Transformer-Based Graph Neural Networks for Alzheimer's and Frontotemporal Dementia Classification Using Filtered EEG Signals

Author: Youssef Mohamed Tageldin

Abstract.

Early detection of neurodegenerative diseases such as Alzheimer's disease (AD) and frontotemporal dementia (FTD) is crucial for timely treatment, yet current diagnostic tools (e.g., MRI, PET) are costly and not widely available. Electroencephalography (EEG) offers a low-cost, non-invasive alternative, but its clinical use is limited by artifacts and signal complexity. In this work, we propose GraphTransformerNet, a transformer-based graph neural network pipeline for classifying AD, FTD, and healthy controls from resting-state EEG recordings. The pipeline first applies a LU-Net–based artifact attenuation model to enhance EEG quality. Features are then extracted in the form of relative band power (across delta, theta, alpha, beta, and gamma bands) and spectral coherence between channel pairs. Each EEG epoch is represented as a graph, with nodes encoding band power features and edges encoding coherence. A multi-head attention–based graph transformer is trained on these graphs to distinguish between diagnostic groups. Using a public OpenNeuro dataset (88 participants: 36 AD, 23 FTD, 29 controls), the model achieved accuracies of 99.1% (AD vs CN), 97.7% (FTD vs CN), and 96.7% (AD vs FTD vs CN) under 10-fold cross-validation. These results highlight the potential of combining EEG with graph learning for scalable, non-invasive dementia screening.

Keywords: EEG, Graph Neural Networks, Alzheimer's Disease, Frontotemporal Dementia, Deep Learning, Brain Connectivity, Transformer

Introduction

Related work

Recent advancements in deep learning have markedly improved diagnostic accuracy for Alzheimer's disease (AD). This section reviews recent studies that have utilized machine learning (ML) techniques to analyze various forms of neuroimaging and electroencephalography (EEG) data. The focus is on how these approaches enhance the diagnosis and prognosis of AD and related neurodegenerative conditions.

[23] conducted a systematic review focusing on predicting the progression from mild cognitive impairment (MCI) to Alzheimer's disease dementia using various neuroimaging data types. The studies reviewed employed magnetic resonance imaging (MRI) and positron emission tomography (PET), with some utilizing magnetoencephalography (MEG). The findings highlighted that combining multiple neuroimaging modalities, such as MRI and PET, generally led to higher classification accuracy, underscoring the value of multimodal approaches in capturing the complex brain changes associated with AD progression.

[24]develops a computer-aided diagnosis system for Alzheimer's disease using EEG data. The methodology involves first filtering the EEG signals with a band-pass elliptic digital filter to remove noise, followed by feature extraction using Discrete Wavelet Transform (DWT). Key features such as logarithmic band power, variance, and kurtosis are extracted and combined into feature vectors.

Nine machine learning algorithms, including SVM, KNN, and Random Forest, were tested to classify these features. The authors reports an accuracy of 99.98 with the KNN model in the problem of AD vs control subjects.

[25]used raw eeg data collected from 141 participants with various cognitive conditions, AD, Mild Cognitive Impairment, and healthy aging subjects. Then the eeg data was pre-processed by using band-pass,FIRand notch filters and sing independent component analysis to filter the data and they used continuous wavelet transform to process the data and they used this data to train a modified AlexNet*citation* architecture that was optimized for 3 class classification the authors reported an accuracy of 98.9%

[26]used EEG data collected from 88 subject with FTD ,AD and CN and then processed the data to extract the relative band power and the spectral coherence and then used this data to train their model **DICE-Net**

Which is a model structured with two parallel blocks one receiving the relative band power and the other receiving the spectral coherence which are then inputted into a multiple layers first a depth wise convolution layer then a positional embedding layer and into a transformer encoding layer and the the outputs of these two branches is concatenated and then passed to a feed forward network

While deep learning has been effective, graph-based machine learning offers a complementary approach that can capture the complex relationships between different brain regions more effectively. EEG data naturally lends itself to graph representations, where nodes represent brain regions or electrodes, and edges represent the interactions between them. Graph-based methods can leverage these relationships to improve the understanding and classification of neurological conditions.

Recent advancements in graph neural networks (GNNs) have been applied to EEG data to exploit these inherent relationships. For example, studies have employed GNNs to model the connectivity between EEG channels, enhancing the ability to detect and classify neurological conditions by incorporating both spatial and temporal dynamics of the data. This shift towards graph-based methods allows for a more nuanced analysis of EEG signals, potentially leading to better diagnostic tools for various neurological disorders.

Among the studies that specifically focus on graph-based methods using eeg data some focus on the classification of epileptic eeg [27] or detecting not healthy eeg signals* EEG-GCNN: Augmenting Electroencephalogram-based Neurological Disease Diagnosis using a Domain-guided Graph Convolutional Neural Network, And [28] classify alzhiemrs deisease by extracting a collection of feature from eeg signal and using each feature they constructa graph and traina gnn model on it the reported an accuracy of 98.4%

Methodology

In this section the proposed methodology and workflow for the filtration and classification of AD and FTD diseases are described and then each part is to be discussed in more detail.

The workflow begins by processing the raw EEG signals then filtering them using a deep learning filtration model then the EEG signals are divided into smaller segments and the features are extracted from these segments, we construct a graph representation of the data. Finally, these graphs are inputted into the Graph-Transformer model, then the model analysis the data and makes predictions on it, then these predications are analysed to measure the performance of the model.

The following Figure 1 provides a visual representation of the entire workflow, from raw EEG data processing to the final Evaluation stage.

