Biomedical Signal Processing and Control 112 (2026) 108390

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Dual-Transformer Cross-Attention Framework for Alzheimer's disease detection via dPTE-Guided EEG channel selection and multi-modal integration☆

Shyamal Y. Dharia [a], Qian Liu [a,c], Stephen D. Smith [a,b], Camilo E. Valderrama [a,d,*]

[a] The University of Winnipeg, Department of Applied Computer Science, 515 Portage Ave, Winnipeg, Manitoba, R3B 2E9, Canada
[b] The University of Winnipeg, Department of Psychology, 515 Portage Ave, Winnipeg, Manitoba, R3B 2E9, Canada
[c] University of Manitoba, Department of Biochemistry and Medical Genetics, Max Rady College of Medicine, 745 Bannatyne Avenue, Winnipeg, Manitoba, R3E 0J9, Canada
[d] University of Calgary, Department of Community Health Sciences, Cumming School of Medicine, 3280 Hospital Drive NW, Calgary, Alberta, T2N 4Z6, Canada

ARTICLE INFO

Dataset link: https://github.com/Shyamal-Dharia/EEG-Phase-Entropy.git

Keywords:
Transformers
Electroencephalogram (EEG)
Deep learning
Phase Transfer Entropy (PTE)
Dementia
Alzheimer's disease

ABSTRACT

Accurate and efficient detection of Alzheimer's Disease (AD) and Frontotemporal Dementia (FTD) poses significant challenges in neuroscience and healthcare. Our proposed AI framework, Dual-Transformer Cross-Attention Network (DTCA-Net), combines deep learning and directed Phase Transfer Entropy (dPTE) to optimize channel selection crucial for the detection of AD and FTD. Our method identified six critical EEG channels (F7, F8, T3, T4, O1, and O2) through dPTE analysis. We propose the DTCA-Net architecture, which fuses dPTE and differential entropy (DE) features via a multi-head cross-attention layer, projecting dPTE into queries and DE into keys/values over the temporal dimension, so that connectivity (dPTE) shifts are explicitly aligned with spectral complexity (DE), yielding richer spatiotemporal representations. By leveraging a reduced set of EEG channels identified via dPTE, DTCA-Net performance is comparable to previous state-of-the-art models. Additionally, we introduce an adaptive post-processing voting mechanism to enhance subject-level predictions. This approach achieves an F1 score of 84.9% for AD vs. control (CN) detection and 66.5% for FTD vs. CN detection. Overall, compared to traditional full-channel utilization, our model demonstrates the practicality of a reduced-channel solution for clinical applications in AD and FTD detection, enhancing the accessibility and cost-effectiveness of EEG-based diagnostics. The code has been released on GitHub.

1. Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder and is the most commonly diagnosed form of dementia in the elderly [1]. According to the World Health Organization (WHO), AD is the sixth leading cause of death among all diseases [2]. AD prevalence is estimated to double every 20 years, reaching 75 million by 2030 and 131 million by 2050 [3]. Notably, 68% of the anticipated global increase in dementia prevalence by 2050 will occur in low- and middle-income countries, where current evidence indicates no decline in the risk of AD and other types of dementia [4].

As the number of individuals affected by dementia continues to rise, the demand for timely and accurate AD and another type of dementia, such as Frontotemporal Dementia (FTD) diagnosis, becomes increasingly critical. Current diagnostic methods, including Magnetic Resonance Imaging (MRI), molecular Positron Emission Tomography (PET) neuroimaging, and cerebrospinal fluid analyses, are commonly used in clinical practice [5]. However, these techniques are often expensive and time-consuming, resulting in delayed diagnoses, typically after significant neurodegeneration has occurred. This highlights the need to explore alternative, faster, and more cost-effective options.

One alternative option for the detection of AD or other type of dementia is Electroencephalography (EEG), a non-invasive neuroimaging technique that captures electrical activity produced by action potentials across various brain regions. This method shows promise in detecting distinct patterns associated with AD, potentially enabling early and more accessible diagnoses [6–8]. Moreover, the observed alterations in brain activity and disruptions in neural networks are key features of both AD and FTD [9]. Consequently, EEG emerges as a valuable

☆ This work was supported by NSERC Discovery Grants (RGPIN-2023-03443, RGPIN-2024-05844, and RGPIN-2024-05575) and the Manitoba Medical Service Foundation (grant number 2025-03).

* Corresponding author at: The University of Winnipeg, Department of Applied Computer Science, 515 Portage Ave, Winnipeg, Manitoba, R3B 2E9, Canada.
E-mail address: c.valderrama@uwinnipeg.ca (C.E. Valderrama).

https://doi.org/10.1016/j.bspc.2025.108390
Received 24 March 2025; Received in revised form 24 June 2025; Accepted 27 July 2025
Available online 20 August 2025
1746-8094/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

diagnostic option for these conditions, offering the temporal resolution necessary to match the speed of cognitive processes and thereby facilitating a deeper understanding of neural network disruptions.

Understanding the connectivity for AD is crucial, as AD is marked by disruptions in intercellular communication, which are critical for maintaining effective neural connectivity. Two key types of neuropathological changes underlie these impairments: first, Beta-amyloid ($A\beta$) protein molecules can cluster to form extracellular deposits that interfere with neuronal functioning and intercellular communication [10] and second, Neurofibrillary tangles consisting of tau proteins that have become twisted into paired helical filaments; these tangles can alter the structure of neuronal axons and increase cell loss [11]. Consequently, these pathological changes could lead to alterations in brain activity. For instance, a recent study [12] found that the $A\beta$ accumulation triggers neuronal hyperactivity and hyperconnectivity in brain regions vulnerable to tau pathology (e.g., temporal lobes). Another study [13] found that greater tau burden was associated with lower segregated and integrated functional connectivity of Default Mode Network (DMN) regions but was related to higher functional connectivity in the salience network. Additionally, [14] found that in FTD, functional connectivity changes extend beyond the salience network and affect the interplay between networks. Moreover, another study [15] found that both AD and FTD subjects showed increased mean Phase Lag Index (PLI) values in the theta frequency band, showing that they have some similar neural connectivity characteristics. It is therefore crucial to explore the connectivity distribution among AD, FTD, and CN individuals.

Recent advances in EEG-based functional connectivity analysis have proven effective not only for AD identification using novel machine learning frameworks such as network-based Takagi–Sugeno–Kang (N-TSK) models [16], but also for understanding brain network modulation in other clinical and therapeutic settings. For example, studies have demonstrated that EEG-derived functional connectivity and spectral power can be modulated by acupuncture stimulation, providing insights into the brain's response to treatment [17], and that periodic-aperiodic EEG measurements can be used to evaluate therapeutic efficacy in clinical applications for neural disorders [18]. To that extent, a recent study [19] introduced a neural-manifold decoder for acupuncture stimulation based on representation learning, achieving 92.42% precision and offering an effective method for revealing the clinical efficacy of acupuncture treatment. These findings underscore the broad utility of functional connectivity analysis in both disease characterization and therapeutic monitoring.

Previous research has leveraged the potential of EEG to develop advanced machine learning models to predict AD [20–23]. While they have achieved suitable results, a question that remains is how such models can be adapted for practical use by physicians. Notably, previous research [20–22] has employed EEG systems with 19 electrodes, which, while effective, can potentially be reduced further. A reduction in the number of channels could lead to the development of more compact and computationally efficient models.

To that extent, we employed the Phase Transfer Entropy (PTE) [24, 25] method, which measures the information flow between EEG channels, and is well-suited for identifying key connectivity patterns in AD, FTD and CN Individuals. Recent research [26] has demonstrated the utility of directed PTE in distinguishing between Attention Deficit Hyperactivity Disorder (ADHD) and healthy children using EEG data, proving its applicability.

In our study, rather than selecting channels based solely on significant group differences, we focus on identifying EEG channels that consistently exhibit strong connectivity across AD, FTD and CN individuals. This strategy aims to find a stable, minimal subset of channels that are informative regardless of group status, addressing a channel selection gap left by previous studies such as [20–23,27]. We aim to determine whether this reduced, cross-group channel set can achieve performance comparable to or better than approaches that utilize all available EEG channels.

In summary, our research makes the following main contributions:

1. We introduced PTE for channel selection across AD, FTD, and CN groups, which showed consistently strong connectivity in five frequency bands.
2. We proposed a Dual-Transformer Cross-Attention Network (DTCA-Net) that integrates PTE-derived connectivity and differential entropy (DE) features from the selected EEG channels to achieve accurate dementia classification.
3. We implement an adaptive post-processing mechanism that optimizes prediction sensitivity by determining an optimal threshold on the training data and applying it to test examples for robust classification of AD, FTD, and CN.

