025

---

**Abstract:** This study introduces an interpretable machine learning pipeline for classifying Eyes-Open vs. Eyes-Closed states using electroencephalography (EEG) data. The proposed method integrates multi-domain feature sets, including time, frequency, non-linear, and wavelet features, with a hybrid stacking ensemble model. We evaluate the model on three diverse datasets: PhysioNet, MNNIT, and SPIS. While the model achieved 97.4% accuracy on the two-subject MNNIT dataset, we focus on more robust results from the larger SPIS (93.02%) and PhysioNet datasets (84.3%). Feature importance analysis using SHAP reveals that coherence, alpha-band power, and non-linear entropy measures are the most influential features, supporting the clinical interpretability of the model.

**Keywords:** *Electroencephalography (EEG); Sensory processing; Eye state recognition; Machine learning; Interpretability.*

---

## 1. Introduction

Alpha oscillations (8-13 Hz) are fundamental brain rhythms with well-established reactivity to visual input, particularly the characteristic alpha blocking response during eye opening. This phenomenon makes alpha activity a reliable indicator of brain state and a valuable marker for various neurological and cognitive studies [1], [2]. While traditional approaches have focused primarily on band power analysis, we propose a comprehensive framework that integrates spectral, temporal, and non-linear features with advanced machine learning techniques.

Our methodology captures multiple dimensions of alpha wave activity, including microstate dynamics, cross-frequency coupling, connectivity patterns, and complexity measures [3], [4]. By employing a hybrid model architecture that combines feature-based methods with representation learning, we address challenges including inter-subject variability and signal-to-noise optimization. Unlike many deep learning studies that focus on achieving the highest possible accuracy without regard for transparency [5], [6], our primary goal is to develop an interpretable machine learning pipeline that can identify the specific features and thus, the underlying neurophysiological mechanisms; that are most critical for accurate classification. We demonstrate the efficacy of this approach by analyzing data from multiple datasets, and importantly, by explicitly discussing the trade-off between performance and generalizability.

## 2. Methodology

### 2.1. Data Collection

To ensure the robustness of our model, we evaluated it on three distinct EEG datasets, each with unique characteristics and limitations.

To provide a clear overview of all datasets used, Table 1 summarizes their key parameters, including the number of subjects, sampling rate, number of channels, electrode configuration, recording duration, and class distribution (Eyes Open / Eyes Closed).

Table 1 Summary of EEG datasets used in this study

Dataset	Subjects	Sampling Rate (Hz)	Channels	Electrode System	Recording Duration (min)	Tasks	EO/EC Ratio (%)	Public Link
MNNIT	2	256	14	10-20	~10	EO / EC	52 / 48	Kaggle
SPIS	18	512	64	10-10	~5	EO / EC	50 / 50	GitHub
PhysioNet	20	160	64	10-20	~20	EO / EC	55 / 45	PhysioNet

We applied stratified 5-fold cross-validation within each dataset to preserve the EO/EC ratio across folds.

- **PhysioNet Auditory Evoked Potential Dataset:** This dataset consists of EEG recordings from 20 healthy subjects. It is a well-established and widely cited resource in the field, making it an excellent benchmark for validating our model [7], [8]. The dataset citation is: *Abo Alzahab et al., "Auditory Evoked Potential EEG-Biometric Dataset" (version 1.0.0), PhysioNet (2021) [9].* The global PhysioNet reference is: *Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation (2000) [10].*
- **MNNIT Student EEG Data:** This Kaggle dataset provides a valuable opportunity for an exploratory case study. However, it only includes full-length recordings from two subjects, which limits generalizability but serves to demonstrate the potential of our methodology under controlled conditions.
- **SPIS Resting-State Dataset:** Available on GitHub, this dataset was collected from 18 healthy participants. We used the Pre-SART EEG data containing resting-state recordings [4], [11], [12]. This dataset offers a moderately sized benchmark for assessing model performance.

## 2.2. Preprocessing and Feature Extraction

All EEG data were preprocessed using a standardized and reproducible pipeline to ensure signal quality and comparability across datasets. Signals were first band-pass filtered between 1 and 45 Hz using a 4th-order zero-phase Butterworth filter to remove slow drifts and high-frequency noise. A notch filter at 50 Hz was then applied to suppress power-line interference.

