Journal of Integrative Neuroscience

J. Integr. Neurosci. **2026**; 25(7): 50737
https://doi.org/10.31083/JIN50737

Original Research

A Multi-Frequency Self-Supervised Fusion Model for EEG-Based Dementia Classification

Jinhua Sheng[1,2,*], Jiaqi Lin[1,2], Qiao Zhang[3,4,5], Zhaozhe Gong[1], Rong Zhang[1,2], Ming Wu[3,4,5], Yuhan Cheng[1,2], Xinrui Zhu[1,2], Zhongyuan Sun[1,2]

[1]School of Computer Science, Hangzhou Dianzi University, 310018 Hangzhou, Zhejiang, China
[2]Key Laboratory of Intelligent Image Analysis for Sensory and Cognitive Health, Ministry of Industry and Information Technology of China, 310018 Hangzhou, Zhejiang, China
[3]Beijing Hospital, 100730 Beijing, China
[4]National Center of Gerontology, 100730 Beijing, China
[5]Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, 100730 Beijing, China
*Correspondence: j.sheng@ieee.org (Jinhua Sheng)
Academic Editor: Bettina Platt
Submitted: 5 February 2026 Revised: 22 April 2026 Accepted: 8 May 2026 Published: 21 July 2026

Abstract

Background: Brain source localization technology enables precise characterization of the spatial distribution of neural activity, serving as a crucial tool for exploring the pathological mechanisms underlying dementia. However, effectively integrating complementary diagnostic information from source localization features across multiple frequency bands remains a major challenge to enhancing classification performance and model interpretability. **Methods:** An attention-based multi-frequency self-supervised fusion model (AM-SSF) is proposed to address this issue. Independent contrastive self-supervised encoders are trained for the θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–48 Hz) frequency bands to learn band-specific latent representations. Then, an attention-guided adaptive fusion module is introduced to dynamically allocate band weights through cross-entropy–based supervised optimization, thereby achieving effective cross-band information integration. Finally, a random forest classifier is employed to evaluate the model's performance in distinguishing Alzheimer's disease (AD) from frontotemporal dementia (FTD). **Results:** Experimental results show that the proposed framework achieves a classification accuracy of 93.1% under five-fold cross-validation, significantly outperforming baseline methods such as single-band self-supervised learning (SSL) and average pooling fusion. Further analysis of the attention weight distributions revealed that the θ and β bands contributed most to model decision-making, providing interpretability regarding frequency-specific effects. **Conclusions:** In summary, the proposed AM-SSF model enhances AD and FTD classification performance while offering valuable insights into the discriminative roles of frequency band features.

Keywords: brain source localization; self-supervised learning; attention mechanism; multi-frequency fusion; Alzheimer's disease; frontotemporal dementia

1. Introduction

Dementia is a neurodegenerative syndrome marked by progressive cognitive decline, posing a substantial burden on global public health systems. Among its subtypes, Alzheimer's disease (AD) and frontotemporal dementia (FTD) represent the most prevalent forms. However, the overlap in early clinical manifestations contrasts sharply with their divergent pathological mechanisms, prognoses, and therapeutic strategies. Consequently, establishing an accurate differential diagnosis between AD and FTD is of critical clinical importance [1].

Electroencephalography (EEG), a non-invasive neuroimaging technique characterized by high temporal resolution and low cost, is widely employed to assess brain functional activity and neurological disorders. In dementia research, quantitative spectral EEG analysis (qEEG) is among the most widely utilized approaches. Specifically, through mathematical analyses such as power spectral density, neural synchrony, and signal complexity, qEEG offers multifaceted perspectives on EEG signals, including frequency, time, and spatial domain information. The quantifiable nature of these measures further facilitates standardized research [2,3]. Moreover, advances in brain source localization techniques have enabled the reconstruction and analysis of cortical neural activity distributions from scalp EEG signals, thereby providing novel insights into the distinct pathophysiological mechanisms underlying AD and FTD [4,5]. Previous studies have demonstrated that neural oscillatory activity across different frequency bands (e.g., δ, θ, α, β) is closely associated with specific cognitive functions and pathological processes. For instance, AD is characterized by pronounced abnormalities in the δ band across widespread cortical regions, particularly the frontal lobe, whereas FTD tends to exhibit distinct anomalies in α activity localized to the anterior orbitofrontal cortex and temporal lobes. In addition, both disorders present distinct β-band activity patterns in the parietal and sensorimotor regions [6]. These findings suggest that complementary diagnostic information resides in the source localization characteristics of different frequency bands. However, effectively integrating these band-specific features while simultaneously enhancing model interpretability remains a critical challenge.

In computational psychiatry, machine learning models have been extensively applied to assist in the diagnosis of dementia. Traditional EEG-based approaches typically rely on manually extracted frequency- or time-domain features, which are then combined with classifiers such as Support Vector Machine (SVM) for diagnostic purposes [7]. Although these methods have achieved moderate success, their performance remains heavily dependent on labor-intensive feature engineering and is limited in capturing the complex, high-dimensional, and nonlinear patterns inherent in EEG signals. Furthermore, the scarcity of high-quality clinical data, resulting from acquisition challenges, privacy restrictions, and the high cost of professional annotation, often leads to limited sample sizes. This scarcity severely constrains both the performance and the generalization capacity of traditional supervised learning models. To address these limitations, self-supervised learning (SSL) models are capable of automatically learning robust and generalizable feature representations by pre-training on unlabeled data. The features obtained through pre-training can subsequently be leveraged for downstream tasks with sparse labels, thereby mitigating the small-sample problem and enhancing overall model performance [8].

Although self-supervised learning has shown promise in EEG signal analysis, existing methods still exhibit clear limitations when handling multi-frequency-band source-localized data. Different EEG frequency bands, such as θ, α, β, and γ, are highly heterogeneous in their physiological mechanisms. Conventional feature-level mixing strategies can easily lead to mutual interference and the masking of band-specific rhythmic information. Therefore, constructing independent self-supervised encoders for each frequency band to achieve feature disentanglement is particularly necessary. Furthermore, given the substantial differences in pathological manifestations across frequency bands between Alzheimer's disease and frontotemporal dementia, traditional concatenation or averaging-based fusion strategies are unable to dynamically highlight key diagnostic information. This necessitates a network equipped with an adaptive attention-weighting mechanism. In addition, purely unsupervised learning without downstream task guidance tends to fit global EEG noise and struggles to precisely capture disease-related pathological features that are directly relevant to diagnosis. Introducing cross-entropy loss from the classification task as explicit supervision can force the attention module to focus on the most clinically discriminative frequency bands, thereby effectively improving model performance. Based on the above analysis, this study proposes an attention-based multi-frequency-band self-supervised fusion model, AM-SSF. The main contributions and objectives of this work can be summarized as follows:

(1) Proposing an Attention-based Multi-frequency Self-supervised Fusion Model (AM-SSF): This model initially employs independent contrastive self-supervised encoders to learn specific latent representations from different frequency bands. Building upon these representations, we design a supervision-guided adaptive fusion module that optimizes attention weights using cross-entropy loss from classification tasks. This approach emphasizes frequency band features most relevant for distinguishing AD from FTD through attention-weighted integration while enhancing model interpretability via visualizable attention weights.

(2) Analyzing Contribution of Different Frequency Bands in Attention Fusion: Through attention weight visualization and statistical analysis, we explore the relative importance of features across frequency bands in downstream diagnostic tasks, thereby providing band-level explanations for understanding model decisions.

(3) Validating the proposed framework's effectiveness in AD and FTD differentiation: Through five-fold cross-validation, comparisons with multiple baseline methods demonstrate that AM-SSF significantly improves classification performance and exhibits strong robustness in small-sample scenarios.

The remainder of this paper is organized as follows. Section 1 provides a review of related work. Section 2 is devoted to the description of the proposed method and model framework, and introduces the dataset together with the data processing workflow. Section 3 presents the experimental setup and reports the results. Section 4 offers a discussion of the findings. The work is finally summarized in Section 5, which also looks ahead to the direction of future research.

1.1 Related Work

1.1.1 Machine Learning for EEG-Based Dementia Analysis

Early studies on EEG-based classification of AD and FTD predominantly relied on manual feature engineering combined with traditional machine learning models. Approaches included: (i) constructing multi-layer resting-state EEG networks across frequency bands and extracting connectivity features from δ−α and δ−β band differences, achieving 81.1% classification accuracy [9]; (ii) combining Multi-Threshold Recurrence Rate Plots (MTRRP) features with support vector machines and leave-one-out cross-validation, achieving 72.88% classification accuracy [10]; and (iii) applying Mann–Whitney U tests and *t*-tests to select time-frequency, connectivity, and complexity features, achieving 87.8% classification accuracy for distinguishing AD from FTD [11]. However, these methods heavily depend on expert knowledge for feature selection and extraction, rendering the process both complex and time-consuming.

