



## An Adaptive Multimodal Meta-Fusion Framework for Alzheimer's Disease and Frontotemporal Dementia Classification using EEG, MRI, and Cognitive Biomarkers

Lalita Wasudev Moharkar <sup>a, b, \*</sup>, Monika Bhagwat <sup>a</sup>

<sup>a</sup> Pillai College of Engineering, Dr. K. M. Vasudevan Pillai Campus, Plot No. 10, Sector 16, New Panvel, Navi Mumbai – 410 206, Maharashtra, India.

<sup>b</sup> Department of Computer Science and Engineering, Xavier Institute of Engineering, Mahim, Opposite S.L. Raheja Hospital, Mahim Causeway, Mahim (West), Mumbai – 400 016, Maharashtra, India

\* Corresponding Author Email: [lalitamoharkar@gmail.com](mailto:lalitamoharkar@gmail.com)

DOI: <https://doi.org/10.54392/irjmt2662>

Received: 14-06-2026; Revised: 29-09-2026; Accepted: 14-10-2026; Published: 21-10-2026



**Abstract:** A single-modality approach to neurodegenerative disease classification is often limited by heterogeneous clinical manifestations, modality-specific feature distributions, and incomplete data availability. This study proposes an Adaptive Multimodal Meta-Fusion (AMMF) framework integrating cognitive-demographic, structural MRI, and EEG biomarkers for classification of the Alzheimer's disease (AD), Frontotemporal Dementia (FTD), and Cognitively Normal (CN) participants. Three modality-specific cohorts from two public repositories were considered: cognitive-demographic data ( $n = 679$ ), structural MRI data ( $n = 595$ ), and EEG data ( $n = 194$ ; BrainLat  $n = 106$  and OpenNeuro  $n = 88$ ). Cognitive-demographic features are prioritized using explainable XGBoost with SHAP-based feature analysis, MRI data are processed using ConvNeXt-Small for spatial representation learning, and EEG signals are analyzed using a Transformer-based architecture for temporal and functional representation learning. For the matched BrainLat cohort, the three modality branches generated an 18-dimensional cognitive-demographic representation, a 128-dimensional EEG representation, and a 768-dimensional MRI representation. The proposed framework then applied adaptive late fusion at the modality-probability level: each modality generated an independent three-class probability vector, and a modality-availability mask was used to renormalize the predefined modality weights of 0.20, 0.30, and 0.50 assigned to the cognitive-demographic, EEG, and MRI modalities, respectively. The resulting weighted class-probability vector was supplied to a Random Forest meta-classifier for final three-class classification. The complete multimodal evaluation was performed on 106 subject-level matched participants with all three modalities. The proposed framework achieved an accuracy of 99.06%, demonstrating the utility of combining complementary cognitive, functional, and structural biomarkers through adaptive probability-level meta-fusion for AD, FTD, and CN classification. A continuity-corrected McNemar test yielded  $\chi^2 = 4.000$  ( $p = 0.0455$ ) for C4 versus C7 and  $\chi^2 = 4.167$  ( $p = 0.0412$ ) for C5 versus C7, indicating statistically significant paired differences, while C6 versus C7 yielded  $\chi^2 = 3.200$  ( $p = 0.0736$ ). Integration of cognitive-demographic, MRI, and EEG modalities for multimodal disease classification. The proposed framework integrates cognitive-demographic, MRI, and EEG modalities for multimodal disease classification by combining heterogeneous features in a unified representation space and applying Random Forest-based meta-classification to the fused modality probabilities.

**Keywords:** Multimodal Learning, Neurodegenerative Disease Classification, EEG Signal Analysis, MRI Deep Learning, Xgboost With SHAP, Convnext-Small Network, Transformer Model, Feature Fusion, Random Forest Classifier

### 1. Introduction

Background: Alzheimer disease (AD), Frontotemporal Dementia (FTD) are important neurodegenerative conditions that affect cognitive and functional abilities. Accurate classification of these conditions is important for clinical assessment and decision-making. Multimodal learning has attracted

increasing interest because of its ability to combine heterogeneous data sources. It provides complementary insights into brain structure, function, and pathological characteristics. Recent deep learning-based multimodal models have demonstrated the ability to learn latent representations from different modalities for improved analysis and prediction of neurodegenerative diseases

[1]. Advances in biomedical data analysis using artificial intelligence and machine learning have improved predictive performance, supporting clinical decision-making [2]. Advances in deep learning have also facilitated hybrid architectures combining convolutional and recurrent neural networks for multimodal healthcare applications [3]. Neuroimaging research also suggests that structural and functional brain connectivity is important for cognitive functioning in neurodegenerative disorders [4]. In addition to neuroimaging, multimodal methods combine cognitive and electrophysiological biomarkers, which can improve the identification of cognitive impairment [5]. Multimodal data fusion has gained increasing importance in intelligent healthcare systems, where AI-based approaches support data integration and clinical decision-making [6]. Attention-based fusion networks have been designed to enhance disease prediction by informative features across modalities [7] and transformer-based models can capture complex interdependencies among heterogeneous data sources [8].

