

A MULTIMODAL MACHINE LEARNING FRAMEWORK FOR EARLY SEIZURE PREDICTION

Anna Suherman

ABSTRACT

Epilepsy is a neurological disorder characterized by recurrent seizures that often occur without warning, posing significant risks to patient safety and quality of life. While current treatments reduce seizure frequency for some individuals, reliable early seizure prediction remains a major challenge. This study developed a machine learning framework for early seizure prediction using electroencephalography (EEG) signals, intended to serve as the foundation for future accessible multimodal seizure prediction systems. EEG recordings from five cases in the publicly available CHB-MIT Scalp EEG Database were segmented into 10-second windows and labeled as preictal (5–30 minutes before seizure onset) or interictal. Statistical, time-domain, and frequency-domain features were extracted from each EEG window and used to train and compare three supervised machine learning models: XGBoost, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Model performance was evaluated using grouped train-test splitting, grouped cross-validation, and the metrics accuracy, precision, recall, F1 score, and ROC-AUC. Among the models evaluated, XGBoost achieved the strongest overall performance, obtaining an accuracy of 76.2%, recall of 84.6%, F1 score of 78.0%, and ROC-AUC of 85.2% on held-out test cases. These findings demonstrate the feasibility of using machine learning to identify preictal EEG patterns for early seizure prediction. Overall, this study provides a foundation for the potential development of more effective seizure prediction systems.

