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)

EXISTING METHODS FOR SEIZURE DETECTION

Seizure Detection Models:

Traditional:

SVM, Random Forest, and Logistic Regression

Signal Processing Techniques:

Wavelet Transform, Spectral Analysis, and Time-Domain Feature extraction

Deep Learning Based:

LSTM and GRU based feature extraction for temporal data

Transformers:

Leverage attention mechanisms

WHY TRANSFORMERS?

Parallel Processing: Leads to high efficiency, crucial for real-time applications such as seizure detection

Attention Mechanism: Long range data relations help identify seizure patterns across multiple time intervals

Adaptable to EEG Data: Provides versatility for complex pattern recognition in spatial & temporal aspects

Transfer Learning: Finetuning pretrained models on EEG data counters less availability of specialized data

Interpretability: Provided by tools such as attention maps and leads to clinical trust and acceptance

LITERATURE SURVEY

EEG-based seizure prediction via Transformer guided CNN, 2022	• Combines CNN and Transformer, achieving high sensitivity and low false rates.
Transformer-Based Epilepsy Detection on Raw EEG Traces for Low-Channel-Count Wearable Continuous Monitoring Devices, 2022:	• Compact transformer using raw data from four channels for real-time EEG seizure detection: low-power, minimal latency, and low false positives.
Seizure Prediction Based on Transformer Using Scalp Electroencephalogram, 2022	• Uses short-time Fourier transform and an attention-driven three transformer tower significantly improves EEG signal analysis for accurate seizure prediction.
Epileptic Seizure Prediction Using Deep Transformer Model, 2021	• Combines Fourier transform and deep learning promising efficient automated EEG-based epilepsy screening in clinical settings.
Epileptic EEG Classification via Graph Transformer Network, 2023	• A hybrid deep learning model leveraging graph neural networks and transformers improves automated epileptic EEG classification by capturing inter-channel relationships and heterogeneous associations.
Unsupervised Multivariate Time-Series Transformers for Seizure Identification on EEG, 2023	• Unsupervised model for EEG seizure identification, offering cost-effective and early epilepsy detection without the need for manual labeling or feature extraction.

THE DATASET – CHB-MIT SCALP EEG DATABASE

Collected at the Children's Hospital Boston, consists of EEG recordings from pediatric subjects with intractable seizures.

Subjects monitored after withdrawal of anti-seizure medication to characterize seizures, assess candidacy for surgical intervention.

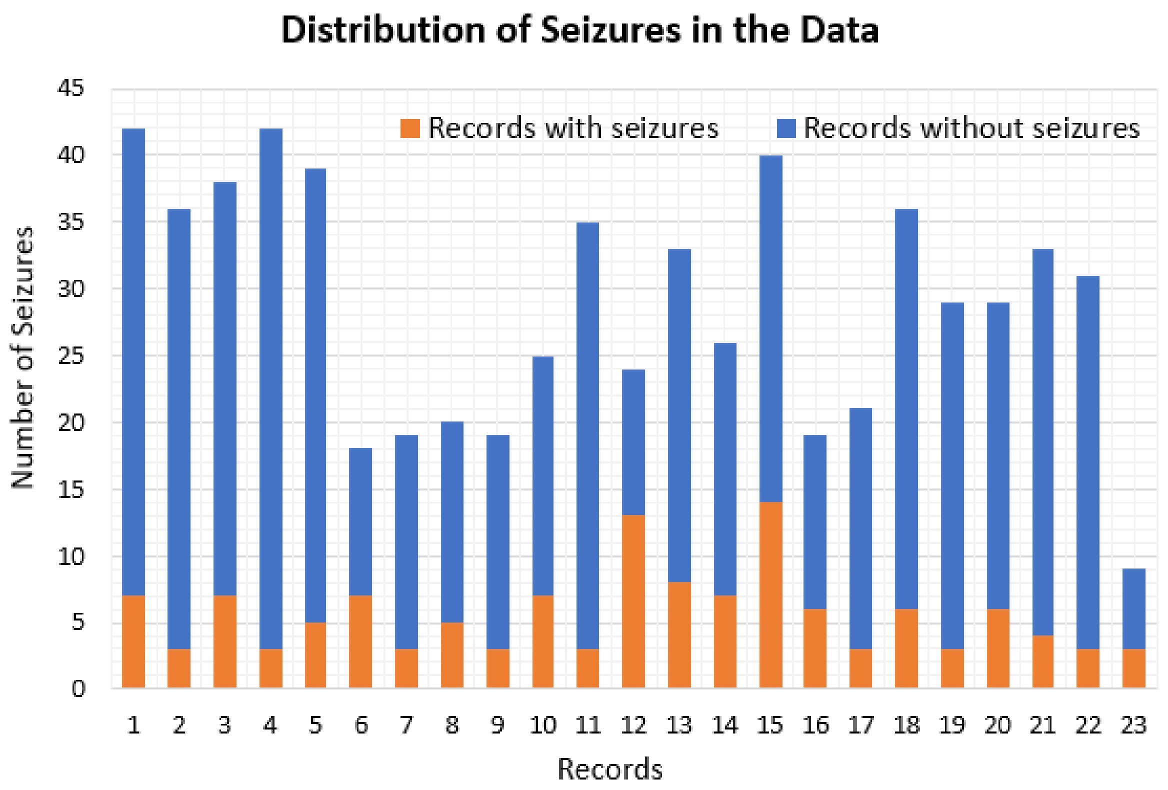
Hardware limitations caused gaps in files, during which no signals were recorded; in most cases, the gaps are 10 seconds or less.

