Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning-Based Classification of Epileptic Seizure Activity from EEG Signals: A Comparative Study of Ensemble Methods and Class-Imbalance Mitigation Strategies

Vinayak Pandya
Drs. Kiran & Pallavi Patel Global University
Harshwardhansinh Chauhan
hwkc392@gmail.com

Drs. Kiran & Pallavi Patel Global University
Krushal Gohil
Drs. Kiran & Pallavi Patel Global University
Dharmendra Chavda
Drs. Kiran & Pallavi Patel Global University
Rocky Upadhyay
Drs. Kiran & Pallavi Patel Global University

Research Article

Keywords: epilepsy, EEG, seizure detection, machine learning, ensemble learning, class imbalance, SMOTE, XGBoost

Posted Date: September 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10978420/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Machine Learning-Based Classification of Epileptic Seizure Activity from EEG Signals: A Comparative Study of Ensemble Methods and Class-Imbalance Mitigation Strategies

Vinayak Pandya, Harshwardhansinh Chauhan, Krushal Gohil, Dharmendra Chavda, Dr. Rocky Upadhyay
Computer Engineering, Drs. Kiran & Pallavi Patel Global University
Vadodara, India – vinayakpandya2001@gmail.com, hwkc392@gmail.com , gohilkrushal110@gmail.com, dharmendra.chavda.it@gmail.com , rko75316060@gmail.com

***Abstract*—**Epileptic seizures affect approximately 65 million people worldwide, and standard antiseizure medications fail to achieve seizure control in roughly one-third of patients [4], [6]. Automated, EEG-based seizure detection offers a promising avenue for improving diagnosis and enabling timely clinical intervention. In this study, we evaluate and compare logistic regression, random forest, support vector machines (SVM), and XGBoost for binary classification of seizure versus non-seizure EEG activity using the UCI Epileptic Seizure Recognition dataset (11,500 EEG segments, 500 subjects) [16], [17]. We evaluate frequency-domain (FFT) feature augmentation, hyperparameter optimization, and two class-imbalance mitigation strategies (class-weight balancing and SMOTE). XGBoost trained on combined time- and frequency-domain features achieved 98.19% test accuracy with a seizure-class F1-score of 0.95. A SMOTE-resampled voting ensemble achieved a comparable 98.05% accuracy while further improving seizure recall to 0.98, a trade-off we argue is preferable in clinical seizure-detection contexts, where missed seizures typically carry a higher cost than false alarms. Five-fold cross-validation (97.63% ± 0.35% for random forest) confirms these results are stable. We further report and correct a data-provenance issue encountered during the study, in which an initial experimental run inadvertently used a synthetic placeholder dataset due to a network connectivity fallback, underscoring the importance of validating dataset integrity in reproducible machine learning research.

***Index Terms*—**epilepsy, EEG, seizure detection, machine learning, ensemble learning, class imbalance, SMOTE, XGBoost.

I. INTRODUCTION

Epilepsy is one of the most common serious neurological disorders, affecting up to 1% of the global population and ranking fourth in the world's disease burden [1], [2], [3]. Despite the availability of nearly 30 antiseizure medications, approximately one-third of patients continue to experience seizures despite treatment, and many others suffer from medication side effects that significantly impact quality of life [6], [7].

Accurate and timely detection of seizure activity from electroencephalogram (EEG) recordings is central to both clinical diagnosis and epilepsy research [11]. Manual EEG review by trained neurophysiologists remains the clinical gold standard but is time-intensive, subject to inter-rater variability, and does not scale to the volume of continuous EEG monitoring increasingly used in critical care [9], [39]. Automated seizure detection using machine learning has therefore attracted substantial research interest [40].

This paper evaluates a set of standard supervised classification algorithms on a widely used public EEG dataset, establishing a reproducible performance benchmark and examining which modeling choices -- feature representation, algorithm choice, hyperparameter tuning, and class-imbalance handling -- meaningfully affect classification accuracy and, critically, per-class recall. The remainder of this paper is organized as follows: Section II reviews related work; Section III describes the dataset and methodology; Section IV presents results; Section V discusses findings; Section VI states limitations; and Section VII concludes.