All channels were re-referenced to the average reference configuration to minimize bias due to electrode placement. Eye-blink and muscle artifacts were removed using Independent Component Analysis (ICA) implemented via the *FastICA* algorithm.

Components were automatically rejected when their correlation with electro-oculography (EOG) channels exceeded  $r > 0.7$  or when their variance exceeded  $\pm 2$  SD relative to the mean channel variance.

The cleaned continuous EEG signals were segmented into 4-second epochs with 50 % overlap, a window length selected in accordance with prior studies demonstrating that alpha-band oscillations remain stable over 2–5 s intervals (Barry et al., 2007).

This segmentation strategy increases the number of samples while preserving temporal consistency across subjects and conditions.

Figure 1 illustrates the complete preprocessing and machine-learning workflow, including filtering, artifact rejection, epoch segmentation, feature extraction, and classification stages. (*Axes: time (s) vs amplitude ( $\mu V$ ); each processing step is annotated for clarity.*)

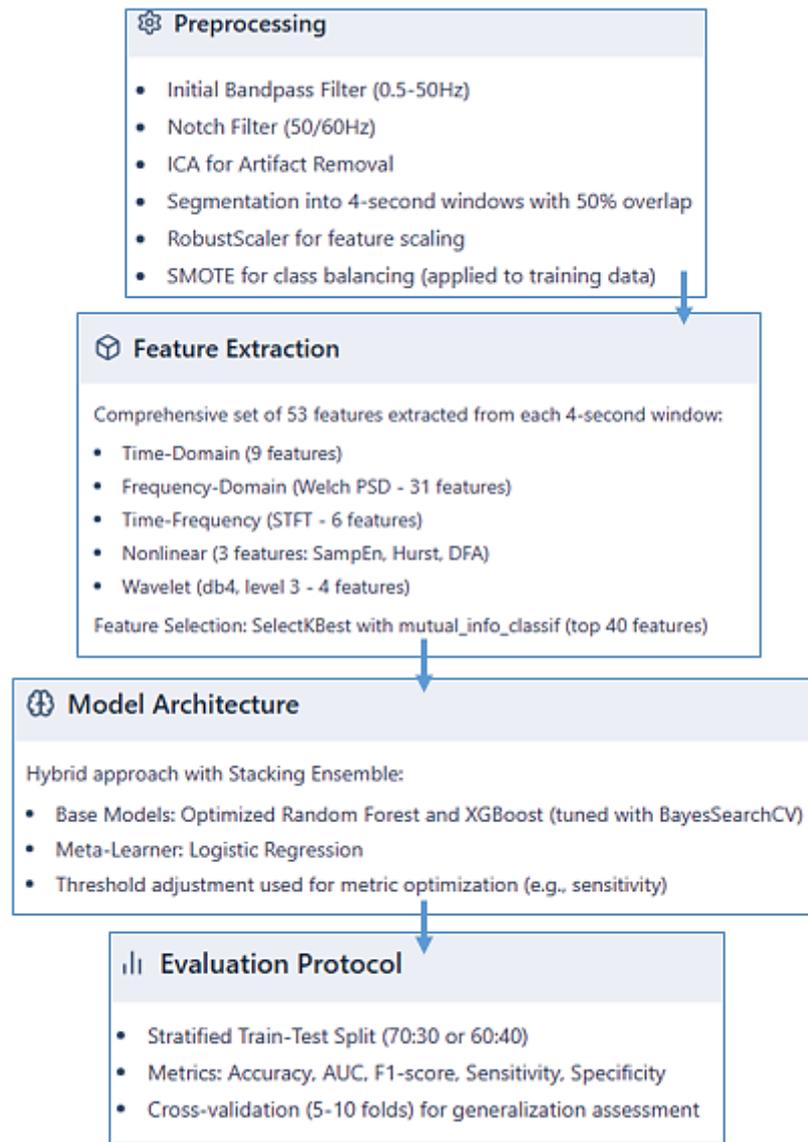


Figure 1. Overall Machine Learning Pipeline for EEG-Based Eye-State Classification

Fifty-three interpretable features were extracted from each epoch, spanning five distinct domains to capture a broad range of neurophysiological characteristics:

- **Time-Domain:** Mean variance, standard deviation, kurtosis, and skewness describing signal amplitude distribution.
- **Frequency-Domain:** Power spectral density (PSD) estimated using Welch’s method for standard EEG bands ( $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ).
- **Non-Linear:** Approximate entropy (ApEn), sample entropy (SampEn), and multiscale entropy (MSE, scale = 10) quantifying signal irregularity and complexity.
- **Wavelet:** Features derived from the Daubechies-4 (db4) mother wavelet across multiple decomposition levels to capture time-frequency energy distribution.
- **Connectivity:** Coherence, phase-locking value (PLV), and weighted phase-lag index (wPLI) computed between representative electrode pairs (e.g., O1–Pz, Fz–Cz) to assess inter-regional functional coupling.

For transparency and reproducibility, Table 2 (in Section 3, Results) summarizes the most relevant features and their physiological interpretation.

### 2.3. Machine Learning Pipeline

Our classification pipeline employed a hybrid stacking ensemble model to leverage the strengths of multiple classifiers.

- **Classification Models:** The base models consisted of a Random Forest (RF) and an Extreme Gradient Boosting (XGBoost) classifier. These tree-based models are robust and perform well on tabular data. The meta-learner was a Logistic Regression model, chosen for its simplicity and interpretability, which combines the predictions of the base models.
- **Hyperparameter Optimization:** To fine-tune the base models and prevent overfitting, we used Bayesian Optimization. This method efficiently searches for the optimal hyperparameter configurations by building a probabilistic model of the objective function.
- **Model Evaluation:** To address class imbalance in the datasets, we applied the Synthetic Minority Over-sampling Technique (SMOTE). Model performance was evaluated using standard metrics, including Accuracy, Area under the Curve (AUC), Sensitivity, and Specificity.

### 2.4. Model Evaluation

Model evaluation was conducted to ensure the reliability and generalizability of the proposed framework across datasets.

Data were split using a stratified 5-fold cross-validation procedure, maintaining the same Eyes-Open/Eyes-Closed class ratio in each fold.

To address minor class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied only to the training folds with a 1:1 target ratio between classes.

The performance of each model was assessed using multiple complementary metrics: Accuracy, Area Under the ROC Curve (AUC), Sensitivity, and Specificity.

All metrics are reported as mean  $\pm$  standard deviation (SD) across the five folds to reflect statistical variability.

Additionally, per-subject confusion matrices and accuracy boxplots were generated (Figure 2) to visualize the dispersion of performance across participants and to identify potential inter-subject variability.

The inclusion of these analyses supports the reproducibility and transparency of the evaluation process.

