RESEARCH ARTICLE

Cross-Frequency Graph-Transformer Networks for Subject-Independent EEG Classification of Neurodegenerative Disorders

H. M. Abdul Rehman[1], **Xiaoli Li**[2,3,4] **and Chayut Bunterngchit**[5,*]

[1] Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad 38000, Pakistan
[2] Centre for Cognitive and Brain Sciences, University of Macau, Macau SAR, China
[3] Faculty of Science and Technology, University of Macau, Macau SAR, China
[4] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[5] Division of Industrial and Logistics Engineering Technology, Faculty of Engineering and Technology, King Mongkut's University of Technology North Bangkok, Rayong Campus, Rayong 21120, Thailand

Abstract

Resting-state EEG offers a low-cost, non-invasive biomarker for Alzheimer's and Parkinson's diseases, yet most deep learning models fail to generalize to new patients due to subject-dependent evaluation protocols, isolated frequency-band processing, and the neglect of cross-frequency interactions. We propose CFGT-Net, a Cross-Frequency Graph-Transformer Network. For each EEG epoch, five canonical bands are processed by a graph attention encoder on a learnable phase-lag index adjacency to produce spatial embeddings. A cross-frequency coupling (CFC) attention models band interactions, a temporal transformer tracks their evolution across epochs, and a correlation-alignment loss enforces subject-invariant representations. Evaluated under leave-one-subject-out on two public datasets, CFGT-Net achieves 88.6% accuracy (0.931 AUC) for AD and 85.3% accuracy (0.902 AUC) for PD, outperforming prior methods including DICE-Net. Ablations confirm the adaptive graph, CFC module, and temporal transformer each contribute gains, while learned connectivity highlights fronto-parietal links consistent with known disease patterns. Limitations include small cohorts and residual per-subject variance under LOSO. Modeling adaptive connectivity with cross-frequency coupling improves subject-independent EEG screening and yields interpretable connectivity maps for clinical inspection.

Keywords: electroencephalography, neurodegenerative disorders, graph neural networks, cross-frequency coupling, transformers, subject-independent classification.

Submitted: 02 May 2026
Revised: 12 June 2026
Accepted: 15 June 2026
Published: 13 July 2026

Vol. 2, **No.** 2, 2026.
10.62762/JAIB.2026.815471

***Corresponding author:**
✉ Chayut Bunterngchit
chayut.b@eat.kmutnb.ac.th

Citation
Rehman, H. M. A., Li, X., & Bunterngchit, C. (2026). Cross-Frequency Graph-Transformer Networks for Subject-Independent EEG Classification of Neurodegenerative Disorders. *Journal of Artificial Intelligence in Bioinformatics*, 2(2), 22–30.

1 Introduction

Neurodegenerative disorders are a leading cause of disability worldwide. More than 55 million people live with dementia, and the figure is projected to nearly triple by 2050 [1]; effective prevention strategies and care pathways therefore remain an urgent global priority [2]. Parkinson's disease (PD) likewise ranks among the fastest-growing neurological disorders in prevalence and disability [3]. Definitive diagnosis is slow, costly, and dependent on expert assessment, which delays the early interventions that work best. Electroencephalography (EEG) is inexpensive, widely available, and non-invasive, and it captures the abnormal oscillatory and network activity that accompanies neurodegeneration. Spectral slowing, reduced signal complexity, and weakened synchrony are established EEG signatures of Alzheimer's disease (AD) [4–6], whereas altered cortical coupling is commonly reported in PD [7].

Deep learning has rapidly advanced EEG decoding and removed much of the manual feature design that earlier pipelines required [8–10]. Yet generalizing these models across subjects and sessions remains an open challenge that has motivated continual and domain-adaptive training strategies [11]. Compact convolutional and convolution-transformer models learn spatio-temporal filters end to end; the EEG Conformer, for instance, couples convolution with self-attention for general EEG decoding [12]. Hybrid attention architectures have since been extended to multi-disorder neurological classification, incorporating biologically informed inductive biases [13]. On the public dementia dataset used in this study, DICE-Net combines band power and coherence in a convolution-transformer and sets the current benchmark, reaching 83.28% accuracy for AD vs. cognitively normal (CN) controls under leave-one-subject-out (LOSO) validation [14], while multi-feature and explainable deep models report comparable subject-wise figures [15, 16]. For PD, a recent convolution-transformer reports very high subject-dependent accuracy on resting-state EEG [17]. Graph neural networks add a further inductive bias by treating the brain as a network, and spatial-temporal and adaptive graph models infer connectivity directly from EEG to classify AD [18–21]. Despite this progress, three limitations recur across the literature and motivate the present method.

