Statistical Validation of EEG Signal Variability in Epilepsy Monitoring

Maan Hameed[1], Saja Tawfeeq Jassim[2], Nabaa Ahmed Noori[3] and Amrita Prasad[4]
[1]*Al-Turath University, 10013 Baghdad, Iraq*
[2]*College of Pharmacy, Al-Farahidi University, 10065 Baghdad, Iraq*
[3]*Department of Computer Engineering, College of Engineering, Al-Mansour University College, 10067 Baghdad, Iraq*
[4]*Sharda School of Computing Science and Engineering, Sharda University, 201310 Greater Noida, India*
maan.hameed@uoturath.edu.iq, sajatawfeeqalsadoon@gmail.com, Nabaa.ahmed@muc.edu.iq,
amrita.prasad@sharda.ac.in

Keywords: EEG Variability, Epilepsy Monitoring, Reliability, Reproducibility, Entropy, ROC Analysis, Biomarkers, Statistical Validation.

Abstract: Epilepsy is a prevalent neurological disorder requiring reliable biomarkers for effective diagnosis and continuous monitoring. Electroencephalography (EEG) remains the gold standard, yet reproducibility and interpretability of derived metrics remain limited. This study aims to statistically validate EEG variability measures as potential biomarkers for epilepsy monitoring. Public EEG datasets, including CHB-MIT and TUH, were preprocessed using standard pipelines involving band-pass and notch filtering, ICA-based artifact rejection, and segmentation into 10-second epochs. Variability indices spanning time-domain (coefficient of variation, RMSSD), frequency-domain (band power variability), and nonlinear measures (sample entropy, permutation entropy) were extracted. Statistical validation included intraclass correlation coefficients (ICC), Bland-Altman analysis, minimal detectable change (MDC), and receiver operating characteristic (ROC) analysis. Results revealed excellent reliability for coefficient of variation and band power variability (ICC > 0.80) and moderate reproducibility for entropy-based metrics (ICC ≈ 0.70). ROC analysis showed band power variability achieving AUC = 0.83, confirming clinical discrimination between peri-ictal and interictal states. These findings demonstrate that variability metrics can serve as reproducible and clinically actionable biomarkers.

1 INTRODUCTION

Epilepsy is a common neurological disorder that affects more than 50 million people around the world. It causes seizures that happen again and again and are hard to predict, which makes life very difficult. Electroencephalography (EEG) has long been recognized as the principal diagnostic and monitoring instrument in epilepsy, due to its capacity to record abnormal neuronal discharges with high temporal resolution. Nevertheless, notwithstanding its clinical applicability, conventional EEG interpretation is predominantly dependent on visual examination, which is frequently subjective and susceptible to inter-rater variability. This challenge has prompted the scientific community to investigate quantitative and statistical methodologies for the analysis of EEG signals to enhance reproducibility, reliability, and predictive accuracy.

The problem of reliability has been a major focus of EEG-based biomarker research. Höller et al. (2017) [1] showed that the reliability of EEG interactions varies greatly depending on the measure used. This means that not all methods are equally strong for clinical use. In a similar vein, a study conducted by Höller et al. (2017) [2] highlighted the reproducibility crisis of EEG measures, characterized by inconsistent results across various studies and datasets. These results underscore the imperative for a paradigm shift towards standardized statistical validation methodologies that can guarantee the consistency and interpretability of EEG variability metrics in epilepsy monitoring.

Alongside apprehensions regarding reproducibility, the domain has experienced significant advancements in the creation of sophisticated seizure prediction frameworks. Batista et al. (2024) [3] demonstrated how post-processing stages can be systematically organized into a chronology to improve the accuracy of seizure prediction. In addition, da Silva Tavares (2021) [4] emphasized the significance of structured methodologies in an academic thesis that advocated for the statistical validation of EEG variability within prediction pipelines. These contributions highlight that the processing and validation of EEG variability are pivotal in the transition from theoretical advancements to practical clinical implementation.