In recent years, deep learning methods have been increasingly applied to automatically learn high-dimensional features, thereby improving classification performance. Convolutional neural network (CNN) can directly extract representations from raw or preprocessed EEG signals [12], for example, by constructing all channels into a three-dimensional feature matrix to investigate the influence of electrode spatial information on classification [13]. More advanced architectures, such as Dual-Input Convolution Encoder Network (DICE-net), integrate convolutional and Transformer modules and utilize frequency-band power and coherence features as inputs. This approach achieved an accuracy of 83.28% in distinguishing AD from CN, thereby demonstrating strong modeling capabilities for complex EEG features [14]. Nevertheless, the limited availability of large-scale annotated EEG datasets continues to impede the clinical adoption of these methods.

1.1.2 Applications of Self-Supervised Learning in EEG Analysis

In recent years, SSL has made significant advances in the fields of natural language processing and computer vision. A notable example is Bidirectional Encoder Representations from Transformers (BERT), which learns language representations from unlabeled text by predicting masked words within sentences [15]. The Simple Framework for Contrastive Learning of Visual Representations (SimCLR) framework employs contrastive learning to maximize the similarity between augmented versions of the same image, thereby producing robust feature representations [16]. Meanwhile, the Masked Autoencoder (MAE) model further employs masked reconstruction tasks to learn efficient representations from large-scale unlabeled images [17]. Collectively, these approaches illustrate that SSL can produce transferable feature representations from unlabeled data, providing effective solutions for scenarios characterized by limited sample sizes and scarce annotations.

Inspired by these successes, researchers have increasingly applied SSL to biomedical signal analysis to address challenges associated with limited labeled data. Banville et al. [18] employed a contrastive learning framework on EEG to obtain cross-task transferable representations. Zhao et al. [19] improved sleep staging performance through multi-view contrastive learning, and Wang et al. proposed a self-supervised pretraining approach that substantially outperformed traditional supervised models in emotion recognition tasks [20]. These studies collectively demonstrate that SSL can extract discriminative latent representations from EEG signals, thereby supporting various downstream tasks. Although self-supervised learning has been successfully applied to EEG signals, systematic integration of multi-frequency features and their application in disease classification tasks are relatively underexplored.

1.1.3 Attention-Based Multi-Feature Fusion

Attention mechanisms were initially proposed by Bahdanau et al. [21] for dynamic alignment in machine translation tasks. Subsequently, the development of the Transformer architecture further illustrated the powerful capability of attention mechanisms in capturing global dependencies and modeling sequential information [22]. Since then, attention mechanisms have been extensively applied in natural language processing, computer vision, and biomedical signal analysis, where they are commonly employed for feature weighting and for improving model interpretability.

In EEG analysis, Chen et al. [23] combined graph convolutional networks with attention mechanisms to effectively model emotion-related brain regions. Aziz et al. [24] introduced attention into temporal convolutional networks to enhance temporal modeling capabilities for brain-computer interface tasks. Similarly, in multimodal medical data analysis, attention mechanisms have been employed to facilitate feature fusion. For instance, Golovanevsky et al. [25] integrated multimodal features using attention mechanisms to achieve substantial improvements in diagnostic performance for Alzheimer's disease.

Traditional machine learning approaches primarily rely on manually engineered features. Although these methods have achieved moderate success in early AD and FTD studies, they are hindered by complex feature extraction procedures and limited generalization capabilities. Deep learning methods can automatically learn high-dimensional representations to enhance classification performance; however, their practical application is often constrained by the scarcity of labeled data. Recently, self-supervised learning has shown considerable potential for mitigating the small-sample problem in EEG analysis, whereas attention mechanisms have proven effective for feature weighting and enhancing interpretability. Existing research has primarily focused on either employing self-supervised learning alone to enhance EEG representations or using attention mechanisms to enhance feature fusion. Exploration that integrates both approaches remains relatively limited, especially regarding self-supervised fusion of multi-band EEG source localization features. Accordingly, this study proposes a novel attention-based multi-band self-supervised fusion framework. By jointly modeling and adaptively weighting EEG features across multiple frequency bands, the framework achieves effective discrimination between AD and FTD while offering interpretability regarding the relative contributions of each frequency band.

2. Materials and Methods

To address the scarcity of high-quality clinical data, this study proposes an attention-based multi-frequency band self-supervised fusion model. The framework is designed to learn latent representations specific to each frequency band and employs a supervision-guided attention

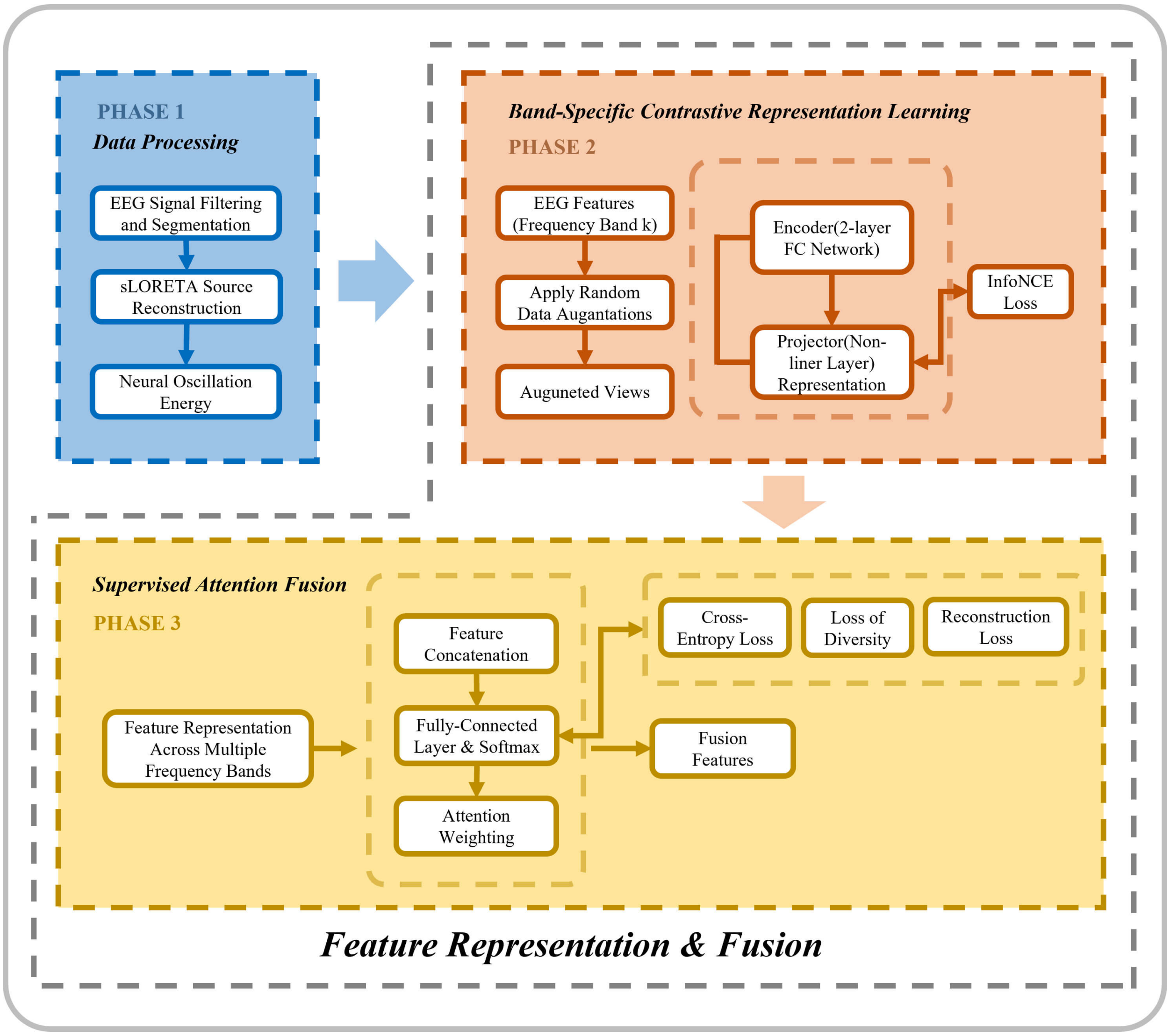

Fig. 1. Overall framework of AM-SSF. The framework comprises three main phases: (1) Data Processing: filtering and segmenting raw EEG signals, followed by sLORETA source reconstruction to extract neural oscillation energy features across frequency bands; (2) Band-Specific Contrastive Representation Learning: applying random data augmentations to each frequency band feature to generate augmented views, which are then mapped into latent space via an encoder (comprising a 2-layer fully connected network) and a projector (non-linear layer), and pre-trained in a self-supervised manner using the InfoNCE loss function; (3) Supervised Attention Fusion: concatenating the multi-band feature representations and computing attention weights through a fully connected layer and Softmax to generate weighted fusion features. This phase adopts a multi-task framework that simultaneously optimizes the supervised cross-entropy loss, unsupervised reconstruction loss, and diversity regularization loss to promote adaptive fusion and improve classification performance. AM-SSF, Attention-based Multi-frequency Self-supervised Fusion Model; EEG, electroencephalography; sLORETA, Standardized Low-Resolution Brain Electromagnetic Tomography; FC, fully connected; InfoNCE, Information Noise-Contrastive Estimation.

mechanism to achieve adaptive feature fusion for classification tasks. The architecture of the model is illustrated in Fig. 1.