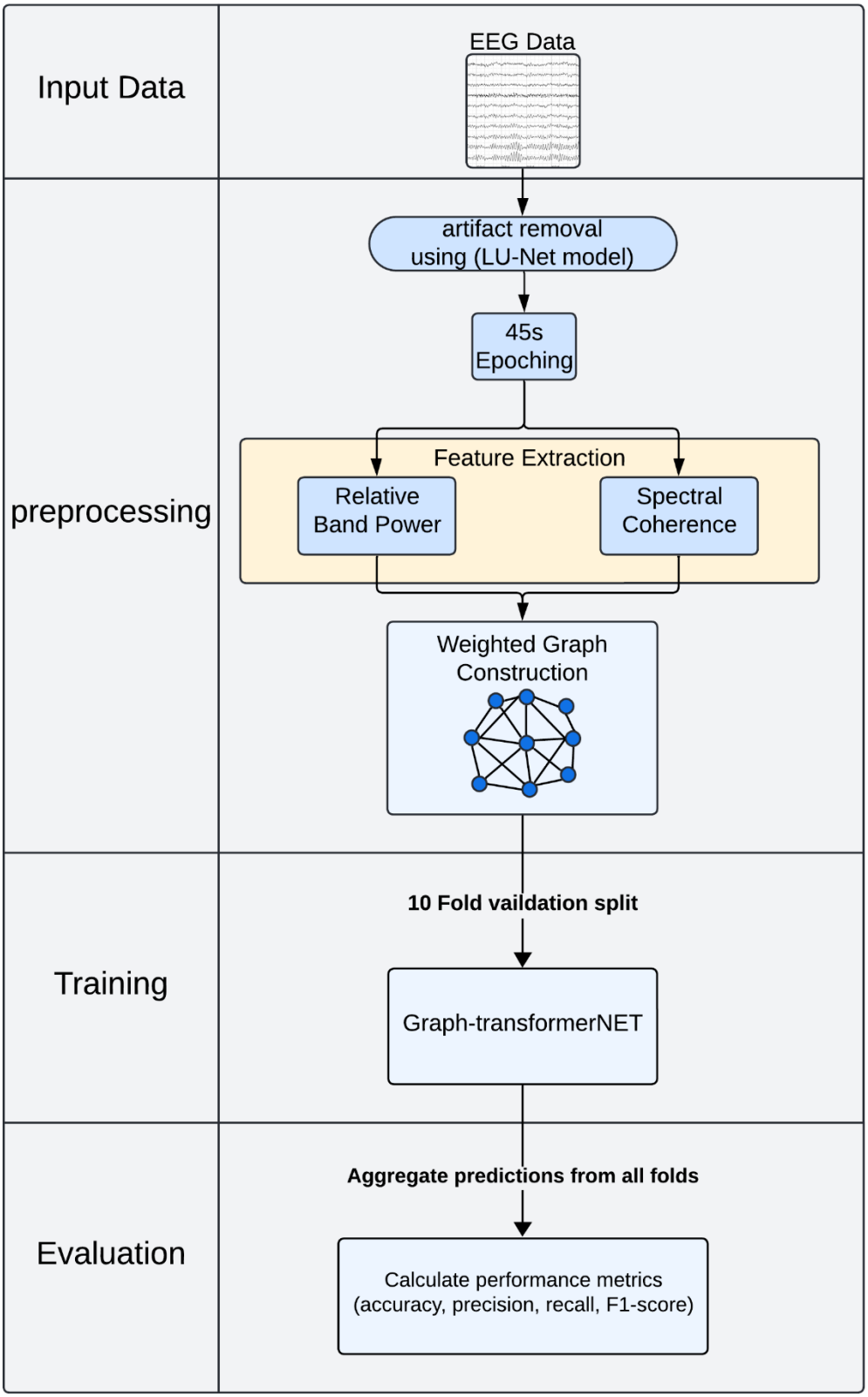

FIGURE 1:WORKFLOW OF THE PROPOSED GRAPH-BASED MACHINE LEARNING MODEL FOR ALZHEIMER'S DETECTION USING EEG DATA.

Data

AD-FTD Classification data

In this work the data from [29] was used which comprises of EEG recordings for 88 subjects divided into three groups: 36 patients diagnosed with Alzheimer's Disease (AD), 23 with frontotemporal dementia (FTD), and 29 cognitively normal controls (CN).

The EEG recordings were obtained during resting-state with eyes closed from all participants. The EEG setup included 19 scalp electrodes positioned according to the 10–20 international system. The sampling rate was set at 500 Hz with a resolution of 10 μV/mm.int the following Table 1 describes the characteristics of the dataset in respect to the groups (AD,CN,FTD)

-

	Gender (F/M)	**Age (mean ± std)**	**MMSE (mean ± std)**	**Total Recording duration (minutes)**	**Mean(range) Recording duration (minutes)**
AD	24/12	66.4±7.9	17.7±4.5	485.5	13.5 (5.1-21.3)
CN	11/18	67.9±5.4	30	276.5	12 (7.9-16.9)
FTD	9/14	63.6±8.2	22.2±2.6	402	13.8 (12.5-16.5)

TABLE 1: CHARACTERISTICS OF THE DATASET

Filtration data

For our filtration model we used the dataset EEGdenoiseNet [30] which was made for training deep learning artifact attenuation models, this dataset contains 4514 clean EEG segments as ground truth, and 3400 pure EOG segments and 5598 pure EMG segments as ocular artifacts and myogenic artifacts respectively the length of each segment was set to 2 sec to insure that the data can be artifact free while maintaining the temporal and spectral characteristics of EEG, as well as EOG and EMG.

This dataset contains both data of both for imaginary and real movement conditions of left and right hand.

Preprocessing

Before the EEG data can be used to train our graph model, it must undergo preprocessing to improve its quality and usability. The first step in this process involves removing artifacts from the EEG data.

To remove artifacts from the EEG data, we developed a custom filtration model based on the LU-Net architecture. This model was trained using the EEGdenoiseNet dataset, which includes clean EEG, EOG, and EMG data. For the training of the artifact attenuation model, simulated noisy data was generated by linearly mixing EEG segments with EOG or EMG segments according to Equation 1:

$$y = x + \lambda . n$$

(1)

Here, y represents the mixed signal of EEG and artifacts (either ocular or myogenic), x denotes the clean EEG segment, n represents the artifacts segment, and λ is a scaling factor that controls the intensity of the artifacts in the mixed signal. By adjusting λ, we can control the signal-to-noise ratio (SNR) of the noisy epoch, as defined in Equation 2:

$$SNR = 10 \log \frac{RMS(x)}{RMS(\lambda . n)}$$

(2)

This artifact attenuation model was used to filter the EEG data collected form patients with AD and FTD, to increase the pool of training sample the EEG records were split into epochs of 45 sec with and then the data was split into 45-second epochs with a 15-second overlap between consecutive segments.

Feature extraction

Following the preprocessing the next step in our workflow is to extract relevant features from theses epochs the following feature were extracted from the data:

Relative band power (RBP)

The RBP is calculated by computing the signal power within a specific frequency as a ratio of the total power of the signal this choice is motivated by the neurophysiological changes associated with AD and FTD.

As in both AD and FTD patients:

- In AD, there is typically an increase in delta (0.5-4 Hz) and theta (4-8 Hz) power, coupled with a decrease in alpha (8-13 Hz) and beta (13-30 Hz) power*refrence*. This "slowing" of EEG rhythms is thought to reflect the loss of cholinergic neurons and the disruption of cortico-cortical connections [31]
- FTD patients often show a different pattern, with less pronounced slowing compared to AD*refrence*. Some studies have reported increased power in higher frequency bands (beta and gamma) in FTD, particularly in frontal regions, which may relate to the characteristic frontal lobe degeneration in this disorder [32]

We get the RBP by first calculating the total power spectral density of the eeg signal we do this by using the Welch's method $P(f)$ which is computed as follows in Eq(3)

$$\mathrm{P(f)} = \frac{1}{\mathrm{K}} \sum_{\mathrm{k=1}}^{\mathrm{K}} |X_k(f)|^2$$

(3)

where:

- $(X_k(f))$is the discrete Fourier transform (DFT) of the $(k) - th$ segment of the signal.

- (k) is the total number of segments.

- Each segment is windowed to reduce spectral leakage.

Then the power in each of the frequency bands which is computed as follows in (4)

$$P_B = \int_{f_1}^{f_2} P(f)\ df \tag{4}$$

where (f_1) and (f_2) are the lower and upper bounds of the band (B).