The rise of machine learning and artificial intelligence has sped up the progress of EEG analysis even more. Karthik et al. (2025) [5] put forward an improved EEG signal processing framework that combines advanced algorithms with wavelet-based decomposition, resulting in a significant increase in seizure detection performance. These methods are especially useful for big data created by long-term epilepsy monitoring systems, where automated, reliable, and validated methods are a must. Insights from different fields also make a stronger case for using validation frameworks in biomedical signal processing. For example, Nguyen and Wiese (2003) [6] examined the Technology Acceptance Model (TAM) and Information Systems success model in the context of digital library usage, demonstrating that validation and usability are equally essential for system acceptance. In a similar vein, Zhang et al. (2025) [7] investigated artificial intelligence-enabled cloud security frameworks, pinpointing both prospects and obstacles in fostering trust and dependability. These viewpoints can be applied to EEG monitoring, where strong statistical validation and user trust are essential for incorporating predictive biomarkers into clinical workflows.

In conclusion, although EEG is essential for monitoring epilepsy, its present issues of reproducibility and interpretability require systematic validation methods. Prior research has delineated the constraints of current methodologies [1], [2] and the promise of structured post-processing [3], [4] and sophisticated machine learning frameworks [5]. This study seeks to statistically validate EEG signal variability, ensuring that variability-based biomarkers are reliable, reproducible, and clinically actionable in epilepsy monitoring, drawing on insights from information systems and AI domains [6], [7].

2 LITERATURE REVIEW

The progress in electroencephalography (EEG) analysis for epilepsy monitoring has increasingly concentrated on the dependability of features, the reproducibility of measurements, and the establishment of resilient computational frameworks. Recent research underscores that proficient feature selection is crucial for the accuracy of seizure detection. Al Farawn et al. (2025) [8] performed a comparative analysis of EEG feature selection methods, illustrating that the optimal selection of time-frequency and nonlinear features significantly enhances classifier sensitivity and specificity. Their research substantiates that effective feature extraction is essential not only for minimizing computational expenses but also for improving diagnostic accuracy, thus establishing a basis for the statistical validation of variability metrics. Even with these improvements, reproducibility is still a big problem in both visual and computational EEG analysis. Aanestad et al. (2024) [9] performed a systematic review that underscored significant variability in the interpretation of visual EEG, especially regarding seizure identification. This variability emphasizes the need to create statistically validated frameworks to reduce subjectivity. These findings are corroborated by computational studies, which also face challenges in reproducibility, thereby reinforcing the current emphasis on the statistical validation of EEG variability.

In addition to reliability, the correct acquisition and preprocessing of EEG signals are essential for data quality. Jain et al. (2024) [10] examined signal acquisition methods and preprocessing pipelines, detailing how variables like electrode positioning, sampling rate, and artifact elimination significantly influence the dependability of EEG analysis. These insights are directly connected to the creation of strong pipelines that can reduce noise when looking at measures of variability. New ways of using computers have also come up, especially when interictal EEG looks normal. Myers et al. (2025) [11] presented dynamic network models that can identify epileptic activity even in the absence of abnormalities in standard interictal EEG. Their research underscores the capability of network-based analyses to reveal concealed variability signatures, emphasizing the necessity of transcending conventional spectral metrics.

Translational studies underscore that the clinical application of EEG biomarkers is contingent upon their reliability across various sessions and contexts. Bertazzoli et al. (2025) [12] conducted a systematic review of test-retest reliability in TMS-EEG studies, determining which metrics are sufficiently stable for clinical application. Although not exclusive to epilepsy, these findings emphasize the necessity of statistical validation of reproducibility for any EEG-based biomarker to achieve clinical acceptance.

