Case Study 3: Final Report
DSDA 310: Senior Sem Capstone
8th December, 2024

Case Study 3:
Detecting Epileptic Seizures from the Electroencephalogram (EEG) Data

Report Compiled By:

Aaditya Joshi' 25
Quantitative Economics & Biochemistry Major, Minor in Finance

Satirtha Saha Protya' 25
Biochemistry, Data Analytics, & Cognitive Science Major

I. PROJECT OVERVIEW:

A. BACKGROUND

Epilepsy is a long-term brain disorder affecting about 50 million people globally, marked by recurring seizures—sudden, involuntary movements that can be partial or affect the whole body, often with loss of consciousness and sometimes bladder or bowel control (World Health Organization, 2024). An epileptic individual goes through epileptic seizures which are not in their control. A seizure is a burst of abnormal electrical activity in the brain that can affect consciousness, behavior, memory, or emotions (Huff & Murr, 2023). In this case study we are observing and focusing more on epileptic seizures. Scholarly articles suggest that An "epileptic seizure" refers to a seizure caused by abnormal brain activity, distinguishing it from non-epileptic events like psychogenic seizures (Stafstrom & Carmant, 2015).

In this fast-paced world, technology plays a major role even in the industry of health sciences. Thus, in order to detect these epileptic seizures health professionals prefer to use the technology of Electroencephalogram (EEG). It is a critical diagnostic tool for epilepsy, especially as advanced imaging techniques have made epilepsy one of the few conditions that consistently require EEG evaluation. It addresses key diagnostic questions: confirming epilepsy, identifying the epileptogenic zone, and assessing treatment effectiveness (Engel, 1984).

B. CASE STUDY OVERVIEW

This case study aims to leverage ML and data science computational techniques to detect seizures using EEG data. Seizures, which result from abnormal electrical activity in the brain, can significantly impact individuals' health and quality of life. Timely and accurate seizure detection is essential for effective intervention and improved clinical outcomes. EEG data, a method for recording brain activity, serves as a valuable tool for identifying seizure events.

However, analyzing EEG data presents notable challenges, including the presence of noise, variations in signal quality, and differences in sampling rates across datasets. This project addresses these complexities by developing models capable of classifying EEG signals as seizure or non-seizure, utilizing both clean, controlled datasets and noisy, real-world data to ensure robustness and practical applicability

This project addresses significant challenges in data science, including preprocessing, feature extraction, and model training with diverse and multi-channel EEG datasets. It emphasizes the importance of transitioning from clean, controlled datasets to noisy, real-world data, testing the robustness and adaptability of AI/ML models in practical healthcare applications. By integrating medical data analysis, signal processing, and machine learning, the project offers a comprehensive and interdisciplinary learning experience.

II. DATA EXPLORATION & PREPROCESSING

Two datasets were examined: the full signal dataset, which contained 500 samples of 23-second recordings, and the short signal dataset, created by segmenting each 23-second signal into 1-second windows, resulting in 11,500 samples (500 x 23 seconds). Each dataset was accompanied by a corresponding label file. The labels were refined by removing outdated labels and assigning binary labels (0 for non-seizure and 1 for seizure) to align with the model's requirements.

The preprocessing phase included several critical steps. EEG signals were segmented into 1-second windows to enhance temporal resolution and allow for consistent feature extraction. Normalization techniques were applied to standardize the data, and any inconsistencies or quality issues, such as noise or variability in the signals, were addressed. Observations on the signal quality guided these adjustments to ensure reliable input for the model. Finally, the cleaned datasets were converted from CSV to Excel format to facilitate further analysis. These steps ensured the data was uniform, well-structured, and suitable for effective feature extraction and model training.