Then the RBP is determined by normalizing the band power in respect to the total power of the signal as follows Eq(5)

$$RBP_B = \frac{P_B}{\sum_i P_i} \tag{5}$$

With frequency bands being used are:

- **Delta (δ)**: 0.5 - 4 Hz

- **Theta (θ)**: 4 - 8 Hz

- **Alpha (α)**: 8 - 13 Hz

- **Beta (β)**: 13 - 30 Hz

- **Gamma (γ)**: 30 - 100 Hz

and the resulting feature is a matrix with the shape $[N, B]$ where N is the Number of samples and B is the number of bands

Spectral Coherence (SC)

SC is measure that provides information about the functional connectivity between the different brain regions it show the degree of synchronization between pairs of Eeg signals in the frequency domain which offers insights into the disruption in synchronization that is associated with AD or FTD [33]

- AD, spectral coherence typically shows a reduction in synchronization between brain regions, especially in the alpha (8-13 Hz) and beta (13-30 Hz) frequency bands. This decline in coherence reflects the breakdown of long-range cortico-cortical connections, which are critical for coordinated brain activity and cognitive functions such as memory and attention. The loss of synchronization in these frequency bands is associated with the overall disruption of functional networks, contributing to the cognitive decline observed in AD patients. These changes in SC are indicative of the widespread neurodegeneration and network disintegration that characterize AD, where the brain's ability to maintain synchronized activity across different regions is compromised [34].
- FTD presents a different pattern of coherence changes compared to AD. While some studies report decreased coherence in certain regions, FTD may also exhibit localized increases in SC, particularly in the frontal lobes. These increases are often seen in the beta and gamma (30-100 Hz) frequency bands. This phenomenon could be due to compensatory mechanisms where remaining brain circuits attempt to maintain function despite the ongoing neurodegeneration. Alternatively, it may reflect disinhibition or aberrant network activity specific to the frontal lobe, which is more severely affected in FTD. This localized increase in coherence could explain some of the behavioral and cognitive symptoms specific to FTD, such as changes in personality and executive function [34].

And to compute the spectral coherence for a pair of signals we need to first compute the PSD for each of the signals which we do by first computing the the wavelet transform for each of the signals, for that we the morlet wavelet to obtain a time-frequency representation of the signals.

The wavelet transform of a signal $x(t)$ is given by Eq(6):

$$W_x(f,t) = \int_{-\infty}^{\infty} x(\tau)\psi^*(\tau - t, f)\ d \tag{6}$$

where:

- $W_x(f,t)$is the wavelet transform of the signal$x(t)$.

- $\psi^*(\tau - t, f)$ is the complex conjugate of the Morlet wavelet centred at time t and frequency f.

Then for each wavelet transformed signal we compute the PSD using Eq(7)

$$P_x(f) = \frac{1}{T}\sum_t |W_x(f,t)|^2 \tag{7}$$

where T is the total number of time points.

Next, we calculated the cross power spectral density between pairs of signals using Eq(8):

$$P_{xy}(f) = \frac{1}{T}\sum_t W_x(f,t) \cdot W_y^*(f,t) \tag{8}$$

where $W_y^*(f,t)$ is the complex conjugate of the wavelet transform of the second signal.

Then the spectral coherence between the two signal is x and y is calculated as in Eq(9)

$$ \tag{9}$$

where:

- $C_{xy}(f)$ is the spectral coherence at frequency f .

$-\left|P_{xy}(f)\right|^2$ is the magnitude squared of the cross power spectral density.

- $P_x(f)$and $P_y(f)$ are the power spectral densities of signals x and y , respectively.

And finally we average the spectral coherence over the frequency bands (Delta, Theta, Alpha, Beta, Gamma) to get the coherence for these bands ,

The average coherence for a frequency band $B = [f_1, f_2]$ is given by Eq(10):

$$\text{Avg}_B = \frac{1}{f_2 - f_1} \int_{f_1}^{f_2} C_{xy}(f)\ df \quad (10)$$

And the resulting feature matrix has a shape of $[N, C, C, B]$

Where N is the number of samples and C is the Number of channels and B is the number of bands

Graph Construction

After extracting the features we create a graph to model the brains's functional connectivity for AD and FTD diagnosis as the graph helps us capture the relations between different brain regions with the nodes representing the specific features of each channel and the edges indicating the connectivity between each pair of channels, Mathematically, the graph $G = (V, E)$ is defined as:

- V is the set of nodes, where each node v_i is defined by the feature vector

$$v_i = \left[\text{RBP}_\delta, \text{RBP}_\theta, \text{RBP}_\alpha, \text{RBP}_\beta, \text{RBP}_\gamma, \text{Channel Index}\right]$$

where:

-RBP_δ to RBP_γ are the relative band power features for the Delta, Theta, Alpha, Beta, and Gamma frequency bands, respectively.

- Channel Index provides spatial information about the EEG channel.

- E is the set of edges, where each edge e_{ij} between nodes i and j is assigned a weight w_{ij}.

The weight w_{ij} of an edge between nodesv_i and v_j is given by:

$$w_{ij} = C_{ij}(B)$$

where $C_{ij}(B)$ is the spectral coherence averaged over a specific frequency band B .

classification

We now describe the architecture of our proposed model, a Graph Neural Network (GNN) that utilizes Transformer-based convolutions to process and classify graph-structured EEG data. The model begins by taking the graph-structured EEG data as input, which is then passed through a series of Transformer-based convolutional layers. The resulting output is fed into a global mean pooling layer, which averages the embeddings across the entire graph. Finally, this output is passed through a feedforward network that generates the model's predictions. The architecture of our model can be seen in the Figure 2

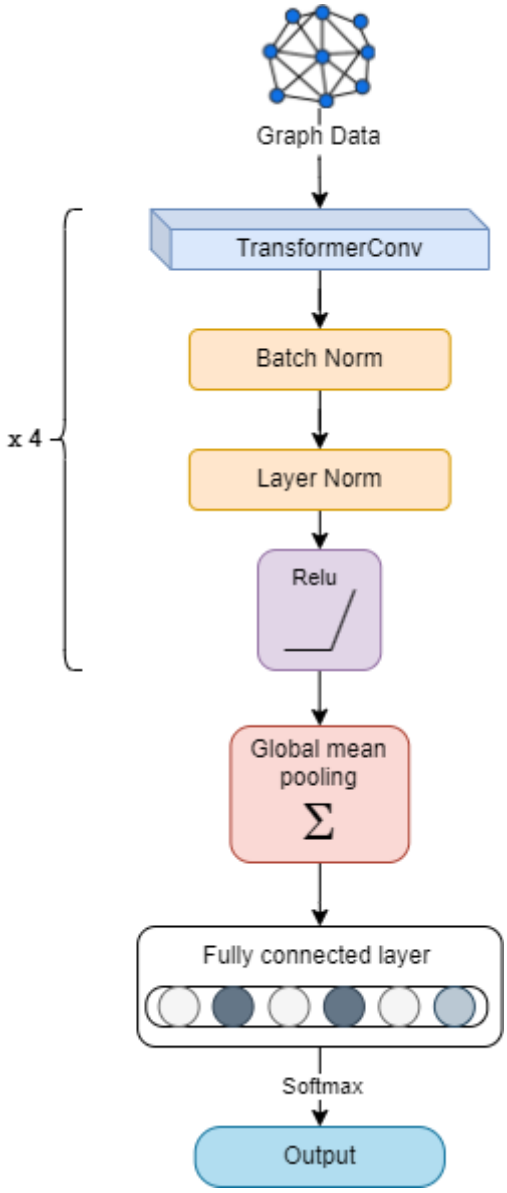

FIGURE 2: ARCHITECTURE OF THE GRAPHTRANSFORMERNET MODEL FOR EEG-BASED AD AND FTD CLASSIFICATION

And to provide a deeper understanding of the model's architecture as we will describe each of the layers into more detail and explaining their role and how they work

Model Architecture Overview

TransformerConv Layers

This model uses four **TransformerConv** layers to process the graph data these layers are based on the works of [35]. Each layer performs the following steps:

.