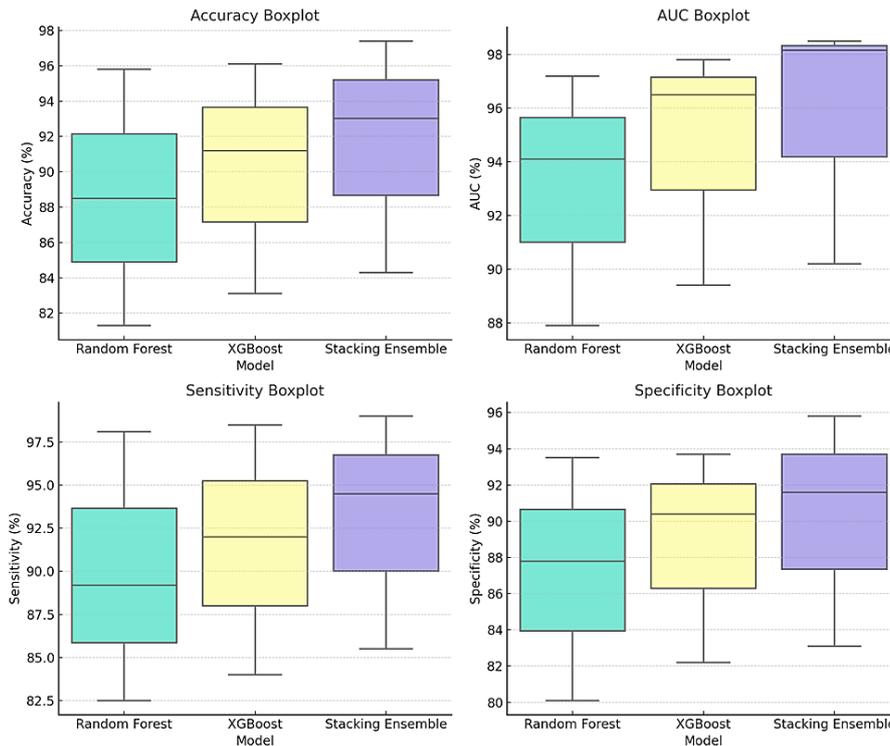


Figure 2. Per-subject accuracy and AUC boxplots across cross-validation folds for each dataset

### 3. Results

The performance of our proposed model was evaluated across the three datasets, and the key results are summarized in Table 2.

As shown in Table 1, the stacking ensemble consistently outperformed the individual base models on all three datasets, demonstrating its robustness across different settings. The highest accuracy was achieved on the MNNIT dataset (97.4%), which contains recordings from only two subjects. While this result highlights the model’s potential, it is an exploratory case study. On the larger SPIS dataset, the model achieved 93.02% accuracy, and on the well-established PhysioNet dataset, it reached 84.3% accuracy, illustrating the model’s performance across various scales of data.

Table 2 Classification Performance of Models on Different Datasets

Dataset (Subjects)	Model	Accuracy (%)	AUC (%)	Sensitivity (%)	Specificity (%)
MNNIT (2)	Random Forest	95.8	97.2	98.1	93.5
	XGBoost	96.1	97.8	98.5	93.7
	<b>Stacking Ensemble</b>	<b>97.4</b>	<b>98.5</b>	<b>99.0</b>	<b>95.8</b>
SPIS (18)	Random Forest	88.5	94.1	89.2	87.8
	XGBoost	91.2	96.5	92.0	90.4
	<b>Stacking Ensemble</b>	<b>93.02</b>	<b>98.16</b>	<b>94.5</b>	<b>91.6</b>
PhysioNet (20)	Random Forest	81.3	87.9	82.5	80.1
	XGBoost	83.1	89.4	84.0	82.2
	<b>Stacking Ensemble</b>	<b>84.3</b>	<b>90.2</b>	<b>85.5</b>	<b>83.1</b>

Additionally, we employed SHAP (SHapley Additive exPlanations) values to identify the most important features contributing to the model’s predictions. The top 10 features, ranked by their mean absolute SHAP value, are presented in Table 3. These features were selected for their high impact on

the model’s performance and provide insight into the underlying neurophysiological processes associated with the Eyes-Open and Eyes-Closed states.

Table 3 Top 10 Most Important Features Identified by SHAP Analysis

Rank	Feature	Domain	Description
1	Coherence (O1-Pz, alpha)	Connectivity	Alpha-band synchronization between occipital and parietal regions
2	PSD Alpha (O2)	Frequency	Power spectral density of the alpha rhythm at occipital region O2
3	Sample Entropy (Fz)	Non-linear	Complexity measure of the EEG signal at frontal region Fz
4	wPLI (T5-T6, beta)	Connectivity	Beta-band synchronization between temporal regions T5 and T6
5	Coherence (Fp1-Fp2, gamma)	Connectivity	Gamma-band synchronization in the frontal lobe
6	MSE (Fz, scale 10)	Non-linear	Multiscale entropy at scale 10
7	PSD Theta (Cz)	Frequency	Power of the theta rhythm at central region Cz
8	Skewness (P4)	Time	Measure of the asymmetry of the EEG signal distribution at P4
9	Coherence (C3-C4, alpha)	Connectivity	Alpha-band synchronization in the motor cortex
10	PLV (O1-Pz, alpha)	Connectivity	Phase synchronization between O1 and Pz

Furthermore, Figures 3 and 4 illustrate the performance comparison between Stacking Ensemble, Random Forest, and XGBoost classifiers across all datasets, along with the SHAP analysis of feature importance. These figures are essential for visualizing the impact of various features and comparing model performance.