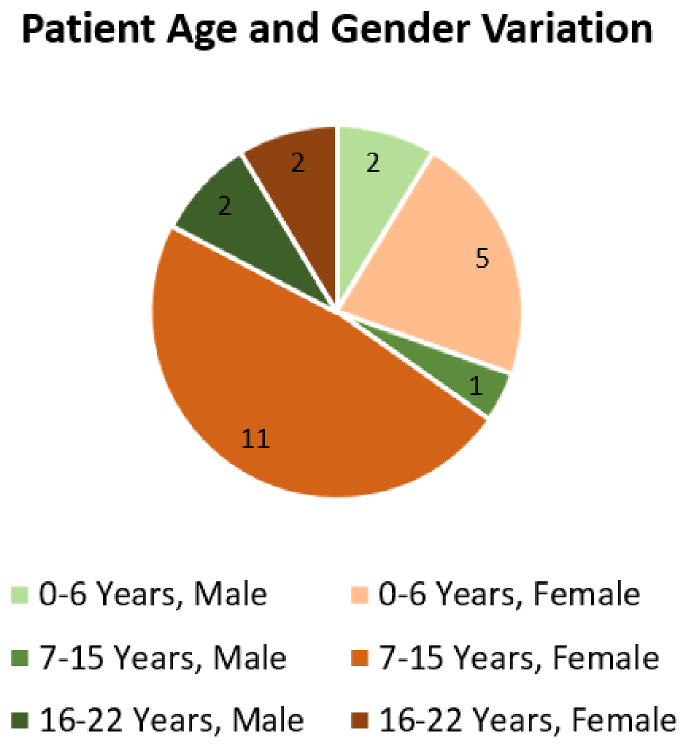
To protect the privacy of the subjects, all protected health information (PHI) in the original files has been replaced with surrogate information.

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 had 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





TWO TRANSFORMER-BASED METHODS

Unsupervised Multivariate Time-Series Transformers

Hybrid Visual Transformer Architecture with Data Uncertainity Learning

MODEL 1: TIME SERIES TRANSFORMER

UNSUPERVISED METHOD

Supervised methods require expert labels indicating EEG segments with seizures

Obtaining large and consistently-labeled EEG datasets is a difficult task

Most EEG datasets are severely imbalanced , thus causing overfitting

Less-explored Method: Reformulate the supervised classification to unsupervised anomaly detection

Anomaly Detection using autoencoder: A model that can be trained using only non-seizure data

