TRANSFORMER-BASED MODEL FOR SEIZURE IDENTIFICATION USING EEG DATA

Alok Panigrahi 200100019

Amruta Parulekar 20D070009

Annie D'Souza 20D070028

Bhavya Kohli 20D070021

Sameep Chattopadhyay 20D070067

Sanjhi Priya 20D070070

I. Introduction

Transformers [1] are a type of deep-learning model that leverage attention mechanisms for vision, language and time-series data processing tasks. This means that long range data relations help identify seizure patterns across multiple time intervals. Transformers allow for parallel processing, which leads to high efficiency, crucial for real-time applications such as seizure detection. They are adaptable to EEG Data, which provides versatility for complex pattern recognition in spatial and temporal aspects. They can also incorporate transfer learning which means fine-tuning pre-trained models on EEG data that counters less availability of specialized data. They provide interpretability with tools such as attention maps and this leads to clinical trust and acceptance.

II. Related works

Traditional techniques for seizure detection using EEG data include support vector machine, random forest, and logistic regression. Support Vector Machine [2] (SVM) is a supervised machine learning algorithm used for classification and regression tasks by finding the optimal hyperplane that maximizes the margin between data points of different classes. Random Forest [3] is an ensemble machine learning algorithm that combines multiple decision trees to improve predictive accuracy and reduce over-fitting in classification and regression tasks. Logistic Regression [4] is a statistical model used in machine learning for binary classification tasks, where it estimates the probability of an input belonging to a particular class.

Signal processing techniques for seizure detection using EEG data include wavelet transforms, spectral analysis, and time-domain feature extraction. Wavelet transforms [5] are mathematical tools used for signal and image processing, enabling both time and frequency domain analysis by decomposing signals into different scales or resolutions. Spectral analysis [6] is a technique for studying the frequency components of a signal, allowing the identification of patterns and characteristics in time series data, audio signals, or images by examining their frequency domain representations. Time domain feature extraction [7] involves analyzing the raw temporal characteristics of data to extract relevant features, such as mean, variance, or time-domain statistics, for various applications, including signal processing and machine learning.

Finally, deep learning-based methods involve networks such as long short-term memory and gated recurrent units that use feature extraction for temporal data. Long Short-Term Memory (LSTM) [8] is a type of recurrent neural network (RNN) architecture designed to capture and remember long-range dependencies in sequential data, making it well-suited for tasks like natural language processing, speech recognition, and time series analysis. Gated Recurrent Unit (GRU) [9] is a variant of recurrent neural network (RNN) that simplifies the architecture of LSTM by using fewer gates, making it computationally more efficient while still being effective for sequential data modeling

Transformer models leverage attention mechanisms and there have been several published studies that utilize these models for the task of seizure detection. [10] combines CNN and Transformer, achieving high sensitivity and low false rates. [11] uses short-time Fourier transform and an attention-driven three transformer tower significantly improves EEG signal analysis for accurate seizure prediction. [12] combines Fourier transform and deep learning promising efficient automated EEG-based epilepsy screening in clinical settings. [13] uses a a hybrid deep learning model leveraging graph neural networks and transformers. It improves automated epileptic EEG classification by capturing inter-channel relationships and heterogeneous associations.

III. Problem Definition

To develop a transformer-based model for accurate and real-time seizure identification using electroencephalogram (EEG) data, with a focus on improving both sensitivity and specificity to minimize false positives and false negatives, ultimately enhancing the clinical utility of seizure detection systems for individuals with epilepsy.