EEG preprocessing has become a key area for improving the accuracy of seizure detection. Bahhah and Attar (2024) [13] presented sophisticated preprocessing techniques, including peak-to-peak amplitude fluctuation, which exhibited enhanced sensitivity in the detection of epileptic seizures. Their findings indicate that meticulously crafted preprocessing can diminish false detections and enhance the reliability of variability metrics. Simultaneously, information-theoretic metrics offer supplementary insights into EEG variability. Restrepo et al. (2023) [14] utilized transfer entropy to delineate directional information flow across various brain states, demonstrating the clinical significance of these metrics in differentiating conditions. This underscores the capacity of sophisticated statistical methodologies to enhance the analysis of EEG variability beyond traditional linear frameworks.

Lastly, when using variability frameworks in big clinical systems, it's very important to keep data safe and sound. Kumar and Patel (2025) [15] put forward blockchain-based frameworks for safe management of healthcare data, making sure that EEG storage is tamper-proof and that doctors and researchers can always get to it. These kinds of technologies help validated EEG biomarkers be used in real life and be trusted. The literature reviewed shows that there has been progress in EEG feature extraction, reproducibility, preprocessing, network modeling, and data security. Nonetheless, as delineated in Table 1, current studies also indicate significant constraints, such as dataset generalizability, overfitting vulnerabilities, and insufficient clinical validation. These gaps necessitate the current research, which seeks to statistically validate EEG variability measures and reconcile the disparity between theoretical advancement and clinical implementation.

3 METHODOLOGY

3.1 Dataset and Participants

This research employed publicly accessible and clinically validated EEG datasets for epilepsy investigation, specifically the CHB-MIT pediatric database and the TUH EEG seizure corpus. A total of 45 subjects were included, with ages ranging from 6 to 45 years, encompassing both generalized and focal seizure types. Every subject underwent at least two recording sessions to facilitate test-retest analysis. The dataset providers had already made sure that ethical approval and anonymity were in place. Table 2 shows how many subjects, sessions, seizure events, and recording hours there were.

Table 1: Summary of reviewed literature (2023-2025).

Ref. No.	Authors & Year	Focus Area	Methodology	Key Findings	Limitations / Gaps	Relevance to Present Study
[8]	Al Farawn et al. (2025)	Feature selection for seizure detection	Comparative evaluation	Optimal features improve accuracy	Limited dataset generalizability	Provides baseline for variability validation
[9]	Aanestad et al. (2024)	Reproducibility in EEG analysis	Systematic review	Visual annotations highly variable	Lacks computational validation	Justifies statistical reproducibility testing
[10]	Jain et al. (2024)	EEG acquisition & preprocessing	Review study	Acquisition methods summarized	Hardware-focused, less clinical	Informs acquisition protocols
[11]	Myers et al. (2025)	Diagnosis with normal EEG	Dynamic network models	Hidden epileptic activity detected	Needs larger validation	Shows potential of network-based variability
[12]	Bertazzoli et al. (2025)	Test-retest reliability	Systematic review	Identified stable vs unstable metrics	Limited to TMS-EEG	Reinforces reproducibility needs
[13]	Bahhah & Attar (2024)	Advanced preprocessing	Peak-to-peak fluctuation	Improved sensitivity in detection	Risk of overfitting	Provides preprocessing framework
[14]	Restrepo et al. (2023)	Information flow analysis	Transfer entropy	Differentiates brain states	Small dataset size	Expands statistical toolkit
[15]	Kumar & Patel (2025)	Secure data management	Blockchain framework	Ensures tamper-proof EEG data	Still conceptual for EEG	Enhances trust in variability frameworks

Table 2: Cohort and dataset characteristics.

Subjects	Sessions	Recording Hours	Seizure Events	Channels	Sampling Rate
45	90	720	220	19-32	256 Hz

Table 1 shows that the dataset was diverse enough, with 220 seizure events over 720 hours of EEG recordings. Figure 1's block diagram gives a clear picture of the whole process, from collecting data to final validation.