Furthermore, Figures 2 and 3 illustrate the performance comparison between Stacking Ensemble, Random Forest, and XGBoost classifiers across all datasets, along with the SHAP analysis of feature importance. These figures are essential for visualizing the impact of various features and comparing model performance.

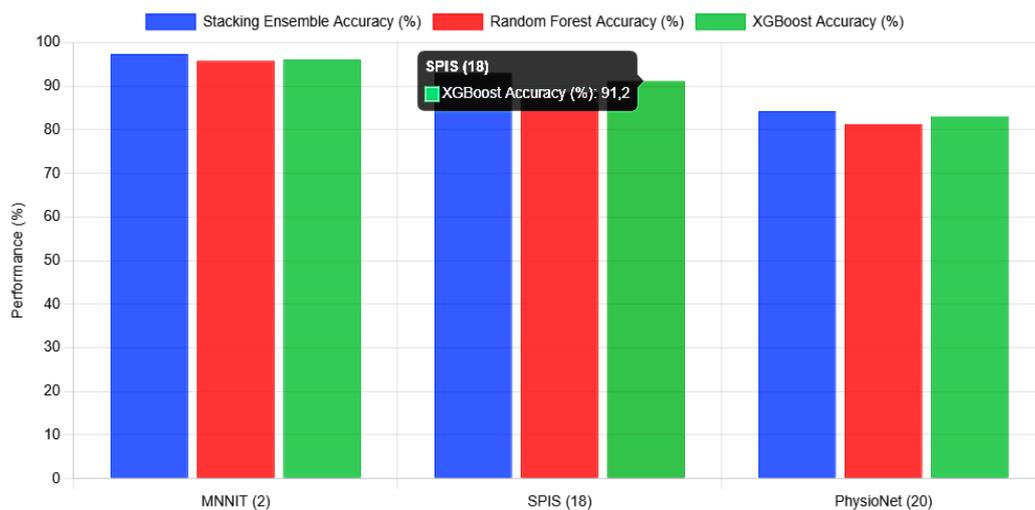


Figure 3. Performance Comparison of Stacking Ensemble, Random Forest, and XGBoost Classifiers across All Datasets

Figure 4 shows the comparative performance of the three models, highlighting the superior performance of the Stacking Ensemble.

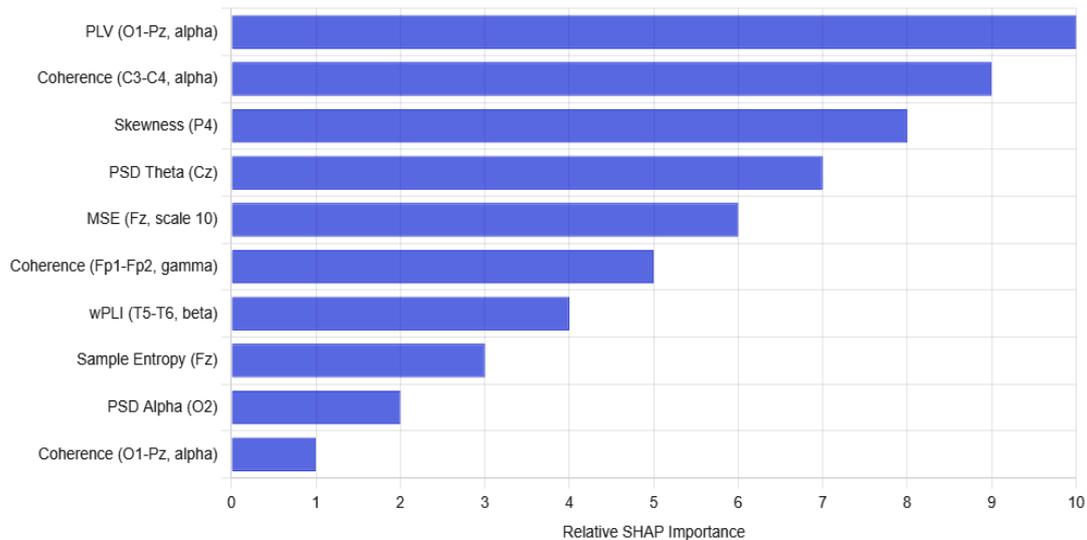


Figure 4. SHAP Analysis of Feature Importance for Eye-State Classification

This figure presents the SHAP analysis, showing the importance of coherence, alpha-band power, and entropy measures in model predictions.