MODEL 1: TIME SERIES TRANSFORMER

AUTOENCODER FOR ANOMALY DETECTION

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

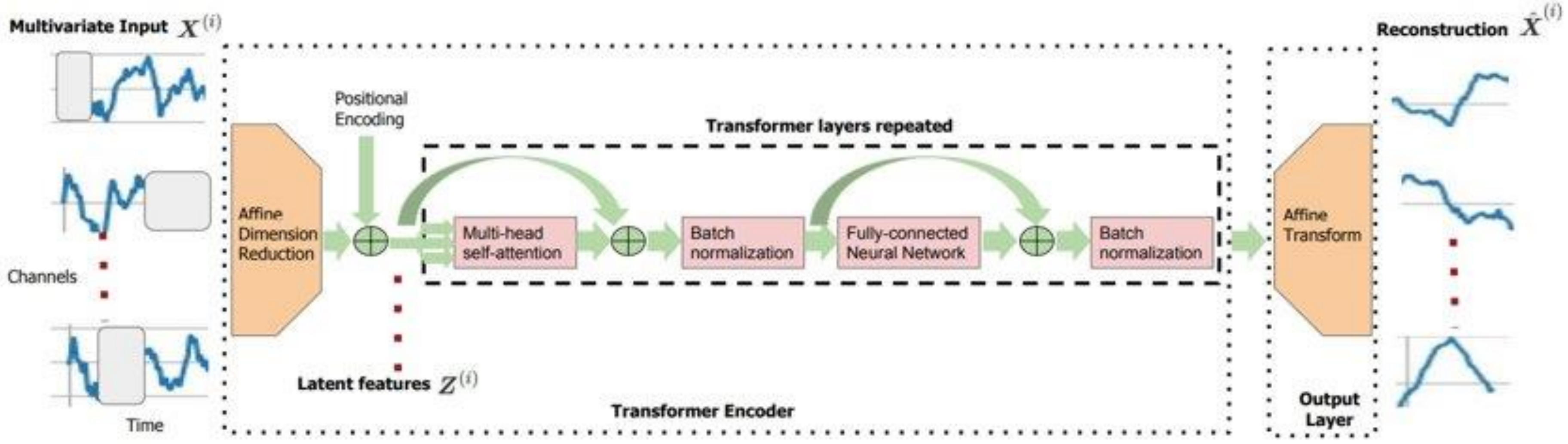
GENERALLY, TRAINED ON RECONSTRUCTION TASKS WITH THE OBJECTIVE OF REDUCING THE RECONSTRUCTION ERROR

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 SHOULD WORK FOR ANOMALY DETECTION

MODEL 1: TIME SERIES TRANSFORMER

TRANSFORMER AUTOENCODER ARCHITECTURE



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

Reference: Potter, İlkay & Zerveas, George & Eickhoff, Carsten & Duncan, Dominique. (2022). Unsupervised Multivariate Time-Series Transformers for Seizure Identification on EEG. 1304-1311. 10.1109/ICMLA55696.2022.00208.

MODEL 1: TIME SERIES TRANSFORMER

DATA PREPARATION

- The training data consists of non-seizure data while the test data contains both seizure and non-seizure
- To prevent data leakage, the train and test data are collected from different subsets of individuals.
- Band-stop Filter: applied to remove the powerline noise at 60 Hz.
- Data is normalized to have zero-mean and unit-variance
- Sliding window: used to create small segments of EEG data
- Partially overlapping windows used, improves temporal localization, but causes data-size doubling

LIGHTER TRAINING

THE EXPERIMENT

THE CHB DATA BY ITSELF IS LARGER THAN 40 GB; SIZE FURTHER INCREASED BY PREPROCESSING.

LARGE AMOUNT OF TRAINING DATA CONSUMES LOT OF TIME & RESOURCES, WHAT IF WE CAN REDUCE THIS?

FOUND LITERATURE ON TRAINING WITH SMALLER SUB SET 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 4 CHANNELS CORRES POND TO REGIONS BEHIND EAR AND CAN EASILY BE CAPTURED BY WEARABLE DEVICES LIKE HEADPHONES.

USING 4 CHANNELS, RATHER THAN 24 ALLOWS US TO TRAIN WITH ONLY 17% OF DATA AND GIVES COMPARABLE PERFORMANCE.

LIGHTER TRAINING

RESULTS, OBSERVATIONS AND CONCLUSIONS

		Accuracy	Precision	Recall	Comments
With Data Leakage	24 channels	0.5	0.999	0.005	Predicted all signals as non-seizure
	4 channels	0.941	0.996	0.904	High accuracy with only 4 channels data
Without Data Leakage	24 channels	0.873	0.961	0.812	Predicted both seizure and non-seizure cases
	4 channels	0.812	0.965	0.775	Lower performance than 24 channel

- 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

MODEL 2: VISION TRANSFORMER

METHOD

DATA UNCERTAINTY LEARNING:

MODELS EACH EEG SAMPLE AS A GAUSSIAN OR LAPLACIAN DISTRIBUTION TO MITIGATE NOISE INTERFERENCE AND ENHANCE ROBUSTNESS.