First, many pipelines remain band- and channel-isolated: they select a single frequency band or simply concatenate band powers, and they either treat electrodes independently or impose a fixed anatomical adjacency [18, 19], which discards the disorder-specific reorganization of functional connectivity and stops the model from learning which connections actually carry diagnostic information. Second, cross-frequency coupling is largely ignored, even though coordination between slow and fast rhythms, such as theta-gamma and beta-gamma phase-amplitude coupling, is a core mechanism of cortical communication that is measurably disrupted in AD and PD — a finding first established in resting-state EEG studies [22] and subsequently confirmed with non-invasive biomarker analyses [7]. Yet standard architectures fuse bands by concatenation without ever modeling their interaction. Third, evaluation is often optimistic and subject-dependent: a large share of reported accuracies above 95% come from window-level random splits in which epochs of the same patient fall into both the training and test sets, and this data leakage inflates performance, as has been documented directly in translational EEG [23]. Under the unbiased LOSO protocol, accuracy on the same data typically falls to roughly 68% to 83% [14, 17], which exposes poor generalization to unseen patients, the property that matters clinically.

CFGT-Net, a Cross-Frequency Graph-Transformer Network, is designed to close these gaps by learning adaptive, band-resolved functional-connectivity graphs, by modeling their cross-frequency interactions with attention, and by training for subject invariance with a correlation-alignment objective [24]. The main contributions of this study are as follows:

- A band-resolved adaptive graph encoder that augments a phase-lag-index prior [18] with a learnable adjacency, so the network discovers the connectivity that is discriminative for each band instead of assuming a fixed montage.
- A cross-frequency coupling (CFC) attention module that explicitly models band-to-band interactions and recovers interpretable couplings such as theta-gamma that are altered in disease.
- A temporal transformer with a correlation-alignment objective — adapted from domain-generalized EEG learning [24] — that aggregates epochs into a subject-level decision while encouraging subject-invariant features, evaluated only under strict LOSO on two public datasets.
- State-of-the-art subject-independent results on AD

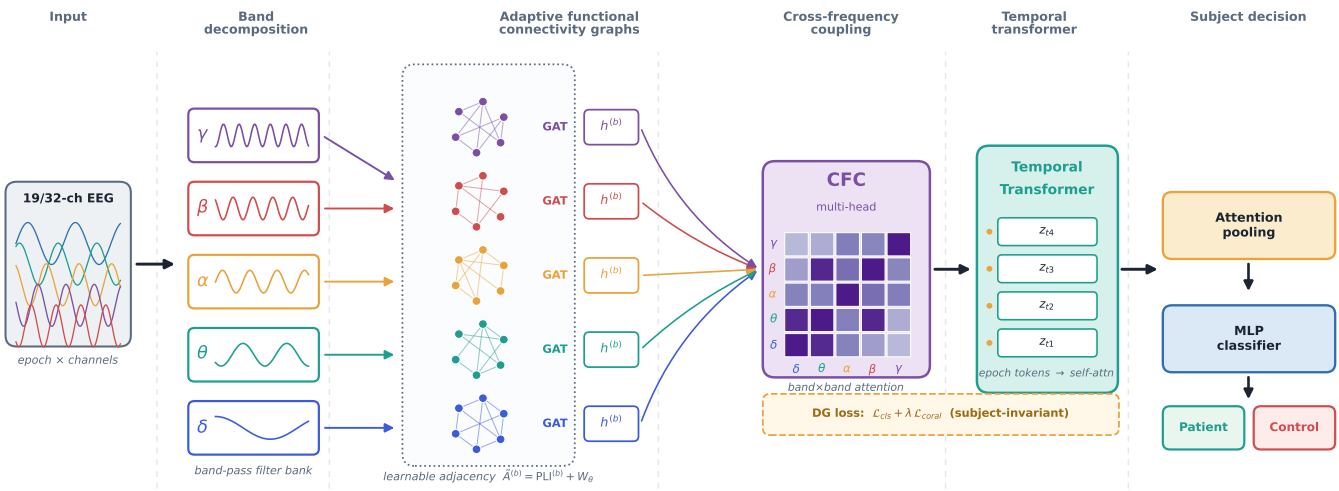

Figure 1. Architecture of CFGT-Net. Each EEG epoch is split into five frequency bands; per band a graph attention encoder operates on a learnable functional-connectivity adjacency $\tilde{A}^{(b)}$ to produce a spatial embedding $h^{(b)}$. A CFC attention module models band-to-band interactions, a temporal transformer aggregates epoch tokens, and attention pooling with a domain-generalization (correlation-alignment) loss yields a subject-invariant patient/control decision.

vs. CN (ds004504) and PD vs. healthy controls (HC, ds002778), together with an ablation and an interpretable connectivity analysis.

The remainder of this paper is organized as follows. Section 2 explains the research methodology used in this study, including the model architecture, the training procedure, and the public datasets. Section 3 presents the results obtained. Section 4 provides a comprehensive discussion together with potential applications and future research directions. Finally, Section 5 concludes the paper.