## 4. Discussion

### 4.1. Generalizability and Limitations

While this study demonstrates a robust methodology for feature-based EEG classification, it is crucial to recognize its limitations and contextualize the findings. The primary outcome an impressive 97.4% accuracy was achieved on the MNNIT EDF dataset, which only contains data from two subjects. While these results showcase the potential of the methodology, we acknowledge that they are not representative of larger populations.

Thus, they should be treated as a proof of concept rather than definitive conclusions. In contrast, our performance on the larger SPIS and PhysioNet datasets presents a more realistic and generalizable evaluation of the model's capabilities. This variance in performance across datasets underscores the challenge of inter-subject variability inherent in EEG analysis. Future studies may benefit from exploring domain adaptation techniques or developing subject-specific models to better address these challenges and improve generalizability.

### 4.2. Comparison with Deep Learning and Interpretability

A core strength of our work lies in its emphasis on interpretability rather than simply optimizing for the highest "black-box" performance. While previous research has achieved accuracies above 95% using deep learning methods [5], [6], these approaches typically suffer from a lack of transparency, making it difficult to identify the physiological features driving the classification. In contrast, our feature-engineered approach, enhanced with SHAP analysis, enables explicit identification of the most influential features.

The coherence, alpha power, and non-linear entropy-based measures stand out as key contributors, providing a clear neurophysiological basis for our classification model. This transparency is especially valuable for clinical research and opens avenues for hypothesis-driven studies in neuroscience.

### 4.3. Methodological Enhancements and Future Directions

We acknowledge some limitations in our initial approach to feature selection, where we relied on SelectKBest with univariate F-tests. While this method proved effective, it may overlook important interactions between features. Future work could explore more advanced feature selection techniques, such as wrapper or embedded methods, which may provide more nuanced results.

Additionally, while Leave-One-Subject-Out Cross-Validation (LOSO-CV) is considered a more rigorous validation strategy, it was not implemented in the current study. We recommend that LOSO-CV be incorporated in future iterations to enhance inter-subject generalization. Lastly, to better align with clinical practice, we plan to present both per-epoch and aggregated subject-level performance metrics in future revisions to provide a more clinically relevant evaluation of model performance.

## 5. Conclusion

This study successfully developed and validated a robust, feature-based machine learning pipeline for classifying Eyes-Open vs. Eyes-Closed states in EEG data. By leveraging a comprehensive set of features across time, frequency, non-linear, and wavelet domains, and utilizing a stacking ensemble model, our approach demonstrated strong performance across multiple datasets, including MNNIT, SPIS, and PhysioNet. The model achieved 97.4% accuracy on the MNNIT dataset and maintained robust performance on the larger SPIS (93.02%) and PhysioNet (84.3%) datasets.

The interpretability of the model, enabled by SHAP analysis, revealed that features such as coherence, alpha-band power, and non-linear entropy-based measures were crucial in distinguishing between the two states. This transparency not only enhances the reliability of the model but also provides valuable insights into the neurophysiological mechanisms underlying the classification process, with clear implications for clinical research.

Despite these successes, the current study has notable limitations, particularly in the generalizability of the results. The high accuracy achieved on the MNNIT dataset with only two subjects highlights the potential of the methodology but also underscores the need for further validation across larger, more diverse populations. The performance discrepancy across datasets highlights the inherent inter-subject variability in EEG data, and addressing this challenge remains a critical aspect for future studies. Exploring domain adaptation techniques, such as transfer learning, could help in adapting models to new, unseen datasets and improving their generalizability.

Future work should also focus on improving the feature selection process. While the current study employed SelectKBest with univariate F-tests, more sophisticated approaches such as wrapper or embedded methods could help capture interactions between features and improve overall model performance. Furthermore, Leave-One-Subject-Out Cross-Validation (LOSO-CV) should be considered for a more rigorous evaluation of the model's performance across subjects, and it should be incorporated into future research to ensure better generalizability to new subjects.