2.1 Band-Specific Contrastive Representation Learning

Traditional supervised learning methods are particularly susceptible to overfitting when applied to small-sample datasets. To address this issue, we perform self-supervised pre-training on features from each frequency band independently before the downstream classification task, aiming to learn feature representations with strong generalization capabilities. This stage aims to allow the model to extract meaningful features from the intrinsic structure of the data, independent of category labels.

EEG features from each frequency band are processed independently by an independent self-supervised encoder. The encoder comprises two components: (1) the encoder network, consisting of two fully connected layers, which maps raw frequency band features into a latent representation space; and (2) the projector, an additional nonlinear layer that transforms the latent representations into a separate space to facilitate contrastive learning tasks.

During training, two types of random data augmentation are applied to each sample within every batch. These augmentations include random scaling, translation, and the addition of noise to the data. The augmented samples form a positive pair, whereas the remaining samples in the batch serve as negative samples. The Information Noise-Contrastive Estimation (InfoNCE) loss is employed as the objective function for the self-supervised task. This loss function seeks to maximize the similarity between positive pairs while minimizing the similarity between positive samples and all negative samples. Consequently, the model learns latent representations that draw similar samples closer and push dissimilar samples farther apart, thereby effectively capturing the intrinsic structure of the data.

$$L_{InfoNCE} = -log \frac{exp\left(\frac{sim(z_i, z_j)}{\tau}\right)}{\sum_{k \neq i} exp\left(sim\left(z_i, z_j\right)/\tau\right)} \tag{1}$$

Here, z_i and z_j denote the projected representations of two augmented views originating from the same original sample, and τ represents the temperature coefficient. The similarity metric is computed using normalized cosine similarity, which is defined as follows:

$$sim\left(z_i, z_j\right) = \frac{z_i \cdot z_j}{||\ z_i\ ||_2||\ z_j\ ||_2} \tag{2}$$

the data augmentation function A($\cdot$) consists of three operations: random scaling, random translation, and the addition of Gaussian noise. For the original feature x, the augmented feature x′ can be expressed as:

$$x' = s \cdot x + \Delta + \epsilon \tag{3}$$

where *s* denotes the scaling factor, Δ represents the translation term, and ϵ indicates Gaussian noise.

2.2 Attention Integration Mechanism for Supervision and Guidance

After obtaining the feature representations for each frequency band, the effective fusion of this information becomes critical for improving classification performance. To this end, we propose a supervision-guided attention fusion mechanism that adaptively assigns weights to different frequency bands and directly optimizes the downstream classification task.

Specifically, the fusion mechanism first concatenates the features f_i from each frequency band into a vector $F = [f_1, \ldots, f_N]$, where N denotes the total number of frequency bands. The attention weights ω_i are computed using a fully connected layer followed by a Softmax activation function:

$$w = Softmax\left(Linear(F)\right) \tag{4}$$

The fused feature f_{fused} is obtained as the weighted sum of features across all frequency bands:

$$f_{fused} = \sum\nolimits_{i=1}^{N} \omega_i \cdot f_i \tag{5}$$

The training objective of this fusion mechanism follows a multi-task learning framework, comprising three components: an unsupervised reconstruction loss, a diversity regularization loss, and a supervised cross-entropy loss. The unsupervised reconstruction loss minimizes the mean squared error (MSE) between the fused feature and each original frequency-band feature, ensuring that the fused representation retains effective information from all bands while preserving critical details. The diversity regularization loss encourages a more balanced distribution of attention weights by incorporating a negative entropy constraint, thereby reducing the risk of over-reliance on any single frequency band. The supervised cross-entropy loss enables end-to-end optimization of the fused features through a classification head, guiding the attention mechanism to adaptively focus on frequency bands that are most informative for distinguishing between AD and FTD. The overall loss function L_{total} is defined as a multi-task learning objective and expressed as:

$$L_{total} = L_{recon} + \alpha \cdot L_{CE} + \beta \cdot L_{div} \tag{6}$$

where α and β are hyperparameters that control the trade-off among the three constituent losses. The unsupervised reconstruction loss L_{recon} is defined as:

$$L_{recon} = \frac{1}{N} \sum\nolimits_{i=1}^{N} \|\ fused - f_i\ \|_2^2 \tag{7}$$

where N denotes the total number of frequency bands, f_{fused} represents the fused feature vector, and f_i corresponds to the original feature vector of the i-th frequency band. The supervised cross-entropy loss L_{CE} is defined as:

$$L_{CE} = -\frac{1}{M}\sum_{j=1}^{M}\left[y_i \log(\hat{y}_j) + (1-y_i)\log(1-\hat{y}_j)\right] \tag{8}$$

where M denotes the batch size, y_i represents the true label of the j-th sample, and $\hat{y}_j$ corresponds to the predicted probability produced by the model. The diversity regularization loss L_{div} is defined as:

$$L_{div} = -\sum_{k=1}^{N} \bar{a}_k \log \bar{a}_k \tag{9}$$

where $\bar{a}_k$ denotes the average attention weight of the k-th frequency band, computed across all samples within a batch.

2.3 Model Evaluation and Analysis

To ensure the reliability of the experimental results, this study employed repeated cross-validation to assess model performance. Subjects were used as the grouping unit during dataset partitioning to ensure that data from the same subject never appeared simultaneously in both training and test sets, thus preventing data leakage. Specifically, 5-fold cross-validation was conducted and repeated 10 times to mitigate biases introduced by random data partitioning. In each training iteration, the model was trained on the training set and evaluated on the corresponding validation set. The final performance metrics were computed as the mean and standard deviation across all iterations, and results are reported as mean ± standard deviation.

For evaluation, five commonly used classification metrics were employed. Accuracy is defined as the ratio of correctly predicted samples (both positive and negative) to the total number of samples, providing an overall measure of model correctness.

$$ACC = \frac{TP+TN}{TP+FN+TN+FP} \tag{10}$$

where TP denotes the number of AD samples correctly classified as AD, FP denotes the number of FTD samples incorrectly classified as AD, FN denotes the number of AD samples incorrectly classified as FTD, and TN denotes the number of FTD samples correctly classified as FTD.

Precision is defined as the ratio of correctly predicted positive samples to the total number of samples predicted as positive, providing a measure of the model's ability to avoid false positives.

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

Recall is defined as the ratio of correctly predicted positive samples to the total number of actual positive samples, providing a measure of the model's ability to identify positive cases.

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

F1 Score is defined as the harmonic mean of precision and recall, providing a balanced measure between the two.

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{13}$$

Specificity is defined as the ratio of correctly identified negative samples to the total number of actual negative samples, providing a measure of the model's ability to correctly exclude non-AD individuals.

$$Specificity = \frac{TN}{TN+FP} \tag{14}$$

These metrics collectively provide a comprehensive evaluation of the model's performance from multiple perspectives.

2.4 Datasets and Data Processing

2.4.1 Dataset

This study utilized a publicly available dataset collected by Miltiadous et al. [26], consisting of 36 patients with AD and 23 patients with FTD. Table 1 shows the mean age, gender distribution, and Mini-Mental State Examination (MMSE) scores for each group. The median disease duration was 25 months (interquartile range (IQR): 24–28.5 months), and no dementia-related comorbidities were reported in the AD group [26]. Details are shown in Table 1.

Table 1. Participant data.