2. Literature review

2.1. EEG-based Alzheimer's disease prediction

Several studies have investigated EEG-based prediction of AD. Miltiadous et al. [20] extracted frequency-band power features from 4-s resting-state EEG epochs in AD, CN, and FTD cohorts, training classifiers to distinguish AD vs. CN and FTD vs. CN. They later extended this work to 30-s windows, incorporating relative power and spectral coherence within a dual-transformer architecture to achieve F1 scores of 84.12% for AD vs. CN and 62.27% for FTD vs. CN [21]. Jain et al. [28] applied a fuzzy-logic and spiking neural network (FLSNN) approach to the same dataset, reporting 96.82% accuracy, and Jiang et al. [29] demonstrated 94.32% accuracy using spectral coherence features across AD, FTD, and CN groups. However, an important consideration is that most previous studies [20–22,27–30] have used 19 EEG channels for AD detection, where each channel captures unique information from across the entire brain, allowing the model to learn a broad range of features. However, using all 19 channels may not be strictly necessary. Also, in practice, setting up an EEG gel-based system for patients typically takes between 12 and 60 min [31]. On the other hand, portable EEG devices usually have fewer channels and could reduce the setup time to less than 20 min [32]. Therefore, a channel selection method to identify the most consistently relevant channels across AD, FTD, and CN groups could help produce faster assessments with higher performance. In this study, we focused on PTE functional connectivity to find the most consistent channels across AD, FTD and CN groups and used those channels only to train the model.

2.2. EEG connectivity measures

There are several methods, other than PTE, that give different information about the connectivity of EEG channels in AD detection. The most often used is spectral coherence, which provides a measure of how consistently two EEG channels maintain the same phase relationship across different frequency bands [33,34]. The spectral coherence has been used by previous work [21,29] to identify AD, FTD and CN groups with high performance. Another method is the Phase Locking Value (PLV) [35–37], which measures the consistency of the phase difference between two signals over time. The core difference is that spectral coherence accounts for both amplitude and phase relationships, while PLV focuses solely on phase information.

In addition to these, other connectivity metrics – such as Amplitude Envelope Correlation (AEC), imaginary coherence, Phase Lag Index (PLI), and weighted Phase Lag Index (wPLI) – have also been used and compared. Notably, Briels et al. [38] demonstrated that both the selected connectivity metric and the analyzed frequency band significantly influence AD-related outcomes. For instance, Engels et al. [39] reported that PLI specifically detects a decline in functional connectivity in the alpha band as AD severity increases. By contrast, PTE offers a fundamentally different approach, measuring the directional flow of information by quantifying how the past state of one signal contributes to predicting the future state of another [24,25]. This

sets PTE apart from symmetric, non-directional functional connectivity measures like spectral coherence, PLV, PLI, wPLI, and imaginary coherence, which primarily assess synchronous relationships without causal inference. While various functional connectivities have been used to identify group-specific patterns in AD, FTD, and CN populations, our study instead leverages PTE to uncover electrodes that consistently exhibit higher directional connectivity across all groups. Our aim is not to highlight group differences, but to identify common channels that maintain prominent causal interactions regardless of clinical condition—an approach better suited for robust channel selection than group-wise comparisons alone.

3. Methods

3.1. Dataset

The dataset [20] used in this study consists of EEG resting-state recordings (with eyes closed) from three distinct groups of older adults: the AD group (36 subjects), the CN group (29 subjects), and the FTD group (23 subjects), totaling 88 subjects. The EEG recordings include signals from 19 scalp electrodes, namely: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2. The sampling rate for all recordings was set at 500 Hz. The duration of recordings varied across the groups. For the AD group, recordings lasted an average of 13.5 min (minimum: 5.1 min, maximum: 21.3 min). For the FTD group, the average recording duration was 12 min (minimum: 7.9 min, maximum: 16.9 min). For the CN group, the average recording duration was 13.8 min (minimum: 12.5 min, maximum: 16.5 min). In total, the dataset includes 485.5 min of recordings from the AD group, 276.5 min from the FTD group, and 402 min from the CN group.

The data were preprocessed by the authors of the dataset [20] which included applying a 0.5–45 Hz bandpass filter, using the Artifact Subspace Reconstruction (ASR) routine to remove persistent or large-amplitude artifacts, and performing Independent Component Analysis (ICA) to eliminate ocular and jaw artifacts. This thorough preprocessing ensured that the data was clean and ready for analysis.

3.2. Phase transfer entropy

We first downsampled all EEG signals to 256 Hz and then performed a 9-level Discrete Wavelet Transform (DWT) with the Daubechies-4 ("db4") wavelet, yielding detail coefficients D1–D9 whose frequency ranges (in Hz) were D1: 0.5–1, D2: 1–2, D3: 2–4, D4: 4–8, D5: 8–16, D6: 16–32, D7: 32–64, D8: 64–128 and D9: 128–256. We reconstructed only those levels corresponding to our frequency bands – D1–D3 for delta (0.5–4 Hz), D4 for theta (4–8 Hz), D5 for alpha (8–16 Hz), D6 for beta (16–32 Hz) and D7 for gamma (32–64 Hz) – then converted each band-limited signal to instantaneous phase via the Hilbert transform.

Following the phase computation, we calculated the delay τ, which represents the average number of time steps between consecutive zero-crossings in the phase signal. Zero-crossings serve as natural markers of phase progression, indicating transitions between oscillation cycles. The delay is calculated as:

$$\tau = \frac{N}{\text{Zero Crossings}} \tag{1}$$

where N is the total number of time points in the signal. The delay τ is essential for segmenting the phase time series into its current state and future state after discretizations. It ensures that the analysis is performed over meaningful time intervals, aligned with the signal's inherent oscillatory behavior. This is important because recent work [40] has found that alterations in oscillatory dynamics can distinguish between AD patients and healthy individuals.

After calculating the delay, we discretized the signals to estimate the joint and marginal probability distributions required for entropy calculations. The phase signal $\phi(t)$ is divided into discrete bins based on a bin size computed using Eq. (2). This bin size is determined according to the Scott's rule [41]:

$$\Delta = c \cdot \frac{\frac{1}{m}\sum_{i=1}^{m} \hat{\sigma}_i(\phi)}{n^{1/3}} \tag{2}$$

where c is a constant (set to 3.49), m is the number of EEG channels (rows in the phase matrix), n is the number of time points (columns in the phase matrix), and $\hat{\sigma}_i(\phi)$ denotes the standard deviation of the ith channel's phase values. The phase signal $\phi(t)$ is then discretized into bins using this bin size as follows:

$$\phi_d(t) = \left\lceil \frac{\phi(t)}{\Delta} \right\rceil \tag{3}$$

The overall EEG signal was converted to discretized phase signals. Further, we divided the discretized phase signals into 10-s windows with 50% overlap, and the joint and marginal probability distributions based on Eq. (6) are estimated within these windows. Each window is now looked at as an independent window and is divided into two, the source signal (x) and target signal (y), based on the delay found from Eq. (1). The joint entropy of two variables, y' (the future state) and y (the current state), is defined as:

$$H(y', y) = -\sum_{y',y} p(y', y) \log p(y', y), \tag{4}$$

where $p(y', y)$ is the joint probability distribution of y' and y. Similarly, the joint entropy for three variables – y', y, and x – is given by:

$$H(y', y, x) = -\sum_{y',y,x} p(y', y, x) \log p(y', y, x). \tag{5}$$

Using these definitions, the PTE from x to y is expressed as:

$$\text{PTE}(x \to y) = H(y', y) + H(y, x) - H(y) - H(y', y, x) \tag{6}$$

Here, $H(y', y)$ represents the entropy of the joint distribution of the future (y') and current (y) states of the target phase, $H(y, x)$ captures the entropy of the joint distribution of the current target phase and the source phase, $H(y)$ is the entropy of the current state of the target phase alone, and $H(y', y, x)$ accounts for the joint entropy of the future target phase, the current target phase, and the source phase. By measuring the directional information flow from x to y, PTE provides insights into the interactions between brain regions, which may differ in conditions such as AD and FTD vs CN.

For the forward direction, the directed PTE was calculated as:

$$\text{dPTE}(x \to y) = \frac{\text{PTE}(x \to y)}{\text{PTE}(x \to y) + \text{PTE}(y \to x)}, \tag{7}$$

which represents the proportion of total information transfer that occurs from x to y. Similarly, for the reverse direction, the directed PTE is computed as:

$$\text{dPTE}(y \to x) = \frac{\text{PTE}(y \to x)}{\text{PTE}(x \to y) + \text{PTE}(y \to x)}. \tag{8}$$

Here, the denominator normalizes the PTE values by the total transfer entropy between x and y, ensuring that the sum of dPTE values for both directions is always equal to 1:

$$\text{dPTE}(x \to y) + \text{dPTE}(y \to x) = 1. \tag{9}$$

This normalization enables a direct comparison of the relative information transfer in each direction, providing a measure of the interaction between x and y signals. Finally, the mean PTE for each subject was calculated by averaging the PTE values across all windows for different frequency bands. The resulting matrices for each subject was square matrices that represent the directional flow between one EEG channel to another for each frequency band.