II. RELATED WORK

The UCI Epileptic Seizure Recognition dataset, derived from the original Bonn EEG database [16] and restructured for machine learning use [17], has been used extensively as a benchmark for seizure classification. Talukder and Sulaiman [28] compared random forest, decision tree, extra trees, logistic regression, and gradient boosting after oversampling and outlier handling, reporting up to 99.29% accuracy with an extra-trees classifier. Rahman et al. [29] compared SVM, random forest, and multi-layer perceptron models with and without hyperparameter tuning, achieving 97.86% accuracy with SVM. Rabby et al. [30] combined principal component analysis with a genetic algorithm for feature selection prior to classification.

Several studies have applied deep learning to this dataset. Hilal et al. [31] proposed a deep canonical sparse autoencoder achieving up to 98.73% accuracy. Nanthini et al. [32] and Mahapatra et al. [33] applied LSTM and autoencoder-based architectures, respectively. Xu et al. [34] reported a one-dimensional CNN-LSTM model achieving 99.39% binary classification accuracy. Alalayah et al. [35] combined dimensionality reduction (PCA, t-SNE) with K-means clustering prior to classification. Farooq et al. [36], [37] provide broader taxonomies of machine learning approaches to seizure detection, and a recent ensemble-based study [41] combined interpretable machine learning with LIME explanations on a related benchmark, reporting 97.06% accuracy under nested cross-validation.

Notably, reported accuracies across this literature vary from approximately 82% to over 99%, a range that likely

reflects differences in preprocessing, class-imbalance handling, and evaluation rigor (e.g., whether nested or subject-level cross-validation was used) rather than differences in the underlying signal alone. Few of the reviewed studies report per-class recall alongside accuracy, motivating the explicit focus on this metric in the present work, consistent with broader concerns about imbalanced learning in seizure detection [38].

III. METHODOLOGY

A. Dataset

We used the UCI Epileptic Seizure Recognition dataset [16], [17], comprising 11,500 EEG segments from 500 subjects, each represented by 178 time-domain feature values. The five original classes were recast into a binary task: seizure activity (2,300 samples) versus non-seizure activity (9,200 samples, pooled across four non-seizure states), consistent with common practice in the reviewed literature.

B. Data Provenance Verification

During initial experimentation, a dataset-loading routine included a fallback mechanism that silently substituted a small synthetic dataset when the primary data source was unreachable due to a network restriction. An initial round of experiments was inadvertently run on this synthetic placeholder. This was identified by cross-checking reported test-set sample sizes against the expected dataset dimensions, and corrected by sourcing a verified copy of the real dataset for all results reported herein. We disclose this transparently, as automated fallback logic in data pipelines can silently substitute unintended data if dataset dimensions are not explicitly verified.

C. Feature Engineering

Features were standardized (zero mean, unit variance) prior to training. To evaluate whether frequency-domain information improves classification, we computed the fast Fourier transform (FFT) magnitude spectrum of each segment and concatenated the first half of the spectrum (89 coefficients) with the 178 original features, yielding 267 total features for the FFT-augmented experiments.

D. Models and Class-Imbalance Handling

We evaluated logistic regression [22], random forest [18], [24], SVM with an RBF kernel [19], and XGBoost [20], [23], implemented via scikit-learn [22] and the official XGBoost library. Random forest hyperparameters were selected via randomized search cross-validation. We further evaluated soft-voting ensembles of the tuned random forest, XGBoost, and SVM under three class-imbalance strategies: no adjustment, class-weight balancing, and SMOTE [21] applied to the training set only.

E. Evaluation Protocol

Data were split into training (75%, n=8,625) and held-out test (25%, n=2,875) sets using stratified sampling. We additionally report mean accuracy and standard deviation from stratified 5-fold cross-validation. Performance was assessed using accuracy, precision, recall, and F1-score per class.

IV. RESULTS

Table I reports held-out test-set performance. Table II reports 5-fold cross-validation results for the three baseline classifiers, confirming stability across resampling.