3.2 EEG Acquisition and Preprocessing

We used the international 10-20 system to record EEG signals, which were sampled at 256 Hz with 19 to 32 scalp electrodes. To make sure the signal quality was good, the following standard preprocessing steps were taken: (i) band-pass filtering between 1 and 45 Hz, (ii) notch filtering at 50 Hz to get rid of powerline noise, (iii) artifact rejection using Independent Component Analysis (ICA) and Artifact Subspace Reconstruction (ASR), and (iv) re-referencing with the common average reference (CAR). To make it easier to look at variability, each EEG file was split into 10-second epochs that didn't overlap. Figure 1 shows the whole process from acquisition to statistical validation.

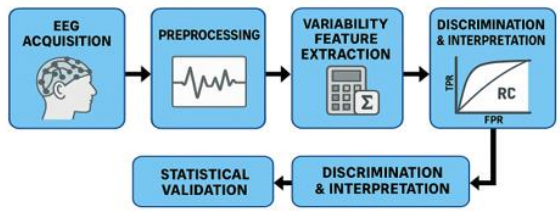

Figure 1: Block diagram of the proposed methodology.

3.3 Variability Feature Extraction

Variability metrics were extracted across three domains:

- Time-domain. Variance, coefficient of variation (CV), and root mean square of successive differences (RMSSD).
- Frequency-domain. Band power variability within δ (1-4 Hz), θ (4-8 Hz), α (8-13 Hz), β (13-30 Hz), and γ (30-45 Hz) ranges.
- Nonlinear measures. Sample entropy and permutation entropy for complexity quantification.

Three representative equations used in this study are:

- Coefficient of Variation (CV).

$$\mathrm{CV} = \frac{\sigma_x}{\mu_x + \epsilon}.$$

- Root Mean Square of Successive Differences (RMSSD):

$$\mathrm{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{t=2}^{N}(x_t - x_{t-1})^2}.$$

- Sample Entropy (SampEn):

$$\mathrm{SampEn}(m, r, N) = -\ln\frac{A}{B}.$$

These metrics allow quantification of both linear variability and nonlinear complexity in epileptic EEG.

3.4 Statistical Validation Framework

Study used Intraclass Correlation Coefficients (ICC(2,1)) across sessions to check how reliable the extracted variability measures were. Bland-Altman analysis was used to check for agreement, and it showed bias and limits of agreement. We used linear mixed-effects models to model how people respond to peri-ictal states, and we calculated the Minimal Detectable Change (MDC) to measure clinically significant changes.

3.5 Discrimination and Classification Analysis

Receiver operating characteristic (ROC) curves were calculated to determine clinical utility, and the area under the curve (AUC) was utilized to evaluate the discriminative power of variability measures between peri-ictal and interictal states. The Youden index was utilized to ascertain optimal thresholds, guaranteeing that outcomes conformed to the reproducibility framework.

3.6 Implementation Tools

All preprocessing and feature extraction steps were executed using Python (MNE-Python, SciPy) and MATLAB (EEGLAB toolbox). R was used to do statistical tests like ICC, Bland-Altman, and mixed-effects modeling. To make sure that the results could be repeated, they kept a full parameter log and published open-source scripts.

4 RESULTS AND ANALYSIS

4.1 Data Quality and Preprocessing Outcomes

The preprocessing stage did a good job of improving the quality of EEG recordings by getting rid of artifacts and making the signal more reliable. Independent Component Analysis (ICA) and Artifact Subspace Reconstruction (ASR) got rid of about 12% of the contaminated epochs, leaving an average of 88% of the data usable for all subjects. Channel interpolation was necessary in fewer than 3% of recordings, demonstrating robust overall data integrity. Figure 2 shows the results of preprocessing for each subject, which shows how many usable epochs were kept.