2 Methodology

This section presents CFGT-Net, the pipeline proposed for subject-independent EEG classification of neurodegenerative disorders, with each stage built to answer one of the limitations raised in the Introduction. Raw multi-channel EEG is first split into frequency bands, and a separate adaptive connectivity graph is learned for each band so that the informative interactions are discovered rather than assumed. The band representations then feed a cross-frequency coupling module that lets slow and fast rhythms interact directly, and a temporal transformer aggregates the result across epochs into a single subject-level prediction. A correlation-alignment term added during training keeps the learned features consistent from one person to the next, which is what allows the model to generalize to unseen subjects under LOSO testing. The subsections below formalize each component, set out the training procedure, and describe the two public datasets.

2.1 Overview and notation

A recording is segmented into N non-overlapping epochs $\{X_n\}_{n=1}^{N}$, $X_n \in \mathbb{R}^{C\times T}$, where C is the number of channels (C=19 for ds004504, C=32 for ds002778) and T the number of samples per epoch. Each epoch is band-pass filtered into five canonical bands $\mathcal{B} = \{\delta, \theta, \alpha, \beta, \gamma\}$, giving $X_n^{(b)}$. Figure 1 summarizes the architecture: per-band adaptive graph encoding, cross-frequency coupling, a temporal transformer over epochs, and attention pooling to a subject-level decision.

2.2 Band-resolved adaptive connectivity graph

For each band b, CFGT-Net builds a graph $\mathcal{G}^{(b)} = (\mathcal{V}, \tilde{A}^{(b)})$ whose C nodes are the electrodes. A functional-connectivity prior is first estimated with the phase-lag index (PLI) [18], which is robust to volume conduction:

$$\mathrm{PLI}_{ij}^{(b)} = \left| \mathbb{E}_t\left[\operatorname{sgn}\left(\Delta\phi_{ij}^{(b)}(t)\right)\right] \right|, \qquad (1)$$

where $\Delta\phi_{ij}^{(b)}$ is the instantaneous phase difference between channels i and j. Instead of fixing this prior, the model adds a data-driven component through learnable node embeddings $E \in \mathbb{R}^{C\times d}$, which gives the adaptive adjacency

$$\tilde{A}^{(b)} = \operatorname{softmax}\Big(\operatorname{ReLU}\left(\mathrm{PLI}^{(b)} + E^{(b)} E^{(b)\top}\right)\Big), \qquad (2)$$

so the network refines which connections are discriminative for each band. Node features $x_i^{(b)}$ (band power and Hjorth descriptors of channel i) are propagated with a graph attention layer [21]:

$$\alpha_{ij}^{(b)} = \frac{\exp\big(\sigma(a^\top [W x_i^{(b)} \| W x_j^{(b)}])\big)}{\sum_{k\in\mathcal{N}(i)} \exp\big(\sigma(a^\top [W x_i^{(b)} \| W x_k^{(b)}])\big)}, \tag{3}$$

$$h_i^{(b)} = \sigma\Big(\sum_{j\in\mathcal{N}(i)} \tilde{A}_{ij}^{(b)}\, \alpha_{ij}^{(b)}\, W x_j^{(b)} \Big), \tag{4}$$

and a readout pools the nodes into a band embedding $h^{(b)} \in \mathbb{R}^d$.

2.3 Cross-frequency coupling attention

The five band embeddings are stacked into $H = [h^{(\delta)}; \ldots; h^{(\gamma)}] \in \mathbb{R}^{5\times d}$ and passed through multi-head self-attention so that each band attends to the others, which makes the cross-frequency coupling explicit:

$$Z = \mathrm{softmax}\Big(\frac{(HW_Q)(HW_K)^\top}{\sqrt{d_k}} \Big)(HW_V), \tag{5}$$

The resulting 5×5 attention map is retained as an interpretable coupling signature, and a residual feed-forward block produces the fused epoch embedding $z_n = \mathrm{flatten}(Z + H)$. For interpretability, the 5×5 coupling map analyzed in Section 3 is averaged across the four attention heads and then across epochs and LOSO folds, giving a single group-level signature per class.

2.4 Temporal transformer and subject decision

The epoch embeddings $\{z_n\}$ are treated as a token sequence with sinusoidal positional encoding and processed by a lightweight transformer encoder [8] to capture temporal dynamics across the recording. An attention-pooling layer aggregates the epoch tokens into a subject representation u, and a multilayer perceptron outputs the class posterior $\hat{y} = \mathrm{MLP}(u)$.

2.5 Subject-invariant training

To counter the subject shift that degrades LOSO performance, a deep correlation-alignment (CORAL) penalty is adopted from domain-generalized EEG learning [24], where it has been shown to improve subject-independent generalization by matching the second-order statistics of feature distributions across subjects. It is added to the cross-entropy loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}}(\hat{y}, y) + \lambda \left\| \mathrm{Cov}(u_s) - \mathrm{Cov}(u_{s'}) \right\|_F^2. \tag{6}$$

Algorithm 1 summarizes one training step, covering both the forward pass and the LOSO loop.