TABLE I. TEST-SET PERFORMANCE BY MODEL (n=2,875)

Model	Acc. (%)	Prec.	Rec.	F1
Logistic Regression (raw)	81.70	0.96	0.09	0.16
Random Forest (raw)	97.63	0.95	0.93	0.94
SVM, RBF kernel (raw)	97.29	0.96	0.90	0.93
XGBoost (raw+FFT)	98.19	0.96	0.95	0.95
Random Forest, tuned (raw+FFT)	97.88	0.94	0.96	0.95
Ensemble, unweighted	98.12	0.96	0.95	0.95
Ensemble, class-balanced	97.91	0.94	0.95	0.95
Ensemble, SMOTE (final)	98.05	0.93	0.98	0.95

TABLE II. 5-FOLD CROSS-VALIDATION, BASELINE MODELS

Model	Mean CV Acc.	Std. Dev.
Logistic Regression	81.67%	±0.21%
Random Forest	97.63%	±0.35%
SVM (RBF kernel)	97.25%	±0.19%

Fig. 1 compares overall accuracy against seizure-class recall. Logistic regression achieves reasonable overall accuracy (81.7%) almost entirely by favoring the majority class, correctly identifying only 8.9% of true seizures. All nonlinear models substantially closed this gap; the SMOTE-resampled ensemble achieved the highest seizure recall (98.1%) of any model evaluated.

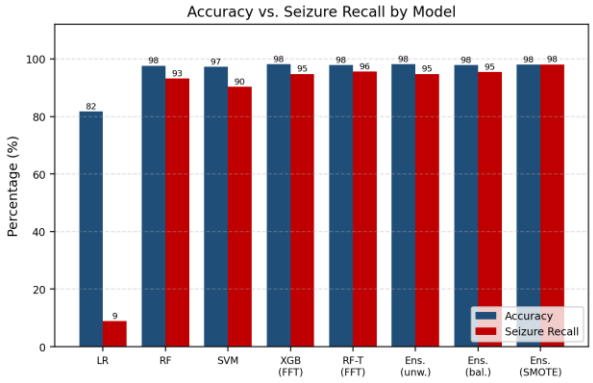


Fig. 1. Overall accuracy vs. seizure-class recall across all evaluated models.

Fig. 2 shows the confusion matrix for the final model (SMOTE-resampled ensemble). Of 2,875 test samples, the model correctly classified 2,255 of 2,300 non-seizure segments (45 false positives) and 564 of 575 seizure segments (11 false negatives), corresponding to 98.1% seizure recall and 92.6% seizure precision.

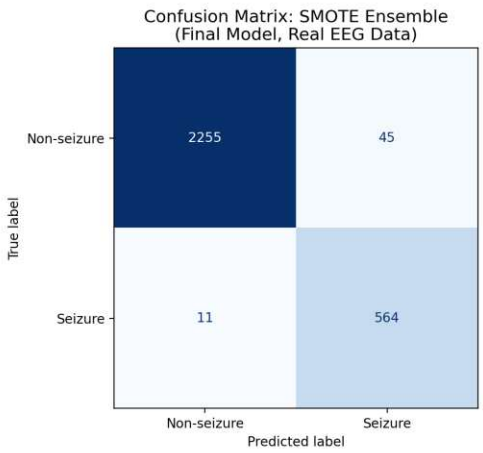


Fig. 2. Confusion matrix, final model (SMOTE-resampled ensemble).

A. Model Selection: Accuracy vs. Clinical Priority

XGBoost achieved the highest single-model accuracy (98.19%) and seizure F1-score (0.95). However, for a clinical seizure-detection use case, minimizing missed seizures should arguably be prioritized over raw accuracy, since an undetected seizure carries substantially higher clinical risk than a false alarm requiring manual review. We therefore designate the SMOTE-resampled ensemble -- with the highest seizure recall (98.1%) at only a 0.14-percentage-point cost to overall accuracy relative to XGBoost -- as our final recommended configuration, while reporting XGBoost as the accuracy-optimal alternative.