3.3. Ranking EEG channels based on dPTE

To rank the top N source and target EEG channels across different frequency bands, we leveraged the dPTE. We first formed an independent selection set of five subjects per class (CN, AD, and FTD), yielding 15 subjects in total. Channel ranking and hyperparameter tuning were conducted exclusively on this subset, which was then excluded from all subsequent model training.

Given a dPTE matrix, where each entry $\text{dPTE}_{i,j}$ quantifies the directional influence from channel i (source) to channel j (target), we computed two influence scores per channel:

For each channel i, the total *source influence* is computed by summing its outgoing connections across a row:

$$T_{\text{source},i} = \sum_{j=1}^{m} \text{dPTE}_{i,j}. \tag{10}$$

Channels with larger row sums exert a stronger influence on others.

Similarly, for each channel j, the total *target influence* is calculated by summing the incoming connections down a column:

$$T_{\text{target},j} = \sum_{i=1}^{m} \text{dPTE}_{i,j}. \tag{11}$$

Channels with larger target values receive more influence from others.

We applied this computation for each frequency band and ranked the channels based on these source and target influences. Specifically, for each subject within each frequency band, we extracted the top N channels showing the greatest row (source) and column (target) sums. We also selected the strongest N channel-to-channel connections by picking the largest $\text{dPTE}_{i,j}$ values.

Next, we aggregated the subject-level rankings to derive a group-level perspective on the most influential channels and their interconnections. In this group-level analysis, we counted the number of subjects in which each channel or connection appeared within the top N. Rather than focusing solely on the strength of individual connections, we emphasized the *consistency of appearance* in the top N across subjects. The rationale is that if a channel frequently ranks among the top N, it consistently exhibits stronger connections relative to other channels, making it a more reliable indicator of network importance. This consistency suggests that the channel plays a significant role across different subjects. Finally, we merged the results from all groups (CN, AD, and FTD) to identify globally influential channels and connections that exhibit robust directional connectivity regardless of clinical condition.

3.4. Feature extraction and preprocessing

Two EEG-derived features, x_1 and x_2, were used to train the model. x_1 utilizes the dPTE from EEG data, where each subject's one-minute EEG data segment was divided into eleven overlapping segments, each with a duration of ten seconds, maintaining a 50% overlap between consecutive segments. The dPTE was computed using DWT, as mentioned in Section 3.2, for each of these segments across five frequency bands. The extracted features from each segment were structured into a multi-dimensional input matrix for our model, represented as $B_1 \times W_1 \times F_1 \times C_1 \times C_1$, where B_1 denotes the batch size, W_1 represents the eleven ten-second windows, F_1 corresponds to the five frequency bands, and $C_1 \times C_1$ is the square matrix of dPTE values for the EEG channels.

x_2 uses differential entropy (DE). The differential entropy is defined as follows:

$$h(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx = \frac{1}{2}\log(2\pi e\sigma^2) \tag{12}$$

where it is assumed that the EEG signals X follow a Gaussian distribution. The entropy of a Gaussian-distributed signal increases logarithmically with the variance σ^2, which reflects the level of uncertainty or variability in the EEG signal. The constant $2\pi e$ in the entropy expression arises from the mathematical properties of the Gaussian distribution. DE is a widely used feature in EEG-based emotion recognition models [42–44] because it captures additional variability information beyond mere power measurements. This additional insight into frequency variability could also be beneficial for analyzing AD and FTD subjects. As previous research [20,21] has shown significant power shifts between groups without quantifying the extent of variability.

The features extracted using DE, with DWT similar to PTE feature extraction, were shaped as $B_2 \times C_2 \times F_2$, where B_2 represents the batch size, C_2 is the number of EEG channels, and F_2 denotes the frequency bands. In contrast to x_1, we did not include an additional W_2 dimension because DE already encapsulates the variance information; adding another dimension would be redundant and would unnecessarily increase the complexity of the data. Finally, the extracted features, x_1 and x_2, provide comprehensive insights into the temporal, spectral, and spatial characteristics of the EEG signals, ensuring an effective representation for our proposed architecture.

Furthermore, feature normalization was performed separately within each cross-validation fold to avoid data leakage. For fold k, $\min_k(x)$ and $\max_k(x)$ denote the minimum and maximum of feature x computed over the training subjects in that fold. We then applied Min–Max scaling to both training and held-out test features as

$$x_{\text{norm}} = \frac{x - \min_k(x)}{\max_k(x) - \min_k(x)}. \tag{13}$$

This ensures that the scaler parameters are never informed by the test data.

3.5. Network architecture

3.5.1. Dual-input transformer encoders

The proposed network architecture, shown in Fig. 1, consists of dual inputs. The first input, x_1, represents PTE features with an input shape defined as $B_1, W_1, F_1 \times C_1 \times C_1$. Here, B_1 denotes the batch size, W_1 represents 11 overlapping 10-s windows that constitute 1 min of EEG data (with 50% overlap), F_1 corresponds to the frequency bands, and $C_1 \times C_1$ is a square matrix representing the PTE. It is important to note that the dimensions $F_1 \times C_1 \times C_1$ are reshaped/flattened, forming the third dimension of the input. The second input, x_2, consists of DE features. In this input, the shape is defined as B_2, C_2, F_2. Additionally, a learnable positional embedding is applied exclusively to x_1 to enable the model to capture temporal variations within the PTE features.

Both inputs are processed through independent transformer encoder layers, enhancing the feature representations of x_1 and x_2 to produce the outputs t_1 and t_2, respectively. The architecture employs two transformer encoders, each utilizing a dropout rate of 0.4, five multi-head attention heads, two transformer encoder layers, and feed forward dimension set to 512. To prevent data leakage, hyperparameters – including dropout – were optimized on a subset of 5 subjects per class for the binary classification task (CN-AD). More details about the model's hyperparameters are detailed in Table 1.

3.5.2. Multi-head cross-attention mechanism

After being processed through the transformer encoders, the features t_1 and t_2 are obtained, shown in Fig. 1, and are further refined using a multi-head cross-attention mechanism [45]. This mechanism combines the outputs using scaled dot-product attention, as shown in the equation below:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{14}$$

where $\mathbf{Q} \in \mathbb{R}^{L_Q \times d_k}$ is the *query* matrix, $\mathbf{K} \in \mathbb{R}^{L_K \times d_k}$ is the *key* matrix, $\mathbf{V} \in \mathbb{R}^{L_K \times d_v}$ is the *value* matrix, and d_k is the dimensionality of the keys and queries, used to scale the dot-product for numerical stability.

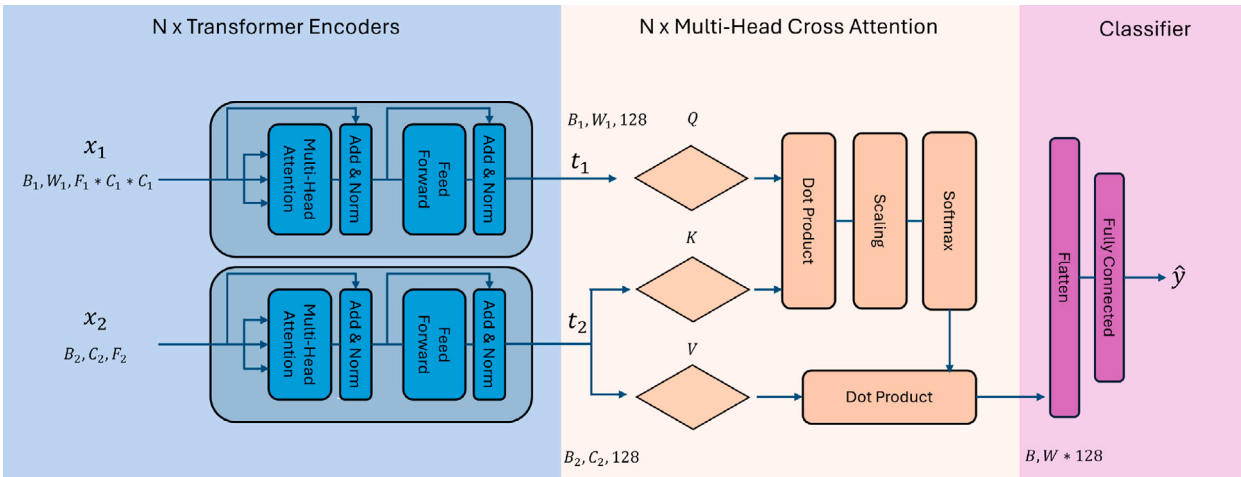


Fig. 1. Overview of the proposed model architecture. The input features x_1 (Phase Transfer Entropy features, with dimensions $B_1, W_1, F_1 \times C_1 \times C_1$) and x_2 (Differential Entropy features, with dimensions B_2, C_2, F_2) are processed through separate stacks of N-layer Transformer Encoders, producing intermediate representations t_1 and t_2, respectively. These representations are fed into the Multi-Head Cross-Attention Mechanism, where t_1 serves as the query Q, and t_2 provides the keys K and values V. The cross-attention output undergoes scaling, softmax, and aggregation, followed by a classifier that includes a flattening step and a fully connected (FC) layer, producing the final predictions $\hat{y}$.

