Biomedical Signal Processing and Control 114 (2026) 109321

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Alzheimer's disease classification from EEG using a multiscale temporal deep network

Simone Zini *, Thomas Barbera, Simone Bianco, Paolo Napoletano

Department of Informatics, Systems and Communication, University of Milano-Bicocca, Viale Sarca 336, Milano, 20126, Italy

ARTICLE INFO

Keywords:
Alzheimer's disease
Deep learning
EEG
Multiscale temporal features

ABSTRACT

Early and accurate diagnosis of Alzheimer's Disease (AD) is crucial for patient care and intervention but remains challenging due to the complexity and variability of clinical data. Electroencephalography (EEG) has emerged as a promising, non-invasive, and cost-effective tool to detect brain activity patterns associated with AD. In this work, we introduce a lightweight Multiscale Temporal Deep Network (MTDNet) that integrates multiple temporal convolutions with recurrent modeling to capture both short- and long-term EEG patterns. Two patient-level classification strategies are also proposed that combine segment-level EEG predictions based on consensus and score aggregation to better align with clinical practice and utility. We evaluate our method on four benchmark EEG datasets (ADSZ, APAVA, ADFTD, BrainLat) where it consistently outperforms state-of-the-art solutions by about 2% at the segment level and by about 6% at the patient level on the most challenging datasets. Unlike recent computationally heavy transformer-based solutions, MTDNet achieves superior accuracy with only 20.5K parameters and 1.8M FLOPs, enabling its deployment in resource-constrained environments. Ablation studies confirm the critical contribution of the multiscale design and show that simple augmentation techniques increase generalization and robustness. Code is available at https://github.com/unimib-islab/MTDNet.

1. Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that affects memory, thinking, and behavior. It is the most common cause of dementia worldwide with 57 million people globally affected by it, as estimated by the World Health Organization in 2021 [1].

The early diagnosis of AD is a challenging and essential task for the management and intervention in the treatment of patients affected by it. Multiple technologies have been adopted for the study of the human brain and the progression of dementia: among the most known and adopted technologies, there are Computed Tomography (CT) [2], Magnetic Resonance Imaging (MRI) [3], Functional Magnetic Resonance Imaging (fMRI) [4], Positron Emission Tomography (PET) [5], Spectroscopy [6], Magnetoencephalography (MEG) [7], etc [8]. However, most of these technologies require a time-consuming and in some cases invasive procedure to collect data from patients, making the entire process complex and slow. In this scenario, Electroencephalography (EEG) emerged as a promising technology for the monitoring and detection of dementia: it represents a fast, non-invasive and reliable solution for data collection, providing a cost-effective alternative for the disease diagnosis procedure [9,10].

In recent years, thanks to the advancements in the Computer Science field with the advent of Machine Learning and Deep Learning techniques, automatic detection and classification of EEG signals for dementia assessment has gained a lot of attention from researches. Multiple works have proven the potential of EEG in combination with machine learning approaches for the early diagnosis of AD and other types of dementia [11–13]. Following these first approaches, multiple new ones have been developed that exploit more and more complex deep learning models [14], based on well-known architectures such as VGG and ResNets, eventually ending up using modern architectures based on transformer models [15–17]. However, such solutions tend to be expensive from the computational point of view, requiring a large amount of data to be trained and high-performance servers for model training and use, making them less suitable for the use on-board of small devices and portable solutions. Even if multiple datasets have been collected alongside the proposal of new models and methods, the limited availability of recordings and the intrinsic high subjectivity of this type of data make the task of dementia assessment still an open challenge [18].

* Corresponding author.
E-mail addresses: simone.zini@unimib.it (S. Zini), thomas.barbera@unimib.it (T. Barbera), simone.bianco@unimib.it (S. Bianco), paolo.napoletano@unimib.it (P. Napoletano).

https://doi.org/10.1016/j.bspc.2025.109321
Received 5 September 2025; Received in revised form 14 November 2025; Accepted 29 November 2025
Available online 3 December 2025

In this context, we propose a lightweight approach for Alzheimer's Disease classification, based on pre-transformers solutions to overcome the limitations related to data availability and computational complexity of recent state-of-the-art approaches. Our contributions are summarized as follows:

- We present a novel lightweight approach based on multiscale temporal feature extraction and Long Short Term Memory (LSTM) modules for binary and multiclass Alzheimer's Disease classification.
- We propose two patient-level classification strategies to aggregate the predictions of EEG segments.
- Experimental results show that our model can outperform state-of-the-art methods for Alzheimer's Disease classification on multiple benchmark datasets, while drastically reducing computational costs and model dimensionality.
- An in-depth comparison of current approaches against the proposed solution is conducted in terms of computational costs and model dimensionality with respect to model accuracy, showing the advantages of our approach.

2. Related works

Methodologies for Alzheimer's disease classification through the analysis of EEG recordings can be divided into two main groups: manual feature extraction-based approaches and end-to-end training-based approaches.

Handcrafted feature extraction. The first group includes approaches that explore how to extract useful, domain-specific features relating to Alzheimer's disease from EEG recordings. Due to the time-sensitive nature of this type of data, feature extraction techniques have primarily focused on the time, frequency and time–frequency domains. These domains offer different perspectives on analyzing brain activity and provide various approaches to analyzing the frequency and temporal synchronization and desynchronization of brain regions.

Most of the existing approaches based on time-related features exploit statistics such as mean, amplitude, variance, kurtosis, skewness and standard deviation [19–21], entropy [21–23], or use principal component analysis (PCA) to extract components directly from the raw signals [24–26].

Frequency-based approaches adopt different feature extraction techniques, such as the computation of the signal power spectral density (PSD) [27–30], the Fourier transform or the Wavelets for the extraction of signal frequency features [31–35], or even more complex features for the desynchronization analysis of neurons such as the Bispectrum [36]. These features are usually extracted from raw EEG signals, sometimes in combination with other features.

Finally, recent approaches tend to combine features from both domains, adopting combinations of power spectral density and time features [37–40], or using raw signals [41], or directly using representations such as continuous wavelet transforms (CWTs), which can decompose signal frequency information while maintaining time coherence and event-related information [11,36].

Other manual feature extraction-based approaches focus on examining the brain connectivity, which describes the interactions between different brain regions. Multiple approaches have been developed, exploiting features such as network resilience, network clustering coefficient, versatility [42–44], or phase locking and phase lag index [27,44,45]. These approaches try to model and detect the different types of spatial and temporal desynchronization that characterize Alzheimer's disease patients recording, in order to classify those cases [46]. Methods analyzing EEG spectral power and functional connectivity have also been applied in neuromodulation contexts [47,48], supporting their general utility as biomarkers of brain state.

Independently from the manual features used, after they have been extracted they are generally classified in a second step using machine learning techniques such as support vector machines (SVMs) [49], K-nearest neighbor (KNN) [50], fuzzy learning models [51] or simple multi-layer perceptrons (MLP) [36].

End-to-end models. The second type of approaches for Alzheimer's disease classification is based on end-to-end model training. These approaches train a neural network to jointly learn feature representations and perform classification directly from raw or preprocessed EEG signals. Thanks to the promising results of deep neural networks in other fields, this research area has expanded rapidly in recent years and it has also been widely explored in the field of Alzheimer's disease classification. Various neural architectures have been investigated, including: Convolutional Neural Networks (CNNs) [13,14,52–55] have been employed to exploit local spatial and temporal patterns within EEG recordings. Some studies convert EEG into 2D representations (e.g., spectrograms) before feeding them to the CNN, while others apply 1D convolutions directly on raw EEG. Recurrent Neural Networks (RNNs) [26,56] have been used to model long-term temporal dependencies. Graph Neural Networks (GNNs) [45,57,58] have been used to explicitly model the spatial structure and connectivity among EEG channels. Transformers models [17] are among the most recent and most performing approaches. Exploiting the self-attention mechanism, they have been used to model complex temporal dependencies.

Despite the advances in this field, the recently proposed solutions still present notable limitations and shortcomings. Although these models achieve good accuracy on multiple datasets, they are generally computationally expensive and require large amounts of training data, which is usually unavailable in the context of Alzheimer's disease classification. Moreover, the large number of parameters and the high dimensionality of these models hinder the possibility of deploying them on portable or usable medical devices, designed to be easily transported and operated outside of traditional medical settings, such as in home care, point-of-care, or mobile clinics. Lastly, to the best of our knowledge, current state-of-the-art approaches only process small temporal samples extracted from EEG recordings of different lengths, focusing the evaluation and analysis on segments without considering the entire patient recording.

Inspired by the need for a methodology to process full recordings and for efficient, lightweight approaches, we propose a new, multiscale temporal approach for short time sequence recordings and two simple, yet versatile and effective, strategies for classifying Alzheimer's disease at the patient level.

3. Proposed methodology

While most machine learning research on EEG-based AD detection focuses on classifying short temporal segments (typically 1-second long), the ultimate goal in real-world applications is to assess the cognitive status of a patient based on an entire EEG recording acquired during an experimental session [59]. Segment-level classification enables convenient benchmarking and helps in overcoming data scarcity by increasing the number of training samples. However, relying solely on per-segment accuracy may not reflect a model's ability to reliably classify patient-level outcomes.

This misalignment between evaluation protocol and clinical utility motivates the need for methods that can infer a label for an entire EEG session by aggregating information from multiple segments. Such an approach allows for evaluating a model's performance in terms that are closer to practical diagnostic scenarios.

Although the core of our proposed method is based on single-segment classification, in this section, we also introduce and evaluate two complementary strategies to perform classification at the recording level. These strategies — majority voting and score averaging — are agnostic to the specific segment classifier and can thus be adopted to compare any segment-based method in terms of patient-level classification performance.