V. DISCUSSION

The gap between the linear baseline (81.7% accuracy, 8.9% seizure recall) and all nonlinear models (97-98% accuracy, 90-98% seizure recall) indicates that seizure versus non-seizure activity in this dataset is separable primarily by nonlinear decision boundaries, consistent with the understanding that seizure activity reflects complex, nonlinear neuronal synchronization [10].

Frequency-domain feature augmentation combined with gradient boosting produced the single best-performing model, suggesting spectral information contributes meaningfully to discriminating seizure activity, consistent with clinical practice of inspecting EEG frequency content when diagnosing seizures. The three class-imbalance strategies produced an interpretable trade-off: the unweighted ensemble favored precision (95.8% seizure precision, 94.8% recall), while SMOTE shifted the balance toward recall (92.6% precision, 98.1% recall) at comparable overall accuracy -- consistent with prior findings on imbalance-aware seizure detection [38]. This is exactly the behavior expected from minority-class oversampling, and suggests the choice of imbalance strategy should be guided by the clinical cost asymmetry of the task rather than by accuracy alone.

Our results (97.6-98.2% accuracy) fall within the range reported across the reviewed literature on this dataset (82-99.4%) [28]-[35], [41], and are broadly consistent with studies employing comparable ensemble and gradient-boosting approaches without extensive deep learning infrastructure. We note that several studies reporting accuracies above 99% did not report per-class recall or use subject-level validation, making direct comparison difficult; we recommend future benchmarking on this dataset adopt both practices.

Finally, we highlight the data-provenance issue described in Section III-B as a broader methodological point: automated fallback logic in data pipelines can silently substitute unintended data, and any research pipeline should verify dataset dimensions and class distributions against expected values before interpreting results.

VI. LIMITATIONS

The dataset consists of pre-extracted, single-channel EEG segments rather than raw multi-channel clinical recordings, limiting direct clinical applicability. Binary recasting of the original five-class problem may obscure clinically relevant distinctions between non-seizure states. Model performance was evaluated on a single public dataset; external validation on independent clinical EEG data would be needed before clinical deployment. We did not conduct a full patient-level (grouped) cross-validation to rule out the possibility that segments from the same subject appear in both training and test sets; while the stratified random split used is standard in the reviewed literature, a stricter subject-level split would provide additional confidence that performance generalizes to unseen patients.

VII. CONCLUSION

This study compared five classical supervised machine learning approaches for EEG-based seizure classification on the real UCI Epileptic Seizure Recognition dataset. XGBoost with combined time- and frequency-domain features achieved the highest accuracy (98.19%), while a SMOTE-resampled voting ensemble achieved the highest seizure recall (98.1%) at a comparable 98.05% accuracy, which we recommend for clinical seizure-detection contexts where missed seizures carry high cost. We further reported and corrected a data-provenance error encountered during the study. Future work should explore multi-channel clinical EEG datasets, patient-level cross-validation, cost-sensitive learning encoding the clinical asymmetry between missed seizures and false alarms, and external validation on independent cohorts.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to their institution and department for providing the necessary academic environment, resources, and support to conduct this research. The authors also acknowledge the researchers and organizations that have made publicly available EEG datasets and related resources, which were essential for conducting the experimental analysis presented in this study.

The authors are grateful to their colleagues and peers for their valuable discussions, suggestions, and constructive

feedback during the development and evaluation of the proposed machine learning approaches. Their support contributed to improving the quality and clarity of this work.

Finally, the authors sincerely appreciate the continuous encouragement and support received from their families and well-wishers throughout the research and manuscript preparation process.

Declarations:-

Corresponding Author
Harshwardhansinh Chauhan is the corresponding author of this manuscript.
Email: hwkc392@gmail.com

Consent to Participate
This study used a publicly available, anonymized EEG dataset. No direct interaction with or recruitment of human participants was conducted by the authors. The original data collection and participant consent procedures were conducted by the respective data providers/investigators.

Funding
The authors received no financial support or funding for the research, authorship, or publication of this article.