Table 1
DTCA-Net hyperparameter search space. Best parameters (in **bold**) were selected using a subset of five subjects per class for CN vs. AD classification, and then applied to CN vs. FTD as well.

Hyperparameter	Value
PTE Branch Input	
d_{model}	**180**
DE Branch Input	
d_{model}	**5**
Common Parameters	
d_{hidden}	256, **512**, 1024
n_{layers}	1, **2**, 3, 4
n_{heads}	1, **5**
d_{output}	64, **128**, 256
p_{dropout}	0.1, 0.2, 0.3, **0.4**
Multi-Head Cross-Attention	
d_{cross}	128
n_{heads}	2, 4, **8**, 16
Training Settings	
$\mathcal{L}$ (Loss Function)	CrossEntropyLoss
η (Learning Rate)	1×10^{-4}
N_{epochs}	100

In our framework, the PTE features are used as **Q**, as they represent directional relationships between EEG channels that *query* additional spectral context. The DE features are assigned to **K** and **V**, providing frequency-domain information that the PTE queries can attend to. This design allows the model to augment causal interaction features (PTE) with complementary spectral details (DE), producing a richer, more informative representation.

The multi-head cross-attention mechanism applies multiple attention heads to jointly attend to different representation subspaces at different positions. The equation for the multi-head attention mechanism is defined as:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)\mathbf{W}^O \tag{15}$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \tag{16}$$

where $\mathbf{W}_i^Q$, $\mathbf{W}_i^K$, and $\mathbf{W}_i^V$ are learnable parameter matrices for each attention head, and $\mathbf{W}^O$ is the output projection matrix.

We used eight heads for the multi-head cross-attention mechanism. The output of the multi-head cross-attention mechanism is then flattened and passed through a dropout layer with a rate of 0.7 to prevent overfitting. We chose this specific dropout rate because previous work [46] demonstrated that a similar setting after their attention fusion module improved the model's generalization on a different EEG classification task. Finally, a classification layer projects the features into a binary classification space, enabling the model to distinguish between the two target classes effectively.

3.6. Performance evaluation method

For performance evaluation, we employed a subject-stratified 10-fold cross-validation strategy, repeated 30 times. The number of experiments was set to 30 to ensure statistical power to perform a confidence interval across the accuracy of the experiments. In this setup, the data from each subject appears exclusively in either the training or the test set, ensuring no data leakage across folds. This design is particularly well-suited for biomedical and physiological datasets, where inter-subject variability is high and the ability to generalize to unseen subjects is critical. By using subject-stratified partitioning and repeated trials, we obtain a more robust and reliable estimate of model performance.

Furthermore, we incorporate a post-processing voting mechanism to consolidate sample-level predictions into a single subject-level decision. Let $\hat{Y}_{s,i}$ represent the predicted label for the ith sample of subject s, where $\hat{Y}_{s,i} = 1$ indicates a prediction of "AD or FTD" and $\hat{Y}_{s,i} = 0$ otherwise. The subject-level prediction $\hat{Y}_s$ is then determined by the following equation:

$$\hat{Y}_s = \begin{cases} 1 & \text{if } \frac{1}{N_s}\sum_{i=1}^{N_s} \hat{Y}_{s,i} \geq \theta, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

where N_s is the total number of samples for subject s, and θ represents the predefined thresholds (20%, 30%, 40%, 50%). This method computes the ratio of a subject's samples classified as "positive" (e.g., AD) and compares it to the threshold θ. The threshold for the test set is determined by selecting the best threshold identified on the training set. By adjusting θ, we can control the trade-off between recall and specificity at the subject-level.

Finally, the subject-level predictions were used to compute evaluation metrics, including accuracy, precision, recall, and F1 score macro. It is important to note that the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) score was calculated at the sample-level, rather than at the subject-level, in order to obtain a more precise

Table 2
Top 5 most frequent outgoing channels.

Channel	Hits	Groups (C/FTD/AD)
O1	39	4/5/4
O2	37	4/5/4
F7	28	2/4/4
T4	27	4/2/4
C3	27	3/4/4

Table 3
Top 5 most frequent incoming channels.

Channel	Hits	Groups (C/FTD/AD)
O1	38	3/5/3
T3	36	5/4/3
O2	32	3/5/4
F7	31	5/4/4
C3	27	4/3/4

Table 4
Top 5 most frequent connections.

Connection	Hits	Groups (C/FTD/AD)
O1 → T4	12	3/2/2
O2 → T4	11	2/3/2
F7 → O2	10	3/2/1
F8 → O2	10	3/3/2
O2 → T3	9	4/2/0

estimate of the true area under the curve. By aggregating predictions at the subject-level for all other metrics, we ensure that the evaluation more accurately reflects the model's performance in a clinically meaningful context—where diagnostic decisions are made on a per-patient basis rather than per individual sample.

3.7. Experimental environment

Our method was implemented using PyTorch (Python version 3.10.13) and trained on an NVIDIA RTX A6000 GPU. The model was trained for 100 epochs with a batch size of 32. We employed the Adam optimizer with a learning rate of 1×10^{-3} and utilized the standard cross-entropy loss function to optimize the model. To address class imbalance, we applied the Synthetic Minority Over-sampling Technique (SMOTE) only to the training split within each fold of our stratified cross-validation; no synthetic samples were ever used in validation or test splits. This ensures that oversampling is fully nested inside each fold and prevents any leakage of information from test subjects. Finally, we evaluated model generalization via 10-fold, subject-stratified cross-validation repeated 30 times with different random splits.

We trained our model using only six EEG channels, selected based on the most consistently and strongly connected nodes identified through the dPTE analysis. The model's hyperparameters, summarized in Table 1, were tuned exclusively on a subset of five subjects used during the dPTE-based channel selection process. These subjects were then excluded from all subsequent model training and evaluation to prevent data leakage and ensure fair assessment.

4. Results and experiments

4.1. Global top-5 EEG connections

The top–5 connections extracted separately for each frequency band and participant group are plotted in Fig. 2. Pooling the per–band results across all five bands yields three aggregate lists, outgoing sources, incoming sinks, and complete source→target pairs, summarized in Tables 2–4. Here, *hits* denotes the number of times a given channel or connection was selected among the top-5 for each subject in a specific frequency band, reflecting its consistency across subjects and bands. The five channels that most frequently acted as sources were O1 (39 hits), O2 (37), F7 (28), T4 (27) and C3 (27). The most frequently targeted channels were O1 (38 hits), T3 (36), O2 (32), F7 (31) and C3 (27). The five edges with the greatest hit counts were O1→T4 (12 hits), O2→T4 (11), F7→O2 (10), F8→O2 (10) and O2→T3 (9); the corresponding group-wise hit distributions (CN /FTD/AD) are reported in Table 4. These five connections involve six unique channels (O1, O2, T3, T4, F7, and F8) and represent the most consistent information–flow pathways across diagnostic groups and frequency bands. Based on these results, the six channels O1, O2, T3, T4, F7, and F8 were selected as inputs for training our proposed DTCA-Net architecture.

4.2. Model performance

DTCA-Net demonstrated strong discrimination between CN and AD participants. On the CN-AD task (Table 5), it achieved a mean accuracy of 85.22% (±12.07) and an F1-score of 84.92% (±12.17), corresponding to 95% confidence intervals of [80.72%, 89.73%] for accuracy and [80.38%, 89.47%] for F1. Precision and recall averaged 87.22% (±11.83) and 85.28% (±12.11), respectively, while the ROC AUC reached 0.83 (±0.01; 95% CI [0.83, 0.84]). In the CN-FTD classification (Table 6), DTCA-Net yielded moderate performance, with a mean accuracy of 67.72% (±11.45; 95% CI [63.45%, 72.00%]) and an F1-score of 66.56% (±11.46; 95% CI [62.28%, 70.83%]). Precision and recall were 67.78% (±12.31) and 70.83% (±14.55), respectively, and the AUC-ROC was 0.63 (±0.03; 95% CI [0.63, 0.65]). These results indicate that DTCA-Net consistently achieves at least 80% accuracy and F1-score in the CN-AD task across folds and repetitions, and between 63% and 71% in the CN-FTD task.