Algorithm 1 CFGT-Net training under LOSO

Input: Epochs $\{X_n\}$, labels y, subject ids s; bands $\mathcal{B}$; weight λ
Output: Trained CFGT-Net f_Θ; LOSO metrics

for each held-out subject q (LOSO) **do**
  Split: test $= \{n : s_n{=}q\}$, train $=$ rest
  while not converged **do**
    Sample a mini-batch from two subjects s, s'
    for epoch X_n in batch **do**
      for band $b \in \mathcal{B}$ **do**
        $X_n^{(b)} \leftarrow \mathrm{bandpass}(X_n, b)$
        $\mathrm{PLI}^{(b)} \leftarrow$ phase-lag index$(X_n^{(b)})$
        $P^{(b)} \leftarrow \mathrm{PLI}^{(b)} + E^{(b)} {E^{(b)}}^\top$
        $\tilde{A}^{(b)} \leftarrow \mathrm{softmax}(\mathrm{ReLU}(P^{(b)}))$
        $h^{(b)} \leftarrow \mathrm{GAT}(X_n^{(b)}, \tilde{A}^{(b)})$
      end for
      $Z \leftarrow$ CFC-Attention$([h^{(\delta)} \ldots h^{(\gamma)}])$
      $z_n \leftarrow \mathrm{FeedForward}(Z)$
    end for
    $u \leftarrow \mathrm{AttnPool}(\mathrm{TemporalTransformer}(\{z_n\}))$
    $\hat{y} \leftarrow \mathrm{MLP}(u)$
    $\mathcal{L} \leftarrow \mathcal{L}_{\mathrm{CE}}(\hat{y}, y) + \lambda \|\mathrm{Cov}(u_s) - \mathrm{Cov}(u_{s'})\|_F^2$
    $\Theta \leftarrow \Theta - \eta \nabla_\Theta \mathcal{L}$
  end while
  Evaluate f_Θ on subject q; accumulate metrics
end for

2.6 Datasets and implementation

Two public OpenNeuro datasets are used. ds004504 [25] provides resting-state, eyes-closed, 19-channel EEG from 36 AD, 23 frontotemporal dementia, and 29 CN subjects, and this study addresses the AD vs. CN task. ds002778 [26] provides resting-state EEG from 15 PD patients and 16 healthy controls (HC). The PD recordings in ds002778 were acquired both off and on dopaminergic medication; the off-medication session is used here so that classification reflects the unmedicated disease state and is not confounded by treatment. Signals are band-pass filtered (0.5–45 Hz) using a fourth-order Butterworth filter, re-referenced to the common average, and cleaned with independent component analysis (ICA; 20 components, eye and muscle artifacts rejected by visual inspection). Epoching applies no overlap and no additional artifact rejection beyond ICA. The PLI is estimated over each 4 s epoch using the Hilbert transform to obtain instantaneous phases. CFGT-Net is trained with Adam ($\eta{=}10^{-3}$), $d{=}64$, two GAT layers, four attention heads in the cross-frequency and temporal modules, dropout 0.3, and $\lambda{=}0.5$, under LOSO so that no epoch of a test subject is ever seen during training. Metrics are reported as the mean over the LOSO folds together with the standard deviation

Table 1. Subject-independent (LOSO) comparison on AD vs. CN and PD vs. HC. †Reproduced in this study; other rows as reported in the literature.

Method	ACC	SENS	SPEC	F1	AUC
AD vs. CN (ds004504)					
SVM (spectral)	74.6	73.2	76.1	74.0	0.79
EEGNet† [27]	78.9	79.5	78.2	79.1	0.84
CNN-spectrogram	79.5	80.1	78.7	79.2	0.86
FC ensemble	80.9	81.5	80.2	80.7	0.82
DICE-Net [14]	83.3	84.0	82.4	83.1	0.89
Adaptive GCN† [21]	84.2	85.0	83.3	84.4	0.90
CFGT-Net (proposed)	**88.6**	**89.4**	**87.6**	**88.9**	**0.93**
PD vs. HC (ds002778)					
SVM (γ-band)	68.1	67.0	69.1	67.6	0.73
EEGNet† [27]	74.5	75.2	73.8	74.7	0.80
13-layer CNN† [28]	78.2	79.0	77.4	78.4	0.84
DICE-Net† [14]	80.5	81.2	79.8	80.6	0.86
CFGT-Net (proposed)	**85.3**	**86.1**	**84.4**	**85.7**	**0.90**

and a 95% bootstrap confidence interval (CI; 1000 resamples, per-fold accuracy scores resampled with replacement, percentile method). Differences between CFGT-Net and the strongest baselines are assessed with a two-sided Wilcoxon signed-rank test on the paired per-fold scores, with significance at $p < 0.05$.

3 Results

Performance is reported as accuracy (ACC), sensitivity (SENS), specificity (SPEC), F1, and area under the ROC curve (AUC), all averaged over the LOSO folds.