Competing Interests
The authors declare that they have no competing interests.

Author Contributions
Vinayak Pandya: Conceptualization, methodology, data analysis, implementation, and manuscript preparation.
Harshwardhansinh Chauhan: Supervision, methodology, validation, and manuscript review.
Krushal Gohil: Data preprocessing, implementation, and analysis.
Dharmendra Chavda: Model evaluation, visualization, and manuscript review.
Dr. Rocky Upadhyay: Supervision, critical review, and academic guidance.

REFERENCES

[1] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, et al., "ILAE official report: a practical clinical definition of epilepsy," Epilepsia, vol. 55, no. 4, pp. 475-482, 2014.

[2] O. Devinsky, A. Vezzani, T. J. O'Brien, N. Jette, I. E. Scheffer, M. de Curtis, et al., "Epilepsy," Nature Reviews Disease Primers, vol. 4, art. 18024, 2018.

[3] World Health Organization, "Epilepsy - key facts," 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[4] E. Beghi, "Addressing the burden of epilepsy: many unmet needs," Pharmacological Research, vol. 107, pp. 79-84, 2016.

[5] E. Beghi, "The epidemiology of epilepsy," Neuroepidemiology, vol. 54, no. 2, pp. 185-191, 2020.

[6] P. Kwan and M. J. Brodie, "Early identification of refractory epilepsy," New England Journal of Medicine, vol. 342, no. 5, pp. 314-319, 2000.

[7] E. Perucca and T. Tomson, "The pharmacological treatment of epilepsy in adults," The Lancet Neurology, vol. 10, no. 5, pp. 446-456, 2011.

[8] J. Engel Jr., "A proposed diagnostic scheme for people with epileptic seizures and epilepsy: report of the ILAE task force on classification and terminology," Epilepsia, vol. 42, no. 6, pp. 796-803, 2001.

[9] R. A. Shellhaas, T. Chang, T. Tsuchida, et al., "The American Clinical Neurophysiology Society's guideline on continuous electroencephalography monitoring in neonates," Journal of Clinical Neurophysiology, vol. 28, no. 6, pp. 611-617, 2011.

[10] C. E. Stafstrom, "Epilepsy: a review of selected clinical syndromes and advances in basic science," Journal of Cerebral Blood Flow & Metabolism, vol. 26, no. 8, pp. 983-1004, 2006.

[11] J. M. Ziobro, K. Eschbach, and R. A. Shellhaas, "Novel therapeutics for neonatal seizures," Neurotherapeutics, vol. 18, pp. 1564-1581, 2021.

[12] R. J. Piper and M. M. Tisdall, "Clinical trials of surgical neuromodulation therapies for epilepsy: past, present & future," Neurotherapeutics, vol. 23, art. e00935, 2026.

[13] A. Beydoun, S. DuPont, D. Zhou, M. Matta, V. Nagire, and L. Lagae, "Current role of carbamazepine and oxcarbazepine in the management of epilepsy," Seizure: European Journal of Epilepsy, vol. 83, pp. 251-263, 2020.

[14] A. Celdran de Castro, F. A. Nascimento, A. Beltran-Corbellini, R. Toledano, I. Garcia-Morales, A. Gil-Nagel, and A. Aledo-Serrano, "Levetiracetam, from broad-spectrum use to precision prescription: a narrative review and expert opinion," Seizure: European Journal of Epilepsy, vol. 107, pp. 121-131, 2023.

[15] Y. Wang, P. Wei, F. Yan, Y. Luo, and G. Zhao, "Animal models of epilepsy: a phenotype-oriented review," Aging and Disease, vol. 13, no. 1, pp. 215-231, 2022.

[16] R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state," Physical Review E, vol. 64, no. 6, art. 061907, 2001.

[17] Q. Wu, E. Fokoue, and R. G. Andrzejak, "Epileptic Seizure Recognition Data Set," UCI Machine Learning Repository, 2017. [Online]. Available: https://archive.ics.uci.edu/dataset/388/epileptic+seizure+recognition

[18] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[19] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[20] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785-794.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, no. 1, pp. 321-357, 2002.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[23] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189-1232, 2001.