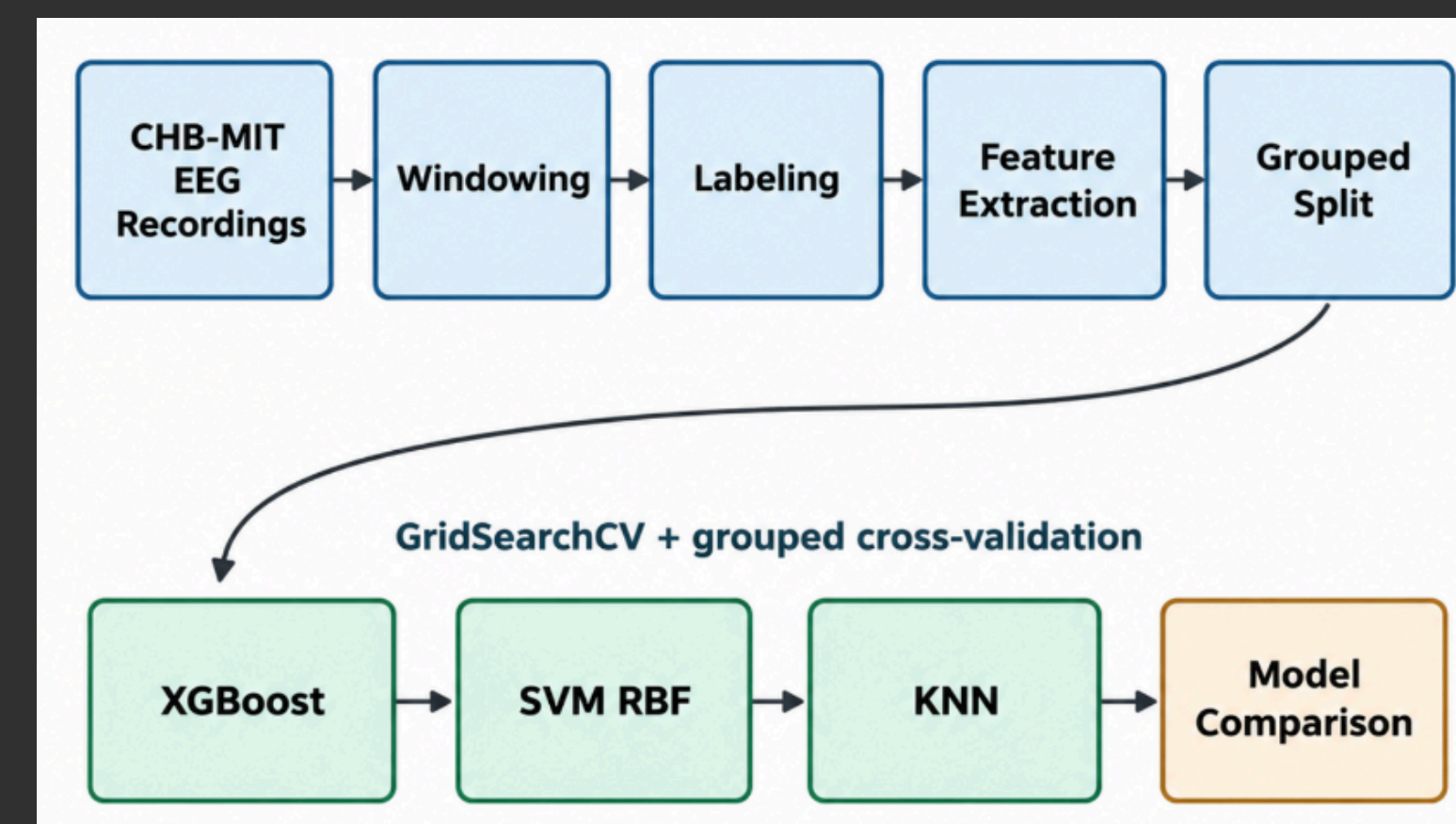
INTRODUCTION

Epilepsy affects 50+ million people worldwide and causes recurrent, unpredictable seizures. 20–33% of patients have drug-resistant epilepsy despite antiseizure medications. Current seizure alert devices detect seizures after they begin, providing little time to prepare.

Therefore, this study evaluates whether machine learning can predict seizures before onset using EEG, supporting the development of more effective seizure monitoring technologies.

METHODS

EEG recordings were divided into 10-second windows. Each window was labeled as preictal (before seizure) or interictal (during seizure). Raw EEG signals were converted into numerical features: statistical, time-domain, frequency-domain. XGBoost, SVM, and KNN were trained on 58,880 raw EEG values and then compared. Final performance was evaluated on 338 EEG windows of held-out patient cases.



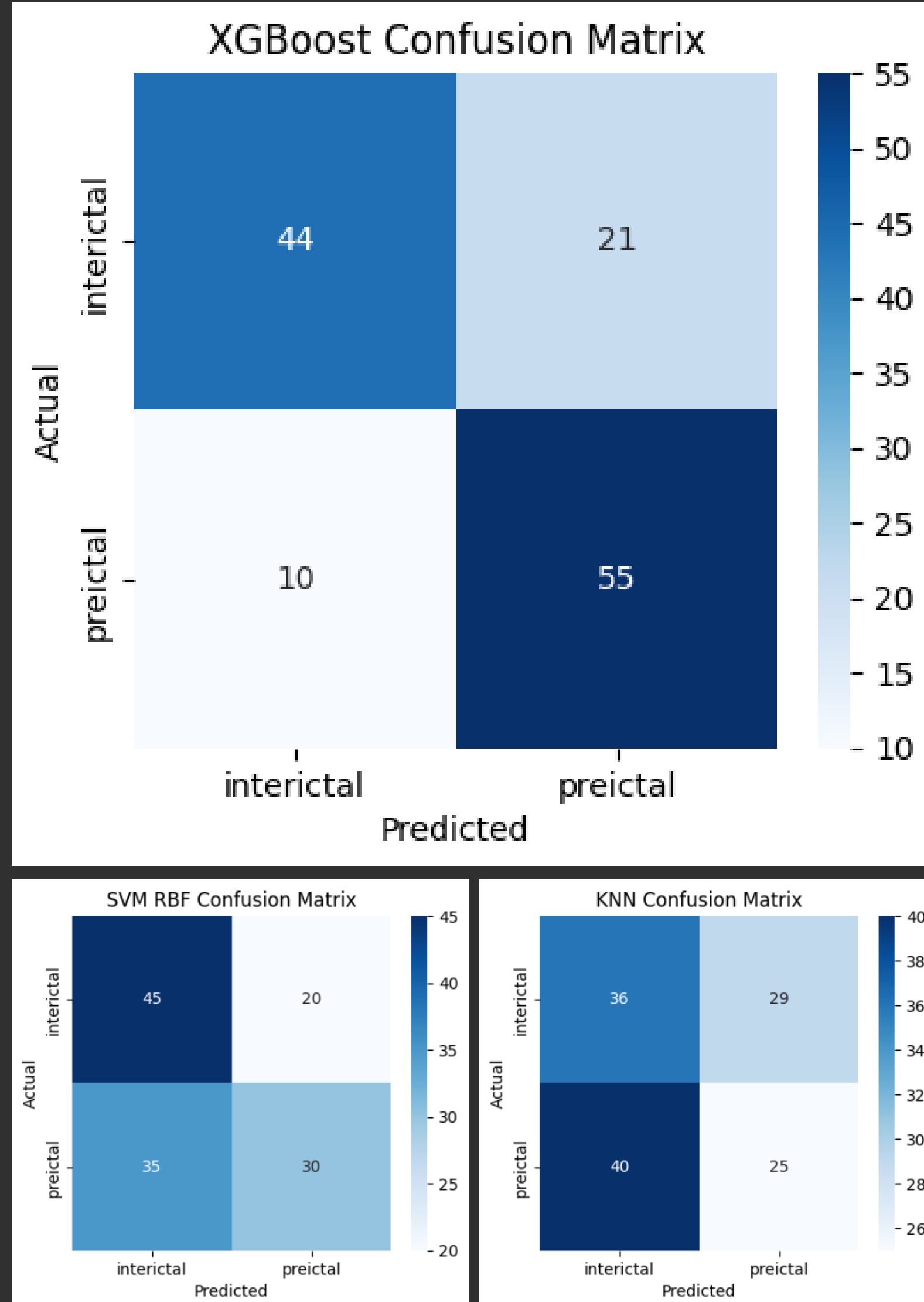
EEG data was grouped so that there was no patient overlap between training and testing, providing a realistic measure of performance on unseen cases.