4.3. Ablation study

To assess the impact of dual-stream fusion, we compared DTCA-Net against PTE-only and DE-only transformer variants. For CN-AD classification (Table 5, Fig. 3), DTCA-Net reduced the AD-CN misclassification rate to 15.9%, versus 17.2% for the PTE transformer and 23.8% for the DE transformer. In CN-FTD classification (Table 6, Fig. 4), DTCA-Net misclassified 28.7% of FTD cases, compared to 31.3% for PTE and 38.9% for DE. These results demonstrate that integrating PTE and DE inputs via cross-attention consistently improves true positive rates and lowers false negatives across both tasks.

4.4. State-of-the-art model comparison

Table 7 summarizes performance on the CN-AD and CN-FTD tasks for DTCA-Net and nine benchmark models evaluated on a similar dataset. On CN-AD, DTCA-Net achieves an F1 score of 84.9% with only six channels, outperforming single-modality methods – LightGBM (77.8%), MLP (75.3%), XGBoost (77.2%), MMSE (EEG) (79.0%) and a CNN-based classifier (77.6%) – and closely matching the multimodal DICE-Net (84.1%). Higher accuracies (96%–97%) reported by FLSNN [28], N-TSK [16] and EEGConvNeXt [27] rely on the full 19 EEG channels.

For CN–FTD, DTCA-Net attains an F1 of 66.5% using six channels, substantially surpassing LightGBM (57.7%), MLP (59.2%), XGBoost (57.4%) and DICE-Net (62.3%), and approaching the CNN-based model (67.8%). EEGConvNeXt achieves 98.6% F1 on CN–FTD with 19 channels. Overall, DTCA-Net offers a favorable trade-off between minimal channel usage, computational efficiency and robust classification performance.

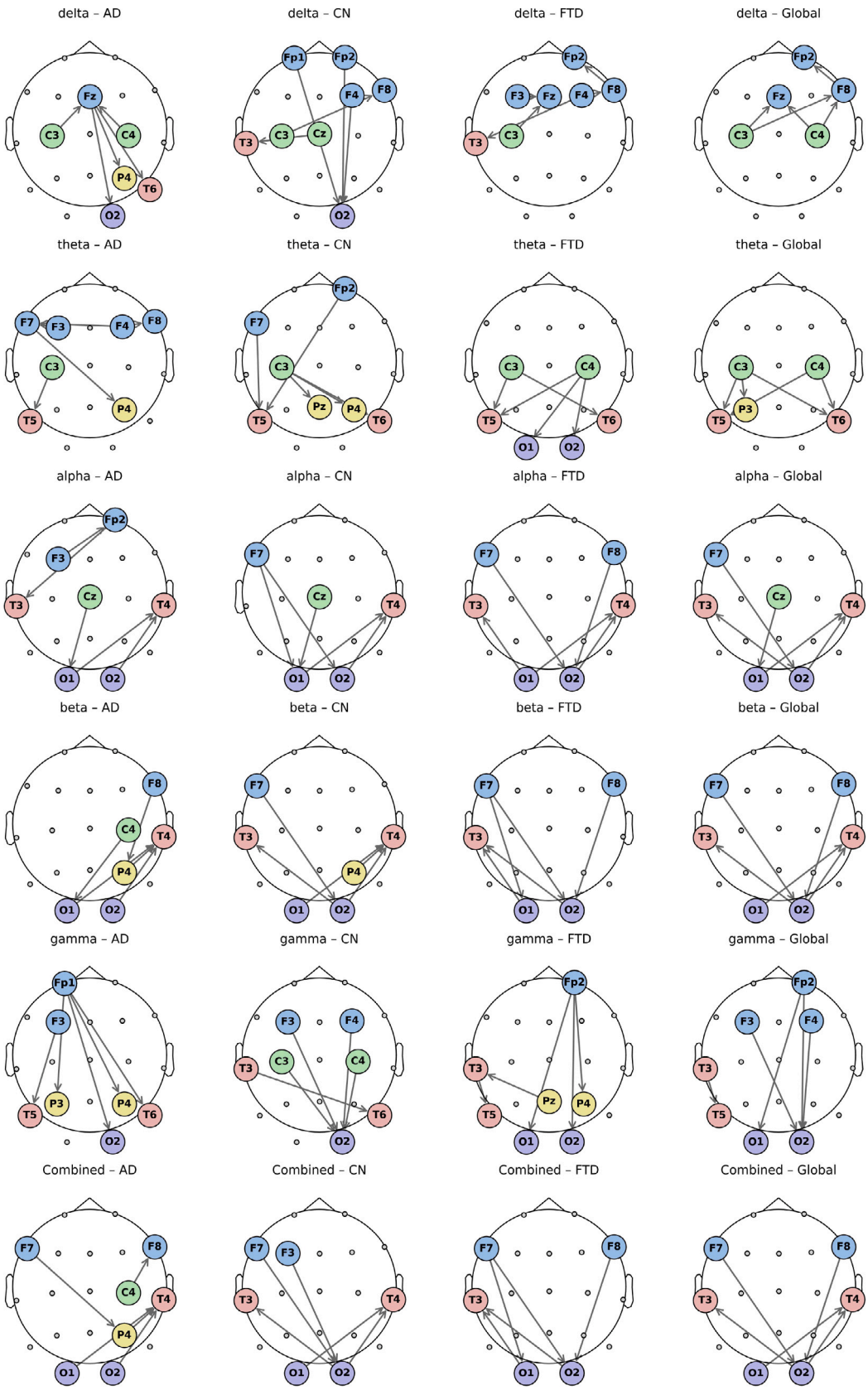


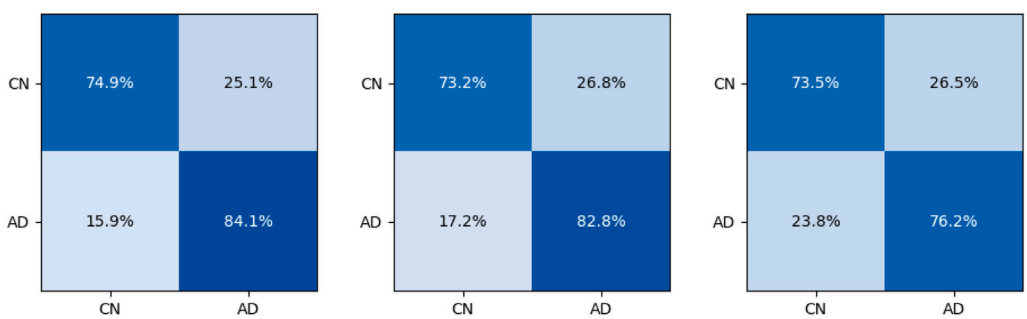
Fig. 2. Top 5 connections identified from the dPTE calculation across all five frequency bands for each group. The first five rows represent the five frequency bands, while the last row shows the global top 5 connections across all frequency bands. The first three columns correspond to each group (AD, CN, and FTD), and the final column represents the top connections across all groups for each frequency band or across all frequency bands.

Table 5
Ablation study of CN-AD for the proposed model (%), each model's first row presents mean and std over 30 repetitions of k-fold cross-validation at subject-level evaluation. The row below it shows the confidence interval in brackets. AUC-ROC is reported at the sample-level.