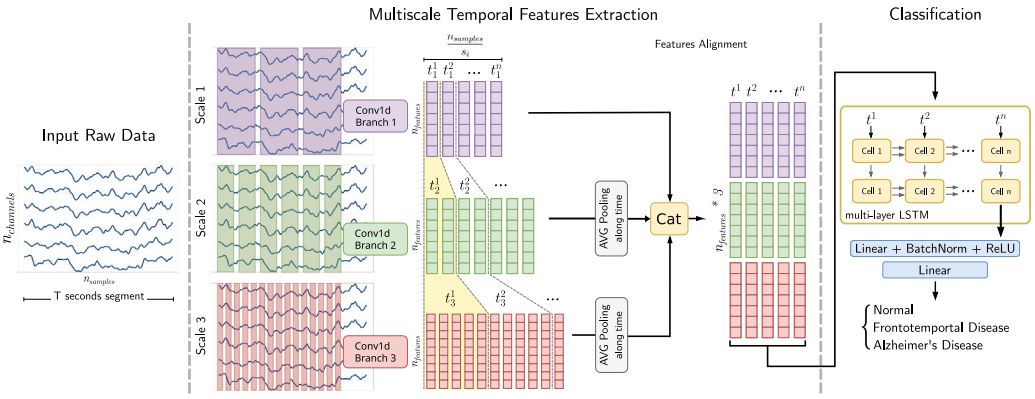

Fig. 1. Overview of the proposed segment-based prediction model. The overall architecture can be divided in two main parts: a Feature extraction (center) and a Classification part (right). The feature extraction is performed at multiple time scales by different convolutional layers. Then the extracted features are aggregated and classified using a LSTM followed by two fully connected layers for the classification in two or three classes, depending on whether the task is binary or ternary classification.

3.1. Proposed segment-classification method

The central motivation behind our proposed method is to effectively capture both the multi-scale temporal patterns present in EEG signals and their sequential dependencies over time. EEG data is inherently complex, exhibiting dynamics at multiple temporal resolutions and contextual patterns that evolve sequentially: high-frequency bands characterized by rapid changes in the signal occurring at a fine temporal granularity (i.e. beta and gamma bands) and low-frequency bands characterized by slower changes in the signal occurring at a coarse temporal granularity (i.e. delta, theta and alpha bands). This multi-scale behavior underlies the rationale of wavelet decomposition-based approaches [31,32] which analyze EEG signals at multiple temporal resolutions, and also inspired the proposed multiscale feature extraction design. Furthermore AD shows its presence in EEG signals not only as static features such as the overall slowing of the EEG spectrum and reduction in frequency components, but also as an alteration of the temporal dynamics such as reduced synchrony or abnormal transitions between brain states [60,61]. By combining temporal convolutions with recurrent modeling, our approach is designed to capture both short-term patterns and long-term dependencies. To address these challenges, we designed a hybrid architecture that integrates two complementary processing components, as illustrated in Fig. 1.

The first component is a multi-temporal convolutional module, which employs temporal convolutions with varying kernel sizes. This module acts as a feature extractor, enabling the model to capture relevant patterns at different time scales — including both fast oscillatory components and slower fluctuations — that are informative for cognitive state characterization.

The second component is a sequential modeling module based on Long Short-Term Memory (LSTM) networks. Applied on top of the convolutional features, this module acts as the classifier, learning temporal dependencies across consecutive EEG windows and integrating contextual information over time to produce a prediction for each segment.

By combining these two modules, our architecture leverages the strengths of both parallel multi-scale feature extraction and recurrent sequence modeling. This joint design enhances the model's ability to discriminate between cognitive states in short EEG segments, providing a more robust foundation for downstream aggregation at the patient level. We called our proposed method MTDNet (Multiscale Temporal Deep Network). The detailed architecture that describes the employed layers and operations is reported in Table 1.

Given an EEG recording of several minutes, the proposed approach operates by dividing the signal into consecutive segments of duration T seconds each. In our implementation, each recording is split into multiple one-second-long segments, which are individually processed by the model and assigned a prediction. During training, each segment inherits the label of the full recording as ground truth, under the assumption that the cognitive condition remains consistent throughout the session. This strategy allows the model to learn from localized temporal patterns while benefiting from a greater number of training instances. Further details on the segmentation and evaluation protocol are provided in Section 4.

Table 1
MTDNet architecture details. The parameters of each layer are here reported.

Scale 1	Scale 2	Scale 3
Conv1D	Conv1D	Conv1D
$n_{\text{features}} = 32/64$	$n_{\text{features}} = 32/64$	$n_{\text{features}} = 32/64$
$k = 10,\ s = 10,\ p = 0$	$k = 5,\ s = 5,\ p = 0$	$k = 2,\ s = 2,\ p = 0$
BatchNorm1D 32/64	BatchNorm1D 32/64	BatchNorm1D 32/64
ReLU	ReLU	ReLU
–	AvgPool1D $k = 2,\ s = 2$	AvgPool1D $k = 5,\ s = 5$
Concatenate		
LSTM input_dim = 96/192 hidden_dim = 16 n_layers = 2		
Linear 16 – 16		
BatchNorm1D 16		
ReLU		
Linear 16 - # classes		

3.1.1. Multiscale temporal feature extractor

Given an input EEG segment $\mathbf{e}$ of duration T seconds, collected at a sampling frequency f, the first component of the proposed model performs multiscale temporal feature extraction using parallel one-dimensional convolutions at multiple scales. Specifically, the architecture operates at three temporal scales: for an input consisting of $T \cdot f$ time samples, multiscale features are extracted using three parallel one-dimensional convolutional layers with kernel sizes of 10, 5, and 2, respectively. The stride for each convolution is set to 10, 5, 2 respectively to ensure non-overlapping feature extraction. Feature extraction is performed channel-wise, enabling the model to capture correlations across EEG channels over short temporal windows. At each scale, the output of the convolution is followed by batch normalization and the application of a ReLU activation function.

The use of this approach for feature extraction allows the model to incorporate multiple branches, each capable of detecting signal characteristics that may occur at different temporal scales. This design is partially inspired by time–frequency analysis, in which a signal is examined to discover how the frequency components change over time. In our case, the convolutional kernels not only detect structures at

specific temporal resolutions within segments of the signal, but also capture inter-channel correlations occurring simultaneously.

The output of this first stage is a set of features extracted at multiple temporal scales, as illustrated in Fig. 1. Given an input with $n_{\text{samples}} = T \cdot f$ time samples, each convolutional branch i produces an output of size $n_{\text{features}} \times \left\lfloor \frac{n_{\text{samples}}}{s_i} \right\rfloor$, where s_i denotes the stride of the ith convolutional layer.

Given the first scale of the model and a convolutional kernel of size $k_i \times n_{\text{channels}}$, the convolutional layer produces a subset of features denoted as t_1^l, where $l \in [1, \ldots, \frac{n_{\text{samples}}}{s_1}]$. The branches corresponding to higher scales, using different kernel sizes, generate a varying number of features corresponding to the same window t_1^l—these are referred to as t_2^l and t_3^l in Fig. 1. For example, considering the first scale with a kernel size of 10, the segment t_1^1 corresponds to 10 time steps. This results in a feature tensor of shape $n_{\text{features}} \times 1$ at the first scale, $n_{\text{features}} \times 2$ at the second scale, and $n_{\text{features}} \times 5$ at the third scale.

In order to use the LSTM in the classification phase, features must be temporally aligned before performing a concatenation along the feature size dimension. To do so, the features obtained from the scales higher than the first one are averaged along the time dimension, with respect to the timestep subset t_i^l to which they belong to. A visual representation of such a relationship is depicted in Fig. 1 highlighted with a dashed yellow area below the feature extraction part. Features are then aligned using the average pooling operation and then concatenated. The final output of this first part is a feature set of dimensions $n_{\text{features}} * 3 \times \left\lfloor \frac{n_{\text{samples}}}{s_i} \right\rfloor$ where each column contains a collection of features for the timestep subset t^l, extracted at different scales.

3.1.2. Temporal feature classification

The features extracted in the previous stage are processed by a Long Short-Term Memory (LSTM) network, which captures temporal dependencies across the aligned feature sequences. Each input to the LSTM consists of a feature vector of dimension $n_{\text{features}} * 3$. We employ a two-layer LSTM with a hidden state dimension of 16. The output of the LSTM is normalized using batch normalization, then passed through a Fully Connected (FC) layer with output dimension 16, followed by another batch normalization layer and a ReLU activation function. Finally, a second FC layer produces the output logits. The number of output logits is either two or three, depending on whether the task is binary or ternary classification.

3.2. Patient-level classification procedure

Since the final objective of our study is the classification of entire patient recordings, we designed a procedure to combine single temporal sample predictions for classifying recordings of a few minutes' length. Specifically, we designed two different strategies to combine predictions obtained from any model able to perform segment-level classification. Fig. 2 depicts an overview of the methodology.

3.2.1. Patient-level classification strategy with consensus

In this case, we classify the entire recording $\mathbf{r}$ adopting majority voting with the predictions obtained from each segment $\mathbf{e}$ obtained from it. Let $\hat{y}_{\mathbf{e}_j}$ be a collection of segment-level predictions—with $j \in [1, \ldots, N_{\mathbf{r}}]$ and $N_{\mathbf{r}}$ the number of segment obtained from recording $\mathbf{r}$; for the recording $\mathbf{r}$, its predicted class $\hat{y}_{\mathbf{r}}$ is obtained with:

$$\hat{y}_{\mathbf{r}} = \underset{c \in C}{\text{argmax}} \sum_{j=1}^{N_{\mathbf{r}}} \mathbb{I}\left(\hat{y}_{\mathbf{e}_j} = c\right) \tag{1}$$

where $c \in [1, \ldots, M]$, with M corresponding to the number of classes of a specific dataset, and $\mathbb{I}$ representing the indicator function. This strategy classifies the entire recording based on the label that is predicted the most by all the segments extracted from it.

