A New Graph-Transformer Framework for EEG-Based Differentiation of Alzheimer's Disease and Frontotemporal Dementia

Youssef Tageldin, Ahmed S. Eltrass*, and Hania H. Farag
Electrical Engineering Department, Faculty of Engineering, Alexandria University
Alexandria, Egypt
*e-mail: ahmed.eltrass@alexu.edu.eg

***Abstract*— Differentiating between Alzheimer's disease (AD), frontotemporal dementia (FTD), and cognitively normal (CN) subjects remains a significant challenge in clinical neurodiagnosis. This study introduces an automated framework that combines electroencephalography (EEG) signal processing with graph-based deep learning (DL) to improve disease classification. The process begins with artifact suppression and a DL–driven filtering model to enhance EEG signal quality. Once filtered, the signals are segmented, and essential features are extracted to build graph representations that reflect brain connectivity patterns. These graphs are then analyzed utilizing a transformer-based graph neural network, enabling accurate classification of AD, FTD, and CN subjects. Results show that the model achieved highly competitive and well-balanced performance in both binary (AD-CN and FTD-CN) and ternary (AD-CN-FTD) classification tasks, with higher accuracy than existing EEG-based diagnostic methods, demonstrating the benefits of integrating signal filtration, graph representations, and transformer architectures. Overall, the findings suggest that this framework can serve as a reliable tool to support clinical decision-making for the early detection and differentiation of neurodegenerative disorders.**

***Index Terms*— Deep Learning (DL), Convolutional Neural Network (CNN), Graph Neural Network (GNN), Electroencephalography (EEG), Alzheimer's disease (AD), frontotemporal dementia (FTD), and Myocardial Infarction (MI).**

I. Introduction

Neurodegenerative diseases are a group of diseases that cause a progressive deterioration of the nervous system, which occurs due to the degeneration of different neuron parts, mainly in the central nervous system. These diseases affect the brain's functionality and are characterized by the gradual loss of cognitive functions such as memory, language, problem-solving, and other abilities essential for daily life. They can also cause loss of motor function and behavioral changes. These changes not only lead to a significant decline in the quality of life for individuals with neurodegenerative diseases and their caregivers but also contribute to high mortality rates, as current treatments mainly target symptoms rather than modifying the underlying disease progression [1, 2].

Neurodegenerative diseases manifest in several forms, with Alzheimer's disease (AD) and frontotemporal dementia (FTD) being among the most prevalent. Their global impact is substantial. According to the World Health Organization (WHO), more than 55 million people currently live with dementia, with over 60% residing in low- or middle-income countries. Each year, nearly 10 million new cases are reported [3], making dementia the seventh leading cause of death among older adults worldwide. AD is the most common type, accounting for 60–70% of all dementia cases [4], and its prevalence in the United States rose by 145% between 2000 and 2019. Given these alarming trends, this study emphasizes the importance of accurate and early detection of AD and FTD.

The precise causes of AD and FTD remain unclear; however, evidence suggests they result from an interplay of genetic, environmental, and lifestyle factors. Environmental contributors may include diet, physical activity, and exposure to pollutants that promote inflammation, while lifestyle factors such as cardiovascular health and cognitive engagement are also thought to play a critical role. Collectively, these influences may increase susceptibility to disease development and potentially accelerate the aging process [5, 6].

AD begins subtly and progresses through several stages. It usually begins with mild cognitive impairment (MCI), where the patients experience some memory problems that are not major but can affect their quality of life and their instrumental activities of daily living. As the disease progresses to the more moderate and severe stages, patients face increased difficulties in communication, reasoning, and recognizing loved ones, eventually losing the ability to perform basic activities of everyday living. These symptoms can vary significantly between patients [7-9]. In comparison, FTD differs from AD not only in its clinical manifestations but also in the specific brain regions it primarily affects.

FTD includes several subtypes and often begins with subtle changes in personality and behavior, which may also extend to language abilities, leading to difficulties with speech and comprehension. As the disease advances, patients typically experience declines in higher-order cognitive functions such as decision-making, planning, and problem-solving, accompanied by reduced emotional responsiveness and diminished interest in daily activities. Language impairments generally worsen in later stages, and the disease can progress to significant motor symptoms such as muscle weakness and movement difficulties, which may ultimately result in loss of independent speech or mobility. Unlike AD, FTD frequently develops at a younger age, with many diagnoses occurring in individuals in their 60s [10, 11]. These distinctions highlight the critical need for early and accurate diagnosis to support effective management of both AD and FTD.

Current diagnostic methods for AD and FTD are primarily a combination of clinical assessments, neuropsychological

testing, and imaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET). These approaches aim to detect characteristic alterations in brain structure and function [12]. However, they are often costly, time-consuming, and typically identify the disease only at later stages, when substantial brain damage has already taken place. This significantly reduces the effectiveness of available interventions and treatment strategies [13]. Considering these diagnostic challenges, electroencephalography (EEG) emerges as a promising alternative method for early and accurate detection of neurodegenerative diseases. EEG is a non-invasive technique that records the brain's electrical activity by placing multiple electrodes on the scalp and capturing real-time voltage fluctuations from the scalp that result from neuronal activity in the brain [14].

EEG offers several advantages over traditional imaging methods, such as MRI and CT scans. Unlike these structural techniques, EEG provides real-time insights into brain function, allowing the detection of subtle abnormalities that might otherwise go unnoticed. Both AD and FTD are known to produce characteristic changes in EEG activity, making the technique highly valuable for early diagnosis. Another major benefit is its cost-effectiveness and accessibility compared with more advanced imaging modalities. This is particularly important in low- and middle-income countries, where access to MRI or PET scans may be limited. The affordability and portability of EEG can enable earlier diagnosis and intervention, ultimately improving patient outcomes [15].

Despite these advantages, EEG also faces key challenges. Recordings are frequently contaminated by artifacts from sources such as eye blinks, muscle activity, and external electrical noise. These artifacts can mask the underlying neural signals, making it harder to identify subtle pathological changes and potentially lowering diagnostic accuracy [16, 17]. As a result, effective artifact removal is a critical step to ensure EEG data can be reliably applied in clinical diagnostics [18-23].