IV. Algorithms used

A. *Unsupervised Multivariate Time-Series Transformers*

Supervised learning methods require expert labels indicating EEG segments with seizures and obtaining large and consistently-labeled EEG datasets is a difficult task. Additionally, most EEG datasets are severely imbalanced , thus

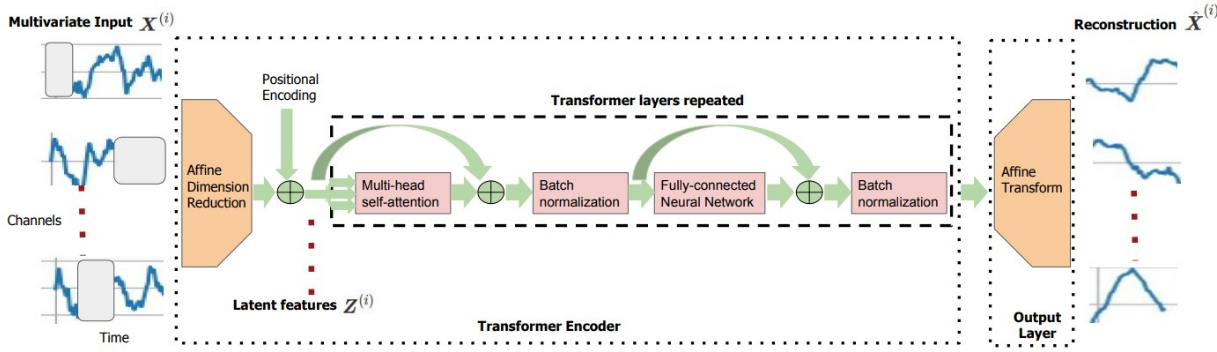

Fig. 1. An auto-encoder consisting of a transformer encoder-decoder pair, designed for multivariate time-series data such as EEG.

causing overfitting. Unsupervised learning is a less explored method for the task of seizure detection. We reformulated the supervised classification to unsupervised anomaly detection by using an auto-encoder.

An auto-encoder [14] consists of a pair of encoder and decoder models, which are jointly trained and optimized. The encoder goes from a high dimensional input space to a low dimensional latent space, while the decoder is vice-versa. Generally, auto-encoders are trained on reconstruction tasks with the objective of reducing the reconstruction error. They have lower reconstruction error for testing when test data is similar to training data. Anomalous data is supposed to have higher error, and thresholding on the normalized absolute error works for anomaly detection. Thus, we used an auto-encoder to create a model that can be trained using only non-seizure data and will be able to predict seizures as anomalies in the test data. Fig. 1 depicts an auto-encoder consisting of a transformer encoder-decoder pair, designed for multivariate time-series data such as EEG.

B. Hybrid Visual Transformer & Data Uncertainity Learning

Electrical activity generated by the brain is miniscule so scalp recorded electrical activity consists of a mix of genuine brain signals combined with lots of noise - termed artifact - generated by other parts of the body, such as heart activity, eye movements and blinks, other facial muscle movements, etc. Data uncertainty learning models each EEG sample as a Gaussian or Laplacian distribution to mitigate noise interference and enhance robustness.

Hybrid visual transformer architecture [15] enhances the processing capability of localized features in the transformer using convolutional neural networks (CNNs). Our model converts raw EEG into a matrix like visual representation. The CNN helps in extracting low-level visual features and patterns, which are then combined with the high-level representations obtained from the transformer. Thus, this fusion leverages the strengths of both approaches: CNNs for spatial feature extraction and Transformers for capturing global dependencies and contextual information. It is particularly useful in seizure detection from EEG data where capturing long-range dependencies is crucial.

V. THE DATASET - CHB-MIT

Collected at the Children's Hospital Boston, the CHB-MIT database [16] consists of EEG recordings from pediatric subjects with intractable seizures. Subjects were monitored after withdrawal of anti-seizure medication to characterize seizures and to assess candidacy for surgical intervention. Hardware limitations caused gaps in files, during which no signals were recorded. In most cases, the gaps are 10 seconds or less. To protect the privacy of the subjects, all protected health information (PHI) in the original files has been replaced with surrogate information. Out of total 23 cases 17 contain exactly one hour of digitized EEG signals, one has two hours data, while five have 4 hours long data. All cases have 23 channel EEG sampled at 256 samples per second with 16-bit resolution. The onsets and ends of 182 seizures are annotated.