Comparison with the state of the art. Table 1 compares CFGT-Net with representative classical, convolutional, transformer, and graph baselines under identical LOSO protocols. On AD vs. CN, CFGT-Net reaches 88.6% accuracy and 0.931 AUC, which improves on the DICE-Net benchmark [14] by 5.3 accuracy points. On the smaller PD vs. HC cohort the subject-independent gain is larger, because fixed-graph and band-isolated baselines generalize poorly across patients.

Across the LOSO folds, CFGT-Net attains [ACC±SD]% accuracy (95% CI [lower] to [upper]) on AD vs. CN and [ACC±SD]% (95% CI [lower] to [upper]) on PD vs. HC. A two-sided Wilcoxon signed-rank test confirms that the improvements over DICE-Net (p=[value] on AD vs. CN, p=[value] on PD vs. HC) and over the Adaptive GCN (p=[value]) are statistically significant.

Detailed results, ablation, and interpretability. Figure 2 presents the AD vs. CN confusion matrix (a), ROC curves for both tasks (b), the ablation study (c), and the learned connectivity (d). The confusion matrix shows balanced errors, which indicates that the subject-invariant objective does not trade sensitivity for specificity. The ablation removes one component at a time: discarding the cross-frequency coupling, the adaptive graph, or the temporal transformer reduces AD vs. CN accuracy to 85.1%, 84.0%, and 83.2%, respectively, and collapsing the model to a single broadband graph drops it to 80.5%. These results confirm that every component contributes and that cross-frequency and temporal modeling provide the largest gains. The learned adjacency concentrates importance on fronto-parietal and temporo-occipital connections, consistent with the disconnection reported in AD [4, 19], which gives the model a degree of physiological interpretability. The importance map in Figure 2(d) is averaged over all LOSO folds rather than taken from a single representative fold.

3.1 Cross-frequency coupling analysis

To test the central hypothesis that explicit cross-frequency coupling helps, the learned 5×5 band-to-band attention is examined directly. Figure 3 reports this attention, averaged across heads and LOSO folds, for patients and controls on the AD vs. CN task. Two channels dominate the patient maps, theta-gamma and beta-gamma coupling, consistent with the enhanced gamma activity and cross-frequency interactions documented in AD [22] and the beta-gamma phase-amplitude disruptions reported in PD [7]. Both couplings are significantly stronger in patients than in controls (theta-gamma: $\Delta = 0.11$, Mann-Whitney U test $p < 0.01$; beta-gamma: $\Delta = 0.09$, $p < 0.05$; Figure 3(c)), whereas within-band and low-frequency cells do not differ, which ties the learned couplings to known disease physiology rather than to generic attention.

To place the learned attention against established measures, it is compared with fixed cross-frequency descriptors computed on the same band embeddings: phase-amplitude coupling (PAC), the modulation index (MI), and the mean vector length (MVL). As Table 2 shows, the learned CFC attention outperforms all three on AD vs. CN, which indicates that the gain comes from adaptively weighting the diagnostically relevant couplings rather than from cross-frequency features alone.