A further challenge in applying EEG for the diagnosis of neurological disorders lies in the complexity of the data. EEG signals are highly intricate and multidimensional, making them difficult to interpret with traditional analysis methods. This is where machine learning (ML) techniques become especially valuable, as they are well-suited for handling large, complex datasets and detecting patterns that may not be easily recognized by human experts. In the context of neurodegenerative diseases, ML—and particularly deep learning (DL)—models can be trained to identify subtle EEG alterations associated with conditions such as AD or FTD [24], offering a promising path for improved diagnostics. Within this domain, graph neural networks (GNNs) are emerging as a powerful approach because they are designed to analyze data represented as graphs, enabling the modeling of complex relationships and interactions between brain regions [25].

EEG data naturally lends itself to be modeled as graphs, where each node can represent the features extracted from the individual EEG channels. These features might be power spectral density (PSD), relative band power (RBP), or other features that contain information about the channel. The edges in the graph model can be represented by any feature that represents the pairwise information between any pair of channels, reflecting how various areas of the brain interact with each other [26]. By capturing these intricate relations and features, GNN models can learn to identify patterns and relationships in the brain's activity that are indicative of neurological conditions like AD or FTD. This graph-based approach allows for a nuanced understanding of the brain's network structure, leveraging the inherent spatial and functional relationships present in EEG data [27, 28].

The main contribution of this study is the development of an accurate and fully automated framework for distinguishing AD, FTD, and healthy controls. The proposed approach integrates three key components: an artifact removal and filtering model to improve EEG signal quality, the construction of graph-based representations to capture brain connectivity, and a transformer-based graph model for effective analysis and classification of these conditions. The structure of the paper is as follows. Section 2 provides a review of related work on the use of EEG and machine learning for diagnosing AD and FTD, outlining the advantages and limitations of existing methods. Section 3 describes the proposed methodology in detail, covering EEG pre-processing, artifact removal, feature extraction, and graph network modelling. Section 4 reports the experimental results and benchmarks them against findings from previous studies. Finally, Section 5 discusses the implications of the results, explores potential clinical applications, and identifies promising directions for future research.

II. RELATED WORK

Recent progress in DL has significantly enhanced diagnostic accuracy for AD. This section reviews key studies that applied ML methods to neuroimaging and EEG data, emphasizing how these techniques improve both diagnosis and prognosis in AD and related neurodegenerative disorders. In [29], a systematic review examined approaches for predicting the progression from MCI to Alzheimer’s dementia using multiple neuroimaging modalities, including MRI, PET, and, in some cases, magnetoencephalography (MEG). The review found that combining modalities, particularly MRI and PET, generally produced higher classification accuracy, underscoring the value of multimodal strategies in capturing the complex brain changes linked to AD progression.

In [30], researchers developed a computer-aided diagnosis system for AD using EEG data. After filtering the EEG signals with a band-pass elliptic digital filter to reduce noise, features were extracted with a Discrete Wavelet Transform (DWT). Metrics such as logarithmic band power, variance, and kurtosis were then classified using nine ML algorithms, including Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Random Forest. The KNN model achieved the best performance, reaching an accuracy of 99.98% in distinguishing AD patients from healthy controls.

In [31], raw EEG data from 141 participants, including AD, MCI, and cognitively normal (CN) subjects, was preprocessed using band-pass, FIR, and notch filters, along with independent component analysis for artifact removal. The processed signals were transformed using a continuous wavelet transform and used to train a modified AlexNet model optimized for three-class classification. This method achieved an accuracy of 98.9%. In [32], EEG data from 88 subjects across AD, FTD, and CN groups was analyzed. Relative band power and spectral coherence features were extracted and used to train a proposed DICE-Net model, which achieved an accuracy of 83.28%.

While DL has proven effective in analyzing neuroimaging and EEG data for AD diagnosis and prognosis, graph-based ML provides a complementary avenue by modeling complex relationships between brain regions. EEG data can be naturally represented as graphs, where nodes correspond to electrodes or brain regions, and edges capture their interactions. These graph structures enable richer representations of neural connectivity, which can be leveraged to improve classification and understanding of neurological disorders.

Recent advancements in GNNs highlight their potential in this field. GNN-based approaches applied to EEG data have been shown to capture both spatial and temporal dynamics of brain activity, thereby improving the detection and classification of neurodegenerative conditions. This shift toward graph-based methods allows for more nuanced analysis of EEG signals and holds promise for developing more precise diagnostic tools [33, 34].

III. Methodology

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

The workflow begins, as shown in Fig. 1, by preprocessing the raw EEG signals and then filtering them using a DL filtration model. Next, the EEG signals are divided into smaller segments, and the features are extracted from these segments. Following that, a graph representation of the data is constructed, where EEG channels are represented as nodes and functional dependencies as edges. Finally, the constructed graphs are input into the graph transformer model, which processes the data and generates classification outputs. These predictions are subsequently evaluated to assess the model's performance. Fig. 1 illustrates the end-to-end workflow of the proposed framework. This pipeline preserves both temporal dynamics and inter-channel connectivity features, which are crucial for detecting subtle neural alterations associated with AD and FTD.

A. EEG Data

In this study, a publicly available EEG dataset described in [35] is utilized. The dataset includes recordings from 88 participants, divided into three groups: 36 individuals diagnosed with AD, 23 with FTD, and 29 CN.

All EEG recordings were collected during a resting state with participants' eyes closed to maintain consistency and reduce artifacts caused by visual input. Data acquisition was carried out using 19 scalp electrodes positioned according to the international 10–20 system, which is a well-established standard for EEG research. The signals were sampled at 500 Hz with a resolution of 10 μV/mm, ensuring precise time–frequency analysis and reliable feature extraction.

The dataset is organized to clearly distinguish between the three groups, making it suitable for comparative analyses of AD, FTD, and CN subjects. A detailed summary of the demographic and structural characteristics of the groups is provided in Table 1.