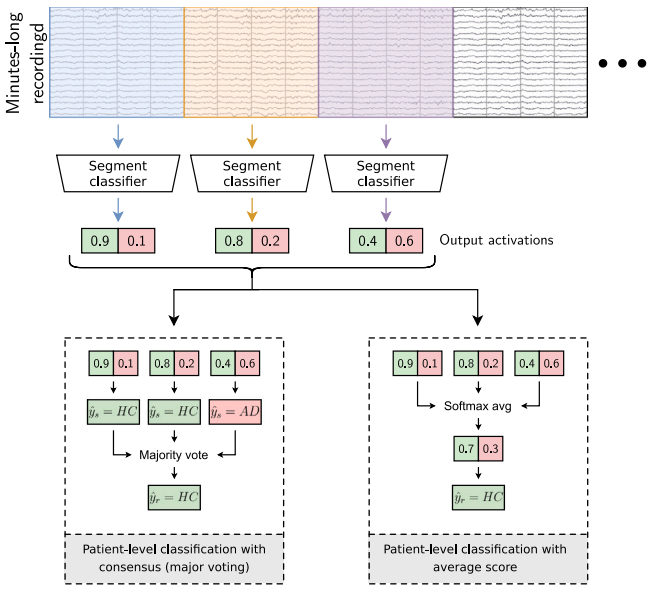

Fig. 2. Schematic representation of the patient-level classification procedure. Given any segment-level classifier (top), an entire recording can be classified adopting one of the two proposed strategies: consensus (left) or average score (right).

3.2.2. Patient-level classification strategy with average score

This strategy performs recording classification by averaging the classification score obtained for each segment $\mathbf{e}$ of a recording $\mathbf{r}$ of a patient. For the averaging procedure, we use the probabilities obtained after the softmax activation $\tilde{y}_{\mathbf{e}_j} = (p_1, p_2, \ldots, p_c)_j$, with $c \in C : [1, M]$, with M equal to the number of classes, as follows:

$$\hat{y}_{\mathbf{r}} = \underset{c}{\text{argmax}} \left(\frac{1}{N_{\mathbf{r}}} \sum_{j=1}^{N_{\mathbf{r}}} \tilde{y}_{\mathbf{e}_j} \right) \tag{2}$$

These two strategies enable us to make predictions based on entire recordings, classifying patients using minutes of recordings rather than single segments of a few seconds. These two strategies can be used independently of the method selected for classifying single segments, which also allows us to fairly compare our proposed approach with others in the state of the art.

4. Experiments

4.1. Experimental setup

To assess the performances of the proposed solution we compared our model with recent state-of-the-art approaches for the classification of Alzheimer's Disease. In particular, we selected five models: TimesNet by Wu et al. [66], Crossformer by Zhang et al. [15], iTransformer by Liu et al. [16], Medformer by Wang et al. [17] and LCADNet by Kachare et al. [54]. All the mentioned architectures have been retrained and tested using the source code provided by Medformer authors[1] and the available source code. Models have been compared using accuracy and macro F1-Score on four different test datasets. In order to have a fair comparison with the work previously done by Wang et al. [17], the models have been retrained with five different seeds (from 41 to 45). The results reported are the average accuracy and F1-Score obtained by averaging the results of each separated run. We also reported the standard deviation, in order to provide insights on models stability across multiple runs. The state-of-the-art approaches

[1] https://github.com/DL4mHealth/Medformer

Table 2
Summary of the characteristics of the selected datasets. ADFTD-Binary is a subsample of the ADFTD original dataset, obtained by excluding the subjects belonging to the FTD class. The total number of training samples refers to the quantity of 1 s-long segments obtained from the recordings of a dataset, i.e. the total number of samples contained in the training and validation splits.

		ADSZ [62]	APAVA [63]	ADFTD [64]	ADFTD-Binary [64]	BrainLat-Binary [65]
No. of classes		2 (HC | AD)	2 (HC | AD)	3 (HC | FTD | AD)	2 (HC | AD)	2 (HC | AD)
No. of subjects		48 (24 | 24)	23 (11 | 12)	88 (36 | 23 | 29)	65 (36 | 29)	67 (32 | 35)
Samp. rate (Hz)		128	256	500[a]	500[a]	512
No. of channels		19	16	19	19	128
Avg rec. length (min.)		0.267	4.324	13.211	13.636	7.41
No. of training samples		586	4536	55 140	41 737	23 635
Age	HC (avg. years ± std)	72 ± N.a.	72.8 ± 6.1	67.9 ± 7.9	67.9 ± 7.9	67.9 ± 8.9
	AD (avg. years ± std)	69 ± N.a.	72.5 ± 8.3	66.4 ± 7.9	66.4 ± 7.9	72.2 ± 7.9
	FTD (avg. years ± std)	–	–	63.6 ± 8.2	–	–
Recording conditions		Resting state eyes open/closed	Resting state eyes closed, awake	Resting state eyes closed	Resting state eyes closed	Resting state eyes closed, awake dim light, attenuated sound

[a] Since the state-of-the-art authors decided to use the ADFTD dataset with a resampling at 256 Hz, we employed their available preprocessed version of the dataset to test their models.

have been retrained employing the hyperparameters set according to the original works, using the open-source code provided by the authors of [17] and [54].

For the evaluation, we adopted the Subject-Independent strategy proposed by Wang et al. [17] splitting the datasets by subject into training, validation and test sets with no overlap.[2] This strategy involves each recording of a subject being part of only one of the splits. The test and validation steps are performed on subjects that were not seen during the training phase, making the overall task more realistic. This approach avoids bias and presents a more challenging task, as the test data may differ significantly from the training data due to variability in the datasets and subject-specific noise.

4.2. Datasets and data preparation

In order to compare the proposed approach with the current state-of-the-art methodologies, we used four well known datasets generally used for the Alzheimer's Disease classification task. Specifically we used the ADSZ dataset by Alves et al. [62], the APAVA dataset by Escudero et al. [63], the ADFTD dataset by Miltiadous et al. [64] and BrainLat dataset by Prado et al. [65]. All of the used datasets are currently publicly available. These datasets are all collections of EEG recordings taken in resting state, and contain Health Control (HC) and Alzheimer's Disease (AD) classes. The number of recording channels vary between the three datasets as well as the sampling frequency originally adopted for the recording procedure. The main datasets characteristics are summarized in Table 2. ADFTD and BrainLat datasets originally present more than two classes. Specifically, ADFTD offer a third class which corresponds to subjects with frontotemporal dementia (FTD), while BrainLat presents three more classes, namely behavioral variant frontotemporal dementia (bvFTD), multiple sclerosis (MS), and Parkinson's disease (PD). Given these very different types of classes, we decided to adopt ADFTD in its original version and a version we called ADFTD-Binary, using only the HC and AD classes, while for the BrainLat dataset we used only its binary version, which we called BrainLat-Binary, obtained by only considering the AD and HC classes.

The proposed approach has been trained on recordings crop of one second each ($T = 1$ s), which we call *segments*: each segment is a $n_{\text{channels}} \times n_{\text{samples}}$ matrix, which is z-score normalized over all the channels, in order to maintain the dynamics between the different channels. Before performing the normalization procedure, we computed the mean value of each channel in order to zero center each signal; then the normalization previously described is applied to the resulting signal. The z-score normalization has been performed using the function provided by the PYTHON SCIKIT-LEARN library. For the training of the state-of-the-art approaches we adopted the data pre-processing described by Wang et al. [17].

In order to provide a fair comparison, the record cropping operation has been performed in the same way as done by the previous existing works: in the case of ADSZ and APAVA datasets, from each recording we extracted segments of $T = 1$ s with an overlap of 0.5 s between each segment. In the case of ADFTD, ADFTD-Binary and BrainLat-Binary instead we extracted samples of $T = 1$ s without overlapping. We obtained a total of 768, 5967, 69794, 53215 and 29788 1 s-long segments from each of the datasets respectively.

4.3. Training details

The proposed model has been developed in PyTorch v2.5.1 and trained on an NVIDIA RTX 4080. The model has been trained for a total of 80 epochs on each dataset, with a starting learning rate of $5e-4$, gradually reduced with an exponential learning rate scheduler with $\gamma = 0.98$. The mini batch size has been set to 256 for all of the datasets except for ADSZ, for which it has been set to 128 (due to the small dimension of the dataset). We used Adam optimizer with weight decay equal to $1e-8$ and adopted the cross entropy loss function. Moreover, we set the n_{features} parameter according to the number of electrodes in the dataset recordings. Specifically, we set $n_{\text{features}} = 64$ for the BrainLat dataset, which has 128 channels, and $n_{\text{features}} = 32$ for all of the other datasets. This configuration was found through empirical testing to strike the best balance between performance and model dimensionality. Additional layer dimensions are described in Table 1.

5. Results and discussion

In this section we comment on the obtained results and discuss the relevance of this work in a clinical setting. Using the patient-level evaluation presented in Section 3.2, we are able to assess the quality of the work in a more realistic scenario, setting the groundwork for clinical applications.

Furthermore, we compare the computational cost of our solution with the state-of-the-art. This is particularly relevant in the context of the remote monitoring and early detection of the illness. In fact, these tasks require portable and low-power hardware [67], which our solution is suited to exploit.

[2] The list of subjects in each split of each dataset is provided at https://github.com/unimib-islab/MTDNet.