Exploratory data analysis was performed on the dataset. Fig. 2 depicts the distribution of records with and without seizures present in the dataset for each of the 23 cases. Fig. 3 shows the segregation of the patients according to their age and gender. The three age groups considered were 0 - 6 years, 7 - 15 years and 16 - 22 years.

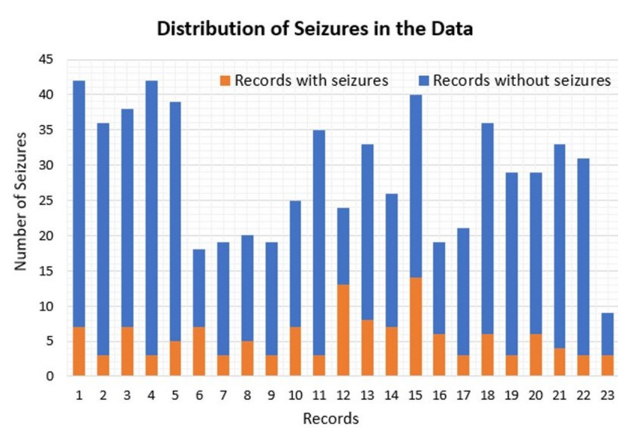

Fig. 2. The distribution of records with and without seizures present in the dataset for each of the 23 cases.

TABLE I
COMPARISON OF THE RESULTS OBTAINED USING THE 4 CHANNEL AND 24 CHANNEL METHODS

	Channels	Accuracy	Precision	Recall	Comments
With data	24	0.5	0.999	0.005	Predicted all signals as non-seizure
Leakage	4	0.941	0.996	0.904	High accuracy with only 4 channels data
Without data	24	0.873	0.961	0.812	Predicted both seizure and non-seizure cases
Leakage	4	0.812	0.965	0.785	Lower performance than 24 channel data

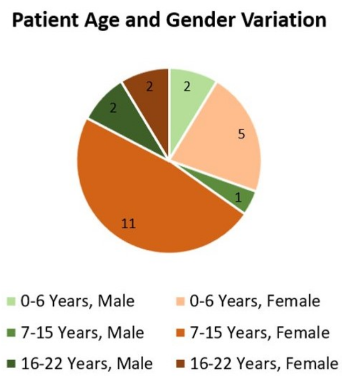

Fig. 3. The segregation of the patients according to their age and gender.

VI. EXPERIMENT 1

A. Hypothesis

The CHB data by itself is larger than 40 GB and its size is further increased by preprocessing. Large amount of training data consumes a lot of time & resources. [17] has a compact transformer which uses raw data from four channels for real-time EEG seizure detection giving low power usage, minimal latency, and low false positives. We hypothesize that we can use 4 out of 23 channels to reduce the amount of training data required while getting an accurate prediction.

B. Data Preprocessing

The training data consisted of non-seizure data while the test data contained both seizure and non-seizure. To prevent data leakage, the train and test data were collected from different subsets of individuals. A band-pass Filter was applied to remove the powerline noise at 60 Hz. Data was normalized to have zero-mean and unit-variance. A sliding window was used to create small segments of EEG data. Partially overlapping windows were used, improving temporal localization, but causing data-size doubling.

C. Model Training

[18] uses an unsupervised model for EEG seizure identification, offering cost-effective and early epilepsy detection without the need for manual labeling or feature extraction. We trained this model on a smaller subset of channels (4 out of 23) F7-T7, T7-P7, F8-T8 and T8-P8 channels chosen due to ease in data collection from these locations. These channels correspond to region behind ears and can be captured by wearable devices like headphones.