III. MODEL & KEY FINDINGS

Initially, we discussed using the Feature Engineering and Random Forest model for this case study however, due to the lack of understanding of EEG and its background, and low accuracy in the outcomes we decided to discontinue with that model. As the outcome of our research and data analysis, we then decided to use the **CNN (convolutional neural network) Model** as this provided us with the most accurate and precise outcome. As Table 2 describes, the CNN model demonstrated exceptional performance in classifying seizure (1) and non-seizure (0) cases, achieving an overall accuracy of 98%. For non-seizure cases, it achieved perfect precision (1.00), high recall (0.98), and an F1-score of 0.99, indicating no false positives and a strong ability to correctly identify non-seizures. For seizure cases, the model attained a precision of 0.91, recall of 0.98, and an F1-score of 0.95, reflecting its capability to identify the majority of seizures with minimal false negatives. The weighted F1-score of 0.98 highlights robust performance even with an imbalanced dataset, making the model a reliable tool for seizure detection in medical applications. The confusion matrix in Figure A for the CNN model further validates its high performance. Out of 460 actual non-seizure cases, 449 were correctly classified, with only 11 false positives. Similarly, for 115 actual seizure cases, 113 were correctly identified, with just 2 false negatives. These results confirm the model's strong recall for both classes (non-seizure: 0.98, seizure: 0.98) and its minimal error rate, making it a reliable tool for accurate seizure detection in medical applications.

Filtration	Chebyshev & 60 Hz
Normalization	Mean & Std from Gaussian fit
# of Conv1D layers	2
Filter in Conv1D	64 & 128 respectively
# of Max pooling 1D layers	2
Pool Size	2
Kernel Size	1
# Dense Layers	2
Activations	ReLu & Sigmoid
Train/Test Split	95%/5%
Epochs	50
Batch Size for training	32

Table 1: CNN Model Parameters

	precision	recall	f1-score	support
0	1.00	0.98	0.99	460
1	0.91	0.98	0.95	115
accuracy			0.98	575
macro avg	0.95	0.98	0.97	575
weighted avg	0.98	0.98	0.98	575

Table 2: CNN Model Outcome

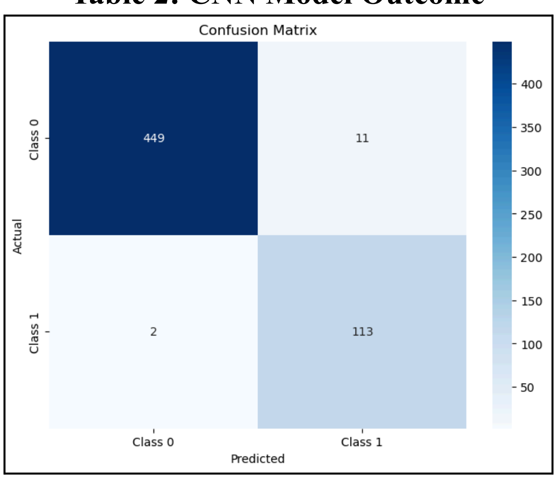

Figure A: Confusion Matrix for Bonn Short Signals

IV. CONCLUSION AND FUTURE RECOMMENDATIONS

Future work should focus on incorporating advanced feature engineering, multi-channel data, and explainability into the model to further enhance its accuracy, robustness, and clinical applicability. Testing on real-world noisy datasets and developing deployment strategies for real-time seizure detection would bridge the gap between research and practical healthcare applications. This interdisciplinary approach highlights the important role of artificial intelligence in advancing the diagnosis and management of epilepsy, ultimately contributing to the development of more personalized and effective treatment strategies.

V. ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Dr. Bergstrom, who generously provided us with the real EEG data and valuable insights that formed the foundation of this study. We also extend our heartfelt thanks for the opportunity to explore the application of artificial intelligence in medical diagnostics, which has greatly enhanced our understanding of interdisciplinary research.

Our deepest appreciation goes to our class professor, Dr.Shende, for their guidance, encouragement, and invaluable support throughout this project. Their expertise and feedback were important in shaping our approach and refining our analysis.

Lastly, we acknowledge the contributions of our peers and the resources made available to us during this study. This collaborative effort has been a truly enriching experience.

BIBLIOGRAPHY

Engel, J. (1984). A Practical Guide for Routine EEG Studies in Epilepsy. *Journal of Clinical Neurophysiology*, *1*(2), 109–142. https://doi.org/10.1097/00004691-198404000-00001.

Huff, J. S., & Murr, N. (2023, February 7). *Seizure*. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK430765/.

Stafstrom, C. E., & Carmant, L. (2015). Seizures and Epilepsy: an Overview for Neuroscientists. *Cold Spring Harbor Perspectives in Medicine*, *5*(6), a022426–a022426. https://doi.org/10.1101/cshperspect.a022426.

World Health Organization. (2024, February 7). *Epilepsy*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/epilepsy.