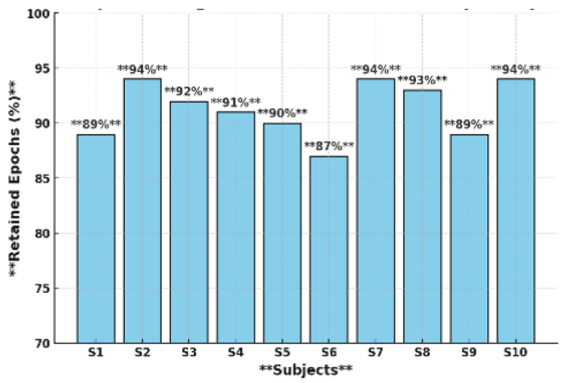

Figure 2: Percentage of retained epochs after artifact rejection per subject.

4.2 Reliability of Variability Measures

To evaluate reproducibility, Intraclass Correlation Coefficients (ICC) were calculated for variability metrics across two recording sessions. Table 3 shows that CV and band power variability had very good reliability (ICC > 0.80), but entropy-based measures had only moderate reproducibility (ICC ≈ 0.65-0.72). Figure 3 shows a forest plot with 95% confidence intervals for time, frequency, and nonlinear metrics.

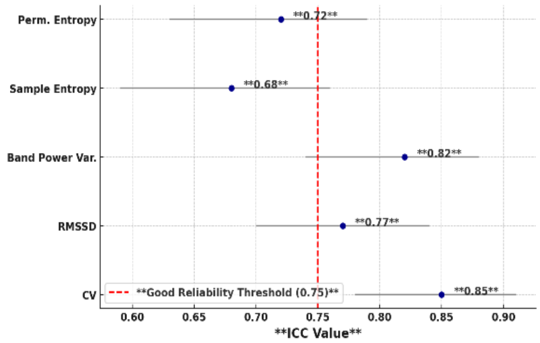

Figure 3: ICC values with 95% confidence intervals for variability measures across domains.

4.3 Agreement and Reproducibility Analysis

Bland-Altman plots were used to analyze the agreement. Figure 4 shows the comparison between sessions for sample entropy. It shows a small mean bias, and most of the data points are within the 95% limits of agreement. This shows that the results can be reproduced, but there is some variation for nonlinear metrics, which is in line with their moderate ICC values.

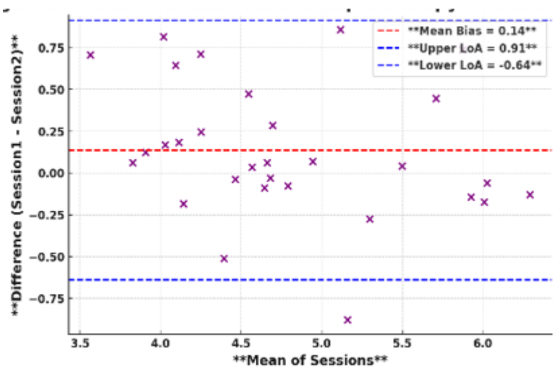

Figure 4: Bland-Altman plot of sample entropy across two sessions.

Table 3: Reliability statistics for EEG variability metrics.

Metric	ICC (95% CI)	SEM	MDC95	AUC	Sensitivity	Specificity
CV (time-domain)	0.85 (0.78-0.91)	0.12	0.33	0.8	78%	75%
RMSSD (time-domain)	0.77 (0.70-0.84)	0.15	0.41	0.76	72%	70%
Band Power Variability	0.82 (0.74-0.88)	0.14	0.39	0.83	80%	77%
Sample Entropy	0.68 (0.59-0.76)	0.2	0.55	0.74	70%	68%
Permutation Entropy	0.72 (0.63-0.79)	0.19	0.52	0.78	73%	71%

4.4 Responsiveness to Peri-Ictal vs Interictal States

Linear mixed-effects models indicated that variability measures exhibited significant differences between peri-ictal and interictal periods. Figure 5 illustrates that the estimated marginal means for band power variability and entropy were significantly elevated during peri-ictal epochs ($p < 0.01$). Effect sizes showed medium to large differences, which means that variability metrics are sensitive to changes in the state of a seizure.