D. Results

Table I compares the results obtained using the 4 channel and 24 channel methods. We also observe that in case of training and testing on the same subset of people, overfitting & data leakage occurs, causing low reconstruction error for of which the model performs poorly. Reducing number of channels causes a decrease in performance but useful for experimentation using lesser computation. Using 4 channels, rather than 24 allows us to train with only 17% of data and gives comparable performance.

VII. EXPERIMENT 2

A. Hypothesis

We hypothesize that EEG data of seizure activity is correlated with gender and age. This is hypothesis is based on the studies [19] and [20] that describe differences in EEG activity in different genders and age groups. Thus, if the model is trained on female data, it should perform better on females than males. Additionally, if the model is trained on patients in one age group, it should perform better on patients of that age group as compared to those much older or younger.

B. Data Preprocessing

We preprocessed the data by applying a 60Hz high pass filter to remove instrument noise. We resampled non-seizure records for maintaining class balance and used 1:5 as the cap. 17 out of 23 channels chosen as these were common in all patients had. Sampling rate was fixed to 256 and sample length was 5 seconds, so final data shape was 17x1280. We performed training on all patients together instead of patient-wise to avoid overfitting. We implemented a patient-wise test-train split instead of a record-wise one to avoid data leakage.

C. Model Training

[21] uses a Hybrid Vision Transformer along with data uncertainty learning for seizure detection in EEG data. The patients were divided into three age groups, 0-6 years, 7-15 years and 16-22 years, based on brain development milestones. They were also divided based on gender. The vision transformer model was trained on females in the 7-15-year age group, with one patient held out for testing purposes.Table II depicts the training set for the experiment.

TABLE II
TRAINING SET

Patient	Age	Gender
3	14	Female
5	7	Female
7	14.5	Female
9	10	Female
11	12	Female
14	9	Female
16	7	Female
17	12	Female
22	9	Female

D. Results

The first test performed was on the 11-year-old female and the 11-year-old male. The second test performed was on 2 females each in the lower and upper age group. Table III shows that the model trained on females of the 7-15 age group, performs better on an 11-year-old female as compared to an 11-year-old male.Table IV shows that the model also tests better on the 11-year-old female as compared to females in other age groups. In general, older females give better results than babies which can be attributed to their pre-developmental brains. We conclude from our observations that EEG data is correlated with gender and age.

TABLE III
GENDER-BASED TEST

Patient	Age	Gender	Loss	Accuracy	Precision	Recall
1	11	Female	**0.1805**	**0.9242**	**0.5682**	**0.9615**
2	11	Male	0.2047	0.8971	0.5588	0.76

TABLE IV
AGE-BASED TEST

Patient	Age	Gender	Loss	Accuracy	Precision	Recall
1	11	Female	**0.1805**	**0.9242**	**0.5682**	**0.9615**
18	18	Female	0.3540	0.8439	0.1270	0.6667
19	19	Female	0.1981	0.9456	0.6735	1
6	1.5	Female	0.6091	0.7698	0.0476	0.1
12	2	Female	0.5291	0.8311	0	0

VIII. COMPARATIVE DISCUSSION

We had initially planned on comparing both methods but can't draw a fair comparison because of different amounts of data used to train.

A. The time-series method

- Unsupervised Learning: does not require seizure data to train, can be trained using normal EEG and can identify seizure abnormalities during inference.
- Heavy to train: had to implement channel reduction and reduce number of patients.
- Directly passes time-series data through the architecture.

B. The vision method

- Supervised Learning: requires similar amounts of labelled EEG and non-EEG data – class balancing implemented.
- Light to train: Training time was low but faced problems with overfitting.
- Uses CNNs on matrix representation of time series data.

IX. FUTURE WORK

In the future, we could try out the 4-channel method for different model architectures and analyse the drop in accuracy vs decrease in data requirements. The results of the second experiment can be used for tasks like synthetic data generation by keeping in mind the age and gender of the target group or ensuring even distribution of data across age groups and genders. Additionally, there weren't enough patients to do extensive tests and holding out more patients from the training set severely impacted training performance and led to overfitting. We hope to use synthetic data or new datasets to do more extensive testing in the future.