	AD (n = 36)	FTD (n = 23)
Sex (m:f)	1:2	14:9
Age	66.4 ± 7.9	63.9 ± 8.2
MMSE	17.75 ± 4.5	22.17 ± 8.22

MMSE, Mini-Mental State Examination; AD, Alzheimer's disease; FTD, frontotemporal dementia; m, male; f, female.

The EEG data were collected in a clinical setting by experienced neurologists using a Nihon Kohden EEG-2100 system (Nihon Kohden Corp., Tokyo, Japan). Signals were recorded according to the international 10–20 system with a total of 19 scalp electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2), with A1–A2 mastoid electrodes serving as the reference. All recordings were conducted under eyes-closed resting-state conditions with a sampling rate of 500 Hz. The duration of EEG recording for each subject was approximately 12–14 minutes.

2.4.2 Data Processing

The raw EEG signals were initially bandpass-filtered between 0.5 and 45 Hz using a Butterworth filter and subsequently re-referenced to the A1–A2 montage. Anomalous EEG segments were then identified and removed using the Artifact Subspace Reconstruction (ASR) method in EEGLab (Swartz Center for Computational Neuroscience, University of California San Diego, La Jolla, CA, USA), with a threshold set at 17 standard deviations. The 19-channel EEG signals were subsequently decomposed into 19 independent components (ICs) using the RunICA algorithm (runica function in EEGLAB, based on the Infomax algorithm). These ICs were automatically classified and filtered with Independent Component Labeler (ICLabel) to remove artifacts related to eye movements and muscle activity.

Electrophysiological source imaging (ESI) was employed for source localization to map EEG signals onto predefined cortical regions of interest (ROIs). For example, at a given time point in an EEG recording, if higher potential activity is observed at scalp electrodes F3 and Fp1, the Standardized Low-Resolution Brain Electromagnetic Tomography (sLORETA) algorithm can solve the inverse problem to map the scalp potential distribution onto cortical space. In doing so, it estimates that the activity most likely originates from neural sources in frontal brain regions, thereby achieving a transformation from scalp potentials to cortical source activity. The preprocessed EEG signals were bandpass-filtered into four frequency bands, θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–48 Hz), and subsequently segmented into non-overlapping 5-second epochs. The sLORETA algorithm was then applied to reconstruct the time series of neural source activity for each brain region within each frequency band. The total neural oscillation energy was quantified by computing the sum of squares of the reconstructed signals. According to signal processing theory, the sum of squares of a time-domain signal is equivalent to the integral of its corresponding frequency-domain power spectrum. Therefore, this energy metric directly quantifies the total oscillatory intensity of each frequency band in every brain region.

Considering that the sLORETA algorithm is highly sensitive to noise and artifacts, and to address the sample imbalance caused by the unequal number of subjects in the AD and FTD groups, we adopted a data-level balancing strategy after rigorously removing artifacts. Specifically, we carefully extracted eight high-quality 5-second segments from the valid resting-state EEG of each AD patient, and thirteen segments from each FTD patient. The final classification dataset comprised a total of 587 instances (288 from the AD group and 299 from the FTD group), achieving a well-balanced distribution between the two groups. For each instance, the features consist of the energy values of 68 brain regions across four frequency bands, resulting in a feature vector of dimensionality 272 (68×4), which was subsequently used for training and evaluation of the classification models.

3. Results

3.1 Experimental Setup

In the AD-FTD discrimination task, the proposed AM-SSF model was systematically evaluated for performance. To ensure reliable and reproducible results, the experiments followed a carefully designed protocol. Independent self-supervised encoders were trained for each frequency band. Each encoder comprised two fully connected layers and incorporated Layer Normalization (LayerNorm), Gaussian Error Linear Unit (GELU) activation, and Dropout to enhance expressive power and robustness. During self-supervised training, data augmentation techniques, including random scaling, translation, and noise perturbation, were applied to improve feature generalization. The output feature dimension for each encoder was set to 128, and training was conducted for 300 epochs.

Different feature fusion methods were subsequently compared. The average pooling method computed the mean of features across all frequency bands at the same dimension. The attention fusion method integrated features from different frequency bands through a weighted attention mechanism. The supervision-guided attention fusion method introduced a cross-entropy loss to further enhance the discriminative power of frequency band weight allocation. The attention network consisted of two fully connected layers, and the weight distribution across frequency bands was obtained through Softmax normalization.

Based on the fused features, a random forest model was employed as the primary classifier. Performance evaluation metrics included accuracy, precision, recall, specificity, and F1-score. To ensure robustness, a repeated group cross-validation strategy was adopted. GroupKFold is a variant of k-fold cross-validation that ensures all samples from the same subject are kept together in either the training or validation set, preventing data leakage across folds. Specifically, GroupKFold was used to perform 5-fold cross-validation, repeated 10 times, ensuring that data from the same subject did not appear simultaneously in both the training and validation sets, thereby effectively preventing data leakage.

To further verify that the grouping strategy did not introduce systematic bias, statistical analyses were conducted at the independent subject level ($n = 59$) to compare the demographic baselines between the training and validation sets in the cross-validation procedure. The results indicated

Table 2. Classification performance of single-band and combined-band models under cross-validation (mean ± standard deviation).

Band	Accuracy	Precision	Recall	Specificity	F1-score
θ	0.905 ± 0.040	0.901 ± 0.047	0.904 ± 0.069	0.906 ± 0.032	0.901 ± 0.046
α	0.687 ± 0.052	0.705 ± 0.110	0.637 ± 0.171	0.747 ± 0.091	0.655 ± 0.104
β	0.857 ± 0.081	0.839 ± 0.119	0.890 ± 0.072	0.827 ± 0.128	0.860 ± 0.085
γ	0.639 ± 0.094	0.642 ± 0.116	0.697 ± 0.097	0.566 ± 0.232	0.656 ± 0.069
θ + β	0.915 ± 0.047	0.898 ± 0.088	0.942 ± 0.051	0.895 ± 0.086	0.916 ± 0.046
Ours	0.931 ± 0.028	0.914 ± 0.042	0.949 ± 0.032	0.921 ± 0.046	0.929 ± 0.032

Note: The results in the table represent the mean ± standard deviation obtained from 10 repetitions of five-fold cross-validation. F1-Score, the harmonic mean of precision and recall (range 0–1).

that, across all folds, there were no statistically significant differences between the two sets in terms of age, MMSE scores, sex distribution, or disease category proportions, as shown in **Supplementary Table 1**. These findings confirm that the training and validation sets achieved good baseline balance, suggesting that the observed improvement in model performance is attributable to the model's effective extraction of disease-related EEG pathological features.

Furthermore, to analyze the impact of data scale on model performance, comparative experiments were conducted using different training set proportions. Training sets comprising 20%, 40%, 60%, and 80% of the data were used, and evaluation was performed under the same five-fold cross-validation protocol.

3.2 Single-Band Self-Supervised Feature Analysis

To assess the independent contributions of each frequency band in self-supervised feature learning, single-band self-supervised learning was conducted separately for the θ, α, β, and γ bands, followed by validation in downstream AD/FTD classification tasks. Accuracy, precision, recall, specificity, and F1-score were used as evaluation metrics. The results indicated notable differences in classification performance among features derived from different frequency bands. The θ band achieved the highest performance in cross-validation, followed by the β band, whereas the α and γ bands exhibited comparatively lower performance.

To further verify the complementarity among different frequency bands, this experiment compared the classification performance of single-band models and combined-band models. The results show that, in the dual-band combination, merging the two best-performing single frequency bands (θ + β) improved the accuracy to 0.915 and achieved a recall of 0.942, demonstrating the complementary value across frequency bands. Ultimately, the complete AM-SSF model integrating features from all four frequency bands achieved the best overall classification performance. Detailed classification performance metrics are presented in Table 2.

To rigorously validate the statistical significance of the performance improvements, Wilcoxon signed-rank tests were conducted for all models. The results indicate that the proposed AM-SSF model significantly outperforms both single-frequency-band models and dual-frequency-band fusion models in terms of accuracy and F1-score. Detailed statistical results, including 95% confidence intervals, test statistic W, and p-values, are provided in **Supplementary Table 2**.

To further investigate this discrepancy, t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction was applied to visualize features learned from individual frequency bands. Fig. 2 shows the distribution of features across frequency bands in the t-SNE space. The results indicate that AD and FTD samples in the θ and β bands display distinct clustering in the feature space, whereas differentiation between classes in the α and γ bands remains relatively ambiguous. Additionally, three clustering metrics, Silhouette, Calinski-Harabasz (CH), and Davies-Bouldin (DB), were computed to quantitatively evaluate class separability. The θ and β bands exhibit relatively high separability, with Silhouette scores of 0.2025 and 0.1542, respectively. In contrast, the α and γ bands show lower metric values, indicating weaker class separability, as summarized in Table 3.