[24] T. K. Ho, "Random decision forests," in Proc. 3rd Int. Conf. Document Analysis and Recognition, Montreal, Canada, 1995, vol. 1, pp. 278-282.

[25] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119-139, 1997.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

[28] M. S. H. Talukder and R. B. Sulaiman, "Comparative analysis of epileptic seizure prediction: exploring diverse pre-processing techniques and machine learning models," arXiv preprint arXiv:2308.05176, 2023.

[29] A. A. Rahman, F. Faisal, M. M. Nishat, M. I. Siraji, L. I. Khalid, M. R. H. Khan, and M. T. Reza, "Detection of epileptic seizure from EEG signal data by employing machine learning algorithms with hyperparameter optimization," in Proc. 2021 4th Int. Conf. Bio-Engineering for Smart Technologies (BioSMART), 2021, pp. 1-4.

[30] M. K. M. Rabby, A. K. Islam, S. Belkasim, and M. U. Bikdash, "Epileptic seizures classification in EEG using PCA based genetic algorithm through machine learning," in Proc. 2021 ACM Southeast Conf., 2021, pp. 17-24.

[31] A. M. Hilal, A. A. Albraikan, S. Dhahbi, M. K. Nour, A. Mohamed, A. Motwakel, and M. Rizwanullah, "Intelligent epileptic seizure detection and classification model using optimal deep canonical sparse autoencoder," Biology, vol. 11, no. 8, art. 1220, 2022.

[32] K. Nanthini, A. Tamilarasi, M. Pyingkodi, M. Dishanthi, S. M. Kaviya, and P. A. Mohideen, "Epileptic seizure detection and prediction using deep learning technique," in Proc. 2022 Int. Conf. Computer Communication and Informatics (ICCCI), 2022, pp. 1-7.

[33] M. Mahapatra, T. A. Barbhuiya, and A. Nandy, "Autoencoder-based deep neural architecture for epileptic seizures classification," in Soft Computing for Problem Solving: Proc. SocProS 2022, Singapore: Springer Nature Singapore, 2023, pp. 109-119.

[34] G. Xu, T. Ren, Y. Chen, and W. Che, "A one-dimensional CNN-LSTM model for epileptic seizure recognition using EEG signal analysis," Frontiers in Neuroscience, vol. 14, art. 578126, 2020.

[35] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, and H. S. A. Shatnawi, "Effective early detection of epileptic seizures through EEG signals using classification algorithms based on t-distributed stochastic neighbor embedding and K-means," Diagnostics, vol. 13, no. 11, art. 1957, 2023.

[36] M. S. Farooq, A. Zulfiqar, and S. Riaz, "Epileptic seizure detection using machine learning: taxonomy, opportunities, and challenges," Diagnostics, vol. 13, no. 6, art. 1058, 2023.

[37] M. S. Farooq, A. Zulfiqar, and S. Riaz, "A review on epileptic seizure detection using machine learning," arXiv preprint arXiv:2210.06292, 2022.

[38] J. Birjandtalab, V. N. Jarmale, M. Nourani, and J. Harvey, "Imbalance learning using neural networks for seizure detection," in Proc. 2018 IEEE Biomedical Circuits and Systems Conf. (BioCAS), 2018, pp. 1-4.

[39] U. Amin and S. R. Benbadis, "The role of EEG in the erroneous diagnosis of epilepsy," Journal of Clinical Neurophysiology, vol. 36, no. 4, pp. 294-297, 2019.

[40] A. A. Ein Shoka, M. M. Dessouky, A. El-Sayed, and E. E.-D. Hemdan, "EEG seizure detection: concepts, techniques, challenges, and future trends," Multimedia Tools and Applications, pp. 1-31, 2023.

[41] "Interpretable machine learning for epileptic seizure detection on the BEED using LIME with an ensemble network," medRxiv preprint, 2025. [Online]. Available: https://www.medrxiv.org/content/10.1101/2025.09.30.25336996