X. CONCLUSION

We implemented two methods for seizure detection using EEG data, an unsupervised method using a time-series transformer, and a supervised method using a hybrid vision transformer and involving data uncertainty learning. For the time-series method, we implemented an additional channel reduction technique by using 4 instead of 23 channels. For the vision method, we tested for age and gender correlation in the prediction of the model. Table V summarizes some of the problems faced by us and the solutions we implemented.

TABLE V
PROBLEMS FACED AND THEIR RESOLUTION

Problem	Solution
Noise in EEG data	Band-pass filtering
Data leakage	Patient-wise test-train split
Large data size	Specific channel selection
Overfitting	Increasing training data & Class balancing

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.

[2] M. Hearst, S. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[3] G. Louppe, "Understanding random forests: From theory to practice," 2015.

[4] J. Peng, K. Lee, and G. Ingersoll, "An introduction to logistic regression analysis and reporting," *Journal of Educational Research - J EDUC RES*, vol. 96, pp. 3–14, 09 2002.

[5] L. Debnath and J.-P. Antoine, "Wavelet transforms and their applications," *Physics Today - PHYS TODAY*, vol. 56, pp. 68–68, 04 2003.

[6] W. Wei, "Spectral analysis," *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, pp. 601–619, 01 2008.

[7] G. Purushothaman, Y. Mohan, and S. Jinisha, "Time domain feature extraction and classification of eeg data for brain computer interface," 05 2012.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 12 1997.

[9] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 2014.

[10] C. Li, X. Huang, R. Song, R. Qian, X. Liu, and X. Chen, "Eeg-based seizure prediction via transformer guided cnn," *Measurement*, vol. 203, p. 111948, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224122011447

[11] J. Yan, J. Li, H. Xu, Y. Yu, and T. Xu, "Seizure prediction based on transformer using scalp electroencephalogram," *Applied Sciences*, vol. 12, no. 9, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/9/4158

[12] A. Bhattacharya, T. Baweja, and S. P. K. Karri, "Epileptic seizure prediction using deep transformer model," *International Journal of Neural Systems*, vol. 32, no. 02, p. 2150058, 2022, pMID: 34720065. [Online]. Available: https://doi.org/10.1142/S0129065721500581

[13] J. Lian and F. Xu, "Epileptic eeg classification via graph transformer network," *International Journal of Neural Systems*, vol. 33, no. 08, p. 2350042, 2023, pMID: 37382113. [Online]. Available: https://doi.org/10.1142/S0129065723500429

[14] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," 2021.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[16] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. circulation [online]. 101 (23), pp. e215–e220." 2000.

[17] P. Busia, A. Cossettini, T. M. Ingolfsson, S. Benatti, A. Burrello, M. Scherer, M. A. Scrugli, P. Meloni, and L. Benini, "Eegformer: Transformer-based epilepsy detection on raw eeg traces for low-channel-count wearable continuous monitoring devices," in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2022, pp. 640–644.

[18] Potter, G. Zerveas, C. Eickhoff, and D. Duncan, "Unsupervised multivariate time-series transformers for seizure identification on EEG," *21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2022.

[19] M. Corsi-Cabrera, J. Ramos, M. Guevara, C. Arce, and S. Gutierrez, "Gender differences in the eeg during cognitive activity," *The International journal of neuroscience*, vol. 72, pp. 257–64, 11 1993.

[20] J. R. Hughes and J. J. Cayaffa, "The eeg in patients at different ages without organic cerebral disease," *Electroencephalography and Clinical Neurophysiology*, vol. 42, no. 6, pp. 776–784, 1977. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0013469477902310

[21] Z. Deng, C. Li, R. Song, X. Liu, R. Qian, and X. Chen, "Eeg-based seizure prediction via hybrid vision transformer and data uncertainty learning," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106401, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197623005857