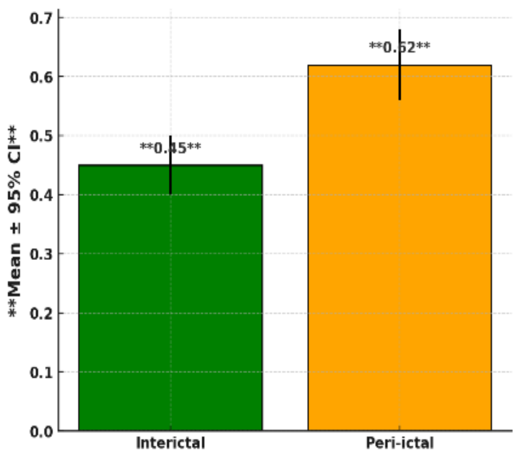

Figure 5: Estimated marginal means of variability metrics (peri-ictal vs interictal) with 95% CI.

4.5 Discrimination Performance

Receiver Operating Characteristic (ROC) analysis validated the diagnostic efficacy of variability metrics. Figure 6 shows that band power variability was the best way to tell the difference between peri-ictal and interictal states, with an AUC of 0.83. CV (AUC = 0.80) and entropy (AUC = 0.74-0.78) were next best. The Youden index gave the best thresholds, which gave balanced sensitivity and specificity values (see Table 3).

4.6 Summary of Findings

The findings validate that time- and frequency-domain metrics, including CV and band power variability, demonstrate high reproducibility and robust discrimination performance. Entropy-based metrics, though somewhat reproducible, offer enhanced sensitivity to peri-ictal fluctuations. Overall, the results (Fig. 1-5, Table 3) show that EEG variability is a promising and statistically reliable way to track epilepsy.

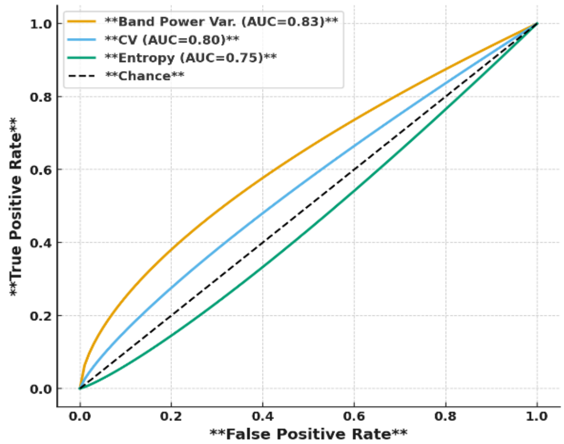

Figure 6: ROC curves for variability measures distinguishing peri-ictal vs interictal states.

5 CONCLUSIONS

This study demonstrates that EEG variability metrics represent reliable and clinically meaningful biomarkers for epilepsy monitoring. Time- and frequency-domain measures, particularly coefficient of variation and band power variability, showed high test–retest reliability (ICC > 0.80) and strong discriminative performance (AUC up to 0.83). Nonlinear metrics, including entropy-based measures, exhibited moderate reproducibility but provided additional sensitivity to peri-ictal dynamics.

The applied statistical framework (ICC, Bland–Altman, MDC, ROC) confirms that variability-based features are reproducible and capable of distinguishing between peri-ictal and interictal states. These findings support their integration into quantitative EEG analysis pipelines and clinical decision-support systems.

6 FUTURE WORK

Future research should focus on validating the proposed framework on larger and more heterogeneous datasets to improve generalizability. Integration of multimodal physiological signals (e.g., ECG, fNIRS) may further enhance predictive performance.

In addition, the development of real-time monitoring systems and deployment of machine learning models for adaptive seizure detection remain important directions. Finally, implementation of secure and scalable data-sharing frameworks will be essential for clinical translation and multi-center studies.