Table 3. Clustering evaluation metrics across different frequency bands.

Band	Silhouette	CH Index	DB Index	Conclusion
θ	0.2025	108.87	1.9091	Best
α	0.0703	29.25	3.7792	Worst
β	0.1542	79.07	2.2698	Second best
γ	0.0918	37.43	3.1667	Poor

CH, Calinski-Harabasz; DB, Davies-Bouldin. CH Index measures the ratio of between-cluster dispersion to within-cluster compactness; higher values indicate better clustering quality. DB Index measures the average similarity between each cluster and its most similar one; lower values indicate better clustering quality.

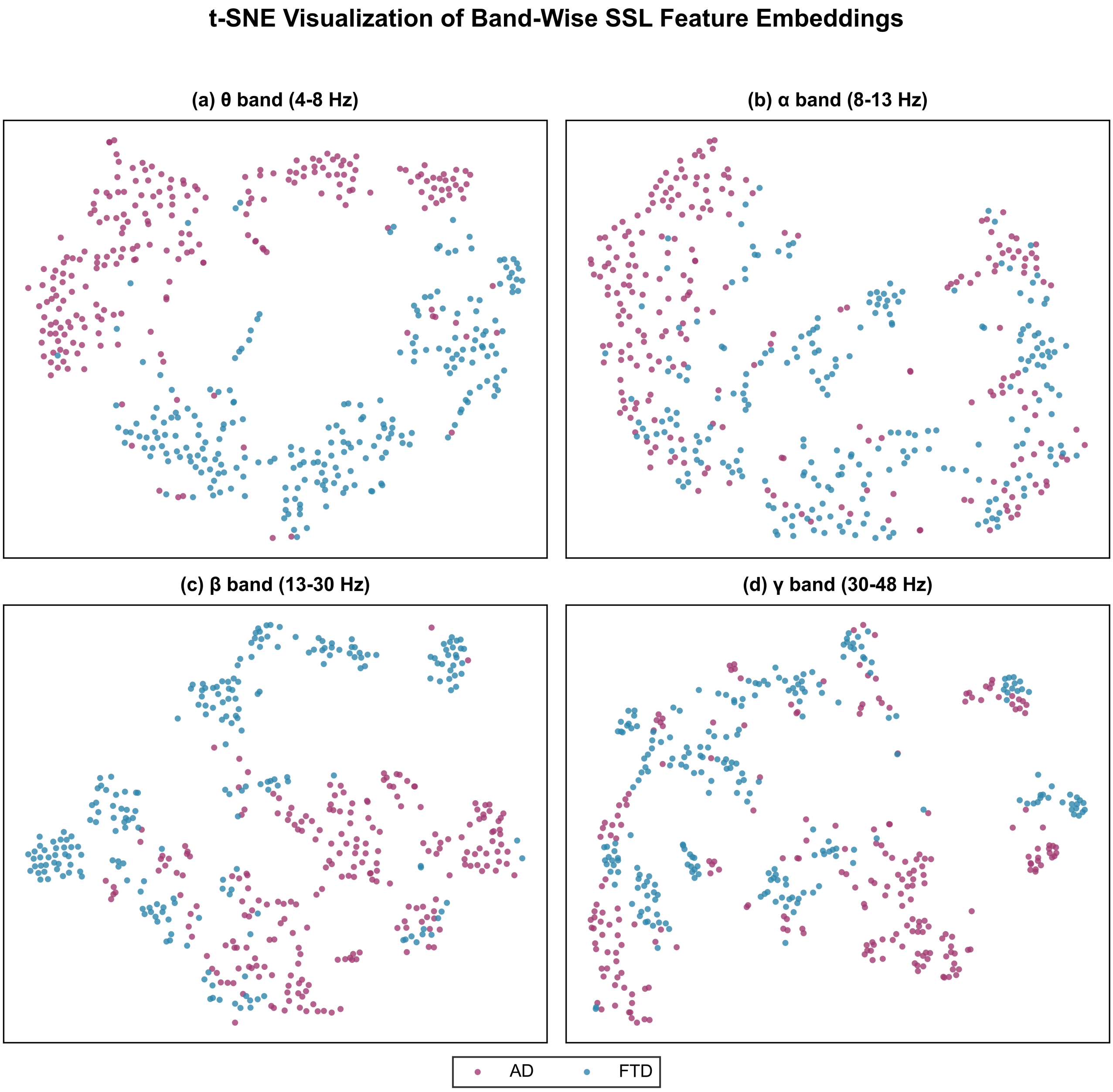

Fig. 2. t-SNE visualization of features learned by self-supervised learning across θ, α, β, and γ frequency bands. (a) distribution of θ-band features in the t-SNE space, (b) distribution of α-band features, (c) distribution of β-band features, (d) distribution of γ-band features. t-SNE, t-Distributed Stochastic Neighbor Embedding; SSL, self-supervised learning.

Although the θ band exhibits superior performance in both classification and feature separability, a single frequency band remains insufficient to comprehensively capture disease-related EEG information. Fusion of features across multiple frequency bands provides more comprehensive discriminative information, thereby enhancing the performance of downstream classification tasks.

3.3 Multi-Frequency Attention Fusion and Baseline Comparison

The effectiveness of multi-frequency feature fusion methods was evaluated based on their performance in downstream classification tasks. This study evaluated model performance using accuracy, precision, recall, specificity, and F1-score, and compared the baseline model, average pooling fusion, attention-based fusion, pretrained model, and our proposed model. As shown in Fig. 3, both average pooling and attention-based fusion outperform the baseline, indicating that multi-frequency-band self-supervised fusion can significantly enhance the discriminative capability of the model. However, their performance is comparable, suggesting that a purely self-supervised attention mechanism provides limited improvement in overall performance. In contrast, the attention-based fusion model

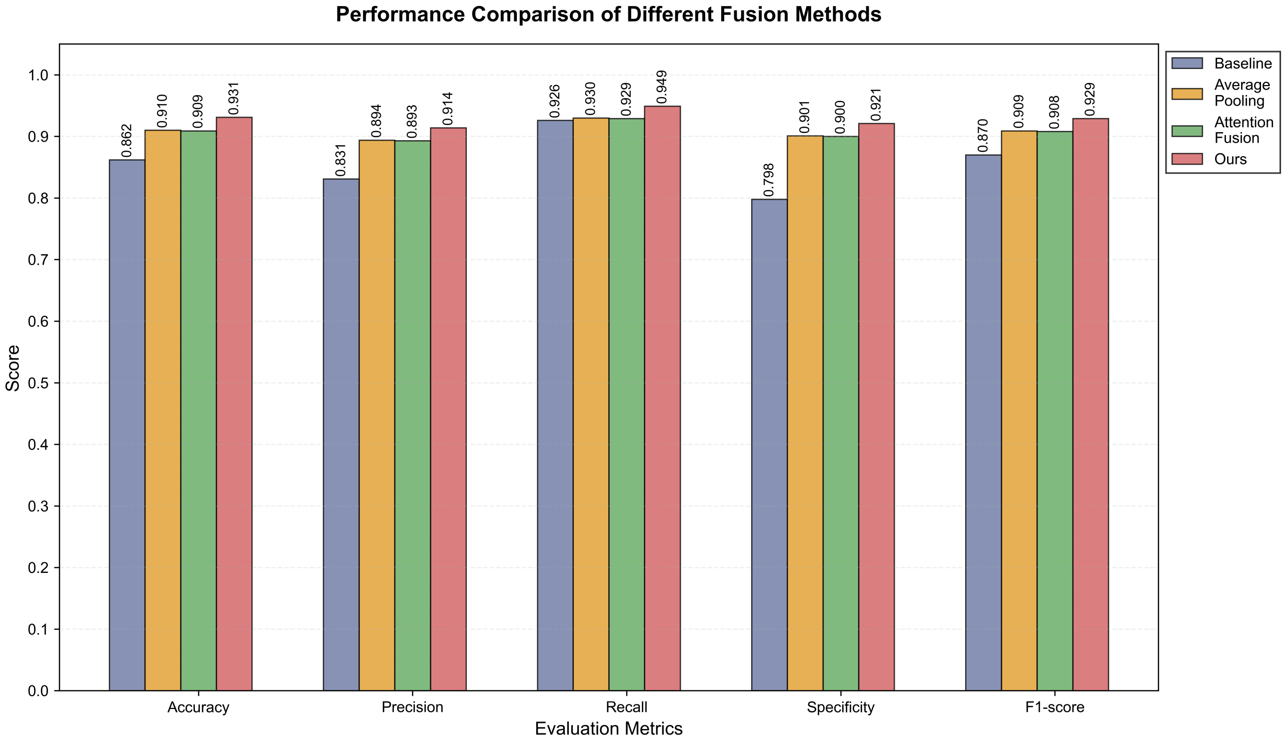

Fig. 3. Performance Comparison of Different Fusion Methods. Classification performance metrics (Accuracy, Precision, Recall, Specificity, and F1-score) of baseline, average pooling, attention fusion, and Ours under 10 repetitions of 5-fold cross-validation.

incorporating cross-entropy supervision significantly improves all evaluation metrics. This indicates that task-guided adaptive attention can effectively steer the allocation of attention weights, thereby strengthening key frequency-band features and improving both the discriminative ability and interpretability of the model.