Table 3
Comparison of the different methodologies with the selected datasets. For each dataset, the average and standard deviation for the accuracy and F1 score are computed using the classification results of five models retrained with five different seeds. Bold for best results, underlined text for second best.

	ADSZ		APAVA		ADFTD-Binary		ADFTD		BrainLat-Binary	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
Segment-level Classification										
TimesNet	70.22 ± 15.75	67.94 ± 17.14	70.23 ± 3.23	68.74 ± 3.62	61.24 ± 1.33	60.93 ± 1.09	49.34 ± 3.49	47.19 ± 2.69	56.82 ± 2.76	56.32 ± 2.76
Crossformer	84.73 ± 0.72	84.57 ± 0.74	74.10 ± 2.34	69.07 ± 3.29	71.72 ± 1.90	71.43 ± 1.94	50.18 ± 1.62	45.22 ± 0.96	54.36 ± 1.58	53.67 ± 1.69
iTransformer	78.79 ± 1.93	78.60 ± 1.85	74.55 ± 1.85	72.30 ± 2.00	73.86 ± 0.69	73.47 ± 0.66	52.61 ± 1.77	46.80 ± 1.26	54.92 ± 1.12	54.61 ± 1.32
Medformer	85.39 ± 3.53	85.35 ± 3.50	76.38 ± 5.93	73.29 ± 7.06	66.31 ± 2.27	66.20 ± 2.32	52.14 ± 1.92	**50.08 ± 1.89**	59.21 ± 2.07	58.74 ± 1.93
LCADNet	81.65 ± 3.26	81.42 ± 3.28	74.16 ± 4.15	72.97 ± 4.55	63.16 ± 1.52	62.34 ± 1.26	52.38 ± 2.19	49.83 ± 1.86	52.00 ± 5.58	45.45 ± 4.30
MTDNet (Ours)	**93.08 ± 0.63**	**92.92 ± 0.63**	**86.00 ± 1.79**	**85.48 ± 1.76**	**74.60 ± 2.01**	**74.39 ± 1.91**	**56.45 ± 2.14**	46.86 ± 0.79	**61.31 ± 1.81**	**61.14 ± 1.85**
Patient-level Classification with Consensus (Major Voting)										
TimesNet	70.00 ± 18.71	62.98 ± 27.42	80.00 ± 11.18	78.67 ± 11.93	64.29 ± 5.05	63.88 ± 5.17	55.79 ± 2.88	52.13 ± 2.35	67.69 ± 6.44	67.01 ± 6.38
Crossformer	100.00 ± 0.00	**100.00 ± 0.00**	75.00 ± 0.00	73.33 ± 0.00	80.00 ± 3.19	79.91 ± 3.24	54.74 ± 6.00	44.08 ± 7.57	61.54 ± 9.42	57.23 ± 12.62
iTransformer	**100.00 ± 0.00**	**100.00 ± 0.00**	75.00 ± 0.00	73.33 ± 0.00	82.86 ± 3.91	82.64 ± 3.81	60.00 ± 4.71	46.49 ± 2.80	64.62 ± 4.21	63.06 ± 4.96
Medformer	96.00 ± 5.48	95.96 ± 5.53	75.00 ± 0.00	73.33 ± 0.00	75.71 ± 3.91	75.35 ± 3.87	55.79 ± 4.71	52.75 ± 5.86	67.69 ± 11.41	67.08 ± 11.13
LCADNet	94 ± 5.48	93.94 ± 5.53	80.00 ± 11.18	78.67 ± 11.93	70.00 ± 3.19	67.63 ± 2.83	61.05 ± 6.00	**57.37 ± 7.93**	55.39 ± 6.44	48.84 ± 3.21
MTDNet (Ours)	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**87.14 ± 5.98**	**86.99 ± 6.04**	**62.11 ± 4.40**	47.75 ± 4.07	**72.31 ± 6.88**	**71.78 ± 7.46**
Patient-level Classification with Average Score										
TimesNet	70.00 ± 18.71	62.98 ± 27.42	80.00 ± 11.18	78.67 ± 11.93	64.29 ± 5.05	63.88 ± 5.17	56.84 ± 4.40	53.02 ± 3.04	67.69 ± 6.44	67.01 ± 6.38
Crossformer	**100.00 ± 0.00**	**100.00 ± 0.00**	75.00 ± 0.00	73.33 ± 0.00	80.00 ± 3.19	79.91 ± 3.24	55.79 ± 6.00	44.88 ± 7.30	61.54 ± 9.42	57.23 ± 12.62
iTransformer	**100.00 ± 0.00**	**100.00 ± 0.00**	75.00 ± 0.00	73.33 ± 0.00	82.86 ± 3.91	82.64 ± 3.81	58.95 ± 4.40	45.71 ± 2.69	63.08 ± 6.44	61.65 ± 6.57
Medformer	94.00 ± 5.48	93.94 ± 5.53	75.00 ± 0.00	73.33 ± 0.00	77.14 ± 5.98	76.74 ± 5.96	56.84 ± 4.40	54.36 ± 5.57	66.15 ± 11.67	65.09 ± 11.90
LCADNet	90.00 ± 0.00	89.90 ± 0.00	80.00 ± 11.18	78.67 ± 11.93	70.00 ± 3.20	67.63 ± 2.83	61.05 ± 6.00	**57.37 ± 7.93**	56.93 ± 6.88	49.90 ± 2.82
MTDNet (Ours)	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**88.57 ± 6.39**	**88.47 ± 6.45**	**62.11 ± 2.35**	48.40 ± 0.99	**72.31 ± 6.88**	**71.78 ± 7.46**

Finally, we present an analysis of the convolutional activations to gain a deeper understanding of the model feature extraction capabilities.

5.1. Comparison with state-of-the-art

Table 3 reports comparisons between state-of-the-art models and the proposed model on each dataset selected. As can be seen, the proposed approach outperforms state-of-the-art models for all of the considered datasets. The most significant results were obtained on the original ADFTD dataset: the proposed model achieved around 3.8% higher accuracy than the second best model. For the two-class datasets, our model outperforms state-of-the-art approaches in terms of both accuracy and F1 score. Generally speaking, our model outperforms all approaches; second place is typically achieved by Medformer or iTransformer at segment-level classification. As shown in Fig. 3, not only does our approach achieve better results than existing methods, it also does it at a lower computational cost and with a smaller dimensionality footprint (see Section 5.2 for more details).

When considering classification strategies for the entire recording, our study shows that ADSZ and APAVA stand out as the simplest datasets due to the very small number of subjects. As can be seen, most models tend to correctly classify all subjects when classifying entire recordings. Despite the lower number of subjects with respect to ADSZ, APAVA appears more challenging. Still, for patient-level classification, our model is the only one able to correctly classify all subjects. ADFTD (both versions) and BrainLat, however, still present a challenge in the classification task. Nevertheless, even for these two datasets, our proposed solution outperforms state-of-the-art methods.

5.2. Analysis on computational costs

To better analyze the efficiency of the proposed solution, a comparison in terms of model dimensionality and computational complexity has been performed. Here we counted the number of parameters of each approach and computed the total amount of Multiply–Accumulate Operations (MACs) Floating Point Operations (FLOPs) and memory usage as measures of models complexity. In particular, MACs and FLOPs are computed using an input segment of 1 s (256 data samples and 19 channels). Table 4 shows the results of such computation, while Fig. 3 represents the tradeoff between FLOPs, number of parameters and accuracy on each of the used datasets. As can be seen TimesNet represents the most cost expensive model, with around 113M parameters and ≈ 58 GFLOPs. Fig. 3 shows also how heavy is this model in comparison to all the others, underlining how expensive the model is considering the overall accuracy achieved over all the datasets. LCADNet is the second best lightweight solution in terms of MACs and FLOPs, however the model present a very high number of parameters and show low accuracies on all of the considered datasets.

Table 4
Comparison of models in terms of number of parameters and computational complexity. For each model are reported, alongside its dimensionality, the Multiply–Accumulate Operations (MACs), Floating Point Operations (FLOPs), and Memory Usage of model parameters, computed on a segment of 1 s (256 data samples) with 19 channels. The lower the better.

	# Params (M)	MACs (G)	FLOPs (G)	Memory Usage (KB)
TimesNet	112.58	28.7925	57.5833	450 318
Crossformer	2.16	0.2445	0.4911	8649
iTransformer	0.84	0.0157	0.0323	3341
Medformer	7.97	0.3341	0.7404	31 895
LCADNet	36.3	0.0148	0.0018	145 200
MTDNet (Ours)	**0.02**	**0.0015**	**0.0018**	**82**

On the other hand, the proposed solution shows a very low impact in terms of model dimensionality and computational complexity, with a total amount of 20.5K of parameters and only 1.5M MACs and 1.8M FLOPs. Despite its lightness, the model shows higher accuracy than most of the state-of-the-art approaches and a much lower computational cost, exhibiting the best tradeoff in terms of accuracy and computational costs. MTDNet shows strong potential for integration within portable and non-invasive BCI systems for Alzheimer's Disease diagnosis. Its lightweight design and high efficiency make it suitable for real-time EEG decoding even on low-power embedded devices, facilitating deployment in home-care or clinical screening environments. The proposed approach aligns with current advances in neural manifold decoding frameworks [68], which demonstrate how low-dimensional EEG representation can effectively capture cognitive impairments and sensorimotor processes.