Multi-Head Attention

- Each node i sends messages to its neighbouring nodes j by computing attention scores $\alpha_{i,j}$, which determine how much focus node i should place on the information from node j.

- The attention score $\alpha_{i,j}$ is computed using the Eq(11):

$$\alpha_{i,j} = \text{softmax}\left(\frac{(W_3 x_i)^\top (W_4 x_j + W_e e_{ij})}{\sqrt{d}}\right) \quad (11)$$

- Here, x_i and x_j are the features of nodes i and j, respectively, e_{ij} is the edge feature, and W_3 , W_4 , and W_e are learned weight matrices. The edge feature e_{ij} is added to the key vector k_j, allowing the model to account for the relationship between the nodes when calculating attention.

Message Passing:

- After computing the attention scores, the model aggregates information from the neighbouring nodes j to update the feature of node i. This aggregation process is described by Eq(12):

$$x_i' = W_1 x_i + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} (W_2 x_j + W_e e_{ij}) \quad (12)$$

- This equation means that node i updates its feature by combining its own transformed feature with a weighted sum of its neighbors' features and the edge features connecting them.

Normalization and Activation:

After message passing, the output is normalized using batch normalization and layer normalization, which stabilize the training process.

The normalized output is then passed through a ReLU activation function and a dropout layer to introduce non-linearity and prevent overfitting.

Global Mean Pooling:

After passing through the TransformerConv layers, the model uses global mean pooling to aggregate the node features into a single graph-level representation shown by the Eq(13). This pooling step compresses the information from all nodes into a fixed-size vector, suitable for classification.

$$\mathbf{h}_{graph} = \frac{1}{N}\sum_{i=1}^{N} h_i \qquad (13)$$

where:

- h_i is the feature vector of node i after the final TransformerConv layer.

- N is the total number of nodes in the graph.

- $\mathbf{h}_{graph}$ is the resulting graph-level feature vector after global mean pooling.

Fully Connected Output Layers:

Finally, the pooled graph representation is passed through a fully connected layer that maps it to the number of classes (e.g., Alzheimer's or non-Alzheimer's). The output is then passed through a log-softmax function to produce the final classification probabilities.

Experimental setup

All of the preprocessing and the model training were run using Kaggle [36]notebooks in python, the notebooks used is equipped with a 2 core CPU and 32 gigs of ram and p100 GPU.

the filtration model was implemented using the TensorFlow [37] library and the processing of the EEG data was done using the MNE and MNE connectivity libraries [38] and the graphs were constructed using networks library and the graph model was implemented using the PyTorch Geometric [39] library

On this setup the time needed to make a prediction using this model is 0.0046 ± 0.0004 sec

Results

In this section, we present the findings of our study on EEG-based classification of Alzheimer's Disease (AD) and Frontotemporal Dementia (FTD). We begin by visualizing the raw EEG data in Figure 3 as gives an overview of the temporal activity captured by the EEG across different channels.

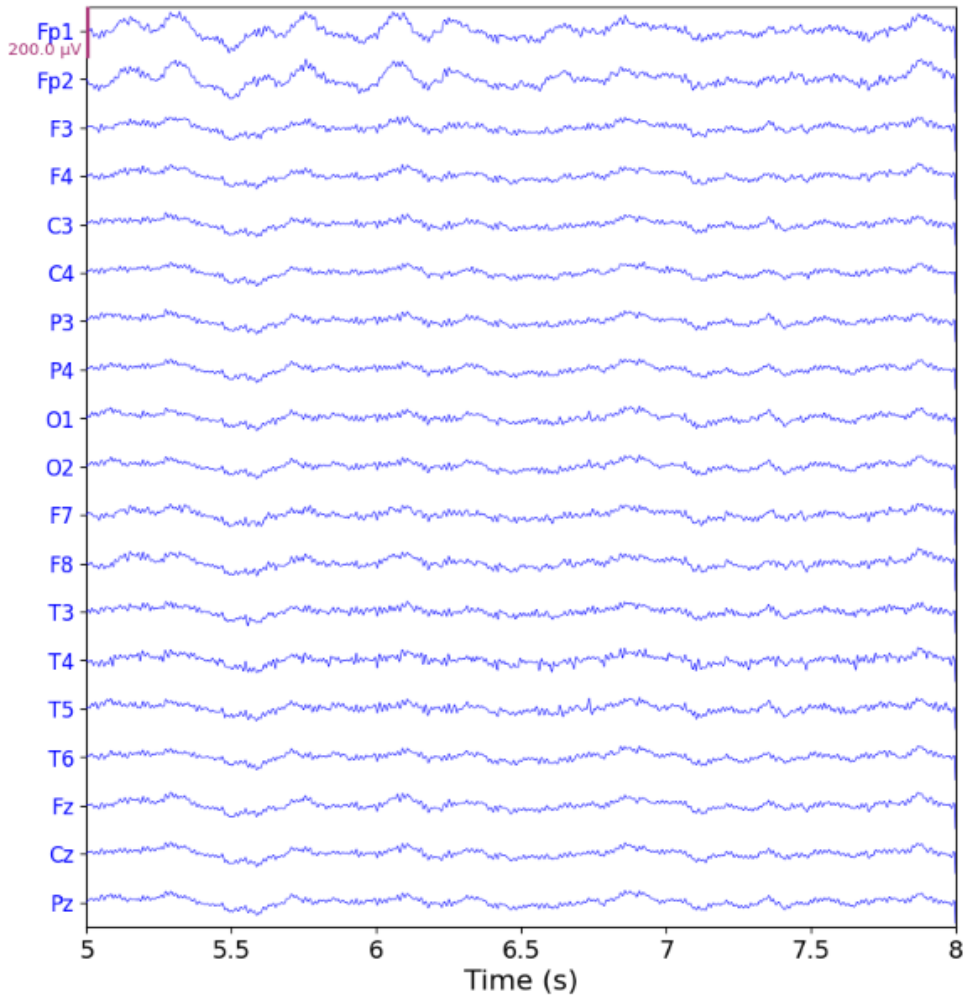

FIGURE 3: SAMPLE OF THE EEG DATASET ,WHERE THE Y-AXIS REPRESENTS THE DIFFERENT EEG CHANNELS AND THE AMPLITUDE OF THE EEG SIGNAL, AND THE X-AXIS REPRESENTS TIME.