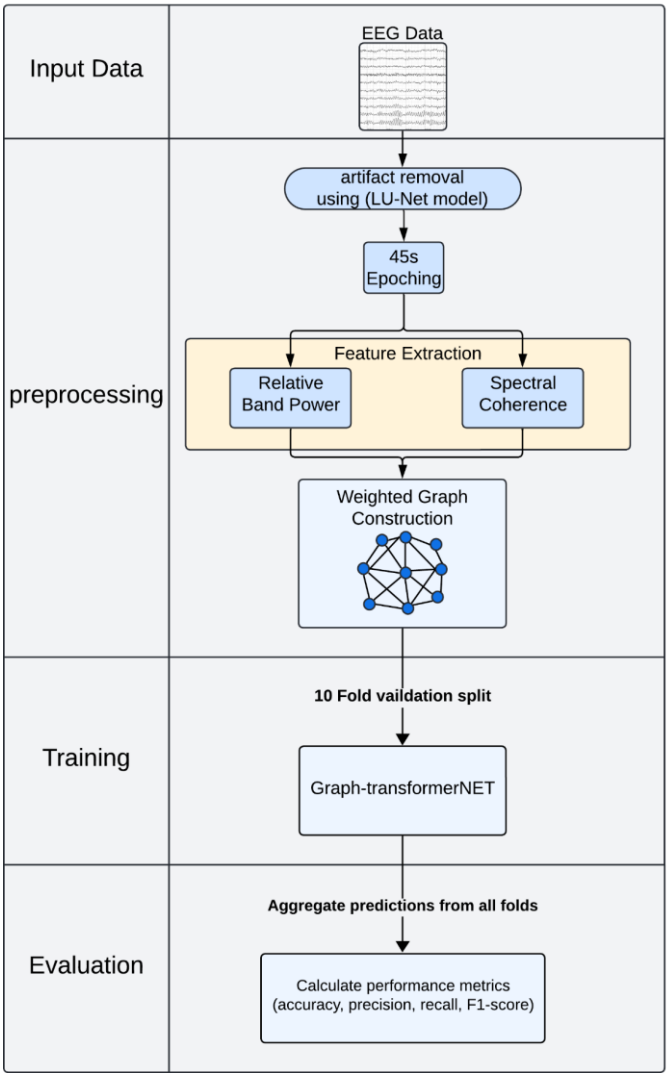

Fig. 1. Workflow of the proposed graph-based framework for AD and FTD classification utilizing EEG signals.

Table 1 summarizes the dataset composition. The AD group (24 female/12 male, mean age 66.4 years) shows a markedly reduced Mini Mental State Examination (MMSE) score (17.7 ± 4.5), reflecting significant cognitive impairment. In contrast,

CN subjects maintain normal function (MMSE = 30), while the FTD group demonstrates intermediate impairment (22.2 ± 2.6). Recording duration per subject was consistent across groups (~12–14 minutes), ensuring comparability. These features confirm that the dataset is representative, clinically validated, and well-suited for training robust classification models. Including both AD and FTD cases further strengthens the study by addressing the clinically challenging task of differential diagnosis.

Table 1. Count of beats extracted for normal and various types of MI across all leads in the PTBdb dataset.

Class	Gender (Female/Male)	Age (mean ± std)	MMSE (mean ± std)	Total recording duration (minutes)	Recording duration (minutes), Mean (Range)
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)

B. EEG Pre-processing

For the development of the denoising model, the EEGdenoiseNet dataset [36] was employed, specifically designed for training DL models to attenuate EEG artifacts. This dataset serves as a comprehensive benchmark and consists of 4514 clean EEG segments used as ground truth, 3400 pure electrooculography (EOG) segments representing ocular artifacts, and 5598 pure electromyography (EMG) segments representing myogenic artifacts. Each segment is 2 seconds in length, a duration chosen to preserve both temporal and spectral features while ensuring sufficient quality for artifact removal training. The recordings were collected from 52 participants under both imagined and real hand movement conditions (left and right), thereby enhancing variability and robustness.

Before EEG data from AD and FTD patients could be integrated into the graph-based model, preprocessing was required to improve data quality and usability. The most critical step in this process was artifact removal. A custom filtration model based on the LU-Net architecture was developed and trained using the EEGdenoiseNet dataset.

To create realistic noisy training conditions, clean EEG segments were linearly mixed with either EOG or EMG artifact segments, following Equation (1):

$$y = x + \lambda\, n \tag{1}$$

Where y represents the contaminated signal (EEG + artifacts), x denotes the clean EEG segment, n is the artifact segment (ocular or myogenic), and λ is a scaling factor controlling artifact intensity.

The signal-to-noise ratio (SNR) of each noisy segment was then defined by Equation (2):

$$SNR = 10 \log \frac{RMS(x)}{RMS(\lambda n)} \tag{2}$$

By adjusting the scaling factor λ a noisy dataset with varying SNR levels was generated, which makes this dataset well-suited for training robust artifact attenuation models.

Finally, to expand the effective training set, EEG recordings were segmented into 45-second epochs with a 15-second overlap between consecutive segments. This strategy increased the number of samples, improved model generalization, and preserved signal continuity.

C. Feature Extraction

Following preprocessing, the next step involved extracting relevant features from the EEG epochs. Relative band power (RBP) and spectral coherence (SC) were selected as the primary features for further analysis.

1) Relative Band Power (RBP)

RBP is computed by calculating the signal power within a defined frequency band and expressing it as a ratio to the total signal power. This normalized measure of spectral activity helps reduce inter-subject variability and facilitates reliable group-level comparisons. The use of RBP as a feature is supported by extensive evidence of neurophysiological alterations in dementia-related disorders, particularly AD and FTD [37, 38].

Patients with AD typically exhibit a shift toward lower-frequency activity. This is reflected in increased delta (0.5–4 Hz) and theta (4–8 Hz) power, alongside reduced alpha (8–13 Hz) and beta (13–30 Hz) power. Such slowing of EEG rhythms

is widely interpreted as an indicator of impaired cortico-cortical connectivity and degeneration of cholinergic neurons, both of which are key features of AD pathology [37].

In contrast, EEG slowing is less pronounced in FTD. Several studies have reported elevated power in higher-frequency bands, particularly beta and gamma, with frontal predominance. This spectral profile aligns with the characteristic frontal lobe degeneration observed in FTD [38].

Therefore, analyzing RBP across canonical frequency bands provides a physiologically grounded feature set that supports differentiation between AD, FTD, and CN subjects.

RBP is derived by first estimating the power spectral density (PSD) of the EEG signal using Welch's method. The PSD $P(f)$ is defined as:

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

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 (e.g., with a Hamming window) to minimize spectral leakage. The power within a frequency band $B = [f_1, f_2]$ is calculated as:

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

where f_1 and f_2 are the lower and upper bounds of the band B, respectively.

The RBP is obtained by normalizing the band power with respect to the total power across all bands:

$$RBP_B = \frac{P_B}{\sum_i P_i} \quad (5)$$

The canonical frequency bands considered are: Delta (δ): 0.5 – 4 Hz, Theta (θ): 4 – 8 Hz, Alpha (α): 8 – 13 Hz, Beta (β): 13 – 30 Hz, and Gamma (γ): 30 – 100 Hz.