To analyze multimodal neuroimaging data to infer relationships between biomarkers and help detect Alzheimer disease and identify the disease subtypes, graph-based learning technologies, such as graph convolutional networks (GCNs), have been used [9]. Moreover, multimodal deep learning methods have been applied in biomedical imaging to assess cognitive impairment [10]. Recent research has focused on incorporating multimodal biomarkers for the diagnosis and monitoring of neurodegenerative disorders. High-order fusion methods that integrate both neuroimaging and clinical data have shown higher classification ability [11], and joint learning methods have demonstrated improved robustness across modalities and better generalization [12]. Besides, classification of genetic information combined with medical images has demonstrated potential to enhance the diagnosis of Alzheimer disease [13]. Bidirectional (behavioral and clinical) multimodal frameworks have also proven to be effective with disease prediction tasks [14]. Moreover, multimodal learning models can enhance predictive performance by modelling relationships among heterogeneous data sources [15]. Altogether, multimodal data integration has a significant effect on diagnostic accuracy as it uses complementary structural data, functional data, and clinical data. This provides a strong framework for integrating complementary structural, functional, and clinical information for AD, FTD, and CN classification.

Despite this progress, most existing multimodal frameworks for AD and FTD classification suffer from three recurring limitations that motivate the present work. First, many studies treat "multimodal fusion" as simple feature concatenation rather than a mechanism that can adapt when one or more modalities are missing, which can occur in real world clinical settings where a full battery of cognitive, MRI, and EEG assessments is

rarely available for every patient. Second, published pipelines rarely verify that the subjects contributing to each modality are the same individuals; when cohorts of different sizes are merged without subject-level correspondence, the resulting classification results can reflect dataset artefacts rather than genuine cross-modal biomarker relationships. Third, EEG-based pipelines that merge clinical metadata with signal-derived features risk inadvertently including diagnosis-related fields among the predictors, which inflates reported accuracy through target leakage rather than through informative biomarker signal. These gaps motivate a framework that is evaluated with explicit subject-level data governance, transparent handling of missing modalities, and a feature set that is checked for leakage before being used for classification.

### 1.1 Aim and Objectives

The aim of this study is to develop and evaluate a multimodal classification framework that integrates cognitive-demographic, structural MRI, and EEG biomarkers for distinguishing AD, FTD, and CN subjects, while building in explicit safeguards against the data-handling problems identified above. The specific objectives are:

- (1) To construct a subject-level matched multimodal dataset in which every fused sample corresponds to a single participant with verified cognitive, MRI, and EEG records, rather than to padded, truncated, or otherwise unmatched records.
- (2) To design and implement a late-fusion mechanism whose behaviour under missing-modality conditions is explicitly specified, implemented, and evaluated rather than assumed.
- (3) To identify and remove any diagnosis-proximal or diagnosis-derived fields from the EEG and cognitive-demographic feature sets prior to model training, so that reported performance reflects genuine biomarker signal rather than target leakage.
- (4) To validate the resulting framework using a leakage-free, participant-level cross-validation protocol with consistent statistical testing.
- (5) To characterize, through class-specific explainability analysis, which biomarkers drive the model's predictions for each diagnostic group.

### 1.2 Related Work

Recent studies have explored multimodal machine learning approaches for detecting neurodegenerative disorders and identifying relevant biomarkers, particularly for Alzheimer's disease.