Figure 4 provides a comparison of the brain Power Spectral Density (PSD) heatmaps across five frequency bands (Delta, Theta, Alpha, Beta, Gamma) for different groups. The heatmaps illustrate the differences in brain activity across regions, highlighting how the spectral characteristics vary between the different groups, and from the figure it is noticeable that There is an observable difference between these classes (AD, FTD, CN), which can be explained by the fact that each of these conditions is associated with distinct brain activation patterns [40].

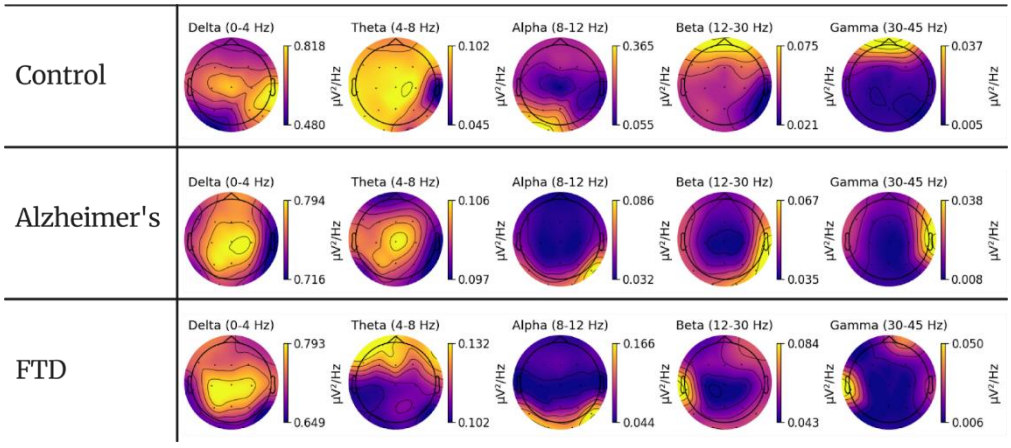

FIGURE 4:COMPARISON OF THE BRAIN PSD HEATMAP ACROSS FREQUENCY BANDS

Figure 5 illustrates a comparison of the Power Spectral Density across EEG channels for different groups. This figure further emphasizes the spectral differences between AD, FTD, and control groups, a significant reduction In the alpha band(8-12 Hz) can be seen in both the AD and FTD groups .

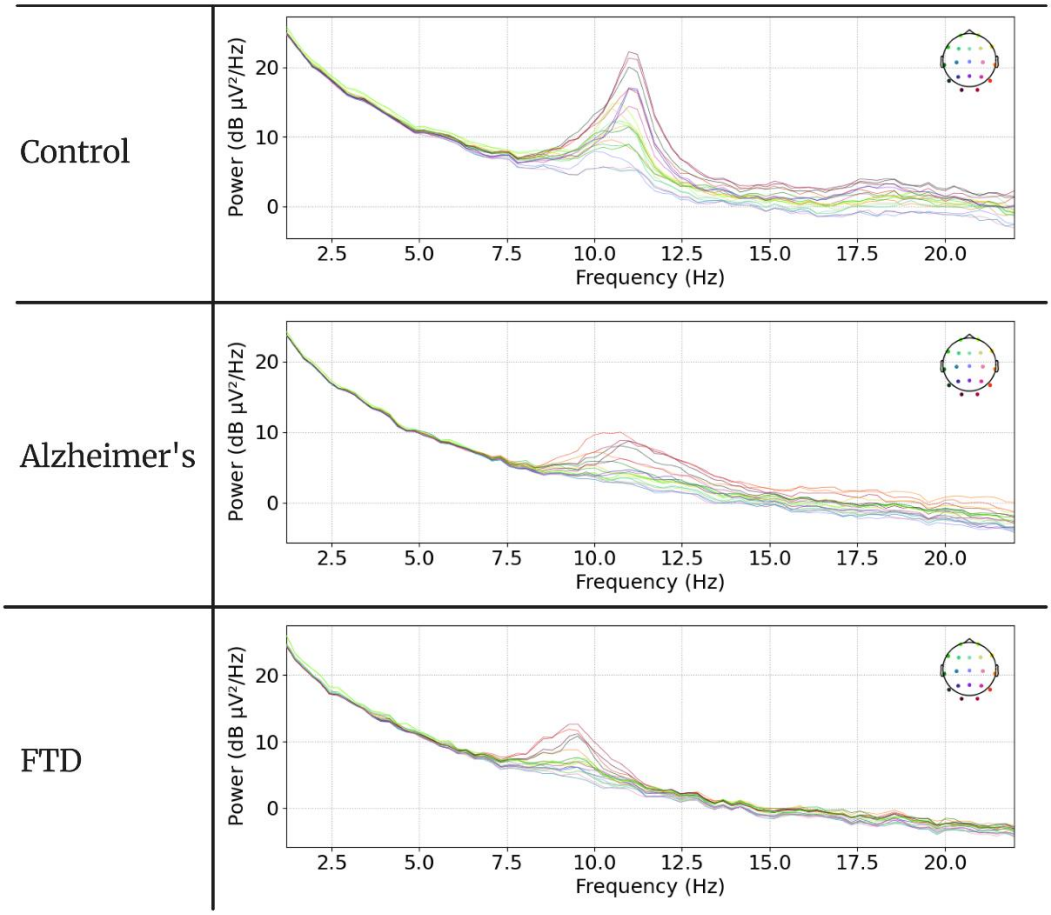

FIGURE 5:COMPARISON OF THE POWER SPECTRAL DENSITY ACROSS EEG CHANNELS

The following Table 2 presents the mean and standard deviation for the extracted features for each power band and group. This statistical summary provides a quantitative measure of the differences in spectral power between AD, FTD, and control groups across the five frequency bands.

Group	Measure	Statistic	Delta	Theta	Alpha	Beta	Gamma
Alzheimer's	SC	mean ± std	0.877 ± 0.134	0.642 ± 0.226	0.514 ± 0.245	0.461 ± 0.225	0.628 ± 0.195
	RBB	mean ± std	0.664 ± 0.028	0.044 ± 0.013	0.020 ± 0.014	0.015 ± 0.009	0.006 ± 0.011
Control	SC	mean ± std	0.933 ± 0.076	0.737 ± 0.174	0.411 ± 0.274	0.388 ± 0.232	0.624 ± 0.186
	RBB	mean ± std	0.652 ± 0.043	0.039 ± 0.014	0.035 ± 0.034	0.017± 0.013	0.005± 0.008
FTD	SC	mean ± std	0.903 ± 0.126	0.710 ± 0.207	0.486 ± 0.261	0.443 ± 0.240	0.613 ± 0.199
	RBB	mean ± std	0.657 ± 0.039	0.045 ± 0.014	0.022 ± 0.020	0.018± 0.018	0.008 ± 0.018

TABLE 2:MEAN AND STANDARD DEVIATION OF EXTRACTED FEATURES

Our GraphTransformerNet model architecture consists of four TransformerConv layers, each followed by batch normalization, layer normalization, ReLU activation, and dropout. The model uses multi-head attention with 10 heads and hidden channels of 128 with the training parameters of the model summarized the in the following Table 3 the dropout rate and the learning rate were chosen using a grid search, and the number of TransformerConv layers , the number of layers for the feed forward network ,the number of hidden channels, the batch size and the weight decay were chosen experimentally by trial and error as they provided the best results