To further evaluate the effectiveness of the proposed network architecture and verify the superiority of multi-frequency-band adaptive fusion over general pretrained models, two deep learning–based pretrained baseline models were introduced for comparison. The experimental results show that the autoencoder and single-SSL models achieved accuracies of 89.20% and 86.58%, respectively, representing improvements over the raw-feature baseline. This indicates that deep pretraining indeed enhances feature representation capability. However, these pretrained models without frequency-band disentanglement still perform significantly worse than the proposed AM-SSF model. This further confirms that single global feature extraction tends to overlook inter-band variability, whereas the multi-frequency-band adaptive fusion mechanism guided by cross-entropy supervision can more effectively integrate core EEG features that are directly relevant to AD/FTD discrimination.

Table 4 presents the average performance and 95% confidence intervals of each method under 10 repetitions of five-fold cross-validation. To rigorously evaluate performance gains, Wilcoxon signed-rank tests were conducted to assess statistical significance among the models. The results show that the proposed AM-SSF model achieves the best performance across all metrics, with an accuracy of 0.931 [0.922, 0.941] and an F1-score of 0.929 [0.919, 0.940]. Statistical tests indicate that AM-SSF exhibits significant differences compared with all competing models in both ACC and F1-score ($p < 0.05$). The complete statistical results and test statistics are provided in **Supplementary Table 2**.

Fig. 4 presents the distribution of attention weights across different frequency bands. The results indicate clear differences in the model's attention allocation across frequency bands. The θ and β bands exhibit higher median attention values and relatively compact distributions, suggesting that the model places greater emphasis on features within these bands. In contrast, the α and γ bands display lower attention values with greater variability, implying a relatively smaller contribution to the classification task. Notably, this observation is consistent with the findings from the single-frequency band feature analysis in Section 3.2, confirming that the θ and β bands possess stronger discriminative power in AD/FTD classification.

3.4 Impact of Data Volume Dimension on Results

To evaluate the model's performance under different training data scales, this study utilized 20%, 40%, 60%, and 80% of the dataset as training sets. Five-fold cross-validation was conducted under identical experimental con-

Table 4. Statistical comparison of fusion strategies based on Accuracy and F1-score.

Method	Accuracy [95% CI]	F1-score [95% CI]	p-value (vs. Ours)
Baseline	0.862 [0.847, 0.876]	0.870 [0.860, 0.881]	<0.001
Single-SSL	0.866 [0.839, 0.893]	0.868 [0.843, 0.894]	<0.001
Autoencoder	0.892 [0.868, 0.916]	0.894 [0.871, 0.917]	<0.001
Average Pooling	0.910 [0.903, 0.918]	0.909 [0.901, 0.918]	0.0089
Attention Fusion	0.909 [0.901, 0.917]	0.908 [0.899, 0.917]	0.0062
Ours	0.931 [0.922, 0.941]	0.929 [0.919, 0.940]	-

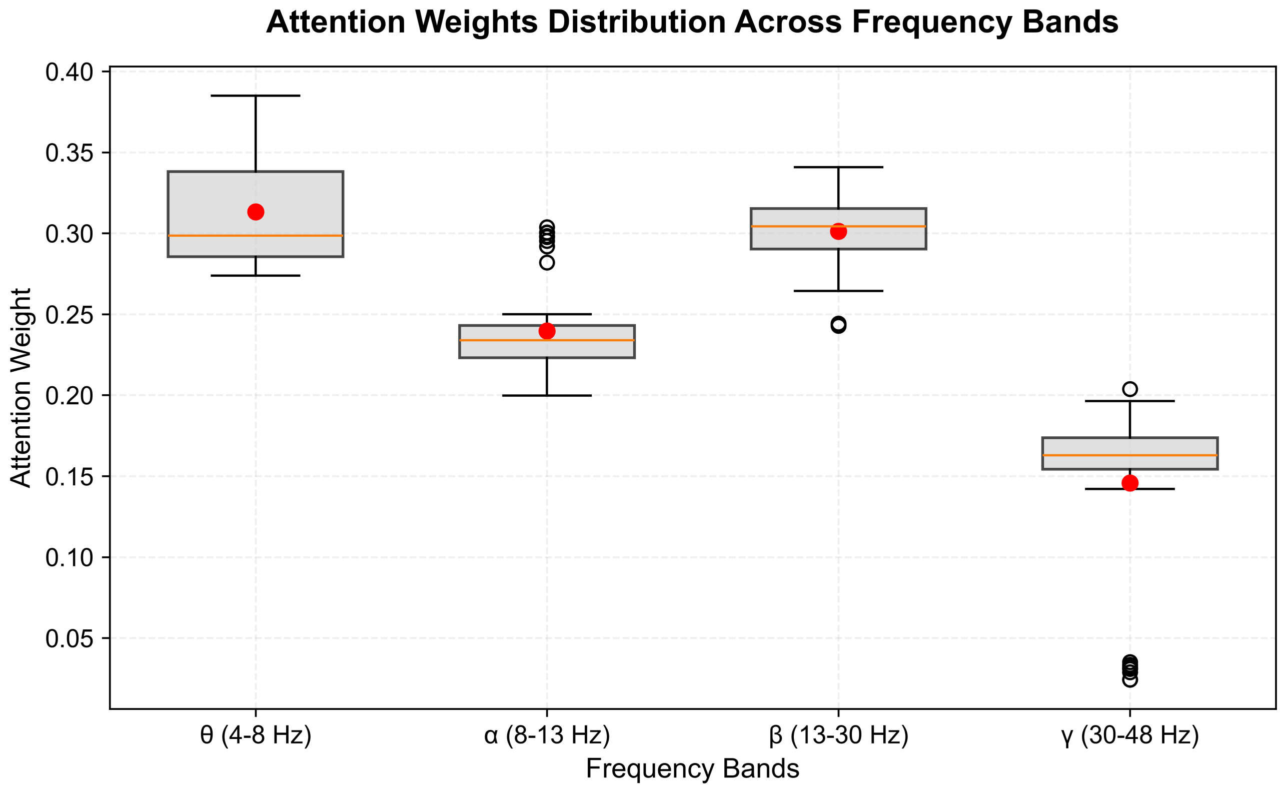

Fig. 4. Attention weight distributions of Ours across θ, α, β, and γ frequency bands.

ditions, and performance was compared with the Baseline method. Table 5 summarizes the classification performance metrics obtained under different training data volumes. The results show that when using 20% of the training data, both models achieve comparable performance. However, the proposed model outperforms the Baseline across all metrics, demonstrating greater stability. As the amount of training data increases, the model performance exhibits a steady upward trend, with notably larger gains over the Baseline at 40% and 60% training data ratios. Overall, the proposed method surpasses the Baseline across all data scales, confirming its effectiveness and robustness.

3.5 Performance Comparison With Existing Studies

To comprehensively evaluate the effectiveness of the proposed model, this study compares the classification performance of AM-SSF with recent methods applied to the same public dataset for AD and FTD discrimination. Table 6 (Ref. [11,27,28]) summarizes the feature extraction techniques, classification models, and key performance metrics of these studies.

Recent related studies have primarily focused on extracting handcrafted features from the sensor space, such as nonlinear dynamics and network connectivity, and combining them with traditional machine learning classifiers for prediction. For example, Rostamikia et al. [11] systematically extracted multidimensional handcrafted features, including time-domain, frequency-domain, nonlinear complexity, and brain network connectivity features, and employed an SVM classifier, achieving a highest accuracy of 87.8% under 10-fold cross-validation. Another study utilized manually extracted Singular Value Decomposition (SVD) entropy features combined with a sliding window technique with 90% overlap and a K-Nearest Neighbors (KNN) model, reporting an accuracy of 91.0% [27]. In addition, a brain network analysis method based on cross-plot

Table 5. Comparison of classification performance between AM-SSF and baseline methods under different training data proportions.