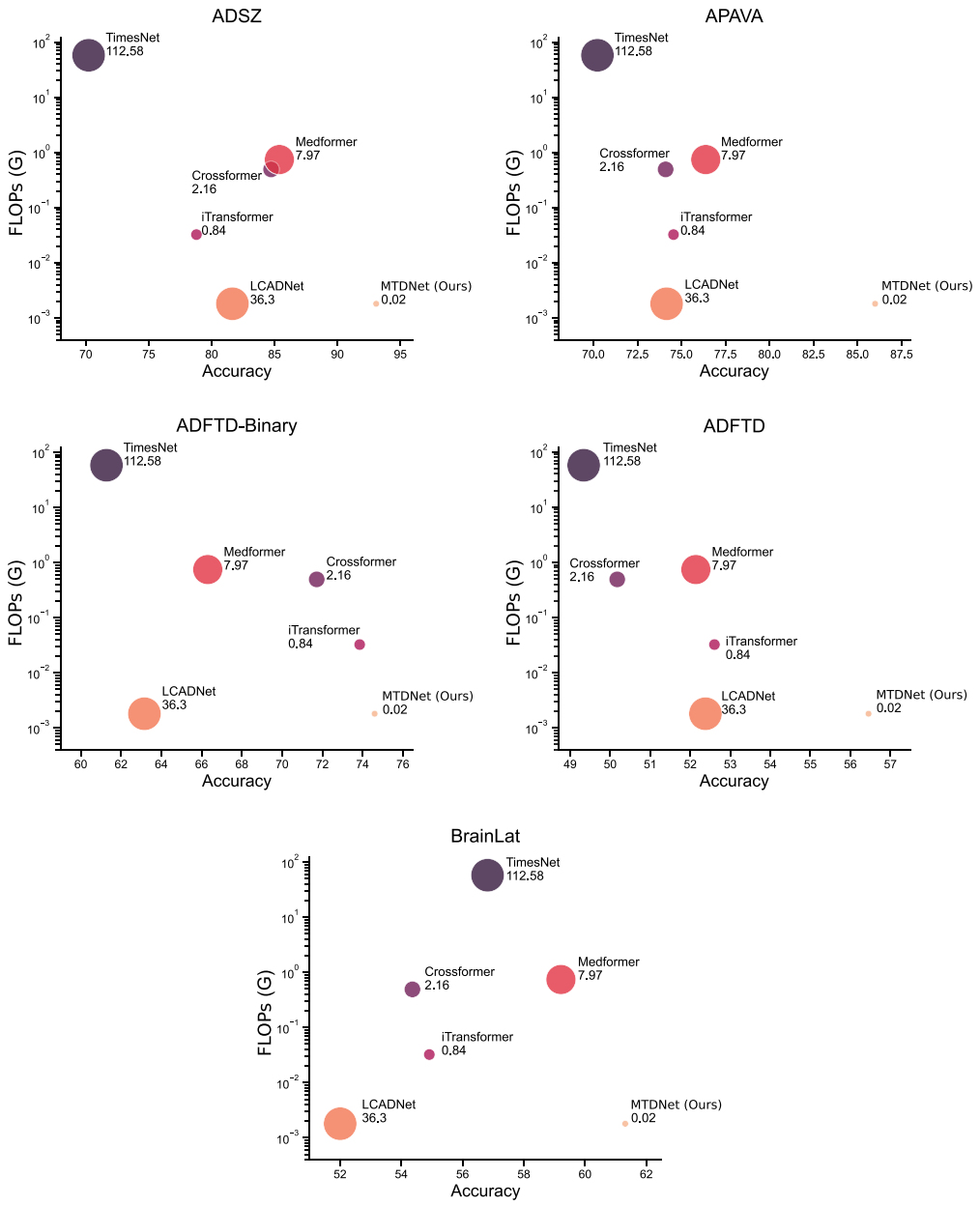

Fig. 3. Comparison of models performance in terms of accuracy and computational complexity measured in Floating Point Operations (FLOPs). The area of each circle is proportional to the number of parameters of each model (Millions of parameters), which is reported below each model name. The best models live in the lower right part of the plots, i.e. low number of parameters and FLOPs while having high accuracy.

5.3. Ablation study on data augmentation

In order to reduce the overfitting risk and to enhance the generalization capabilities of the model, we adopted simple data augmentation steps in the training phase. Specifically on each T second segment three augmentation functions are randomly applied: (i) time flipping, (ii) electrode channel shuffling, and (iii) dropout on input values (random zero values). These three transformations could occur all at the same time: each transformation has a probability of 50% to be applied to the input before feeding them to the model during training phase. In order to determine the impact of the different data augmentation strategies we performed an ablation study by retraining the model with different combination using ADFTD-binary datasets. We chose this one since is the largest one considered in terms of total recording length. The model has been retrained with each combination with five different seeds as done for the main experiment. Results of the ablation in terms of accuracy and F1-Score are reported in Table 5.

As can be seen, the use of augmentation techniques brings a huge benefit to the final trained model, giving a boost of $\approx .6\%$ in accuracy and $\approx 8\%$ in F1-Score. While all the techniques brought benefits in the training procedure, the highest improvements are given by the *Channel Shuffling* procedure. Even if the use of the *Flipping* and *Input dropout* techniques produces a marginal improvement, we used them in combination with the *Channel Shuffling*, combining the benefits coming from all of the augmentations techniques.

5.4. Ablation study on scales contribution

To assess the contribution given by the different scales defined in the proposed architecture, we performed an ablation study to determine their impact when used alone and in different combinations. The results are presented in Table 6.

We trained the proposed model on both versions of the ADFTD dataset with different combinations of scales used for the Temporal

Table 5
Ablation study on data augmentation techniques: Accuracy and F1-Score values obtained by retraining the model on the ADFTD-Binary dataset with every combination of Flip, Channel shuffling and Input dropout.

Ablation with ADFTD-Binary				
Flip	Channel shuffling	Input dropout	Accuracy	F1-Score
–	–	–	66.61 ± 2.11	65.92 ± 2.42
✓	–	–	67.72 ± 2.44	66.78 ± 2.94
–	✓	–	73.21 ± 2.35	72.95 ± 2.37
–	–	✓	67.83 ± 1.57	67.00 ± 1.87
✓	✓	–	72.41 ± 1.02	72.19 ± 1.04
✓	–	✓	68.99 ± 2.12	68.37 ± 2.44
–	✓	✓	73.17 ± 2.19	72.89 ± 2.29
✓	✓	✓	**74.60 ± 2.01**	**74.39 ± 1.91**

Table 6
Ablation study on the impact of features extracted at different scales: Accuracy and F1-Scores of different combinations of scales on ADFTD dataset (both in the binary and ternary versions).

Scales			ADFTD-Binary		ADFTD	
1	2	3	ACC	F1	ACC	F1
✓	–	–	73.95 ± 4.55	73.58 ± 4.73	54.28 ± 2.47	46.76 ± 2.38
–	✓	–	70.46 ± 3.26	70.17 ± 3.52	52.79 ± 3.42	44.97 ± 3.29
–	–	✓	68.48 ± 2.73	68.10 ± 2.81	52.20 ± 2.88	44.63 ± 1.22
✓	✓	–	71.94 ± 4.19	71.74 ± 4.11	50.18 ± 2.40	44.06 ± 2.46
–	✓	✓	73.07 ± 1.88	72.54 ± 2.19	54.29 ± 1.98	47.33 ± 1.80
✓	–	✓	73.97 ± 2.44	73.82 ± 2.37	55.34 ± 1.63	**47.45 ± 1.55**
✓	✓	✓	**74.60 ± 2.01**	**74.39 ± 1.91**	**56.45 ± 2.14**	46.86 ± 0.79

Features Extraction phase. The first three rows correspond to the three versions of the model where a single scale at the time is used, while the others are different combinations of scales at the same time. Finally, the last row reports the results of the proposed model which uses all of the scales simultaneously.

As can be seen in Table 6, the combination of multiple scales together brings benefits in terms of accuracy, F1-Score and also model stability. Among the single-scale models, the first and largest scale is the one that achieves the best results, at the expense of overall stability, as shown by the highest standard deviation values. In general, the adoption of multiple scales mitigates such phenomenon, in particular when using the third scale in combination with the other two. The combination of scales brings overall improvements in both performances and stability. More prominent improvements can be observed on the ADFTD version of the dataset, where the combination of the first and third scales brings the best results among the observed versions. This phenomenon demonstrate the actual contribution of very different scales in the features extraction procedure. Finally, the results obtained by the combination of all the considered scales shows the best results overall, confirming the contribution of each scale to the final trained model. A final note on the F1-Score obtained by the final model on ADFTD: as can be seen the model achieved a slightly lower score, while drastically reducing standard deviation. This confirms how applying multiple scales can improve model stability (same improvement can be also noticed with ADFTD-Binary) while improving also generalization capabilities, as demonstrated by the results obtained on multiple datasets.

5.5. Discussion on multi-scale layers activations

In order to analyze the impact of using a multiscale approach further and improve our understanding of the EEG data, we examined the features extracted by the first three convolutions. These features were taken after the alignment procedure and before the concatenation operation was performed. This analysis was performed on the test set of the ADFTD-Binary dataset, as it is the most challenging of the datasets considered.

For each scale, we applied the Fast Fourier Transform (FFT) to the three sets of features, each of size $n_{features} \times t^n$, using the Scipy v1.14.1 function, in order to study the frequency components of each feature. We then computed the average frequency from each FFT coefficient array extracted from the activations of each convolutional kernel. Fig. 4 shows the boxplots obtained for each scale and kernel, divided by label, with all features extracted from each sequence. Fig. 5 also shows the boxplot obtained using all the samples in different combinations based on classification accuracy. As can be seen, the extracted features generally differ greatly between the two classes, demonstrating the model's ability to discriminate between them. Overall, features in healthy patients tend to appear within certain ranges with low standard deviation values. In contrast, features appear much less frequently in patients with Alzheimer's disease and show a high standard deviation. Generally, we can assume that patients with Alzheimer's disease exhibit features similar to those of healthy individuals, but these features tend to manifest with a notably lower frequency compared to the other class. The multiscale temporal feature extraction captures both high-frequency oscillatory alterations (α and β) and slow-wave dynamics (δ and θ), which are known biomarkers of cognitive impairment [69]. Differences can also be observed between scales.