RESULTS

XGBoost outperformed SVM and KNN in accuracy, precision, recall, F1, and ROC-AUC, thus showing that it is better suited to predicting seizures based on preictal EEG patterns.

Model Performance on Held-Out CHB-MIT Cases					
Model	Accuracy	Precision	Recall	F1	ROC-AUC
XGBoost	0.762	0.724	0.846	0.780	0.852
SVM RBF	0.577	0.600	0.462	0.522	0.604
KNN	0.469	0.463	0.385	0.420	0.475

84.6% recall suggests preictal EEG data contains sufficient detectable patterns that can provide early seizure warnings.



DISCUSSION

XGBoost achieved the strongest performance in accuracy, precision, recall, F1, and ROC-AUC. False negatives are particularly important because they represent missed preictal windows: XGBoost correctly identified 55 of 65 preictal windows, missing only 10, resulting in an 84.6% recall. The relatively high detection of preictal windows reinforces its potential for early seizure prediction and supports the research question that machine learning can distinguish preictal from interictal EEG patterns. However, the model still produced false positives and false negatives, and the dataset was limited to 338 windows from five cases. Therefore, testing with more patients and additional datasets would be necessary to determine whether the model generalizes effectively.

CONCLUSION

This study demonstrated the feasibility of using machine learning for early seizure prediction, with XGBoost achieving the strongest performance. Future research could evaluate the model using larger patient populations and incorporate multimodal physiological signals. Overall, this research could contribute to more accessible and proactive seizure monitoring and management.

REFERENCES

- World Health Organization, "Epilepsy," Feb. 7, 2024. Available: <https://www.who.int/news-room/fact-sheets/detail/epilepsy>.
- Epilepsy Foundation, "Drug Resistant Epilepsy," 2018. Available: <https://www.epilepsy.com/treatment/medicines/drug-resistant-epilepsy>.
- Mayo Clinic, "Epilepsy - Symptoms and Causes," 2026. Available: <https://www.mayoclinic.org/diseases-conditions/epilepsy/symptoms-causes/syc-20350093>.
- J. Guttag, "CHB-MIT Scalp EEG Database," PhysioNet, version 1.0.0, 2010. doi: 10.13026/C2K01R.
- A. H. Shoeb, Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment, Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
- T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
- F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
- 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.
- A. V. Perez-Sanchez et al., "Artificial Intelligence-Based Epileptic Seizure Prediction Strategies: A Review," AI, vol. 6, no. 10, p. 274, Oct. 2025. doi: 10.3390/ai6100274.