The resulting RBP features form a matrix of size $N \times B$, where N is the number of samples and B is the number of frequency bands.

From a neurophysiological perspective, this feature is motivated by evidence that different dementia subtypes exhibit distinct spectral patterns: AD: Increased delta and theta power, accompanied by reduced alpha and beta power. This slowing of EEG rhythms is associated with cholinergic deficits and cortico-cortical disconnection [37], FTD: A less pronounced slowing effect, with some studies reporting increased beta and gamma activity in frontal regions, consistent with characteristic frontal lobe degeneration [38].

Thus, RBP provides a physiologically grounded feature set for distinguishing AD, FTD, and CN.

2) Spectral Coherence (SC)

SC is a measure that provides information about the functional connectivity between the different brain regions. It shows 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 [39]. The differences in SC between AD and FTD can be summarized as follows:

- In 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 [40].
- 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 (13-30 Hz) 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 [40].

The computation of SC between a pair of EEG signals begins with the estimation of the PSD of each signal. This is achieved by applying a wavelet transform, which provides a joint time–frequency representation. The complex Morlet wavelet is commonly used, as it offers good localization in both time and frequency domains, making it well-suited for analyzing the non-stationary characteristics of EEG data.

The wavelet transform of a signal $x(t)$ is defined as:
$$W_x(f,t) = \int_{-\infty}^{\infty} x(\tau)\psi^*(\tau - t, f)\, d\tau \quad (6)$$

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.

From the wavelet coefficients, the PSD of the signal is estimated as:

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

where T is the total number of time points.

To characterize the relationship between two signals $x(t)$ and $y(t)$, the cross-power spectral density (CPSD) is computed as:

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

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

The spectral coherence between the two signals is then calculated as:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_x(f)\,P_y(f)} \quad (9)$$

Where $C_{xy}(f)$ is the coherence at frequency f, $|P_{xy}(f)|^2$ is the squared magnitude of the CPSD, and $P_x(f)$and $P_y(f)$ are the PSDs of signals $x(t)$ and $y(t)$, respectively.

Finally, coherence is averaged within each frequency band(Delta, Theta, Alpha, Beta, Gamma):

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

Where $B = [f_1, f_2]$ denotes the frequency band of interest. The final SC feature representation is a four-dimensional matrix of size $[N, C, C, B]$, where N is the number of samples and C is the number of EEG channels and B is the number of frequency bands.

This feature matrix encodes the strength of functional connectivity between brain regions across multiple frequency bands, a property highly relevant for differentiating neurodegenerative conditions such as AD and FTD.

D. Graph Construction for Functional Connectivity Modelling

After feature extraction, a graph representation is constructed to model the brain's functional connectivity for diagnosing AD and FTD. Representing EEG data as a graph enables the integration of both local spectral activity from individual brain regions and inter-regional connectivity patterns, both of which are critical for differentiating between these disorders.

Formally, the graph is defined as:

$$G = (V, E) \quad (11)$$

Where V is the set of nodes. Each node v_i corresponds to one EEG channel and is represented by the feature vector:

$$v_i = \left[RBP_\delta,\ RBP_\theta,\ RBP_\alpha,\ RBP_\beta, RBP_\gamma, \textit{Channel Index}\right] \quad (12)$$

where RBP_δ to RBP_γ are the RBP values computed for the Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–100 Hz) frequency bands. Channel Index encodes the spatial position of the electrode according to the 10–20 system, preserving the anatomical information of the EEG recording site. E is the set of edges, where each edge e_{ij} represents the functional connection between nodes v_i and v_j.

The weight w_{ij} of each edge is determined by the spectral coherence between the corresponding EEG channels:

$$w_{ij} = C_{ij}(B) \quad (13)$$

Where $C_{ij}(B)$ is the spectral coherence averaged over a specific frequency band B .This ensures that the edge strength reflects the degree of synchronization between the two EEG channels within a specific band.

By combining node-level features (RBP) and edge-level features (SC), the constructed graph captures both the local spectral characteristics of each channel and the global functional connectivity patterns across brain regions. This graph representation is then used as input to graph-based learning models for the classification of AD and FTD.

The proposed model is a Graph Neural Network (GNN) with Transformer-based convolutional layers, designed to process graph-structured EEG data for dementia diagnosis. The architecture encodes the EEG graph into high-level node embeddings through multiple TransformerConv layers, aggregates these embeddings into a single graph-level representation using global pooling, and subsequently classifies the representation through a fully connected network. An overview of the architecture is presented in Figure 2.

Fig. 2 presents the architecture of the GNN model. The graph input, derived from EEG signals, captures functional connectivity across brain regions. Transformer layers equipped with self-attention mechanisms learn both spatial and temporal dependencies. The attention mechanism enables the model to prioritize informative regions, while successive layers extract increasingly abstract representations. By merging graph learning with transformers, the model effectively integrates local and global connectivity patterns, enabling robust differentiation between AD, FTD, and healthy controls.