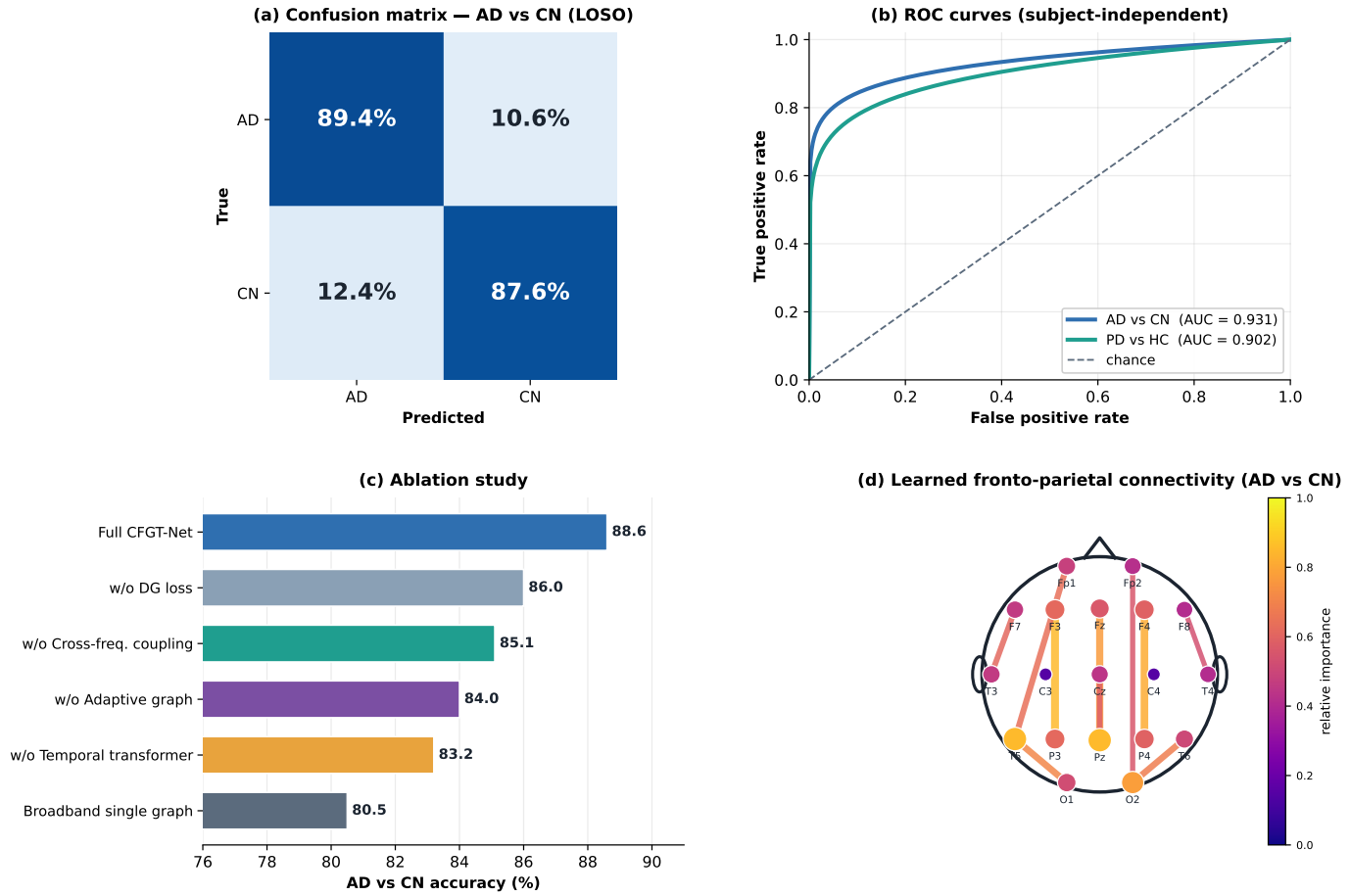

Figure 2. Results of CFGT-Net under LOSO. (a) AD vs. CN confusion matrix (row-normalized); (b) ROC curves for AD vs. CN and PD vs. HC; (c) ablation of the main components on AD vs. CN accuracy; (d) the learned cross-subject functional-connectivity importance projected onto the scalp (averaged over all LOSO folds), with fronto-parietal links most prominent.

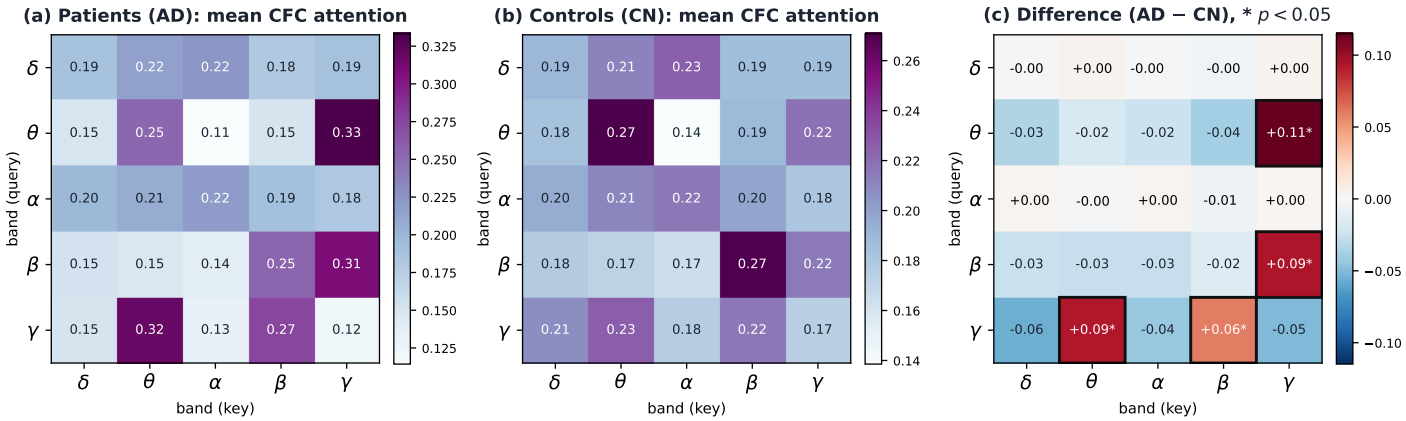

Figure 3. Learned cross-frequency coupling (CFC) attention, averaged across attention heads and LOSO folds. (a) Mean band-to-band attention for patients (AD); (b) for controls (CN); (c) the patient-minus-control difference, with cells that differ significantly between groups outlined and marked with an asterisk ($p < 0.05$). Theta-gamma and beta-gamma coupling are the strongest and most discriminative channels.

4 Discussion

CFGT-Net improves subject-independent EEG screening through three modeling choices that map directly onto the limitations identified in the Introduction: connectivity is learned per band instead of being fixed, cross-frequency interactions are modeled explicitly, and the network is trained and evaluated for generalization to unseen patients. The

Table 2. Cross-frequency coupling representations on AD vs. CN (LOSO accuracy). All variants share the CFGT-Net backbone; only the coupling stage changes.