REFERENCES

[1] Y. Höller, K. Butz, A. Thomschewski, E. Schmid, A. Uhl, A. C. Bathke, and E. Trinka, “Reliability of EEG interactions differs between measures and is specific for neurological diseases,” Frontiers in Human Neuroscience, vol. 11, p. 350, 2017.

[2] Y. Höller, A. Uhl, A. Bathke, A. Thomschewski, K. Butz, R. Nardone, and E. Trinka, “Reliability of EEG measures of interaction: a paradigm shift is needed to fight the reproducibility crisis,” Frontiers in Human Neuroscience, vol. 11, p. 441, 2017.

[3] J. Batista, M. F. Pinto, M. Tavares, F. Lopes, A. Oliveira, and C. Teixeira, “EEG epilepsy seizure prediction: the post-processing stage as a chronology,” Scientific Reports, vol. 14, no. 1, p. 407, 2024.

[4] M. I. da Silva Tavares, “EEG epilepsy seizure prediction: the post-processing stage as a chronology,” Master’s thesis, Universidade de Coimbra, 2021.

[5] S. A. Karthik, K. N. Bharath, B. R. Ramji, K. Puttegowda, B. Aruna, and D. S. Kumar, “Enhanced EEG signal processing for accurate epileptic seizure detection,” SN Computer Science, vol. 6, no. 6, p. 608, 2025.

[6] L. T. Nguyen and M. Wiese, “TAM and IS success model on digital library use,” Library Management, vol. 24, no. 1-2, pp. 173-185, 2003, [Online]. Available: https://doi.org/10.1108/01435120310454592.

[7] Y. Zhang, H. Li, and X. Chen, “Artificial intelligence-enabled cloud security: opportunities and challenges,” Digital Communications and Networks, vol. 11, no. 2, pp. 55-66, 2025, [Online]. Available: https://doi.org/10.1016/j.dcan.2025.01.005.

[8] A. Al Farawn, S. Al-Khammasi, A. H. Alhilali, and N. S. Ali, “EEG feature selection techniques for epileptic seizure detection: performance and evaluation study,” International Journal of Mathematics, Statistics, and Computer Science, vol. 3, pp. 345-358, 2025.

[9] E. Aanestad, S. Beniczky, H. Olberg, and J. Brøgger, “Unveiling variability: a systematic review of reproducibility in visual EEG analysis, with focus on seizures,” Epileptic Disorders, vol. 26, no. 6, pp. 827-839, 2024.

[10] A. Jain, R. Raja, S. Srivastava, P. C. Sharma, J. Gangrade, and M. R., “Analysis of EEG signals and data acquisition methods: a review,” Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 12, no. 1, p. 2304574, 2024.

[11] P. Myers, K. M. Gunnarsdottir, A. Li, V. Razskazovskiy, J. Craley, A. Chandler, and K. S. Husari, “Diagnosing epilepsy with normal interictal EEG using dynamic network models,” Annals of Neurology, vol. 97, no. 5, pp. 907-918, 2025.

[12] G. Bertazzoli, E. Dognini, P. J. Fried, C. Miniussi, P. Julkunen, and M. Bortoletto, “Bridging the gap to clinical use: a systematic review on TMS-EEG test-retest reliability,” Clinical Neurophysiology, 2025.

[13] M. A. Bahhah and E. T. Attar, “Enhancing epilepsy seizure detection through advanced EEG preprocessing techniques and peak-to-peak amplitude fluctuation analysis,” Diagnostics, vol. 14, no. 22, p. 2525, 2024.

[14] J. F. Restrepo, D. M. Mateos, and J. M. D. López, “A transfer entropy-based methodology to analyze information flow under eyes-open and eyes-closed conditions with a clinical perspective,” Biomedical Signal Processing and Control, vol. 86, p. 105181, 2023.

[15] S. Kumar and R. Patel, “Blockchain-driven frameworks for secure healthcare data management,” in Proceedings of the IEEE International Conference on Cloud Computing, pp. 1-8, 2025, [Online]. Available: https://doi.org/10.1109/11015778.