**Table 1.** Comparative summary of reviewed multimodal neurodegenerative-disease studies

| Ref.  | Study                                                  | Method                                                                     | Dataset                                | Major Findings                                                                                                                            | Limitations                                                                                     | Research Gap                                                                                           |
|-------|--------------------------------------------------------|----------------------------------------------------------------------------|----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| [1,6] | Liu <i>et al.</i> (2024);<br>Chen <i>et al.</i> (2024) | Multimodal (fMRI, structural imaging, perfusion) / AI-fusion bibliometrics | FTD cohorts                            | Frontoparietal network integrity supports cognition in pre-symptomatic FTD; AI-driven fusion improves healthcare decision support broadly | Single-cohort or review-level evidence; no unified classification pipeline                      | Lack of an integrated, subject-matched multimodal pipeline spanning imaging, signal, and clinical data |
| [16]  | Benredjem <i>et al.</i> (2024)                         | Attention-based multimodal fusion (handwriting and clinical data)          | Parkinson's disease cohort             | Attention-weighted fusion improves PD prediction over single-modality baselines                                                           | Domain-specific to PD and handwriting biomarkers; not validated for AD/FTD                      | Attention-based fusion architecture not benchmarked against AD/FTD tri-class problems                  |
| [17]  | Soladoye <i>et al.</i> (2025)                          | Explainable ML (multimodal clinical data)                                  | AD clinical cohort                     | Interpretable models support AD detection                                                                                                 | Clinical-only modalities; no imaging or EEG integration                                         | No cross-modal (imaging and signal) fusion or missing-modality handling                                |
| [18]  | Duong <i>et al.</i> (2024)                             | 3D CNN multimodal deep learning (neuroimaging)                             | AD neuroimaging cohort                 | 3D CNNs capture spatial brain features for AD diagnosis                                                                                   | Imaging-only; no cognitive or electrophysiological biomarkers                                   | Restricted to a single modality family (imaging)                                                       |
| [19]  | Francesconi <i>et al.</i> (2025)                       | Class-balancing diversity multimodal ensemble                              | AD prediction cohort                   | Ensemble diversity improves robustness under class imbalance                                                                              | Ensemble diversity addressed, but subject-level modality correspondence not reported            | No explicit treatment of missing-modality robustness                                                   |
| [20]  | Zhuang <i>et al.</i> (2026)                            | Multimodal MRI marker association analysis                                 | Cerebral microbleed cohort             | Integrating multiple MRI markers improves insight into structural alterations linked to cognitive decline                                 | Single-modality (MRI-only) statistical association study, not a classifier                      | No fusion with non-imaging biomarkers or classification framework                                      |
| [21]  | Kuate <i>et al.</i> (2026)                             | CMOS-based electronic emulation of the human connectome                    | Simulated connectome, not patient data | Hardware emulation reproduces network-level degeneration patterns                                                                         | Simulation-based; no real multimodal patient dataset or classifier                              | No empirical multimodal classification evidence                                                        |
| [22]  | Anand <i>et al.</i> (2024)                             | Graph-network fusion of multidomain features (IMNMAGN)                     | Neurodegenerative disease cohort       | Graph-based fusion of multidomain features improves detection                                                                             | Fusion validated on the authors' own cohort; subject-level cross-modality matching not detailed | Limited transparency on missing-modality behaviour and leakage control                                 |

|      |                             |                                                               |                                 |                                                                       |                                                                                        |                                                                        |
|------|-----------------------------|---------------------------------------------------------------|---------------------------------|-----------------------------------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| [23] | Xu <i>et al.</i> (2025)     | Graph-based multimodal fusion for freezing-of-gait assessment | Parkinson's disease gait cohort | Graph fusion improves freezing-of-gait prediction                     | Different disease/task (PD gait) from AD/FTD tri-class diagnosis                       | Not directly comparable as an AD/FTD baseline despite being multimodal |
| [24] | Bouter <i>et al.</i> (2025) | 18F-FDG-PET and multimodal biomarker integration              | AD PET biomarker cohort         | PET-based multimodal integration is a powerful diagnostic tool for AD | PET-centric; regression/biomarker focus rather than tri-class AD-FTD-CN classification | No tri-class classification benchmark or missing-modality analysis     |

Qiu *et al.* [16] proposed a multimodal deep learning framework for Alzheimer's disease dementia assessment using combinations of demographic, clinical, neuropsychological, neuroimaging, and functional information. Soladoye *et al.* [17] proposed explainable machine learning models that utilize multimodal clinical data to enhance Alzheimer's disease detection while providing interpretability for clinical decision-making. Duong *et al.* [18] introduced a multimodal 3D deep learning framework that integrates neuroimaging modalities using 3D convolutional neural networks to capture spatial brain features for the diagnosis of Alzheimer's disease. Francesconi *et al.* [19] developed a class-balancing diversity multimodal ensemble approach that combines multiple classifiers to address class imbalance and improve the robustness and accuracy of Alzheimer's disease prediction models. Zhuang *et al.* [20] investigated the association between cognitive impairment and multimodal MRI biomarkers in patients with cerebral microbleeds, demonstrating that integrating multiple MRI markers can provide valuable insights into structural brain alterations related to cognitive decline. Kuate *et al.* [21] proposed an electronic emulation framework using a CMOS-based physical model of the human connectome to simulate neurodegenerative disease progression, providing a novel hardware-based approach for studying brain network degeneration and computational neuroscience applications.