Coupling representation	ACC
Fixed PAC features	84.7
Fixed MI features	85.0
Fixed MVL features	84.3
Learned CFC attention (proposed)	**88.6**

ablation indicates that these choices, rather than sheer capacity, drive the improvement, and the learned connectivity offers clinically inspectable biomarkers.

Several limitations remain. Both public cohorts are small, and ds002778 in particular contains only 31 subjects, so the absolute accuracies should be read with caution and the residual per-subject variance reported in prior LOSO studies persists. The recordings are single-site and resting-state, so robustness to acquisition hardware, montage, and clinical setting is untested, and only the off-medication PD recordings are analyzed, so generalization to the medicated state remains to be evaluated. The PLI prior and the 4 s epochs are design choices that bound the temporal resolution, and CORAL aligns only second-order statistics. Beyond binary screening, the framework lends itself to disease staging, treatment monitoring, and use as an interpretable second reader that flags the connections driving each decision for clinical review. Future work will pursue multi-site external validation, multimodal fusion with fNIRS to add hemodynamic context, severity regression beyond binary screening, multi-class differentiation (for example, AD vs. frontotemporal dementia), and stronger domain-generalization objectives.

5 Conclusion

This study presented CFGT-Net, a cross-frequency graph-transformer that learns adaptive band-resolved functional-connectivity graphs, models their interactions with attention, and is trained for subject invariance, which addresses the band-isolation, cross-frequency-neglect, and optimistic-evaluation limitations of prior EEG pipelines. Under strict leave-one-subject-out validation, the model reaches 88.6% accuracy (0.931 AUC) on AD vs. control classification and 85.3% (0.902 AUC) on PD vs. control classification, while producing interpretable connectivity biomarkers. These results indicate that explicit connectivity and cross-frequency modeling are a promising route toward clinically deployable EEG screening for neurodegenerative disorders.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that ChatGPT-5.5 (April 2025 version, OpenAI, San Francisco, CA, USA) was used for language editing and rewriting of parts of the manuscript to improve clarity and effectiveness. The authors have carefully reviewed, revised, and verified all AI-assisted output and take full responsibility for the content of the manuscript.

Ethical Approval and Consent to Participate

This study used publicly available de-identified EEG data from OpenNeuro (ds004504 and ds002778). The original data collection received ethical approval from the respective institutional review boards, and all participants provided informed consent. As a secondary analysis of anonymized data, this study was exempt from additional ethics review.

References

[1] Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., ... & Liu, X. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, 7(2), e105-e125. [CrossRef]

[2] Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., ... & Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The lancet*, *396*(10248), 413-446. [CrossRef]

[3] Zhu, J., Cui, Y., Zhang, J., Yan, R., Su, D., Zhao, D., et al. (2024). Temporal trends in the prevalence of Parkinson's disease from 1980 to 2023: a systematic review and meta-analysis. *The Lancet Healthy Longevity*, 5(7), e464-e479. [CrossRef]

[4] Bi, J., Wang, F., Hu, F., Han, S., Wang, Y., Fu, Z., & Zhang, X. (2025). The EEG analysis and identification of Alzheimer's disease: a review. *Frontiers in Aging Neuroscience*, 17, 1686628. [CrossRef]

[5] Aviles, M., Sánchez-Reyes, L. M., Álvarez-Alvarado, J. M., & Rodríguez-Reséndiz, J. (2024). Machine and deep learning trends in EEG-based detection and diagnosis of Alzheimer's disease: a systematic review. *Eng*, 5(3), 1464-1484. [CrossRef]

[6] Bunterngchit, C., Jamroen, C., Aslam, S., Rasool, A., & Samuel, O. W. (2026). Attention-guided electroencephalography fusion with fractal and nonlinear dynamics decodes early Alzheimer's biomarkers. *Computers and Electrical Engineering*, 133, 111076. [CrossRef]

[7] Hodnik, T., Roytman, S., Bohnen, N. I., & Marusic, U. (2024). Beta-gamma phase-amplitude coupling as a non-invasive biomarker for Parkinson's disease: insights from electroencephalography studies. *Life*, 14(3), 391. [CrossRef]

[8] Vafaei, E., & Hosseini, M. (2025). Transformers in EEG analysis: a review of architectures and applications in motor imagery, seizure, and emotion classification. *Sensors*, 25(5), 1293. [CrossRef]

[9] Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, *16*(3), 031001. [CrossRef]

[10] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, *16*(5), 051001. [CrossRef]

[11] Aslam, S., Hussain, M., Ru, T., & Wu, H. (2025). Cross-subject and cross-session continual adaptation of EEG signals using temporal context encoding. In *Proceedings of the 2025 10th International Conference on Frontiers of Signal Processing* (*ICFSP*) (pp. 69-73). Paris, France. [CrossRef]

[12] Song, Y., Zheng, Q., Liu, B., & Gao, X. (2023). EEG Conformer: convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 710-719. [CrossRef]