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DTCA	85.22 ± 12.07	87.22 ± 11.83	85.28 ± 12.11	84.92 ± 12.17	0.83 ± 0.01
	[80.72, 89.73]	*[82.80, 91.64]*	*[80.76, 89.80]*	*[80.38, 89.47]*	*[0.83, 0.84]*
PTE	78.67 ± 10.42	80.83 ± 11.47	78.89 ± 10.48	78.44 ± 10.32	0.78 ± 0.01
	[74.78, 82.56]	*[76.55, 85.12]*	*[74.97, 82.80]*	*[74.59, 82.30]*	*[0.77, 0.78]*
DE	70.78 ± 6.31	71.81 ± 7.88	71.67 ± 7.64	70.76 ± 6.27	0.80 ± 0.02
	[68.42, 73.13]	*[68.86, 74.75]*	*[68.81, 74.52]*	*[68.42, 73.10]*	*[0.80, 0.81]*

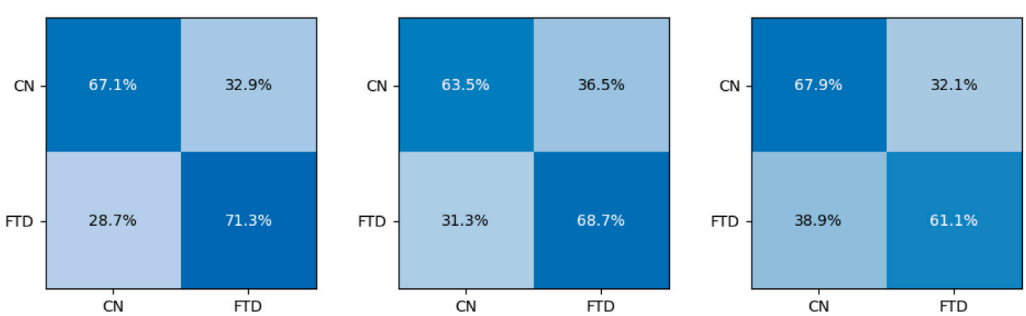
Table 6
Ablation study of CN-FTD for the proposed model (%), each model's first row presents mean and std over 30 repetitions of k-fold cross-validation at subject-level evaluation. The row below it shows the confidence interval in brackets. AUC-ROC is reported at the sample-level.

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DTCA	67.72 ± 11.45	67.78 ± 12.31	70.83 ± 14.55	66.56 ± 11.46	0.63 ± 0.03
	[63.45, 72.00]	*[63.18, 72.37]*	*[65.40, 76.27]*	*[62.28, 70.83]*	*[0.63, 0.65]*
PTE	64.61 ± 10.13	64.72 ± 10.69	68.33 ± 14.34	63.44 ± 9.60	0.59 ± 0.02
	[60.83, 68.39]	*[60.73, 68.71]*	*[62.98, 73.69]*	*[59.86, 67.03]*	*[0.59, 0.61]*
DE	61.67 ± 5.53	60.28 ± 5.96	61.39 ± 8.74	60.06 ± 5.42	0.66 ± 0.02
	[59.60, 63.73]	*[58.05, 62.50]*	*[58.13, 64.65]*	*[58.03, 62.08]*	*[0.66, 0.67]*



(a) Dual Transformer (b) PTE Transformer (c) DE Transformer

Fig. 3. Confusion matrices of transformer models for CN-AD classification.



(a) Dual Transformer (b) PTE Transformer (c) DE Transformer

Fig. 4. Confusion matrices of transformer models for CN-FTD classification.

5. Discussion

5.1. Importance of selected channels

The channels selected through dPTE analysis were O1, O2, T3, T4, F7, and F8—primarily located in the occipital, temporal, and frontal regions. These selections were based on data from only five subjects per class (AD, CN, and FTD), due to the limited dataset. However, the chosen regions align with prior findings that emphasize their relevance in both AD and FTD. For instance, in the occipital region (O1 and O2), [15] reported that AD patients exhibited increased theta-band node degree along with decreased alpha-band clustering coefficient and local efficiency—a pattern not found in FTD. Our analysis similarly revealed consistent activity in the occipital region, where channels O1 and O2 frequently acted as both sources and sinks (see Table 2). Notably, all five FTD subjects showed high hit counts in these channels (5/5), suggesting their potential significance.

Moreover, it is well-established that the frontal and temporal regions are primarily affected in both AD and FTD [48–50]. In our analysis, the selected frontal channels (F7 and F8) and temporal channels (T3 and T4) demonstrated intensified activity via dPTE channel selection, further supporting their importance. Complementary to our

Table 7
Model performance comparison for CN-AD and CN-FTD classification tasks.

Model	Evaluation Protocol	#Ch	CN-AD		CN-FTD	
			Acc	F1	Acc	F1
LightGBM [21]	LOOCV	19	76.2	77.8	69.1	57.7
MLP [21]	LOOCV	19	73.6	75.3	69.9	59.2
XGBoost [21]	LOOCV	19	75.5	77.2	69.2	57.4
DICE-Net [21]	LOOCV	19	83.3	84.1	75.0	62.3
FLSNN [28]	Hold-out (60/20/20)	19	97.1	97.1	–	–
MMSE (EEG) [47]	10-fold CV	19	79.0	–	80.3	–
EEGConvNeXt [27]	LOOCV	19	96.3	96.3	98.6	98.6
N-TSK [16]	5-fold CV	16	97.3	–	–	–
CNN-Based [30]	LOOCV	19	79.4	77.6	72.8	67.8
DTCA-Net (Ours)	Repeated 10-fold CV (30×)	6	85.2	84.9	67.7	66.5

findings, [51] reported that better cognitive performance in AD patients is associated with higher EEG power in frontal and temporal areas. Furthermore, reduced EEG activity in these regions has been linked to lower MMSE scores, which are often used to evaluate executive functions, attention, and working memory [52].

While the selection of these channels based on a limited sample size (n = 5 per class) restricts generalizability, the consistency of our results with previous literature suggests that these regions are promising for further investigation. Thus, although our current findings should be interpreted cautiously due to sample size constraints, they still provide valuable insight into the potential role of specific brain regions in differentiating AD and FTD via EEG analysis.

5.2. Comparative performance analysis

Our model uses post-processing voting and makes predictions at the subject-level, whereas prior studies operate at the sample-level. This, together with our use of PTE and DE, pose challenges to direct comparison. However, we noticed that prior CN–AD and CN–FTD approaches fall into two groups. The first group comprises large or complex architectures – EEGConvNeXt [27], FLSNN [28], DICE-Net [21], and N-TSK [16] – that used 19 EEG channels and reported high accuracies (83.3%–98.6%) under single hold-out or LOOCV protocols. The second group includes classical machine-learning baselines – LightGBM, XGBoost, and shallow MLP models [21] – which, while computationally efficient, generally plateau in the 73.6%–79.4% accuracy range on the same benchmarks. By contrast, DTCA-Net uses only six strategically selected electrodes and connectivity-aware PTE and DE features, yet achieves 85.2% accuracy (84.9% F1) on CN–AD and 67.7% (66.5% F1) on CN–FTD under a strict subject-stratified 10-fold CV repeated 30 times. In this way, DTCA-Net closes much of the gap to heavyweight models while retaining the low channel count and computational efficiency of classical models.

5.3. AD and FTD detection

Our proposed model achieved an F1 Score of 84.9% and 66.5% for detecting AD-CN and FTD-CN, respectively. A key strength of our approach lies in its ability to predict these conditions using only a reduced set of EEG channels and a 1 min signal window. This is a significant extension of prior work, where studies utilized shorter windows, such as 4 s [20] and 30 s [21], with 19 EEG channels compared to our 6 EEG channels. We have demonstrated that reliable predictions can be achieved with longer signal durations and fewer channels, which makes our method more practical for clinical applications. By employing dPTE, our method captures causal functional connectivity between brain regions, offering dynamic feature representations that go beyond traditional electrode-level features like DE or PSD, which primarily reflect localized brain activity rather than directed causal interactions. Furthermore, we opted for a 1 min window to take full advantage of sequential models like transformers, which are capable of learning from extended sequences through positional encoding. This enables the model to capture long-range dependencies that are difficult for other models to detect when using shorter sequences. Additionally, longer sequences can reveal more meaningful temporal patterns related to underlying neural connectivity, a feature that may be critical for detecting complex neurological conditions. However, this approach introduces a trade-off. The use of longer signal windows reduces the number of available training samples, potentially limiting the amount of data the model can learn from. Despite this challenge, our results demonstrate that the model effectively captures long-range dependencies through the PTE transformer branch in DTCA-Net. Also, our fusion strategy – combining multi-head attention with spectral features (DE) – enhances the model's capacity to distinguish between target classes.

Although our model does not match the performance of state-of-the-art methods such as N-TSK [16], FLSNN [28], and EEGConvNeXt [27], it outperforms baseline classical machine-learning models and several advanced deep-learning architectures – including DICE-Net [21], MMSE (EEG) [47], and CNN-based frameworks [30] – on CN-AD classification and achieves competitive results on CN-FTD tasks. Furthermore, by reducing the number of EEG channels required, our approach enhances clinical feasibility by lowering hardware complexity and simplifying data-collection procedures.

5.4. Limitations and future work

Our study faces some limitations. First, channel selection was based on a very small sample (N = 5 per class), which may hinder the generalizability of the resulting electrode subset to a larger population. Nevertheless, the channels identified with the subjects achieved a good performance on the remaining subjects of the dataset, thus showing their potential for discriminating between different AD states. Moreover, the identification of occipital, temporal, and frontal channels as relevant channels is consistent with previous AD and FTD research [48–52]. Second, we validated our dPTE-based selection on a single dataset; to strengthen our approach, future work will use one independent cohort for channel selection and a separate dataset for performance evaluation. Finally, our analysis is confined to resting-state EEG. Future studies should (1) extend the framework to diverse demographic cohorts [53] to develop more generalizable channel sets, and (2) apply it to task-based EEG – where cognitive processes are actively engaged – using larger datasets such as PEARL [54], particularly in cognitively healthy, middle-aged individuals genetically at risk for AD.