Fig. 5 illustrates a significant comparison between the "All" case, which includes all segments, and the "HC Wrong" case, where healthy subjects are misclassified as Alzheimer's patients. Analyzing these two cases reveals that the most noticeable characteristic is the shift in statistics when considering samples that were incorrectly classified as patients with Alzheimer's disease. This shift in the mean and median values represents a lower overall frequency of appearance of the features for these specific samples. Note that in the "AD Wrong" case, which shows the distribution of AD class samples incorrectly classified as healthy, there is no increase in mean or median frequencies.

One possible explanation for this phenomenon is that, when considering a one-second time sample, for example, the brain activity of the two types of patient may be the same, regardless of their disease status. These specific segments would not contain useful information for the classification task, resulting in poorer classification accuracy and incorrect training outcomes. Therefore, some samples are irrelevant for classifying a patient's condition. If these samples could be identified, they could be excluded from the classification pipeline, thereby improving the training procedure, inference, and potentially leading to significant performance enhancements.

6. Conclusions

In this work we proposed a novel and lightweight approach for Alzheimer's Disease classification based on EEG signals, called MTDNet, leveraging a multiscale temporal feature extractor combined with a simple LSTM-based classifier. Our proposed method is specifically designed to address the challenges posed by the limited data availability and the high computational cost of current state-of-the-art deep learning models, such as Transformer-based architectures. To better align the evaluation protocol with clinical utility we also introduced two patient-level classification strategies based on aggregating the predictions from EEG segments, that allow a more robust diagnosis.

Through extensive experiments on four publicly available datasets, we demonstrated that our model achieves competitive or superior performance compared to recent state-of-the-art methods with an accuracy improvement of $\sim 2\%$ at the segment classification level and $\sim 6\%$ at the patient-level on the most complex datasets such as ADFTD and BrainLat. Furthermore, unlike recent transformer-based approaches, which are computationally intensive, MTDNet achieves higher accuracy with just 20.5K parameters and 1.8M FLOPs. Our model exhibits remarkable stability and generalization capabilities across datasets, even when compared to more complex and resource-intensive approaches.

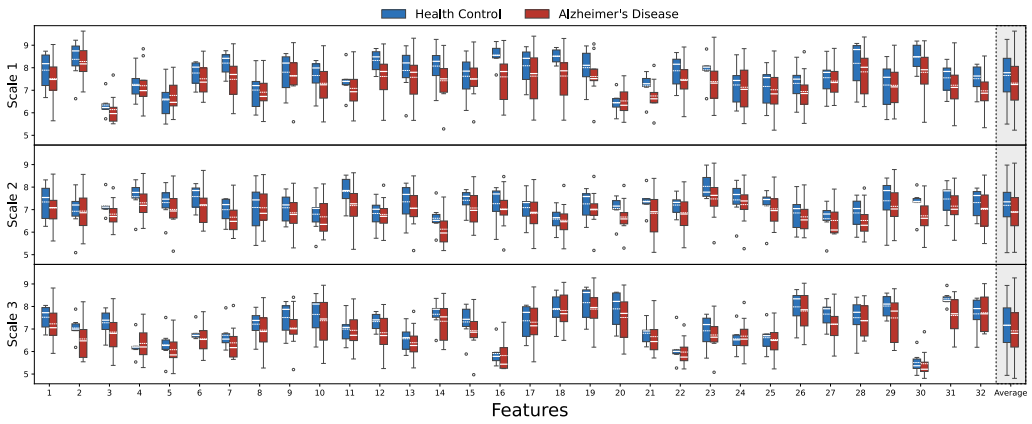

Fig. 4. Analysis of the frequency of the activations per kernel and scale. For each scale, for each convolutional kernel, it is shown a boxplot computed over the FFT transform of the activations for all the sample of each class. The boxplot shows the mean and average values respectively with a full and a dotted white line. This visualization is computed over the ADFTD-Binary dataset. The last column shows the average over all the activations, regardless of the kernel.

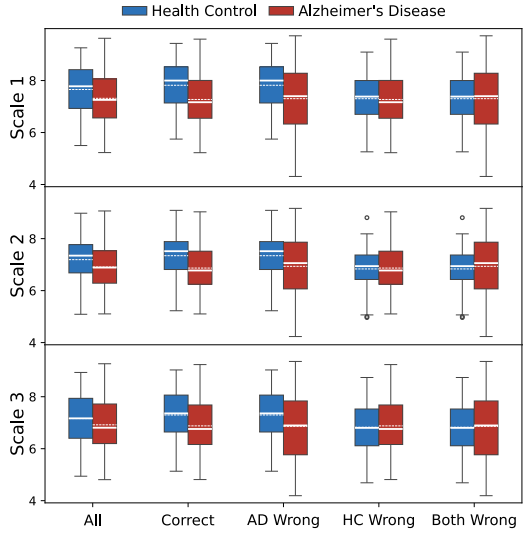

Fig. 5. Analysis of average activation frequencies per scale. The analysis is reported for each scale, with samples selected based on their classification. All: all samples are used; Correct: Only correctly classified samples are used; AD Wrong: Only AD class samples incorrectly classified as healthy are used; HC Wrong: Only HC class samples incorrectly classified as Alzheimer's patients are used; Both Wrong: Only incorrectly classifies samples are used.

The ablation studies conducted further confirmed the importance of the multiscale feature extraction strategy and highlighted the contribution of simple yet effective data augmentation techniques in improving model performance and preventing overfitting. Our findings underline that combining multiscale convolutional feature extraction with temporal modeling is a powerful and efficient strategy for EEG-based Alzheimer's Disease classification. The biggest shortcoming of the current approach relates to the patient-level evaluation, which, in its current state, is effective but still very basic. Guided by the analyses in Section 5.5, future work will focus on introducing and studying strategies that focus on the actual dynamics of the signals in the recording, with the aim of developing more specific strategies for patient-level classification, such as segment selection and weighted voting systems.

CRediT authorship contribution statement

Simone Zini: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thomas Barbera:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Simone Bianco:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Paolo Napoletano:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Paolo Napoletano reports financial support was provided by PNC0000003. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially funded by the National Plan for NRRP Complementary Investments (PNC, established with the decree-law 6 May 2021, n. 59, converted by law n. 101 of 2021) in the call for the funding of research initiatives for technologies and innovative trajectories in the health and care sectors (Directorial Decree n. 931 of 06-06-2022) - project n. PNC0000003 - AdvaNced Technologies for Human-centrEd Medicine (project acronym: ANTHEM).[3] This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

Data availability

Data used is already available on the web.

[3] https://fondazioneanthem.it/

References

[1] Dementia — who.int, 2025, https://www.who.int/news-room/fact-sheets/detail/dementia, (Accessed 04 April 2025).

[2] V. Berti, A. Pupi, L. Mosconi, Pet/ct in diagnosis of dementia, Ann. New York Acad. Sci. 1228 (1) (2011) 81–92, http://dx.doi.org/10.1111/j.1749-6632.2011.06015.x.

[3] S.L. Risacher, A.J. Saykin, J.D. West, L. Shen, H.A. Firpi, B.C. McDonald, Baseline mri predictors of conversion from mci to probable ad in the adni cohort, Curr. Alzheimer Res. 6 (4) (2009) 347–361, http://dx.doi.org/10.2174/156720509788929273.

[4] M. Machulda, H. Ward, B. Borowski, J. Gunter, R. Cha, P. O'Brien, R. Petersen, B. Boeve, D. Knopman, D. Tang-Wai, R. Ivnik, G. Smith, E. Tangalos, C. Jack Jr., Comparison of memory fmri response among normal, mci, and alzheimer's patients, Neurology 61 (4) (2003) 500–506, http://dx.doi.org/10.1212/01.WNL.0000079052.01016.78.

[5] K. Ishii, Pet approaches for diagnosis of dementia, Am. J. Neuroradiol. 35 (11) (2014) 2030–2038, http://dx.doi.org/10.3174/ajnr.A3695.

[6] M. Firbank, R. Harrison, J. O'Brien, A comprehensive review of proton magnetic resonance spectroscopy studies in dementia and parkinson's disease, Dement. Geriatr. Cogn. Disord. 14 (2) (2002) 64–76, http://dx.doi.org/10.1159/000064927.

[7] J. Bosboom, D. Stoffers, E. Wolters, C. Stam, H. Berendse, Meg resting state functional connectivity in parkinson's disease related dementia, J. Neural Transm. 116 (2) (2009) 193–202, http://dx.doi.org/10.1007/s00702-008-0132-6.

[8] I. Bazarbekov, A. Razaque, M. Ipalakova, J. Yoo, Z. Assipova, A. Almisreb, A review of artificial intelligence methods for alzheimer's disease diagnosis: Insights from neuroimaging to sensor data analysis, Biomed. Signal Process. Control. 92 (2024) http://dx.doi.org/10.1016/j.bspc.2024.106023.

[9] N.K. Al-Qazzaz, S.H.B.M. Ali, S.A. Ahmad, K. Chellappan, M.S. Islam, J. Escudero, Role of eeg as biomarker in the early detection and classification of dementia, Sci. World J. 2014 (2014) http://dx.doi.org/10.1155/2014/906038.

[10] E. Perez-Valero, M.A. Lopez-Gordo, C. Morillas, F. Pelayo, M.A. Vaquero-Blasco, A review of automated techniques for assisting the early detection of alzheimer's disease with a focus on eeg, J. Alzheimer's Dis. 80 (4) (2021) 1363–1376, http://dx.doi.org/10.3233/JAD-201455.