Parameter	Value
Dropout Rate	0.2
Learning Rate	0.0001
Weight decay	0.0001
Batch Size	512

TABLE 3:TRAINING PARAMETERS

A robust validation strategy was employed to ensure the reliability and generalizability of our results. To prevent overfitting, early stopping was utilized with a patience of 60 epochs, allowing the model's training to be monitored. Additionally, a 10-fold cross-validation strategy was implemented to ensure a thorough evaluation of the model. In this approach, the dataset was divided into 10 folds, with the model trained and validated on each fold. The predictions from each fold were then aggregated, and the overall performance metrics were calculated across all folds to provide a comprehensive assessment of the model's effectiveness. To further enhance the robustness of our results and account for the potential variability in model initialization and training, the following Figure 6 shows an example of the training curves for one of the training folds

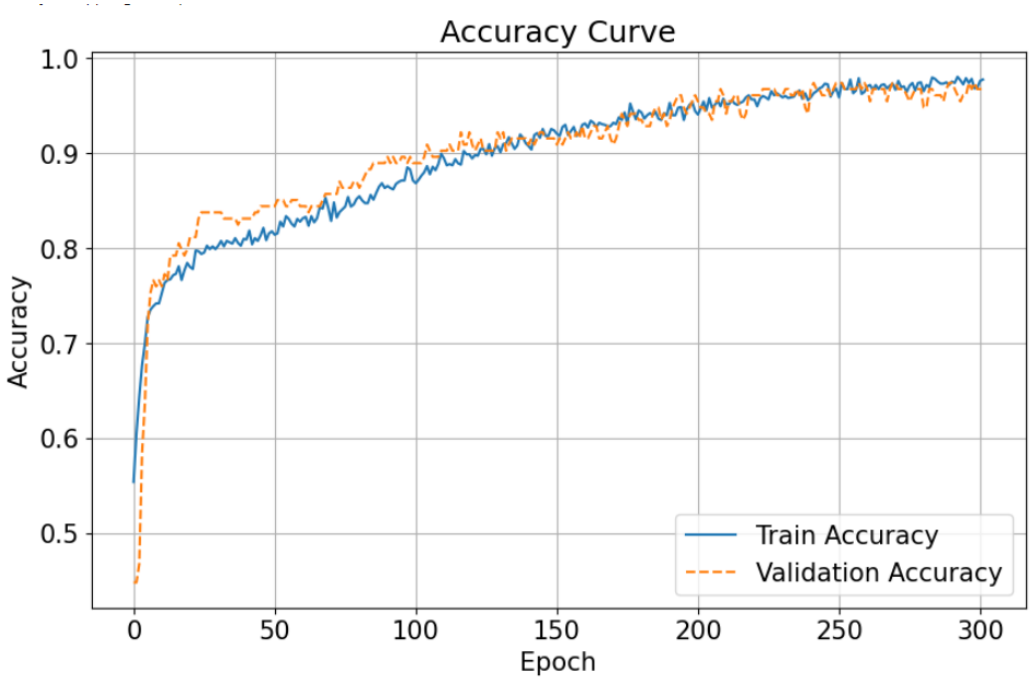

FIGURE 6:TRAINING CURVE FOR FOLD 8 OUT OF 10

The following Table 4 presents the results of the model in the three tested cases: AD-CN, FTD-CN, and AD-FTD-CN. The performance is reported for the cleaned data to showcase the ability of our moder to learn from the data.

Case	Accuracy	Sensitivity	Specificity
AD vs CN	99.1±0.9%	99.1±1.2%	99.1±1.1%
FTD vs CN	97.7±0.9%	97.7±0.8%	97.87±0.8%
AD vs FTD vs CN	96.7±0.8%	96.5±1%	96.6±0.9%

TABLE 4: CLASSIFICATION PERFORMANCE OF THE GRAPHTRANSFORMER MODEL FOR AD, FTD, AND CN DISCRIMINATION USING CLEANED DATA

The model's performance on cleaned EEG data is highly robust, with strong results across all classification tasks. For the AD vs. CN case, the model achieved 99.1% accuracy, sensitivity, and specificity, indicating consistent and balanced performance.

In the FTD vs. CN classification, accuracy was 97.7%, with a similar sensitivity and specificity. For the ternary classification (AD vs. FTD vs. CN), the model maintained a high accuracy of 96.7%, with sensitivity and specificity around 96.5% and 96.6%, respectively. These results highlight the model's effectiveness and ability in distinguishing between AD, FTD, and CN specially in the challenging task of ternary classification.

To further evaluate the effectiveness of the preprocessing steps, we compared the model's performance on both cleaned and uncleaned EEG data in Table 5. The cleaning process, which involved artifact removal using a LU-NET model trained on the EEGdenoiseNet dataset

Case	Raw Data Accuracy	Filtered Data Accuracy
AD vs CN	97.6±1.7%	99.1±0.9%
FTD vs CN	96.7±2.6%	97.7±0.9%
AD vs FTD vs CN	96±1.5%	96.7±0.8%

TABLE 5:CLASSIFICATION PERFORMANCE OF THE GRAPHTRANSFORMERNET MODEL COMPARING BETWEEN THE RAW DATA AND THE FILTERED DATA ACCURACY

The comparison between the raw and filtered data highlights the significant impact of data preprocessing on model performance. For the AD vs. CN case, the accuracy improved from 97.6% with raw data to 99.1% with filtered data, showing a substantial enhancement in classification performance. Similarly, in the FTD vs. CN case, the accuracy increased from 96.7% to 97.7%, and in the AD vs. FTD vs. CN scenario, the accuracy improved from 96% to 96.7%. These improvements emphasize the importance of filtering and artifact removal in boosting the model's accuracy and ensuring more reliable detection of Alzheimer's and frontotemporal dementia.

In Table 6 we compare our work with other studies in the field, our model demonstrates competitive performance. Specifically, the accuracy achieved in the AD-CN and FTD-CN classifications is on par with or exceeds that of other state-of-the-art methods. This comparison highlights the strength of our approach in utilizing both spectral features and graph-based learning.

Study	year	cohorts	stimuli	methodology	Accuracy	sensitivity	specificity
[28]	2022	20 AD 20 CN	Resting state	edge-filtering methods and graph convolutional networks (GCNs)	92%	97.4%	86.7%
[41]	2021	30 AD 35 CN	-	Hjorth parameter and entropy on SVM model	81%	69.8%	83.5%
[26]	2023	36 AD 29 CN	Resting state	RBP, SCC, Dual-input-convolutional encoder	83.28	78.81	87.9
[42]	2019	63 AD 63 CN	Resting state	Mexican Hat CWT and Bispectrum estimation and MLP Model	92.95	-	-
[24]	2022	63 AD 63 CN	Resting state	DWT+LBP,STD,VAR,KUR,AR,RMS and NO	99.98	99.93	99.98
[43]	2023	38AD 17CN	Transcranial magnetic stimulation-TMS	time-domain and statistical EEG features, trained Random Forest classifier	92.05	96.15	87.94
Our Work	2024	36 AD 29 CN	Resting state	RBP,SC,convolutional transformer graph network	99.1	99.1	99.1

TABLE 6: COMPARISON OF OUR WORK WITH OTHER WORKS IN THE FIELD

In comparison to existing studies, our work demonstrates superior performance in classifying Alzheimer's disease (AD) from cognitively normal (CN) individuals using EEG data [28] achieved an accuracy of 92% using edge-filtering methods and graph convolutional networks (GCNs), with sensitivity and specificity of 97.4% and 86.7%, respectively. [41] employed Hjorth parameters and entropy on an SVM model, resulting in a lower accuracy of 81%, with sensitivity at 69.8% and specificity at 83.5%. [26] reported an accuracy of 83.28% with a dual-input convolutional encoder, highlighting the challenges of achieving high sensitivity and specificity concurrently.