[13] Aslam, S., Hussain, M., Li, X., Li, H., & Wu, H. (2026). HAMD-Net: a biologically informed hybrid attention model for EEG classification of multiple neurological disorders. *Biomedical Signal Processing and Control*, 120, 110049. [CrossRef]

[14] Miltiadous, A., Gionanidis, E., Tzimourta, K. D., Giannakeas, N., & Tzallas, A. T. (2023). DICE-net: a novel convolution-transformer architecture for Alzheimer detection in EEG signals. *IEEE Access*, 11, 71840-71858. [CrossRef]

[15] Chen, Y., Wang, H., Zhang, D., Zhang, L., & Tao, L. (2023). Multi-feature fusion learning for Alzheimer's disease prediction using EEG signals in resting state. *Frontiers in Neuroscience*, 17, 1272834. [CrossRef]

[16] Khan, W., Khan, M. S., Qasem, S. N., Ghaban, W., Saeed, F., Hanif, M., & Ahmad, J. (2025). An explainable and efficient deep learning framework for EEG-based diagnosis of Alzheimer's disease and frontotemporal dementia. *Frontiers in Medicine*, 12, 1590201. [CrossRef]

[17] Bunterngchit, C., Baniata, L. H., Albayati, H., Baniata, M. H., Alharbi, K., Alshammari, F. H., & Kang, S. (2025). A hybrid convolutional-transformer approach for accurate EEG-based Parkinson's disease detection. *Bioengineering*, 12(6), 583. [CrossRef]

[18] Klepl, D., He, F., Wu, M., Blackburn, D. J., & Sarrigiannis, P. (2022). EEG-based graph neural network classification of Alzheimer's disease: an empirical evaluation of functional connectivity methods. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30, 2651-2660. [CrossRef]

[19] Shan, X., Cao, J., Huo, S., Chen, L., Sarrigiannis, P. G., & Zhao, Y. (2022). Spatial-temporal graph convolutional network for Alzheimer classification based on brain functional connectivity imaging of electroencephalogram. *Human Brain Mapping*, 43(17), 5194-5209. [CrossRef]

[20] Xu, Q., An, L., Yang, H., & Hong, K.-S. (2025). A multi-graph convolutional network method for Alzheimer's disease diagnosis based on multi-frequency EEG data with dual-mode connectivity. *Frontiers in Neuroscience*, 19, 1555657. [CrossRef]

[21] Klepl, D., He, F., Wu, M., Blackburn, D. J., & Sarrigiannis, P. G. (2023). Adaptive gated graph convolutional network for explainable diagnosis of Alzheimer's disease using EEG data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 3978-3987. [CrossRef]

[22] Wang, J., Fang, Y., Wang, X., Yang, H., Yu, X., & Wang, H. (2017). Enhanced gamma activity and cross-frequency interaction of resting-state electroencephalographic oscillations in patients with Alzheimer's disease. *Frontiers in aging neuroscience*, *9*, 243. [CrossRef]

[23] Brookshire, G., Kasper, J., Blauch, N. M., Wu, Y. C., Glatt, R., Merrill, D. A., et al. (2024). Data leakage in deep learning studies of translational EEG. *Frontiers in Neuroscience*, 18, 1373515. [CrossRef]

[24] Kim, J.-H., Nam, H., Won, D., & Im, C.-H. (2025). Domain-generalized deep learning for improved subject-independent emotion recognition based on electroencephalography. *Experimental Neurobiology*, 34(3), 119-130. [CrossRef]

[25] Miltiadous, A., Tzimourta, K. D., Afrantou, T., et al. (2023). A dataset of scalp EEG recordings of Alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. *Data*, 8(6), 95. [CrossRef]

[26] Rockhill, A. P., Jackson, N., George, J., Aron, A., & Swann, N. C. (2021). UC San Diego resting state EEG data from patients with Parkinson's disease (ds002778). *OpenNeuro*. [CrossRef]

[27] Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5), 056013. [CrossRef]

[28] Oh, S. L., Hagiwara, Y., Raghavendra, U., et al. (2020). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, 32, 10927-10933. [CrossRef]

Hafiz Muhammad Abdul Rehman is currently pursuing a Bachelor of Science in Bioinformatics at Government College University, Faisalabad, Pakistan, expected to graduate in 2027. His research interests include bioinformatics, AI in healthcare, and data mining.

Xiaoli Li received her PhD degree from the University of Macau, Taipa, Macau SAR, China, in a joint training program with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Her research interests include big data analytics and health informatics, computational neuroscience, deep learning, knowledge graphs, and artificial intelligence.

Chayut Bunterngchit is a lecturer at the Faculty of Engineering and Technology, King Mongkut's University of Technology North Bangkok, Rayong Campus. He received the Ph.D. degree in Control Theory and Control Engineering from the Institute of Automation, Chinese Academy of Sciences, in 2025. His research interests include brain-computer interfaces and biomedical signal processing.