6. Conclusions

Our study demonstrates the effectiveness of using dPTE to identify significant EEG channels, F7, F8, T3, T4, O1, and O2, in the frontal, temporal, and occipital regions for AD and FTD detection, aligning closely with previous findings. By employing our proposed model with a multi-head cross-attention mechanism, which enhances data representation by fusing connectivity (dPTE) and spectral complexity (DE), we achieved reliable predictions using a reduced set of six EEG channels with 84.9% F1-macro for CN vs AD and 66.5% for CN vs FTD. Our

method was robustly validated using Subject-stratified 10-fold cross-validation, repeated 30 times, yielding upper-bound AUC-ROC values of 0.84 and 0.63 for the two tasks, respectively, confirming robust discriminatory power. By reducing channel count, our approach could lower EEG acquisition time and patient burden, demonstrating the practical value of targeted channel selection for scalable, EEG-based AD or dementia screening.

CRediT authorship contribution statement

Shyamal Y. Dharia: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Qian Liu:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Stephen D. Smith:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis. **Camilo E. Valderrama:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is publicly available and code is uploaded on GitHub: https://github.com/Shyamal-Dharia/EEG-Phase-Entropy.git.

References

[1] D.V. Puri, S.L. Nalbalwar, A.B. Nandgaonkar, J.P. Gawande, A. Wagh, Automatic detection of alzheimer's disease from eeg signals using low-complexity orthogonal wavelet filter banks, Biomed. Signal Process. Control. 81 (2023) 104439, http://dx.doi.org/10.1016/j.bspc.2022.104439.

[2] A. Atri, The Alzheimer's disease clinical spectrum: Diagnosis and management, Med. Clin. North Am. 103 (2) (2019) 263–293, http://dx.doi.org/10.1016/j.mcna.2018.10.009, Neurology for the Non-Neurologist.

[3] Y. Ding, Y. Chu, M. Liu, Z. Ling, S. Wang, X. Li, Y. Li, Fully automated discrimination of Alzheimer's disease using resting-state electroencephalography signals, Quant. Imaging Med. Surg. 12 (2) (2021) URL https://qims.amegroups.org/article/view/80770.

[4] 2024 Alzheimer's disease facts and figures, Alzheimer' s & Dement. 20 (5) (2024) 3708–3821, http://dx.doi.org/10.1002/alz.13809, arXiv:https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/alz.13809.

[5] B. Dubois, H.H. Feldman, C. Jacova, S.T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P.J. Visser, P. Scheltens, Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria, Lancet Neurol. 6 (8) (2007) 734–746, http://dx.doi.org/10.1016/S1474-4422(07)70178-3.

[6] N. Houmani, F. Vialatte, E. Gallego-Jutglà, G. Dreyfus, V.-H. Nguyen-Michel, J. Mariani, K. Kinugawa, Diagnosis of Alzheimer's disease with electroencephalography in a differential framework, PloS One 13 (3) (2018) e0193607, http://dx.doi.org/10.1371/journal.pone.0193607.

[7] Y. Shim, D.W. Yang, S. Ho, Y.J. Hong, J.H. Jeong, K.H. Park, S. Kim, M.J. Wang, S.H. Choi, S.W. Kang, Electroencephalography for early detection of Alzheimer's disease in subjective cognitive decline, Dement. Neurocognitive Disord. 21 (4) (2022) 126, http://dx.doi.org/10.12779/dnd.2022.21.4.126.

[8] S.-K. Kim, H. Kim, S.H. Kim, J.B. Kim, L. Kim, Electroencephalography-based classification of Alzheimer's disease spectrum during computer-based cognitive testing, Sci. Rep. 14 (1) (2024) 5252, http://dx.doi.org/10.1038/s41598-024-55656-8.

[9] C. Moral-Rubio, P. Balugo, A. Fraile-Pereda, V. Pytel, L. Fernández-Romero, C. Delgado-Alonso, A. Delgado-Álvarez, J. Matias-Guiu, J.A. Matias-Guiu, J.L. Ayala, Application of machine learning to electroencephalography for the diagnosis of primary progressive aphasia: A pilot study, Brain Sci. 11 (10) (2021) http://dx.doi.org/10.3390/brainsci11101262.

[10] T.L. Spires-Jones, B.T. Hyman, The intersection of amyloid beta and tau at synapses in Alzheimer's disease, Neuron 82 (4) (2014) 756–771, http://dx.doi.org/10.1016/j.neuron.2014.05.004.

[11] J.-P. Brion, Neurofibrillary tangles and Alzheimer's disease, Eur. Neurol. 40 (3) (1998) 130–140, http://dx.doi.org/10.1159/000007969, arXiv:https://karger.com/ene/article-pdf/40/3/130/2712779/000007969.pdf.

[12] S.N. Roemer-Cassiano, F. Wagner, L. Evangelista, B.-S. Rauchmann, A. Dehsarvi, A. Steward, A. Dewenter, D. Biel, Z. Zhu, J. Pescoller, M. Gross, R. Perneczky, M. Malpetti, M. Ewers, M. Schöll, M. Dichgans, G.U. Höglinger, M. Brendel, S. Jäkel, N. Franzmeier, Amyloid-associated hyperconnectivity drives tau spread across connected brain regions in Alzheimer's disease, Sci. Transl. Med. 17 (782) (2025) eadp2564, http://dx.doi.org/10.1126/scitranslmed.adp2564.

[13] E. Guzmán-Vélez, I. Diez, D. Schoemaker, E. Pardilla-Delgado, C. Vila-Castelar, J.T. Fox-Fuller, A. Baena, R.A. Sperling, K.A. Johnson, F. Lopera, J. Sepulcre, Y.T. Quiroz, Amyloid-β and tau pathologies relate to distinctive brain dysconnectomics in preclinical autosomal-dominant alzheimer's disease, Proc. Natl. Acad. Sci. 119 (15) (2022) e2113641119, http://dx.doi.org/10.1073/pnas.2113641119.

[14] L.K. Ferreira, O. Lindberg, A.F. Santillo, L.-O. Wahlund, Functional connectivity in behavioral variant frontotemporal dementia, Brain Behav. 12 (12) (2022) e2790, http://dx.doi.org/10.1002/brb3.2790.

[15] S. Wu, P. Zhan, G. Wang, X. Yu, H. Liu, W. Wang, Changes of brain functional network in Alzheimer's disease and frontotemporal dementia: a graph-theoretic analysis, BMC Neurosci. 25 (1) (2024) 30, http://dx.doi.org/10.1186/s12868-024-00877-w.

[16] 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) (2019) 60–71, http://dx.doi.org/10.1109/TFUZZ.2019.2903753.

[17] 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.

[18] 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.

[19] 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. Heal. Informatics 29 (6) (2025) 4147–4160, http://dx.doi.org/10.1109/JBHI.2025.3530922.

[20] 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, frontotemporal dementia and healthy subjects from routine EEG, Data 8 (6) (2023) http://dx.doi.org/10.3390/data8060095.

[21] A. Miltiadous, E. Gionanidis, K.D. Tzimourta, N. Giannakeas, A.T. Tzallas, DICE-Net: A novel convolution-transformer architecture for Alzheimer detection in EEG signals, IEEE Access 11 (2023) 71840–71858, http://dx.doi.org/10.1109/ACCESS.2023.3294618.

[22] 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. 12 (11) (2022) http://dx.doi.org/10.3390/app12115413.

[23] 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.

[24] M. Lobier, F. Siebenhühner, S. Palva, J.M. Palva, Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions, NeuroImage 85 (2014) 853–872, http://dx.doi.org/10.1016/j.neuroimage.2013.08.056, New Horizons for Neural Oscillations.

[25] A. Hillebrand, P. Tewarie, E. van Dellen, M. Yu, E.W.S. Carbo, L. Douw, A.A. Gouw, E.C.W. van Straaten, C.J. Stam, Direction of information flow in large-scale resting-state networks is frequency-dependent, Proc. Natl. Acad. Sci. 113 (14) (2016) 3867–3872, http://dx.doi.org/10.1073/pnas.1515657113, arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1515657113.

[26] A. Ekhlasi, A. Motie Nasrabadi, M.R. Mohammadi, Classification of the children with ADHD and healthy children based on the directed phase transfer entropy of EEG signals, Front. Biomed. Technol. 8 (2021) http://dx.doi.org/10.18502/fbt.v8i2.6515.