Data volume	Method	Accuracy	Precision	Recall	F1-score
20%	Ours	0.764 ± 0.210	0.719 ± 0.385	0.698 ± 0.401	0.682 ± 0.365
	Baseline	0.738 ± 0.148	0.639 ± 0.367	0.717 ± 0.393	0.647 ± 0.339
40%	Ours	0.862 ± 0.115	0.874 ± 0.181	0.884 ± 0.122	0.857 ± 0.109
	Baseline	0.786 ± 0.149	0.806 ± 0.202	0.869 ± 0.189	0.795 ± 0.130
60%	Ours	0.877 ± 0.066	0.876 ± 0.100	0.858 ± 0.059	0.864 ± 0.068
	Baseline	0.645 ± 0.095	0.599 ± 0.103	0.694 ± 0.089	0.640 ± 0.087
80%	Ours	0.931 ± 0.033	0.914 ± 0.066	0.949 ± 0.032	0.929 ± 0.037
	Baseline	0.862 ± 0.051	0.831 ± 0.094	0.926 ± 0.051	0.870 ± 0.036

AM-SSF, Attention-based Multi-frequency Self-supervised Fusion Model.

Table 6. Performance comparison of methods for AD vs. FTD classification on the same dataset.

Reference	Feature Space	Classifier/Model	Accuracy	F1-score
Ranjan and Kumar (2024) [28]	Sensor space (CPTE network parameters)	RF/KNN	0.876	-
Rostamikia et al. (2024) [11]	Sensor space (Multi-dimensional hand-crafted features)	SVM	0.878	-
Lal et al. (2024) [27]	Sensor space (SVD entropy)	KNN	0.910	0.915
Ours	Source space (Self-supervised features)	Attention fusion + RF	0.931	0.929

AD, Alzheimer's disease; FTD, frontotemporal dementia; CPTE, Cross-plot transition entropy; RF, Random Forest; KNN, K-Nearest Neighbors; SVM, Support Vector Machine; SVD, Singular Value Decomposition.

transition entropy (CPTE) achieved an overall accuracy of 87.6% in distinguishing between the two groups [28].

In contrast, the proposed AM-SSF model achieves a highest accuracy of 93.1%, demonstrating the advantage of deep self-supervised representation learning over traditional handcrafted feature engineering. This study overcomes the physical limitations of the sensor space by employing deep networks to extract neural energy features in the source space, effectively reducing spatial blurring caused by volume conduction effects. Furthermore, the introduced multi-frequency-band self-supervised fusion mechanism eliminates the need for manually selecting specific frequency bands based on prior knowledge, enabling adaptive integration of high-dimensional discriminative information across the full spectrum.

4. Discussion

This study proposes an attention-based multi-frequency self-supervised fusion model designed to classify AD and FTD using EEG signals. The model learns disease-specific representations from distinct EEG frequency bands through independent self-supervised encoders. Under the supervision of the classification task, the model employs an attention mechanism to adaptively integrate cross-frequency features, thereby enhancing its ability to capture disease-related information. Experimental results demonstrate that the proposed AM-SSF model significantly outperforms both single-band models and conventional fusion methods in overall classification performance while maintaining high robustness under small-sample conditions.

4.1 Analysis of Single-Band Feature Self-Supervised Representations

First, the performance of individual frequency bands was evaluated within the SSL framework. The results indicated that self-supervised features from the θ and β bands achieved the best classification performance, with average accuracies of 90.5% and 85.7%, respectively. In contrast, the α and γ bands achieved relatively lower accuracies, ranging from 60% to 70%. These findings reveal significant differences in the contributions of various frequency bands to dementia classification, with the θ and β bands exhibiting stronger discriminative power. Further evidence from clustering metrics and t-SNE visualization corroborated these findings, showing superior category separation of θ and β band features in the latent space. This discovery is consistent with existing neurophysiological evidence on abnormal θ and β oscillations in AD and FTD [29,30], suggesting that subsequent multi-band fusion strategies should prioritize these critical frequency bands.

4.2 The Role of Multi-Band Self-Supervised Learning and Attention Fusion in Downstream AD/FTD Classification

In the AD/FTD classification task, although the baseline method exhibited some discriminative capability, its overall performance remained limited. After introducing multi-frequency self-supervised features, the model achieved a significant improvement in classification performance, indicating complementary information across frequency bands. Notably, both average pooling and attention fusion achieved comparable performance; however, average pooling could not differentiate the relative importance of frequency bands, whereas the attention mechanism adaptively adjusted the weight distribution. When further combined with cross-entropy supervision, the model achieved the best overall performance (accuracy = 0.931). Specifically, the θ and β frequency bands received higher attention weights, consistent with single-band results and supported by existing neuroscientific evidence. This outcome highlights the advantage of the attention mechanism in feature fusion and demonstrates the model's ability to focus automatically on the most relevant frequency band features.

4.3 Model Validity Under Small Sample Conditions

Under small-sample conditions, the proposed model maintained robust performance (accuracy = 76.4%) and outperformed the baseline in both precision and F1-score. This demonstrates that self-supervised pretraining, together with attention fusion, effectively mitigates performance degradation under limited data conditions. As the training dataset expanded, model performance progressively improved, reaching its peak at 80% data coverage (accuracy = 93.1%). The proposed method consistently outperformed the baseline across all data proportions, confirming its robustness and stability under small-sample scenarios. This characteristic is particularly valuable for EEG research scenarios with limited data availability and high annotation costs.

4.4 Computational Complexity

In practical applications of medical artificial intelligence, whether an auxiliary diagnostic model can be deployed in real clinical environments is largely constrained by computational complexity. To this end, we evaluated the computational cost of the proposed AM-SSF model. In terms of theoretical computational scale, the model avoids large-scale deep matrix operations. The total number of trainable parameters in the deep learning component is 350.34 K, and the floating-point operations per single forward pass (FLOPs) amount to 0.35 M. In terms of practical runtime efficiency, the average inference time per sample is 54.06 milliseconds, which meets the requirements of auxiliary diagnosis in real clinical scenarios.

4.5 Limitations

Nevertheless, several limitations should be acknowledged. First, this study currently relies on a single-center cohort. In real-world settings, EEG acquisition devices and subject population characteristics vary inherently across different clinical centers, which may introduce unknown effects on the stability of multi-frequency-band self-supervised features. Therefore, future work should incorporate multi-center cohorts to validate cross-domain generalization ability. Second, the current attention mechanism only reveals feature importance at the frequency-band level and has not yet established precise spatial mapping relationships between abnormal frequency bands and specific affected brain regions or dynamic neural circuits. Finally, the subjects included in this study are primarily typical AD and FTD patients, and the model's discriminative ability has not been fully evaluated in clinically common mixed dementia cases or patients at very early disease stages. Future work will focus on integrating EEG source localization techniques to construct a spatio-temporal-frequency multi-dimensional interpretable fusion framework, and to validate and optimize the model in real clinical cohorts with more complex pathological types. Meanwhile, considering the efficient optimization of feature fusion weights, advanced swarm intelligence optimization algorithms may be explored in the future to further enhance the model's generalization performance in complex environments [31].

5. Conclusions

In this study, a model termed AM-SSF is proposed for the classification of AD and FTD. The model enhances feature representation through self-supervised pretraining and employs an attention mechanism to adaptively weight information across distinct frequency bands, thereby enabling the effective integration of discriminative features. Experimental results demonstrate that AM-SSF significantly outperforms baseline methods across varying training data proportions. Its highest accuracy reaches 93.1%, with an F1-score of 92.9%, and it maintains robustness even under small-sample conditions. Furthermore, the attention fusion module with cross-entropy supervision not only enhances classification performance but also improves the interpretability of frequency-band weighting. The model assigns higher attention weights to the θ and β bands, which aligns with previously reported EEG rhythm abnormalities in AD and FTD, thereby supporting the biological plausibility and interpretability of the proposed approach.

Availability of Data and Materials

Data used in preparation of this article are obtained from the OpenNeuro dataset (https://openneuro.org/datasets/ds004504/versions/1.0.6). Additional information can be provided by corresponding author upon reasonable request.

Author Contributions

JS designed the project and supervised the overall research; JL collected data, performed experiments and analysis; QZ and ZG co-designed the research; RZ, MW, YC, XZ, and ZS participated in data collection or analysis. JL and JS wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and have agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

We thank all participants who participated in our study.

Funding

This work was supported by Key Program of Natural Science Foundation of Zhejiang Province under Grant No. LZ24F010007 and the National Natural Science Foundation of China under Grant No. 62271177.