[11] F.C. Morabito, M. Campolo, C. Ieracitano, J.M. Ebadi, L. Bonanno, A. Bramanti, S. De Salvo, N. Mammone, P. Bramanti, Deep convolutional neural networks for classification of mild cognitive impaired and alzheimer's disease patients from scalp eeg recordings, in: 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow, RTSI 2016, Institute of Electrical and Electronics Engineers Inc., 2016, http://dx.doi.org/10.1109/RTSI.2016.7740576.

[12] E. Capecci, Z.G. Doborjeh, N. Mammone, F. La Foresta, F.C. Morabito, N. Kasabov, Longitudinal study of alzheimer's disease degeneration through eeg data analysis with a neucube spiking neural network model, in: Proceedings of the International Joint Conference on Neural Networks, 2016-October, Institute of Electrical and Electronics Engineers Inc., 2016, pp. 1360–1366, http://dx.doi.org/10.1109/IJCNN.2016.7727356.

[13] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, F.C. Morabito, A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings, Neurocomputing 323 (2019) 96–107, http://dx.doi.org/10.1016/j.neucom.2018.09.071.

[14] M. j. Kim, Y.C. Youn, J. Paik, Deep learning-based eeg analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset, NeuroImage 272 (2023) http://dx.doi.org/10.1016/j.neuroimage.2023.120054.

[15] Y. Zhang, J. Yan, Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting, in: 11th International Conference on Learning Representations, ICLR 2023, International Conference on Learning Representations, ICLR, 2023.

[16] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, M. Long, Itransformer: Inverted transformers are effective for time series forecasting, in: 12th International Conference on Learning Representations, ICLR 2024, International Conference on Learning Representations, ICLR, 2024.

[17] Y. Wang, N. Huang, T. Li, Y. Yan, X. Zhang, Medformer: A multi-granularity patching transformer for medical time-series classification, in: Advances in Neural Information Processing Systems, Neural information processing systems foundation, 2024, p. 37.

[18] T. Barbera, J. Burger, A. D'Amelio, S. Zini, S. Bianco, R. Lanzarotti, P. Napoletano, G. Boccignone, J.L. Contreras-Vidal, On using ai for eeg-based bci applications: Problems, current challenges and future trends, Int. J. Hum.-Comput. Interact. (2025) http://dx.doi.org/10.1080/10447318.2025.2561185.

[19] J. Yin, J. Cao, S. Siuly, H. Wang, An integrated mci detection framework based on spectral–temporal analysis, Int. J. Autom. Comput. 16 (6) (2019) 786–799, http://dx.doi.org/10.1007/s11633-019-1197-4.

[20] C. Ieracitano, N. Mammone, A. Bramanti, S. Marino, A. Hussain, F.C. Morabito, A time-frequency based machine learning system for brain states classification via eeg signal processing, in: Proceedings of the International Joint Conference on Neural Networks, 2019-July, Institute of Electrical and Electronics Engineers Inc., 2019, http://dx.doi.org/10.1109/IJCNN.2019.8852240.

[21] M.S. Safi, S.M.M. Safi, Early detection of alzheimer's disease from eeg signals using hjorth parameters, Biomed. Signal Process. Control. 65 (2021) http://dx.doi.org/10.1016/j.bspc.2020.102338.

[22] T.M. Rutkowski, M.S. Abe, T. Komendzinski, M. Otake-Matsuura, Older adult mild cognitive impairment prediction from multiscale entropy eeg patterns in reminiscent interior image working memory paradigm, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2021-January, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 6345–6348, http://dx.doi.org/10.1109/EMBC46164.2021.9629480.

[23] W. Fruehwirt, P. Zhang, M. Gerstgrasser, D. Grossegger, R. Schmidt, T. Benke, P. Dal-Bianco, G. Ransmayr, L. Weydemann, H. Garn, M. Waser, M. Osborne, G. Dorffner, Bayesian gaussian process classification from event-related brain potentials in alzheimer's disease, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10259 LNAI, 2017, pp. 65–75, http://dx.doi.org/10.1007/978-3-319-59758-4_7.

[24] G. Biagetti, P. Crippa, L. Falaschetti, S. Luzzi, C. Turchetti, Classification of alzheimer's disease from eeg signal using robust-pca feature extraction, in: Procedia Computer Science, vol. 192, Elsevier B.V., 2021, pp. 3114–3122, http://dx.doi.org/10.1016/j.procs.2021.09.084.

[25] I. Chouvarda, L. Mpaltadoros, I. Boutziona, G.N. Tsakonas, M. Tsolaki, K. Diamantaras, Exploring classification in open and closed eyes eeg data for people with cognitive disorders, in: Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - BIOSIGNALS, 2022, pp. 298–305, http://dx.doi.org/10.5220/0011010100003123, SciTePress.

[26] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, S. Luzzi, C. Turchetti, Eeg-based alzheimer's disease recognition using robust-pca and lstm recurrent neural network, Sensors 22 (10) (2022) http://dx.doi.org/10.3390/s22103696.

[27] D. Klepl, F. He, M. Wu, D.J. Blackburn, P. Sarrigiannis, Eeg-based graph neural network classification of alzheimer's disease: An empirical evaluation of functional connectivity methods, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2651–2660, http://dx.doi.org/10.1109/TNSRE.2022.3204913.

[28] G. Tavares, R. San-Martin, J.N. Ianof, R. Anghinah, F.J. Fraga, Improvement in the automatic classification of alzheimer's disease using eeg after feature selection, in: Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 2019-October, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 1264–1269, http://dx.doi.org/10.1109/SMC.2019.8914006.

[29] D. Pirrone, E. Weitschek, P. Di Paolo, S. De Salvo, M.C. De Cola, Eeg signal processing and supervised machine learning to early diagnose alzheimer's disease, Appl. Sci. (Switzerland) 12 (11) (2022) http://dx.doi.org/10.3390/app12115413.

[30] R. Cassani, T.H. Falk, Alzheimer's disease diagnosis and severity level detection based on electroencephalography modulation spectral 'patch' features, IEEE J. Biomed. Health Informat. 24 (7) (2020) 1982–1993, http://dx.doi.org/10.1109/JBHI.2019.2953475.

[31] K. Alsharabi, Y. Bin Salamah, A.M. Abdurraqeeb, M. Aljalal, F.A. Alturki, Eeg signal processing for alzheimer's disorders using discrete wavelet transform and machine learning approaches, IEEE Access 10 (2022) 89781–89797, http://dx.doi.org/10.1109/ACCESS.2022.3198988.

[32] G. Fiscon, E. Weitschek, M.C. De Cola, G. Felici, P. Bertolazzi, An integrated approach based on eeg signals processing combined with supervised methods to classify alzheimer's disease patients, in: Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 2750–2752, http://dx.doi.org/10.1109/BIBM.2018.8621473.

[33] G. Fiscon, E. Weitschek, G. Felici, P. Bertolazzi, S. De Salvo, P. Bramanti, M.C. De Cola, Alzheimer's disease patients classification through eeg signals processing, in: IEEE SSCI 2014-2014 IEEE Symposium Series on Computational Intelligence - CIDM 2014: 2014 IEEE Symposium on Computational Intelligence and Data Mining, Proceedings, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 105–112, http://dx.doi.org/10.1109/CIDM.2014.7008655.

[34] D.V. Puri, J.P. Gawande, P.H. Kachare, I. Al-Shourbaji, Optimal time-frequency localized wavelet filters for identification of alzheimer's disease from eeg signals, Cogn. Neurodynamics 19 (1) (2025) http://dx.doi.org/10.1007/s11571-024-10198-7.

[35] D. Puri, P. Kachare, S. Khare, I. Al-Shourbaji, A. Jabbari, A. Alameen, Hybrid reptile-snake optimizer based channel selection for enhancing alzheimer's disease detection, J. Bionic Eng. 22 (2) (2025) 884–900, http://dx.doi.org/10.1007/s42235-024-00636-x.

[36] C. Ieracitano, N. Mammone, A. Hussain, F.C. Morabito, A novel multi-modal machine learning based approach for automatic classification of eeg recordings in dementia, Neural Netw. 123 (2020) 176–190, http://dx.doi.org/10.1016/j.neunet.2019.12.006.

[37] A. Miltiadous, K.D. Tzimourta, N. Giannakeas, M.G. Tsipouras, T. Afrantou, P. Ioannidis, A.T. Tzallas, Alzheimer's disease and frontotemporal dementia: A robust classification method of eeg signals and a comparison of validation methods, Diagnostics 11 (8) (2021) http://dx.doi.org/10.3390/diagnostics11081437.

[38] D. Geng, C. Wang, Z. Fu, Y. Zhang, K. Yang, H. An, Sleep eeg-based approach to detect mild cognitive impairment, Front. Aging Neurosci. 14 (2022) http://dx.doi.org/10.3389/fnagi.2022.865558.

[39] N. Sharma, M. Kolekar, K. Jha, Y. Kumar, Eeg and cognitive biomarkers based mild cognitive impairment diagnosis, IRBM 40 (2) (2019) 113–121, http://dx.doi.org/10.1016/j.irbm.2018.11.007.

[40] B. Oltu, M.F. Akşahin, S. Kibaroğlu, A novel electroencephalography based approach for alzheimer's disease and mild cognitive impairment detection, Biomed. Signal Process. Control. 63 (2021) http://dx.doi.org/10.1016/j.bspc.2020.102223.

[41] M. Kashefpoor, H. Rabbani, M. Barekatain, Supervised dictionary learning of eeg signals for mild cognitive impairment diagnosis, Biomed. Signal Process. Control. 53 (2019) http://dx.doi.org/10.1016/j.bspc.2019.101559.