HYBRID VISUAL TRANSFORMER ARCHITECTURE:

ENHANCES THE PROCESSING CAPABILITY OF LOCALIZED FEATURES IN THE TRANSFORMER USING CNNS

The model converts raw EEG into a matrix like visual representation. The CNN can help in extracting low-level visual features and patterns, which are then combined with the high-level representations obtained from the transformer. DUL is used for noise-robustness.

Reference: Zhiwei Deng, Chang Li, Rencheng Song, Xiang Liu, Ruobing Qian, Xun Chen, EEG-based seizure prediction via hybrid vision transformer and data uncertainty learning, Engineering Applications of Artificial Intelligence, Volume 123, Part C, 2023, 106401, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2023.106401.

MODEL 2: VISION TRANSFORMER

ALTERATIONS MADE IN PREPROCESSING AND TRAINING

1. Preprocessed the data by applying a 60Hz high pass filter to remove instrument noise.
2. Resampled non-seizure records for maintaining class balance and used 1:5 as the cap.
3. 17 out of 23 channels chosen as these were common in all patients had.
4. Sampling rate was fixed to 256, sample length was 5sec, so final data shape was 17x1280.
5. We performed training on all patients together instead of patient-wise to avoid overfitting.
6. We implemented a patient-wise test-train split instead of a record-wise one to avoid data leakage.

AGE AND GENDER-BASED ANALYSIS

HYPOTHESIS

We hypothesize that EEG data of seizure activity is correlated with gender and age.

This is hypothesis is based on the studies 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.

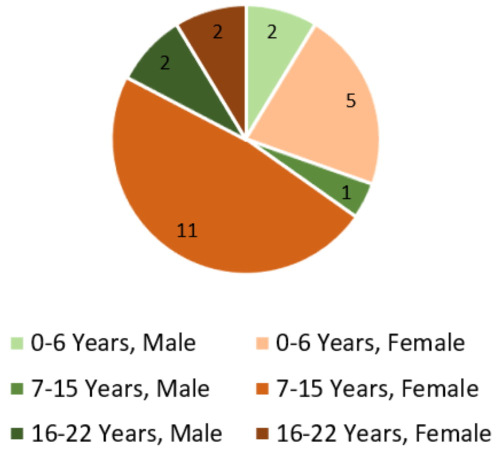
- References:

1. Corsi-Cabrera M, Ramos J, Guevara MA, Arce C, Gutiérrez S. Gender differences in the EEG during cognitive activity. Int J Neurosci. 1993 Oct;72(3-4):257-64. doi: 10.3109/00207459309024114. PMID: 8138380
2. John R. Hughes, Juan J. Cayaffa,The EEG in patients at different ages without organic cerebral disease,Electroencephalography and Clinical Neurophysiology,Volume 42, Issue 6,1977,Pages 776-784,ISSN 0013-4694,doi

AGE AND GENDER-BASED ANALYSIS

THE EXPERIMENT

Patient Age and Gender Variation



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.

Test 1: Test on 11-year-old female and 11-year-old male
Test 2: Test on 2 females each in the lower and upper age group

AGE AND GENDER-BASED ANALYSIS
RESULTS

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

Test 1

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

Test 2

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

AGE AND GENDER-BASED ANALYSIS

CONCLUSION

EEG data is correlated with gender and age.

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.

It 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.

These results 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

There weren't enough patients to do extensive tests and holding out more patients from the training set severely impacts training performance and leads to overfitting.

Reference: Carrle Friedrich Philipp, Hollenbenders Yasmin, Reichenbach Alexandra, "Generation of synthetic EEG data for training algorithms supporting the diagnosis of major depressive disorder ", Frontiers in Neuroscience, Volume 17, 2023 DOI: 10.3389/fnins.2023.1219133

COMPARISON BETWEEN THE TWO METHODS

TIME-SERIES

- **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

VISION

- **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

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

CONCLUSION: OUR CONTRIBUTIONS

TECHNIQUES

TIME SERIES METHOD:

- Channel reduction

VISION METHOD:

- Age and gender analysis

PROBLEMS AND SOLUTIONS

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