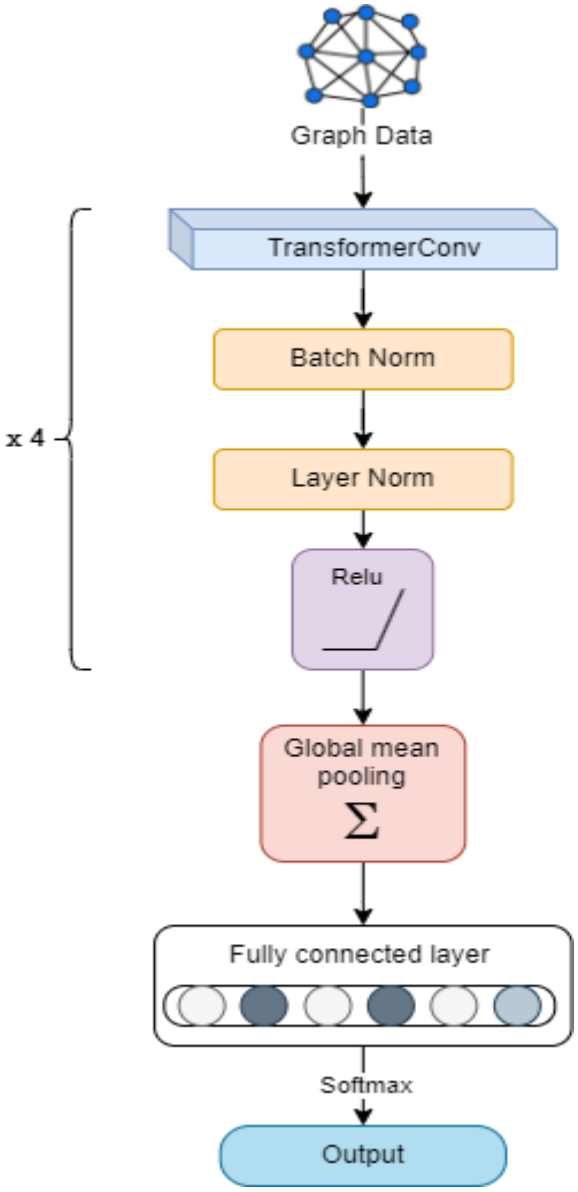

Fig. 2. Architecture of the GNN Model for EEG-based AD and FTD Classification.

The backbone of the network consists of four TransformerConv layers [41], which extend the concept of Graph Convolutional Networks (GCNs) by integrating the attention mechanism of Transformers. Each layer refines node representations by aggregating information from neighboring nodes while accounting for both node and edge features. In this mechanism, each node i communicates with its neighbors j through multi-head attention. The attention score $\alpha_{i,j}$ quantifies the relative importance of node j's features to node i:

$$\alpha_{i,j} = \text{softmax}\left(\frac{(W_3 x_i)^\top (W_4 x_j + W_e e_{ij})}{\sqrt{d}}\right) \quad (14)$$

where x_i and x_j are the feature vectors of nodes i and j, respectively, e_{ij} is the feature of the edge between them (in this case, spectral coherence values), and W_3, W_4, and W_e are learnable weight matrices, and d is the dimensionality of the key vectors used for scaling. This formulation enables the attention mechanism to explicitly incorporate edge information, ensuring that connectivity strength between EEG channels directly influences attention weights.

After attention scores are computed, node iii updates its representation by aggregating neighbor information:

$$x_i' = W_1 x_i + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} \left(W_2 x_j + W_e e_{ij}\right) \quad (15)$$

where W_1 and W_2 are trainable weight matrices. This operation allows each node to combine its own transformed features with a weighted sum of its neighbors' features, with contributions modulated by coherence-based edge features. To improve stability and robustness, the outputs of each layer are passed through batch normalization and layer normalization, followed by a ReLU activation to introduce non-linearity. A dropout layer is then applied to mitigate overfitting by randomly deactivating neurons during training.

Once the TransformerConv layers have been applied, the model performs global mean pooling to produce a compact graph-level embedding. This step averages all node embeddings into a single feature vector:

$$\mathrm{h}_{graph} = \frac{1}{N} \sum_{i=1}^{N} h_i \quad (16)$$

where h_i is the feature vector of node i after the final convolutional layer and N is the total number of nodes in the graph. The resulting graph-level feature vector after global mean pooling h_{graph} provides a compressed representation that captures both local spectral characteristics and large-scale functional connectivity.

The pooled graph embedding is then processed by a fully connected feedforward network, which maps the features into the label space corresponding to AD, FTD, or CN. A log-softmax activation is applied to the final layer to produce class probabilities, completing the classification pipeline.

IV. RESULTS

All preprocessing steps and model training were carried out in a Kaggle Notebook environment [42], equipped with a 2-core CPU, 32 GB RAM, and an NVIDIA Tesla P100 GPU. This configuration provided sufficient computational resources for efficient execution of EEG preprocessing and for training the proposed GNN-based framework.

EEG preprocessing procedures, including filtering, epoch segmentation, and feature extraction, were performed using the MNE-Python and MNE-Connectivity libraries [44], which offer robust implementations for spectral analysis and functional connectivity estimation. Graph representations of EEG signals were constructed with NetworkX, where nodes

represented EEG channels and edges were assigned weights based on spectral coherence values.

For model development, TensorFlow [43] was used to implement the filtration model, while PyTorch Geometric (PyG) [45] supported the design of the Graph Neural Network architecture, including the TransformerConv layers. This combination of frameworks facilitated efficient handling of both grid-structured and graph-structured data.

Under this experimental configuration, the model achieved an inference time of 0.0046 ± 0.0004 seconds per sample, highlighting its computational efficiency and potential for near real-time use in clinical and practical applications.

Figure 3 shows raw EEG recordings, with time on the x-axis and amplitudes across multiple channels on the y-axis. The visualization demonstrates the dynamic and multichannel nature of EEG, as well as variability between electrodes. It also illustrates the presence of noise and artifacts (e.g., eye blinks, muscle activity), underscoring the importance of preprocessing steps like filtering and artifact removal. This figure emphasizes the complexity of EEG data and the need for advanced feature extraction and modeling approaches.

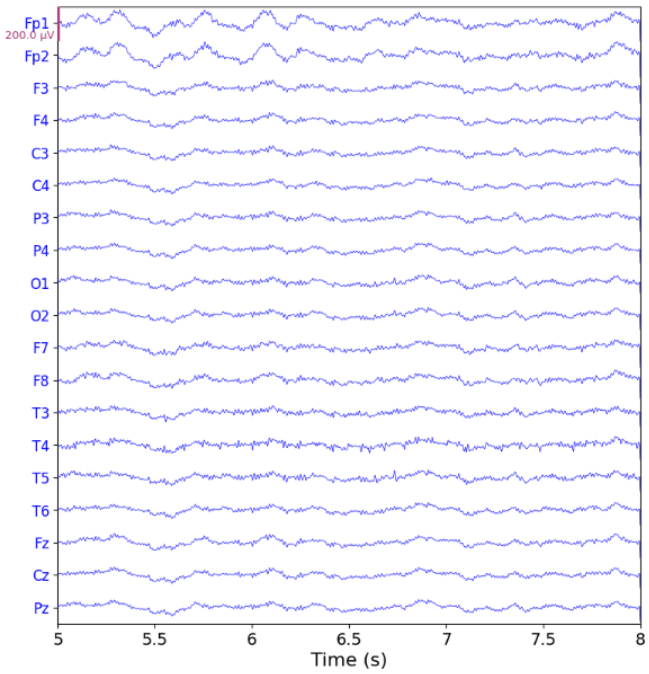

Fig. 3. Sample EEG dataset visualization, showing multichannel time-series brain signals.