[42] F. Duan, Z. Huang, Z. Sun, Y. Zhang, Q. Zhao, A. Cichocki, Z. Yang, J. Sole-Casals, Topological network analysis of early alzheimer's disease based on resting-state eeg, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (10) (2020) 2164–2172, http://dx.doi.org/10.1109/TNSRE.2020.3014951.

[43] X. Li, C. Yang, P. Xie, Y. Han, R. Su, Z. Li, Y. Liu, The diagnosis of amnestic mild cognitive impairment by combining the characteristics of brain functional network and support vector machine classifier, J. Neurosci. Methods 363 (2021) http://dx.doi.org/10.1016/j.jneumeth.2021.109334.

[44] N. Youssef, S. Xiao, M. Liu, H. Lian, R. Li, X. Chen, W. Zhang, X. Zheng, Y. Li, Y. Li, Functional brain networks in mild cognitive impairment based on resting electroencephalography signals, Front. Comput. Neurosci. 15 (2021) http://dx.doi.org/10.3389/fncom.2021.698386.

[45] X. Shan, J. Cao, S. Huo, L. Chen, P.G. Sarrigiannis, Y. Zhao, Spatial–temporal graph convolutional network for alzheimer classification based on brain functional connectivity imaging of electroencephalogram, Hum. Brain Mapp. 43 (17) (2022) 5194–5209, http://dx.doi.org/10.1002/hbm.25994.

[46] G. Pfurtscheller, F. Lopes Da Silva, Event-related eeg/meg synchronization and desynchronization: Basic principles, Clin. Neurophysiol. 110 (11) (1999) 1842–1857, http://dx.doi.org/10.1016/S1388-2457(99)00141-8.

[47] H. Yu, X. Wu, L. Cai, B. Deng, J. Wang, Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (5) (2018) 977–986, http://dx.doi.org/10.1109/TNSRE.2018.2828143.

[48] H. Yu, F. Li, J. Liu, D. Liu, H. Guo, J. Wang, G. Li, Evaluation of acupuncture efficacy in modulating brain activity with periodic-aperiodic eeg measurements, IEEE Trans. Neural Syst. Rehabil. Eng. 32 (2024) 2450–2459, http://dx.doi.org/10.1109/TNSRE.2024.3421648.

[49] J.C. McBride, X. Zhao, N.B. Munro, C.D. Smith, G.A. Jicha, L. Hively, L.S. Broster, F.A. Schmitt, R.J. Kryscio, Y. Jiang, Spectral and complexity analysis of scalp eeg characteristics for mild cognitive impairment and early alzheimer's disease, Comput. Methods Programs Biomed. 114 (2) (2014) 153–163, http://dx.doi.org/10.1016/j.cmpb.2014.01.019.

[50] S. Siuly, O.F. Alcin, E. Kabir, A. Sengur, H. Wang, Y. Zhang, F. Whittaker, A new framework for automatic detection of patients with mild cognitive impairment using resting-state eeg signals, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (9) (2020) 1966–1976, http://dx.doi.org/10.1109/TNSRE.2020.3013429.

[51] H. Yu, X. Lei, Z. Song, C. Liu, J. Wang, Supervised network-based fuzzy learning of eeg signals for alzheimer's disease identification, IEEE Trans. Fuzzy Syst. 28 (1) (2020) 60–71, http://dx.doi.org/10.1109/TFUZZ.2019.2903753.

[52] F. Li, S. Matsumori, N. Egawa, S. Yoshimoto, K. Yamashiro, H. Mizutani, N. Uchida, A. Kokuryu, A. Kuzuya, R. Kojima, Y. Hayashi, R. Takahashi, Predictive diagnostic approach to dementia and dementia subtypes using wireless and mobile electroencephalography: A pilot study, Bioelectricity 4 (1) (2022) 3–11, http://dx.doi.org/10.1089/bioe.2021.0030.

[53] Ö.K. Cura, G.C. Yilmaz, H.S. Ture, A. Akan, Deep time-frequency feature extraction for alzheimer's dementia eeg classification, in: TIPTEKNO 2022 - Medical Technologies Congress, Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022, http://dx.doi.org/10.1109/TIPTEKNO56568.2022.9960155.

[54] P. Kachare, D. Puri, S.B. Sangle, I. Al-Shourbaji, A. Jabbari, R. Kirner, A. Alameen, H. Migdady, L. Abualigah, Lcadnet: a novel light cnn architecture for eeg-based alzheimer disease detection, Phys. Eng. Sci. Med. 47 (3) (2024) 1037–1050, http://dx.doi.org/10.1007/s13246-024-01425-w.

[55] D.V. Puri, P.H. Kachare, S.B. Sangle, R. Kirner, A. Jabbari, I. Al-Shourbaji, M. Abdalraheem, A. Alameen, Leadnet: Detection of alzheimer's disease using spatiotemporal eeg analysis and low-complexity cnn, IEEE Access 12 (2024) 113888–113897, http://dx.doi.org/10.1109/ACCESS.2024.3435768.

[56] S. Sridhar, V. Manian, Eeg and deep learning based brain cognitive function classification, Computers 9 (4) (2020) 1–18, http://dx.doi.org/10.3390/computers9040104.

[57] T. Barbera, S. Zini, S. Bianco, P. Napoletano, Lightweight graph neural network for dementia assessment from eeg recordings, in: 8th IEEE International Forum on Research and Technologies for Society and Industry Innovation, RTSI 2024 - Proceeding, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 190–195, http://dx.doi.org/10.1109/RTSI61910.2024.10761763.

[58] D. Klepl, F. He, M. Wu, D.J. Blackburn, P. Sarrigiannis, Adaptive gated graph convolutional network for explainable diagnosis of alzheimer's disease using eeg data, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2023) 3978–3987, http://dx.doi.org/10.1109/TNSRE.2023.3321634.

[59] E. Sibilano, A. Brunetti, D. Buongiorno, M. Lassi, A. Grippo, V. Bessi, S. Micera, A. Mazzoni, V. Bevilacqua, An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state eeg, J. Neural Eng. 20 (1) (2023) http://dx.doi.org/10.1088/1741-2552/acb96e.

[60] F.C. Morabito, M. Campolo, D. Labate, G. Morabito, L. Bonanno, A. Bramanti, S. De Salvo, A. Marra, P. Bramanti, A longitudinal eeg study of alzheimer's disease progression based on a complex network approach, Int. J. Neural Syst. 25 (2) (2015) http://dx.doi.org/10.1142/S0129065715500057.

[61] E.P. Scheijbeler, W. de Haan, C.J. Stam, J.W.R. Twisk, A.A. Gouw, Longitudinal resting-state eeg in amyloid-positive patients along the alzheimer's disease continuum: considerations for clinical trials, Alzheimer's Res. Ther. 15 (1) (2023) http://dx.doi.org/10.1186/s13195-023-01327-1.

[62] C.L. Alves, A.M. Pineda, K. Roster, C. Thielemann, F.A. Rodrigues, Eeg functional connectivity and deep learning for automatic diagnosis of brain disorders: Alzheimer's disease and schizophrenia, J. Phys.: Complex. 3 (2) (2022) http://dx.doi.org/10.1088/2632-072X/ac5f8d.

[63] J. Escudero, D. Abásolo, R. Hornero, P. Espino, M. López, Analysis of electroencephalograms in alzheimer's disease patients with multiscale entropy, Physiol. Meas. 27 (11) (2006) http://dx.doi.org/10.1088/0967-3334/27/11/004.

[64] A. Miltiadous, K.D. Tzimourta, T. Afrantou, P. Ioannidis, N. Grigoriadis, D.G. Tsalikakis, P. Angelidis, M.G. Tsipouras, E. Glavas, N. Giannakeas, A.T. Tzallas, A dataset of scalp eeg recordings of alzheimer's disease, Front. Dement. Heal. Subj. from Routine Eeg, Data 8 (6) (2023) http://dx.doi.org/10.3390/data8060095.

[65] P. Prado, V. Medel, R. Gonzalez-Gomez, A. Sainz-Ballesteros, V. Vidal, H. Santamaría-García, S. Moguilner, J. Mejia, A. Slachevsky, M.I. Beherens, D. Aguillon, F. Lopera, M.A. Parra, D. Matallana, M.A. Maito, A.M. Garcia, N. Custodio, A.Á. Funes, S. Piña-Escudero, A. Birba, S. Fittipaldi, A. Legaz, A. Ibañez, The brainlat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds, Sci. Data 10 (1) (2023) http://dx.doi.org/10.1038/s41597-023-02806-8.

[66] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, M. Long, Timesnet: Temporal 2d-variation modeling for general time series analysis, in: 11th International Conference on Learning Representations, ICLR 2023, International Conference on Learning Representations, ICLR, 2023.

[67] S. Frey, M. Guermandi, S. Benatti, V. Kartsch, A. Cossettini, L. Benini, Biogap: A 10-core fp-capable ultra-low power iot processor, with medical-grade afe and ble connectivity for wearable biosignal processing, in: 2023 IEEE International Conference on Omni-Layer Intelligent Systems, COINS 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1–7, http://dx.doi.org/10.1109/COINS57856.2023.10189286.

[68] H. Yu, F. Zeng, D. Liu, J. Wang, J. Liu, Neural manifold decoder for acupuncture stimulations with representation learning: An acupuncture-brain interface, IEEE J. Biomed. Health Informatics 29 (6) (2025) 4147–4160, http://dx.doi.org/10.1109/JBHI.2025.3530922.

[69] J. Jeong, Eeg dynamics in patients with alzheimer's disease, Clin. Neurophysiol. 115 (7) (2004) 1490–1505, http://dx.doi.org/10.1016/j.clinph.2004.01.001.