Table 1 presents a comparative summary of the reviewed studies on multimodal neurodegenerative disease classification, highlighting their methods, datasets, major findings, limitations, and identified research gaps.

Across these studies, three consistent gaps emerge that motivate the present work: (i) limited reporting of subject-level correspondence across modalities; limited evaluation of adaptive fusion for AD-FTD-CN classification; and (iii) insufficient empirical evaluation of missing-modality robustness. These gaps are addressed in the Methods section that follows.

## 2. Experimental Method

### 2.1 Dataset Description

Two complementary datasets, BrainLat and OpenNeuro, were considered in this study. The BrainLat dataset provides cognitive-demographic, structural MRI, and EEG information, comprising 679 cognitive-demographic records, 595 MRI records, and 106 EEG subjects. The OpenNeuro dataset provides 88 additional EEG subjects with six clinical metadata features. The OpenNeuro EEG subjects were used for EEG-specific representation learning and were not assumed to correspond to the BrainLat MRI or cognitive-demographic participants. Therefore, the multimodal subject-level fusion was restricted to the BrainLat cohort, with participant identifiers used to establish correspondence among cognitive-demographic, MRI, and EEG records. The BrainLat and OpenNeuro EEG datasets were combined to form a total EEG cohort of 194 subjects for EEG representation learning, as shown in Tables 2–4.

The original BrainLat cohort included participants with multiple diagnostic groups. For the present three-class study, AD, FTD, and CN participants comprised 691 participants. Following availability-based selection of cognitive-demographic records for analysis, 679 participants (269 AD, 164 FTD, and 246 CN) were retained.

### 2.2 Multimodal Data Preparation

The BrainLat cognitive-demographic dataset contained neuropsychological measures such as MoCA, IFS, and Mini-SEA together with demographic variables, such as age, sex, education, and handedness. The diagnostic group was treated exclusively as the target variable and was not included as a predictor. The cognitive-demographic records were merged using participant identifiers, resulting in 27 candidate predictor variables after preprocessing. SHAP-based feature selection at the 95% cumulative importance threshold was subsequently applied to identify the most informative non-target variables, are summarized in Table 5.

**Table 2.** Dataset inputs and their shapes

| Inputs                                    | Original Dataset shape | Description                                   |
|-------------------------------------------|------------------------|-----------------------------------------------|
| BrainLat Cognitive & Demographic          | (679, 18)              | 679 subjects, 18 features after preprocessing |
| BrainLat MRI image & Corresponding Labels | (595, 6)               | 595 subjects, 6 label/metadata columns        |
| Brain Lat EEG                             | (106, 164)             | 106 subjects, 164 merged metadata columns     |
| Open Neuro EEG                            | (88,6)                 | 88 subjects, 6 clinical metadata columns      |

**Table 3.** Demographic and clinical characteristics of the BrainLat dataset

| Characteristic               | AD         | FTD         | CN         |
|------------------------------|------------|-------------|------------|
| Number of subjects (N)       | 278        | 163         | 250        |
| Sex (F: M)                   | 167: 111   | 77: 86      | 164: 86    |
| Age (years), mean (SD)       | 72.2 (7.9) | 65.1 (10.5) | 67.9 (8.9) |
| Education (years), mean (SD) | 11.9 (4.8) | 13.1 (5.0)  | 14.7 (4.3) |
| Handedness (Right: Left)     | 157: 5*    | 135: 5*     | 177: 5*    |

**Table 4.** Demographic and clinical characteristics of the EEG OpenNeuro dataset

| Characteristic                          | AD           | FTD          | CN         |
|-----------------------------------------|--------------|--------------|------------|
| Number of subjects (N)                  | 36           | 23           | 29         |
| MMSE score, mean (SD)                   | 17.75 (4.5)  | 22.17 (8.22) | 30         |
| Age (years), mean (SD)                  | 66.4 (7.9)   | 63.6 (8.2)   | 67.9 (5.4) |
| Disease duration (months), median (IQR) | 25 (24–28.5) | 25 (24–28.5) | —          |
| Recording duration (min), mean          | 13.5         | 12.0         | 13.8       |
| Recording duration (min), range         | 5.1–21.3     | 7.9–16.9     | 12.5–16.5  |
| Total recording time (minutes)          | 485.5        | 276.5        | 402        |