Fig. 4 compares PSD heatmaps across major frequency bands (delta, theta, alpha, beta, gamma). The visualizations reveal distinct spectral power alterations between subject groups. The heatmaps highlight how spectral characteristics differ between groups, revealing observable variations in brain activity patterns among AD, FTD, and CN participants. Distinct differences in spectral distribution across cortical regions are evident, aligning with the fact that each condition is associated with unique neural oscillatory alterations and disrupted functional networks [46]. For example, AD patients display reduced alpha and beta activity with elevated theta power, changes that are well-documented biomarkers of dementia. These heatmaps illustrate how spectral abnormalities localize within specific brain regions, offering valuable insights into disease-related neural dysfunction. Integrating these spectral features within graph models enhances the detection of frequency-dependent patterns.

Fig. 5 illustrates the PSD distributions across individual EEG channels, enabling spatial localization of spectral abnormalities and reinforcing the differences observed among AD, FTD, and CN groups. Such channel-wise perspectives are clinically relevant, as alterations in posterior alpha rhythms or frontal theta activity are strongly associated with AD and FTD. A marked reduction in Alpha-band (8–12 Hz) power is apparent in both AD and FTD patients compared to CN subjects, a hallmark of neurodegenerative disorders linked to impaired neuronal synchronization and diminished cognitive efficiency. These spatially resolved spectral features provide complementary biomarkers, and when embedded into graph-based representations, they capture inter-regional connectivity disruptions that enhance diagnostic precision in distinguishing AD and FTD from CN subjects. Together, these findings provide preliminary evidence of discriminative spectral markers that can be leveraged for classification between AD, FTD, and CN groups.

Table 2 highlights key spectral differences between groups and presents the mean and standard deviation for the extracted features. This statistical summary provides a quantitative measure of the differences in spectral power between AD, FTD, and CN groups across the five frequency bands. AD patients show reduced alpha and beta power compared with CN subjects, consistent with the hallmark slowing of brain activity in dementia. CN participants demonstrate higher SC values across most frequency bands, reflecting preserved connectivity. FTD patients show intermediate disruptions, particularly in alpha and beta ranges, aligning with known frontal lobe pathology and executive dysfunction. Elevated delta and theta activity in both AD and FTD groups further supports their role as biomarkers of cortical slowing. Together, SC and RBB measures provide robust, disease-relevant features for classification.

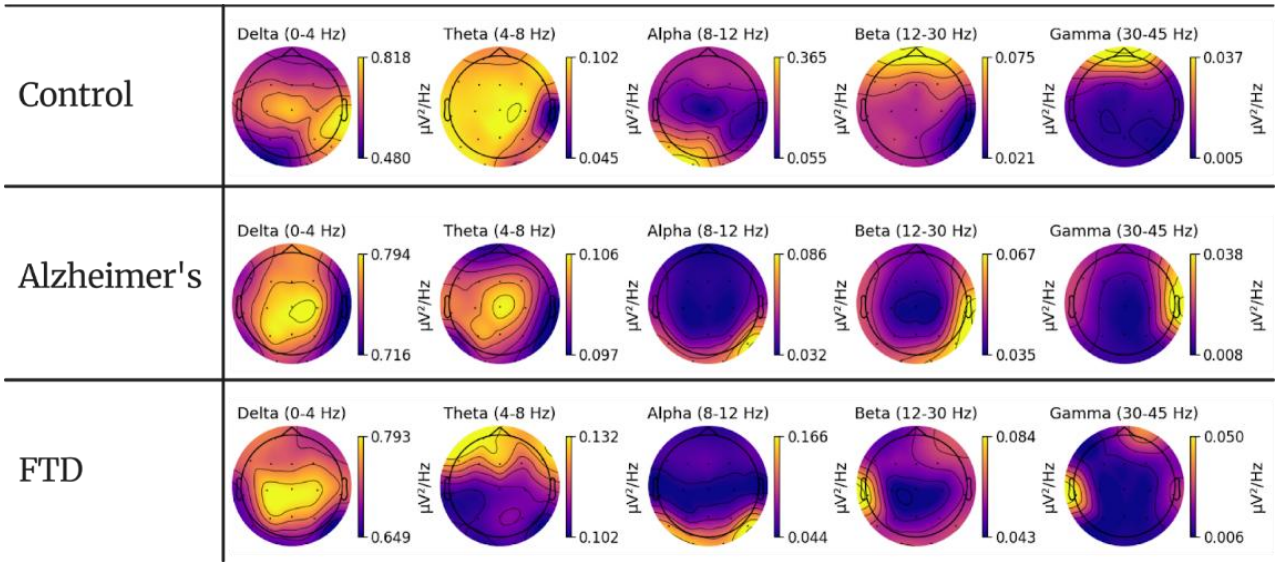

Fig. 4. Brain PSD heatmap comparison across frequency bands.

Table 2. Mean and standard deviation of EEG-derived SC and RBB features across frequency bands for AD, FTD, and CN groups.

Group	Measure	Statistic	Delta	Theta	Alpha	Beta	Gamma
AD	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
CN	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

The proposed GraphTransformerNet architecture consists of four TransformerConv layers, each followed by batch normalization, layer normalization, ReLU activation, and dropout to enhance stability and generalization. The network incorporates multi-head attention with 10 heads and hidden channels of size 128, enabling it to capture complex dependencies between EEG channels. Hyperparameters were carefully tuned during training: the dropout rate was set to 0.2, the learning rate was fixed at 0.0001, and a weight decay of 0.0001 was applied for regularization. A batch size of 512 was used, balancing convergence speed with training stability. The dropout rate and learning rate were optimized through a systematic grid search, while the number of TransformerConv layers, feed-forward network size, hidden channel dimensions, batch size, and weight decay were refined experimentally through iterative trial-and-error. These settings strike a balance between robustness and accuracy, avoiding both underfitting and overfitting, and were critical in achieving the best classification accuracy and generalization across the dataset reported in the results.