[27] M. Acharya, R.C. Deo, P.D. Barua, A. Devi, X. Tao, EEGConvNeXt: A novel convolutional neural network model for automated detection of alzheimer's disease and frontotemporal dementia using EEG signals, Comput. Methods Programs Biomed. 262 (2025) 108652, http://dx.doi.org/10.1016/j.cmpb.2025.108652.

[28] S. Jain, R. Srivastava, Electroencephalogram (EEG) based fuzzy logic and spiking neural networks (FLSNN) for advanced multiple neurological disorder diagnosis, Brain Topogr. 38 (3) (2025) 33, http://dx.doi.org/10.1007/s10548-025-01106-1.

[29] R. Jiang, X. Zheng, J. Sun, L. Chen, G. Xu, R. Zhang, Classification for Alzheimer's disease and frontotemporal dementia via resting-state electroencephalography-based coherence and convolutional neural network, Cogn. Neurodynamics 19 (1) (2025) 46, http://dx.doi.org/10.1007/s11571-025-10232-2.

[30] K. Stefanou, K.D. Tzimourta, C. Bellos, G. Stergios, K. Markoglou, E. Gionanidis, M.G. Tsipouras, N. Giannakeas, A.T. Tzallas, A. Miltiadous, A novel CNN-based framework for Alzheimer's disease detection using EEG spectrogram representations, J. Pers. Med. 15 (1) (2025) http://dx.doi.org/10.3390/jpm15010027, URL https://www.mdpi.com/2075-4426/15/1/27.

[31] D. Hoffman, I. Haislip, C. Cool, Estimated time to properly apply electroencephalogram (EEG) electrodes: A survey (p11-1.009), Neurology 102 (2024) 6212, http://dx.doi.org/10.1212/WNL.0000000000206322, 7_supplement_1.

[32] A. Lau-Zhu, M.P. Lau, G. McLoughlin, Mobile EEG in research on neurodevelopmental disorders: Opportunities and challenges, Dev. Cogn. Neurosci. 36 (2019) 100635, http://dx.doi.org/10.1016/j.dcn.2019.100635.

[33] M. Murias, S.J. Webb, J. Greenson, G. Dawson, Resting state cortical connectivity reflected in EEG coherence in individuals with autism, Biol. Psychiatry 62 (3) (2007) 270–273, http://dx.doi.org/10.1016/j.biopsych.2006.11.012, Mechanisms of Circuit Dysfunction in Neurodevelopmental Disorders.

[34] J.S. Bendat, A.G. Piersol, Random data: analysis and measurement procedures, John Wiley & Sons, 2011, http://dx.doi.org/10.1002/9781118032428.

[35] J.-P. Lachaux, E. Rodriguez, J. Martinerie, F.J. Varela, Measuring phase synchrony in brain signals, Hum. Brain Mapp. 8 (4) (1999) 194–208.

[36] F. Mormann, K. Lehnertz, P. David, C. E. Elger, Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients, Phys. D: Nonlinear Phenom. 144 (3) (2000) 358–369, http://dx.doi.org/10.1016/S0167-2789(00)00087-7.

[37] S. Aydore, D. Pantazis, R.M. Leahy, A note on the phase locking value and its properties, NeuroImage 74 (2013) 231–244, http://dx.doi.org/10.1016/j.neuroimage.2013.02.008.

[38] C.T. Briels, D.N. Schoonhoven, C.J. Stam, H. de Waal, P. Scheltens, A.A. Gouw, Reproducibility of EEG functional connectivity in Alzheimer's disease, Alzheimer's Res. Ther. 12 (2020) 1–14, http://dx.doi.org/10.1186/s13195-020-00632-3.

[39] M.M. Engels, C.J. Stam, W.M. van der Flier, P. Scheltens, H. de Waal, E.C. van Straaten, Declining functional connectivity and changing hub locations in Alzheimer's disease: an EEG study, BMC Neurol. 15 (2015) 1–8, http://dx.doi.org/10.1186/s12883-015-0400-7.

[40] D. Klepl, F. He, M. Wu, M.D. Marco, D.J. Blackburn, P.G. Sarrigiannis, Characterising Alzheimer's disease with EEG-based energy landscape analysis, IEEE J. Biomed. Heal. Informatics 26 (3) (2022) 992–1000, http://dx.doi.org/10.1109/JBHI.2021.3105397.

[41] D.W. Scott, On optimal and data-based histograms, Biometrika 66 (3) (1979) 605–610, http://dx.doi.org/10.2307/2335182.

[42] R.-N. Duan, J.-Y. Zhu, B.-L. Lu, Differential entropy feature for EEG-based emotion classification, in: 2013 6th International IEEE/EMBS Conference on Neural Engineering, NER, 2013, pp. 81–84, http://dx.doi.org/10.1109/NER.2013.6695876.

[43] W.-B. Jiang, X.-H. Liu, W.-L. Zheng, B.-L. Lu, SEED-VII: A multimodal dataset of six basic emotions with continuous labels for emotion recognition, IEEE Trans. Affect. Comput. 16 (2) (2025) 969–985, http://dx.doi.org/10.1109/TAFFC.2024.3485057.

[44] W. Liu, J.-L. Qiu, W.-L. Zheng, B.-L. Lu, Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition, IEEE Trans. Cogn. Dev. Syst. 14 (2) (2022) 715–729, http://dx.doi.org/10.1109/TCDS.2021.3071170.

[45] Y.-H.H. Tsai, S. Bai, P.P. Liang, J.Z. Kolter, L.-P. Morency, R. Salakhutdinov, Multimodal transformer for unaligned multimodal language sequences, in: Proceedings of the Conference. Association for Computational Linguistics. Meeting, vol. 2019, NIH Public Access, 2019, p. 6558, http://dx.doi.org/10.18653/v1/P19-1656.

[46] S.Y. Dharia, C.E. Valderrama, S.G. Camorlinga, Multimodal deep learning model for subject-independent EEG-based emotion recognition, in: 2023 IEEE Canadian Conference on Electrical and Computer Engineering, CCECE, 2023, pp. 105–110, http://dx.doi.org/10.1109/CCECE58730.2023.10289007.

[47] Y. Chen, H. Wang, D. Zhang, L. Zhang, L. Tao, Multi-feature fusion learning for alzheimer's disease prediction using EEG signals in resting state, Front. Neurosci. Volume 17 - 2023 (2023) http://dx.doi.org/10.3389/fnins.2023.1272834.

[48] L.L. Beason-Held, J.O. Goh, Y. An, M.A. Kraut, R.J. O'Brien, L. Ferrucci, S.M. Resnick, Changes in brain function occur years before the onset of cognitive impairment, J. Neurosci. 33 (46) (2013) 18008–18014, http://dx.doi.org/10.1523/JNEUROSCI.1402-13.2013.

[49] L. Chouliaras, J.T. O'Brien, The use of neuroimaging techniques in the early and differential diagnosis of dementia, Mol. Psychiatry 28 (10) (2023) 4084–4097, http://dx.doi.org/10.1038/s41380-023-02215-8.

[50] S. Ribarič, Detecting early cognitive decline in Alzheimer's disease with brain synaptic structural and functional evaluation, Biomedicines 11 (2) (2023) http://dx.doi.org/10.3390/biomedicines11020355.

[51] A. Azargoonjahromi, H. Nasiri, F. Abutalebian, Resting-state EEG reveals regional brain activity correlates in Alzheimer's and frontotemporal dementia, MedRxiv (2024) 2008–2024, http://dx.doi.org/10.1101/2024.08.05.24311520.

[52] M. Torabinikjeh, V. Asayesh, M. Dehghani, A. Kouchakzadeh, H. Marhamati, S. Gharibzadeh, Correlations of frontal resting-state EEG markers with MMSE scores in patients with Alzheimer's disease, Egypt. J. Neurol. Psychiatry Neurosurg. 58 (1) (2022) 31, http://dx.doi.org/10.1186/s41983-022-00465-x.

[53] K. Rezaee, M. Zhu, Diagnose Alzheimer's disease and mild cognitive impairment using deep CascadeNet and handcrafted features from EEG signals, Biomed. Signal Process. Control. 99 (2025) 106895, http://dx.doi.org/10.1016/j.bspc.2024.106895, URL https://www.sciencedirect.com/science/article/pii/S1746809424009534.

[54] P. Dzianok, E. Kublik, PEARL-neuro database: EEG, fMRI, health and lifestyle data of middle-aged people at risk of dementia, Sci. Data 11 (1) (2024) 276, http://dx.doi.org/10.1038/s41597-024-03106-5.