[42] achieved a comparable accuracy of 92.95% using Mexican Hat CWT and Bispectrum estimation combined with an MLP model, though sensitivity and specificity were not fully reported [24] achieved almost perfect results with 99.98% accuracy, sensitivity, and specificity, using a combination of DWT+LBP, STD, VAR, KUR, AR, RMS, and NO features, although such high results may be explained by them using a different dataset.

[43] utilized transcranial magnetic stimulation (TMS) data and a Random Forest classifier, achieving an accuracy of 92.05%, with sensitivity and specificity of 96.15% and 87.94%, respectively.

Our work outperforms these studies with an accuracy of 99.1%, maintaining a balanced sensitivity and specificity of 99.1% each. This superior performance is achieved using resting-state EEG data, relative band power (RBP), spectral coherence (SC), and a convolutional transformer graph network, demonstrating the effectiveness of our approach in accurately classifying AD from CN individuals.

Discussion and future work

In this section, we will discuss the results of our work, the limitations, and possibilities for future research. From the start of our workflow, there is considerable room for experimentation. For example, the workflow can be tested with datasets for other neurological disorders to validate its effectiveness and robustness in classifying these conditions. Additionally, it could be extended to tasks such as the classification of emotional states or the classification of heart diseases using ECG data structured as graphs.

Regarding changes to the workflow itself, experimentation with various variables is possible. One area of focus could be the length of segments, as the lengths used in the literature vary significantly, with no consensus on the optimal length for a specific task. There is also room for experimentation with the architecture used in the filtration model. Although we tested multiple architectures, most yielded poor performance, indicating the need for further exploration.

Feature extraction is one of the areas with the most potential for experimentation and improvement. The features used in the literature are highly varied, with numerous possible combinations. In our testing, we explored several common statistical features, such as Shannon entropy, Hjorth parameters, and spectral entropy for channel-wise features, as well as phase locking value, mutual information, and Pearson correlation for pair-wise features. However, we settled on the features we used because including all the aforementioned features did not lead to improved performance and, in some cases, even resulted in lower performance depending on the mix of features. Therefore, a more rigorous and automated methodology for feature selection is needed, rather than relying on trial and error.

Finally, more ablation experiments can be conducted by mixing different types of layers to enhance the model architecture.

References

[1] Z. D. Zhou and A. H. Kihara, "Neurodegenerative Diseases: Molecular Mechanisms and Therapies.," in *International journal of molecular sciences*, Switzerland, 2023.

[2] D. G. Gadhave, V. V. Sugandhi, S. K. Jha, S. N. Nangare, G. Gupta, S. K. Singh, K. Dua, H. Cho, P. M. Hansbro and K. R. Paudel, "Neurodegenerative disorders: Mechanisms of degeneration and therapeutic approaches with their clinical relevance," *Ageing Research Reviews,* vol. 99, p. 102357, 2024.

[3] W. H. O. (WHO), "Dementia WHO Fact Sheet," March 2023. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/dementia.

[4] Alzheimer's Association, "2021 Alzheimer's Disease Facts and Figures," *Alzheimer's & Dementia,* vol. 17, pp. 327-406, 2021.

[5] R. J. Boyd, D. Avramopoulos, L. L. Jantzie and A. S. McCallion, "Neuroinflammation represents a common theme amongst genetic and environmental risk factors for Alzheimer and Parkinson diseases," *Journal of Neuroinflammation,* vol. 19, p. 223, 2022.

[6] R. A. Armstrong, "Risk factors for Alzheimer's disease," *Folia Neuropathologica,* vol. 57, p. 87–105, 2019.

[7] E. Joe and J. M. Ringman, "Cognitive symptoms of Alzheimer\textquoterights disease: clinical management and prevention," *BMJ,* vol. 367, 2019.

[8] K. Jekel, M. Damian, C. Wattmo, L. Hausner, R. Bullock, P. J. Connelly, B. Dubois, M. Eriksdotter, M. Ewers, E. Graessel, M. G. Kramberger, E. Law, P. Mecocci, J. L. Molinuevo, L. Nygård, M. G. M. Olde-Rikkert, J.-M. Orgogozo, F. Pasquier, K. Peres, E. Salmon, S. A. M. Sikkes, T. Sobow, R. Spiegel, M. Tsolaki, B. Winblad and L. Frölich, "Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review," *Alzheimer's Research & Therapy,* vol. 7, p. 17, 2015.

[9] V. Crowell, A. Reyes, S. Q. Zhou, M. Vassilaki, S. Gsteiger and A. Gustavsson, "Disease severity and mortality in Alzheimer's disease: an analysis using the U.S. National Alzheimer's Coordinating Center Uniform Data Set," *BMC Neurology,* vol. 23, p. 302, 2023.

[10] A. J. Walker, S. Meares, P. S. Sachdev and H. Brodaty, "The differentiation of mild frontotemporal dementia from Alzheimer's disease and healthy aging by neuropsychological tests.," *International psychogeriatrics,* vol. 17, no. 1, pp. 57-68, March 2005.

[11] D. Bathgate, J. S. Snowden, A. Varma, A. Blackshaw and D. Neary, "Behaviour in frontotemporal dementia, Alzheimer's disease and vascular dementia," *Acta neurológica scandinavica,* vol. 103, p. 367–378, 2001.

[12] M. A. DeTure and D. W. Dickson, "The neuropathological diagnosis of Alzheimer's disease," *Molecular Neurodegeneration,* vol. 14, p. 32, 2019.

[13] N. L. Foster, J. L. Heidebrink, C. M. Clark, W. J. Jagust, S. E. Arnold, N. R. Barbas, C. S. DeCarli, R. S. Turner, R. A. Koeppe, R. Higdon and S. Minoshima, "FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease.," *Brain : a journal of neurology,* vol. 130, no. Pt 10, pp. 2616-35, October 2007.

[14] D. Pirrone, E. Weitschek, P. Di Paolo, S. De Salvo and M. C. De Cola, "EEG Signal Processing and Supervised Machine Learning to Early Diagnose Alzheimer's Disease," *Applied Sciences,* vol. 12, 2022.

[15] R. Cassani, M. Estarellas, R. San-Martin, F. J. Fraga and T. H. Falk, "Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment.," *Disease markers,* vol. 2018, p. 5174815, 2018.