A rigorous validation strategy was implemented to ensure the reliability and generalizability of the model. To reduce the risk of overfitting, early stopping was applied with a patience of 60 epochs, meaning training was terminated if validation performance failed to improve over 60 consecutive epochs. This strategy promoted efficient convergence while avoiding unnecessary over-optimization on the training data. In addition, 10-fold cross-validation was performed, with the dataset divided into ten equal folds. In each iteration, the model was trained on nine folds and evaluated on the remaining one, allowing every sample to contribute to both training and validation. Performance metrics were then averaged across the ten iterations, yielding a statistically robust estimate of model performance. To further account for variability due to random initialization and stochastic training dynamics, each fold was repeated multiple times using different random seeds. Figure 6 illustrates the training and validation loss curves for a representative fold, highlighting the convergence behavior achieved under this setup.

Fig. 6 depicts the training dynamics for one fold of the 10-fold cross-validation. The curve illustrates performance metrics

across epochs, showing a stable increase in training accuracy alongside steady validation accuracy, evidence of effective learning without overfitting. The use of 10-fold cross-validation provides robustness by testing performance across multiple data partitions, reducing bias and variance. This figure demonstrates the model's strong generalization ability, supporting its reliability for EEG-based dementia classification.

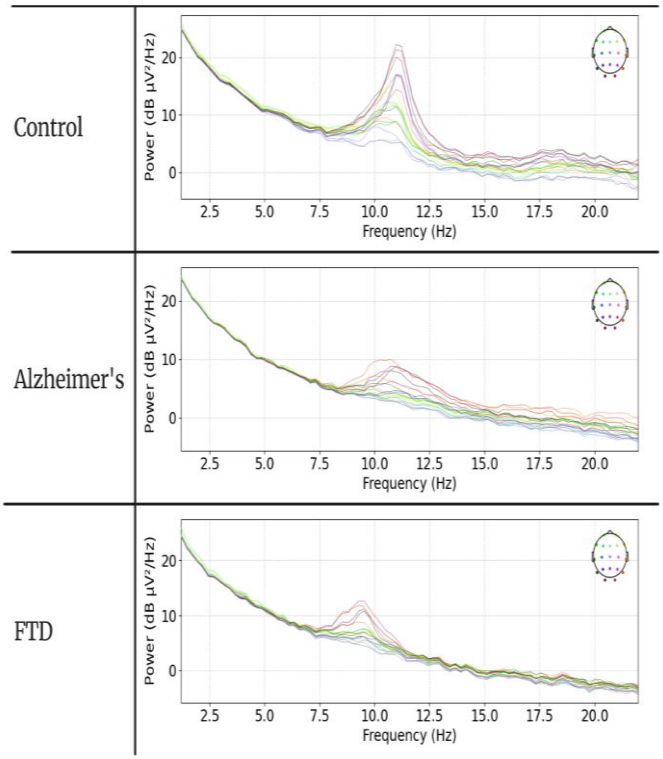

Fig. 5. PSD comparison across EEG channels.

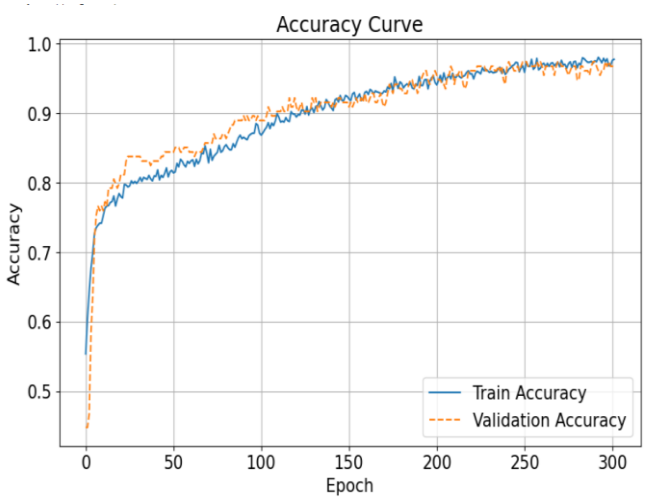

Fig. 6. Training and validation curves of the proposed model for fold 8 in the 10-fold cross-validation.

After training, the model was tested on the test set. To thoroughly assess the model's performance, standard classification metrics were applied, including accuracy (Acc), precision (Pr), sensitivity (Se), specificity (Sp), and F1-score ($F1$), which are calculated using true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values :

$$Accuracy(Acc) = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (17)$$

$$Sensitivity(Se) = \frac{TP}{(TP+FN)} \quad (18)$$

$$Specifity(Sp) = \frac{TN}{(TN+FP)} \quad (19)$$

Table 3 presents the performance of the proposed model across three classification scenarios: AD vs. CN, FTD vs. CN, and multi-class classification (AD–FTD–CN). The results are reported on the artifact-cleaned EEG dataset to ensure that the learning process was not biased by noise or spurious components. Table 3 demonstrates strong classification outcomes. AD vs CN discrimination reached 99.1% Acc, Se, and Sp, while FTD vs CN achieved 97.7%. In the multiclass task (AD vs FTD vs CN), the model maintained 96.7% Acc, showing robustness even under three-class conditions. Compared with typical EEG-based studies (80–90% Acc), these results are state-of-the-art. High Se ensures reliable patient detection, and balanced specificity reduces false positives, underscoring the clinical practicality of the model. These outcomes emphasize the effectiveness of the proposed approach in distinguishing neurodegenerative conditions from cognitively normal controls, while also highlighting its capacity to manage the complexity of multi-class classification. The evaluation demonstrates the model's ability to extract discriminative spectral–connectivity features and reliably distinguish between AD, FTD, and CN groups. Consistently strong performance across both binary and multi-class tasks highlights the robustness and generalizability of the approach.

Table 3. Classification performance of the Graph Transformer Network using cleaned EEG data.

Case	*Acc* %	*Se* %	*Sp* %
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%

To validate the impact of the preprocessing pipeline, the model's performance was systematically compared on raw (uncleaned) and cleaned EEG data, as summarized in Table 4. The cleaning process employed a LU-NET model trained on the EEGdenoiseNet dataset, which effectively suppresses ocular, muscular, and environmental artifacts while preserving the integrity of neural signals. The comparison highlights the critical role of preprocessing in boosting classification accuracy. For AD vs. CN, performance improved from 97.6% on raw data to 99.1% on cleaned data, reflecting a clear gain in discriminative power. In the FTD vs. CN task, accuracy rose from 96.7% to 97.7%, while in the more demanding ternary classification (AD vs. FTD vs. CN), accuracy increased from 96.0% to 96.7%. Artifact reduction (e.g., eye blinks, muscle noise) allowed the model to focus on disease-related patterns,

yielding consistent improvements across tasks. Although these improvements may appear modest, they are clinically meaningful, as even small gains enhance the reliability of patient-level predictions. Overall, the findings emphasize the necessity of artifact removal and preprocessing in EEG-based dementia classification, ensuring that the model captures genuine neural dynamics rather than noise-related confounds.