Conflicts of Interest

The authors declare no conflicts of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/JIN50737.

References

[1] Mollah SA, Nayak A, Barhai S, Maity U. A comprehensive review on frontotemporal dementia: its impact on language, speech and behavior. Dementia & Neuropsychologia. 2024; 18: e20230072. https://doi.org/10.1590/1980-5764-DN-2023-0072
[2] Yu M, Gouw AA, Hillebrand A, Tijms BM, Stam CJ, van Straaten ECW, et al. Different functional connectivity and network topology in behavioral variant of frontotemporal dementia and Alzheimer's disease: an EEG study. Neurobiology of Aging. 2016; 42: 150–162. https://doi.org/10.1016/j.neurobiolaging.2016.03.018
[3] Yuan Y, Zhao Y. The role of quantitative EEG biomarkers in Alzheimer's disease and mild cognitive impairment: applications and insights. Frontiers in Aging Neuroscience. 2025; 17: 1522552. https://doi.org/10.3389/fnagi.2025.1522552
[4] Caso F, Cursi M, Fanelli G, Falautano M, Leocani L, Comi G, et al. EEG spectral analysis and low-resolution brain electromagnetic tomography (LORETA) in diagnosis of frontotemporal dementia and differences with Alzheimer's disease and healthy subjects. Clinical Neurophysiology. 2011; 122: S20–S21. https://doi.org/10.1016/S1388-2457(11)60066-7
[5] Ranjan S, Tripathi A, Shende H, Badal R, Kumar A, Yadav P, et al. Deep learning-based classification of dementia using image representation of subcortical signals. BMC Medical Informatics and Decision Making. 2025; 25: 113. https://doi.org/10.1186/s12911-025-02924-w
[6] Nishida K, Yoshimura M, Isotani T, Yoshida T, Kitaura Y, Saito A, et al. Differences in quantitative EEG between frontotemporal dementia and Alzheimer's disease as revealed by LORETA. Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology. 2011; 122: 1718–1725. https://doi.org/10.1016/j.clinph.2011.02.011
[7] Ge Y, Yin J, Chen C, Yang S, Han Y, Ding C, et al. An EEG-based framework for automated discrimination of conversion to Alzheimer's disease in patients with amnestic mild cognitive impairment: an 18-month longitudinal study. Frontiers in Aging Neuroscience. 2025; 16: 1470836. https://doi.org/10.3389/fnagi.2024.1470836
[8] Chien HYS, Goh H, Sandino CM, Cheng JY. MAEEG: Masked auto-encoder for EEG representation learning. arXiv. 2022. https://doi.org/10.48550/arXiv.2211.02625 (preprint).
[9] Si Y, He R, Jiang L, Yao D, Zhang H, Xu P, et al. Differentiating Between Alzheimer's Disease and Frontotemporal Dementia Based on the Resting-State Multilayer EEG Network. IEEE Transactions on Neural Systems and Rehabilitation Engineering : a Publication of the IEEE Engineering in Medicine and Biology Society. 2023; 31: 4521–4527. https://doi.org/10.1109/TNSRE.2023.3329174
[10] Zheng H, Xiong X, Zhang X. Multi-Threshold Recurrence Rate Plot: A Novel Methodology for EEG Analysis in Alzheimer's Disease and Frontotemporal Dementia. Brain Sciences. 2024; 14: 565. https://doi.org/10.3390/brainsci14060565
[11] Rostamikia M, Sarbaz Y, Makouei S. EEG-based classification of Alzheimer's disease and frontotemporal dementia: a comprehensive analysis of discriminative features. Cognitive Neurodynamics. 2024; 18: 3447–3462. https://doi.org/10.1007/s11571-024-10152-7
[12] Zhang C, Kim YK, Eskandarian A. EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. Journal of Neural Engineering. 2021; 18: 046014. https://doi.org/10.1088/1741-2552/abed81
[13] Fan C, Yang B, Li X, Zan P. Temporal-frequency-phase feature classification using 3D-convolutional neural networks for motor imagery and movement. Frontiers in Neuroscience. 2023; 17: 1250991. https://doi.org/10.3389/fnins.2023.1250991
[14] Miltiadous A, Gionanidis E, Tzimourta KD, Giannakeas N, Tzallas AT. DICE-Net: A novel convolution-transformer architecture for Alzheimer detection in EEG signals. IEEE Access. 2023; 11: 71840–71858. https://doi.org/10.1109/ACCESS.2023.3294618
[15] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics. 2019.
[16] Chen T, Kornblith S, Norouzi M, Hinton GH. A simple framework for contrastive learning of visual representations. In 37th international conference on machine learning (ICML). PMLR. 2020.
[17] He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. In IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE. 2022.
[18] Banville H, Chehab O, Hyvärinen A, Engemann DA, Gramfort A. Uncovering the structure of clinical EEG signals with self-supervised learning. Journal of Neural Engineering. 2021; 18: 10.1088/1741–2552/abca18. https://doi.org/10.1088/1741-2552/abca18
[19] Zhao C, Wu W, Zhang H, Zhang R, Zheng X, Kong X. Sleep Stage Classification Via Multi-View Based Self-Supervised Contrastive Learning of EEG. IEEE Journal of Biomedical and Health Informatics. 2024; 28: 7068–7077. https://doi.org/10.1109/JBHI.2024.3432633
[20] Wang X, Ma Y, Cammon J, Fang F, Gao Y, Zhang Y. Self-Supervised EEG Emotion Recognition Models Based on CNN. IEEE Transactions on Neural Systems and Rehabilitation Engineering : a Publication of the IEEE Engineering in Medicine and Biology Society. 2023; 31: 1952–1962. https://doi.org/10.1109/TNSRE.2023.3263570
[21] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv. 2014. https://doi.org/10.48550/arXiv.1409.0473 (preprint).
[22] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In neural information processing systems (NIPS 2017). Curran Associates, Inc. 2017.
[23] Chen W, Liao Y, Dai R, Dong Y, Huang L. EEG-based emotion recognition using graph convolutional neural network with dual attention mechanism. Frontiers in Computational Neuroscience. 2024; 18: 1416494. https://doi.org/10.3389/fncom.2024.1416494

[24] Aziz MZ, Yu X, Guo X, He X, Huang B, Fan Z. BCINetV1: Integrating Temporal and Spectral Focus Through a Novel Convolutional Attention Architecture for MI EEG Decoding. Sensors (Basel, Switzerland). 2025; 25: 4657. https://doi.org/10.3390/s25154657

[25] Golovanevsky M, Eickhoff C, Singh R. Multimodal attention-based deep learning for Alzheimer's disease diagnosis. Journal of the American Medical Informatics Association : JAMIA. 2022; 29: 2014–2022. https://doi.org/10.1093/jamia/ocac168

[26] Miltiadous A, Tzimourta KD, Afrantou T, Ioannidis P, Grigoriadis N, Tsalikakis DG, et al. A dataset of EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects. OpenNeuro. 2023; 1: 10.3390. https://doi.org/10.18112/openneuro.ds004504.v1.0.6

[27] Lal U, Chikkankod AV, Longo L. A Comparative Study on Feature Extraction Techniques for the Discrimination of Frontotemporal Dementia and Alzheimer's Disease with Electroencephalography in Resting-State Adults. Brain Sciences. 2024; 14: 335. https://doi.org/10.3390/brainsci14040335

[28] Ranjan S, Kumar L. Exploring brain network organization in alzheimer disease and frontotemporal dementia: A cross-plot transition entropy approach. arXiv. 2024. https://doi.org/10.48550/arXiv.2409.16823 (preprint).

[29] Lindau M, Jelic V, Johansson SE, Andersen C, Wahlund LO, Almkvist O. Quantitative EEG abnormalities and cognitive dysfunctions in frontotemporal dementia and Alzheimer's disease. Dementia and Geriatric Cognitive Disorders. 2003; 15: 106–114. https://doi.org/10.1159/000067973

[30] Giustiniani A, Danesin L, Bozzetto B, Macina A, Benavides-Varela S, Burgio F. Functional changes in brain oscillations in dementia: a review. Reviews in the Neurosciences. 2022; 34: 25–47. https://doi.org/10.1515/revneuro-2022-0010

[31] Zhong R, Wang Z, Hussien AG, Houssein EH, Al-Shourbaji I, Elseify MA, et al. Success History Adaptive Competitive Swarm Optimizer with Linear Population Reduction: Performance benchmarking and application in eye disease detection. Computers in Biology and Medicine. 2025; 186: 109587. https://doi.org/10.1016/j.compbiomed.2024.109587