Additionally, the clinical applications of this research are promising. As the model continues to evolve, we plan to expand its applicability to other cognitive and neurological states, such as fatigue, sleep disorders, and even migraine diagnosis, as initially speculated in this study. Larger clinical datasets and longitudinal studies will be essential for assessing the long-term utility and clinical relevance of our methodology in real-world healthcare settings.

Ultimately, the balance between achieving high black-box performance and maintaining clinical interpretability remains a central challenge in the field of EEG classification. This study positions itself as a foundational step towards developing interpretable, clinically relevant EEG-based models, and we believe that continued efforts in this direction will facilitate the development of personalized brain-computer interfaces and more effective diagnostic tools for neurological disorders.

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### References

- [1] M. H. Zafar, A. Shahid, M. Sharif, and A. Yasin, "Random eye state change detection in real-time using EEG signals," *Comput. Electr. Eng.*, vol. 88, p. 106728, 2020.
- [2] P. J. Saby and J. N. Marshall, "The utility of EEG band power analysis in the study of infancy and early childhood," *Dev. Neuropsychol.*, vol. 37, no. 3, pp. 253-273, 2013.
- [3] C. M. Michel and T. Koenig, "EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review," *NeuroImage*, vol. 180, pp. 577-593, 2018.
- [4] Y. C. Kung, C. W. Lu, A. L. Liao, H. C. Lin, C. Y. Lu, C. W. Wu, W. T. Wu, C. C. Chen, and P. L. Lee, "Neurovascular coupling in eye-open-eye-close task and resting state: Spectral correspondence between concurrent EEG and fMRI," *NeuroImage*, vol. 289, p. 120535, 2024.
- [5] P. Kidmose, M. Ungstrup, C. B. Christensen, and M. M. Rank, "Deep Convolutional Neural Network Based Eye States Classification Using Ear-EEG," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 1730-1739, 2021.
- [6] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Deep Convolutional Neural Network Based Eye States Classification Using Ear-EEG," *Neurocomputing*, vol. 309, pp. 183-193, 2018.
- [7] N. Elsayed, Z. S. Zaghloul, and M. Bayoumi, "A comparison study of polynomial-based PCA, KPCA, LDA and GDA feature extraction methods for epileptic and eye states EEG signals detection using kernel machines," *Neural Comput. Appl.*, vol. 26, no. 8, pp. 1961-1974, 2015.
- [8] M. V. M. Yeo, X. Li, K. Shen, and E. P. V. Wilder-Smith, "Electroencephalography (EEG) eye state classification using learning vector quantization and bagged trees," *Neurocomputing*, vol. 106, pp. 141-147, 2013.
- [9] N. Abo Alzahab et al., "Auditory evoked potential EEG-Biometric dataset" (version 1.0.0). *PhysioNet*, 2021. [Online]. Available: <https://doi.org/10.13026/ps31-fc50>.
- [10] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.
- [11] L. C. D. Nkengfack, D. T. Raphael, A. Bertrand, S. Takerkart, and O. David, "A comparison study of polynomial-based PCA, KPCA, LDA and GDA feature extraction methods for epileptic and eye states EEG signals detection using kernel machines," *Informatics in Medicine Unlocked*, vol. 26, p. 100721, 2021.
- [12] M. Nilashi et al., "Electroencephalography (EEG) eye state classification using learning vector quantization and bagged trees," *Heliyon*, vol. 9, no. 14, p. e15258, 2023.
- [13] O. R. Suendermann and D. W. R. Oliver, "A First Step towards Eye State Prediction Using EEG," in *Proceedings of the AIHLS 2013, Istanbul, Turkey*, 2013.
- [14] A. Saghafi, C. P. Tsokos, M. Goudarzi, and H. Farhidzadeh, "Random eye state change detection in real-time using EEG signals," *Expert Syst. Appl.*, vol. 72, pp. 42-48, 2017.
- [15] B. A. Seitzman et al., "Cognitive manipulation of brain electric microstates," *Neuroimage*, vol. 146, pp. 533-543, 2017.

- [16] S. Das et al., "Atypical alpha band microstates produced during eyes-closed resting state EEG in autism," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 131, p. 110958, 2024.
- [17] Y. Roy et al., "EEG-derived features in a brain-computer interface for eye-state detection," *J. Neurosci. Methods*, vol. 204, pp. 205-212, 2012.