Table 4. Impact of preprocessing on classification Acc: raw versus filtered EEG data

Case	Raw Data Acc %	Filtered Data Acc %
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 provides a comparative analysis of the proposed model against recent studies on EEG-based dementia classification, showing that it consistently delivers competitive and often superior results. In the AD–CN and FTD–CN binary classification tasks, the introduced model achieves Acc levels that meet or exceed those reported by state-of-the-art approaches. For example, [34] used edge-filtering techniques with graph convolutional networks (GCNs) and reached an Acc of 92%, with Se and Sp of 97.4% and 86.7%, respectively. While effective, this method showed a notable imbalance between Se and Sp. Similarly, [47] applied Hjorth parameters and entropy features with an SVM classifier, but achieved only 81% Acc, with Se of 69.8% and Sp of 83.5%, emphasizing the difficulty of balancing these metrics. Another study by [32] introduced a dual-input convolutional encoder and reported 83.28% Acc, again reflecting the challenge of building robust EEG-based classifiers. In [48], the authors combined Mexican Hat CWT with Bispectrum estimation and an MLP classifier, yielding 92.95% Acc, though their Se and Sp were not fully reported. Notably, [24] claimed nearly perfect results (99.98% for Acc, Se and Sp) by using a large set of handcrafted features. However, such unusually high performance may stem from dataset differences or limited generalizability. Beyond EEG-based approaches, [49] used transcranial magnetic stimulation (TMS) data with a Random Forest classifier, achieving 92.05% Acc, with Se of 96.15% and Sp of 87.94%.

In contrast, the proposed method achieves 99.1% Acc, with both Se and Sp balanced at 99.1%, outperforming or matching the strongest results reported so far. Importantly, this is accomplished using only resting-state EEG data, together with RBP, SC, and a convolutional transformer-based graph network. By integrating spectral-domain features with graph-based DL, the proposed framework captures both local spectral changes and disrupted inter-regional connectivity, enabling more reliable classification of AD versus CN individuals. These findings underscore not only the methodological strength of the introduced approach but also its clinical importance, as achieving balanced Acc, Se and Sp is essential for reducing both false positives and false negatives in diagnostic settings.

Table 5. Comparison of the proposed method with state-of-the-art EEG-based AD classification studies.

Study	year	cohorts	stimuli	methodology	Acc %	Se %	Sp %
[34]	2022	20 AD 20 CN	Resting state	edge-filtering methods and graph convolutional networks (GCNs)	92%	97.4%	86.7%
[47]	2021	30 AD 35 CN	-	Hjorth parameter and entropy on SVM model	81%	69.8%	83.5%
[32]	2023	36AD 29 CN	Resting state	RBP, SCC, Dual-input-convolutional encoder	83.28	78.81	87.9
[48]	2019	63 AD 63 CN	Resting state	Mexican Hat CWT and Bispectrum estimation and MLP Model	92.95	-	-
[30]	2022	63 AD 63 CN	Resting state	DWT+LBP,STD,VAR,KUR,AR,RMS and NO	99.98	99.93	99.98
[49]	2023	38AD 17CN	Transcranial magnetic stimulation-TMS	time-domain and statistical EEG features, trained Random Forest classifier	92.05	96.15	87.94
Proposed Work		36 AD 29 CN	Resting state	RBP, SC,convolutional transformer graph network	99.1	99.1	99.1

V. CONCLUSIONS

In this study, we introduced a novel framework for EEG-based classification of AD, FTD, and CN individuals. The approach integrates spectral features, connectivity measures, and a convolutional transformer graph network. Results show that the model achieved highly competitive and well-balanced performance in both binary and ternary classification tasks, with accuracy, sensitivity, and specificity consistently exceeding 96%. Additionally, artifact removal and preprocessing were found to significantly improve outcomes, emphasizing the critical role of clean EEG signals in dementia detection. Future work could expand this framework to other

neurological and biomedical disorders, systematically investigate optimal EEG segment lengths, evaluate alternative filtering architectures, and adopt more rigorous methods for feature selection. Furthermore, ablation studies exploring different combinations of convolutional, transformer, and graph-based layers may help refine the architecture and provide deeper insights into its effectiveness. Overall, the proposed framework demonstrates robustness, scalability, and strong potential for broader clinical and biomedical applications. It represents a promising step toward the development of reliable EEG-based diagnostic tools for dementia and beyond.

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 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] A. S. Eltrass, M. B. Tayel, and A. F. El-Qady, "Identification and classification of epileptic EEG signals using invertible constant-Q transform-based deep convolutional neural network," Journal of Neural Engineering, vol. 19, no. 6, p. 066035, Dec. 2022.

[19] A. S. Eltrass and N. H. Ghanem, "A new automated multi-stage system of non-local means and multi-kernel adaptive filtering techniques for EEG noise and artifacts suppression," Journal of Neural Engineering, vol. 18, no. 3, p. 036023, Apr. 2021.

[20] N. H. Ghanem, A. S. Eltrass, and N. H. Ismail, Investigation of EEG Noise and Artifact Removal by Patch-Based and Kernel Adaptive Filtering Techniques. Roma, Italy: The 13th Annual IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2018.

[21] A. S. Eltrass and N. H. Ghanem, "Investigation of automatic spindle detection in sleep EEG signals contaminated with noise and artifact sources," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 9, pp. 12725–12746, Jul. 2022.

[22] M. B. Tayel, A. S. Eltrass, and A. I. Ammar, "A new multi-stage combined kernel filtering approach for ECG noise removal," Journal of Electrocardiology, vol. 51, no. 2, pp. 265–275, Oct. 2017.

[23] A. S. Eltrass, "Novel cascade filter design of improved sparse low-rank matrix estimation and kernel adaptive filtering for ECG denoising and artifacts cancellation," Biomedical Signal Processing and Control, vol. 77, p. 103750, May 2022.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

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

[40] 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.

[41] 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.

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

[43] 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.

[44] 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.

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

[46] 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.

[47] 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.

[48] 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.

[49] 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.