[16] M. Sazgar and M. G. Young, "EEG Artifacts," in *Absolute Epilepsy and EEG Rotation Review: Essentials for Trainees*, Cham, Springer International Publishing, 2019, p. 149–162.

[17] X. Jiang, G.-B. Bian and Z. Tian, "Removal of Artifacts from EEG Signals: A Review," *Sensors,* vol. 19, 2019.

[18] G. Fiscon, E. Weitschek, A. Cialini, G. Felici, P. Bertolazzi, S. De Salvo, A. Bramanti, P. Bramanti and M. C. De Cola, "Combining EEG signal processing with supervised methods for Alzheimer's patients classification," *BMC Medical Informatics and Decision Making,* vol. 18, p. 35, 2018.

[19] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, "Geometric Deep Learning: Going beyond Euclidean data," *IEEE Signal Processing Magazine,* vol. 34, p. 18–42, July 2017.

[20] G. Vecchiato, A. Susac, S. Margeti, F. De Vico Fallani, A. G. Maglione, S. Supek, M. Planinic and F. Babiloni, "High-Resolution EEG Analysis of Power Spectral Density Maps and Coherence Networks in a Proportional Reasoning Task," *Brain Topography,* vol. 26, p. 303–314, 2013.

[21] K. Raeisi, M. Khazaei, P. Croce, G. Tamburro, S. Comani and F. Zappasodi, "A graph convolutional neural network for the automated detection of seizures in the neonatal EEG," *Computer Methods and Programs in Biomedicine,* vol. 222, p. 106950, 2022.

[22] A. Demir, T. Koike-Akino, Y. Wang, M. Haruna and D. Erdogmus, "EEG-GNN: Graph Neural Networks for Classification of Electroencephalogram (EEG) Signals," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021.

[23] S. Grueso and R. Viejo-Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review.," *Alzheimer's research & therapy,* vol. 13, no. 1, p. 162, September 2021.

[24] K. AlSharabi, Y. Bin Salamah, A. M. Abdurraqeeb, M. Aljalal and F. A. Alturki, "EEG Signal Processing for Alzheimer's Disorders Using Discrete Wavelet Transform and Machine Learning Approaches," *IEEE Access,* vol. 10, pp. 89781-89797, 2022.

[25] C. J. Huggins, J. Escudero, M. A. Parra, B. Scally, R. Anghinah, A. V. L. D. Araújo, L. F. Basile and D. Abasolo, "Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing," *Journal of Neural Engineering,* vol. 18, p. 046087, June 2021.

[26] A. Miltiadous, E. Gionanidis, K. D. Tzimourta, N. Giannakeas and A. T. Tzallas, "DICE-Net: A Novel Convolution-Transformer Architecture for Alzheimer Detection in EEG Signals," *IEEE Access,* vol. 11, pp. 71840-71858, 2023.

[27] J. Zhou, X. Zhang and Z. Jiang, "Recognition of Imbalanced Epileptic EEG Signals by a Graph-Based Extreme Learning Machine," *Wireless Communications and Mobile Computing,* vol. 2021, p. 5871684, 2021.

[28] D. Klepl, F. He, M. Wu, D. J. Blackburn and P. Sarrigiannis, "EEG-Based Graph Neural Network Classification of Alzheimer's Disease: An Empirical Evaluation of Functional Connectivity Methods," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 30, pp. 2651-2660, 2022.

[29] A. Miltiadous, K. D. Tzimourta, T. Afrantou, P. Ioannidis, N. Grigoriadis, D. G. Tsalikakis, P. Angelidis, M. G. Tsipouras, E. Glavas, N. Giannakeas and A. T. Tzallas, *"A dataset of EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects",* OpenNeuro_alzhimers_dataset, 2023.

[30] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li and Q. Liu, "EEGdenoiseNet: a benchmark dataset for deep learning solutions of EEG denoising," *Journal of Neural Engineering,* vol. 18, p. 056057, October 2021.

[31] D. V. Moretti, C. Babiloni, G. Binetti, E. Cassetta, G. D. Forno, F. Ferreric, R. Ferri, B. Lanuzza, C. Miniussi, F. Nobili, G. Rodriguez, S. Salinari and P. M. Rossini, "Individual analysis of EEG frequency and band power in mild Alzheimer's disease," *Clinical Neurophysiology,* vol. 115, pp. 299-308, 2004.

[32] L. E. Hughes, P. J. Nestor, J. R. Hodges and J. B. Rowe, "Magnetoencephalography of frontotemporal dementia: spatiotemporally localized changes during semantic decisions.," *Brain : a journal of neurology,* vol. 134, no. Pt 9, pp. 2513-22, September 2011.

[33] C. J. Stam and E. C. W. van Straaten, "The organization of physiological brain networks," *Clinical Neurophysiology,* vol. 123, pp. 1067-1087, 2012.

[34] M. Pievani, W. de Haan, T. Wu, W. W. Seeley and G. B. Frisoni, "Functional network disruption in the degenerative dementias.," *The Lancet. Neurology,* vol. 10, no. 9, pp. 829-43, September 2011.

[35] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang and Y. Sun, *Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification,* 2021.

[36] Kaggle, *Kaggle Notebooks: A Platform for Data Science and Machine Learning Experiments,* 2024.

[37] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever,

K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,* 2015.

[38] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen and M. S. Hämäläinen, "MEG and EEG Data Analysis with MNE-Python," *Frontiers in Neuroscience,* vol. 7, p. 1–13, 2013.

[39] M. Fey and J. E. Lenssen, "Fast Graph Representation Learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[40] F. G. Metzger, B. Schopp, F. B. Haeussinger, K. Dehnen, M. Synofzik, A. J. Fallgatter and A.-C. Ehlis, "Brain activation in frontotemporal and Alzheimer's dementia: a functional near-infrared spectroscopy study," *Alzheimer's Research & Therapy,* vol. 8, p. 56, 2016.

[41] M. S. Safi and S. M. M. Safi, "Early detection of Alzheimer's disease from EEG signals using Hjorth parameters," *Biomedical Signal Processing and Control,* vol. 65, p. 102338, 2021.

[42] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain and F. C. Morabito, "A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing,* vol. 323, pp. 96-107, 2019.

[43] A.-M. Tăuţan, E. P. Casula, M. C. Pellicciari, I. Borghi, M. Maiella, S. Bonni, M. Minei, M. Assogna, A. Palmisano, C. Smeralda, S. M. Romanella, B. Ionescu, G. Koch and E. Santarnecchi, "TMS-EEG perturbation biomarkers for Alzheimer's disease patients classification," *Scientific Reports,* vol. 13, p. 7667, 2023.

[44] W. Mumtaz, S. Rasheed and A. Irfan, "Review of challenges associated with the EEG artifact removal methods," *Biomedical Signal Processing and Control,* vol. 68, p. 102741, 2021.

[45] N. Pathak, S. K. Vimal, I. Tandon, L. Agrawal, C. Hongyi and S. Bhattacharyya, "Neurodegenerative Disorders of Alzheimer, Parkinsonism, Amyotrophic Lateral Sclerosis and Multiple Sclerosis: An Early Diagnostic Approach for Precision Treatment," *Metabolic Brain Disease,* vol. 37, p. 67–104, 2022.