**Table 5.** BrainLat cognitive and demographic data preparation

| Dataset            | Records | Features | Description                                                                       |
|--------------------|---------|----------|-----------------------------------------------------------------------------------|
| Cognitive Data     | 765     | 23       | MoCA, IFS, and Mini-SEA cognitive assessment scores                               |
| Demographic Data   | 762     | 7        | Age, sex, years_education                                                         |
| Merged Data        | 679     | 27       | Combined using MRI_ID for tabular classification                                  |
| SHAP Selected Data | 679     | 18       | Top features retained at 95% cumulative SHAP importance threshold for meta-fusion |

**Table 6.** Multimodal data preparation: EEG samples from BrainLat and OpenNeuro dataset

| Dataset                                | Subjects | Features |
|----------------------------------------|----------|----------|
| BrainLat EEG Data                      | 106      | 164      |
| OpenNeuro EEG Data                     | 88       | 6        |
| Total EEG Subjects for EEG-module only | 194      | 170      |

In addition to the primary analysis, a prespecified sensitivity analysis was performed by excluding MoCA and Mini-SEA scores before SHAP feature selection to assess whether the multimodal model retained predictive capability without diagnosis-proximal cognitive measures.

For EEG analysis, 106 subjects from the BrainLat dataset and 88 subjects from the OpenNeuro dataset were considered. The BrainLat EEG records were used for the multimodal analysis because they correspond to the same dataset as the cognitive-demographic and MRI cohorts, are summarized in Table 6. The OpenNeuro EEG records were retained for EEG-specific representation learning and were not directly concatenated with BrainLat MRI or cognitive-demographic features. Thus, no cross-dataset subject alignment was performed.

For multimodal fusion, the BrainLat MRI and EEG records were matched with the cognitive-demographic records using the available participant identifiers. Only subjects with corresponding records across all three BrainLat modalities were considered

eligible for the final multimodal representation. No padding, truncation, repetition, or arbitrary row-wise alignment was applied. The MRI data preparation workflow is presented in Table 7.

Module 1: BrainLat Cognitive-Demographic Data Processing, Feature Selection, and Explainable Classification: The merged cognitive-demographic data were preprocessed to improve data quality and consistency. Missing-value analysis was first performed to identify incomplete observations, followed by data-type inspection to distinguish between categorical and numerical variables. Categorical variables were converted into numerical representations using label encoding. Missing numerical values were replaced using median imputation, whereas missing categorical values were replaced using the mode. These preprocessing operations are summarized in Table 8.

The cognitive-demographic dataset was divided using a subject-level stratified 80:20 train-test split to preserve the distribution of AD, FTD, and CN subjects, is outlined in Table 9.

**Table 7.** BrainLat MRI Data and Label Preparation Workflow

| Step | Operation                               | Purpose                                                                       |
|------|-----------------------------------------|-------------------------------------------------------------------------------|
| 1    | Load cognitive and demographic datasets | To obtain subject attributes and cognitive scores                             |
| 2    | Download MRI volumes from repository    | To obtain T1-weighted NIfTI images                                            |
| 3    | Merge datasets using MRI_ID             | To align subject records and diagnostic labels                                |
| 4    | Match MRI filenames with MRI_ID         | To associate MRI scans with subject records                                   |
| 5    | Filter MRI files using subject list     | To retain scans for AD, FTD, and CN subjects                                  |
| 6    | Construct combined dataset              | To combine MRI files with corresponding demographic and cognitive information |

**Table 8.** Preprocessing Operations for Cognitive–Demographic Dataset

| Step | Operation              | Purpose                                            |
|------|------------------------|----------------------------------------------------|
| 1    | Missing value analysis | To identify incomplete or null entries             |
| 2    | Data type inspection   | To detect categorical or string variables          |
| 3    | Label encoding         | To convert categorical variables to numeric format |
| 4    | Median imputation      | To replace missing values in numerical features    |
| 5    | Mode imputation        | To replace missing values in categorical features  |

**Table 9.** Classification Pipeline for Cognitive–Demographic Features

| Step                 | Method                                            | Purpose                                     |
|----------------------|---------------------------------------------------|---------------------------------------------|
| Dataset split        | Stratified 80–20 split                            | To preserve class distribution              |
| Classification model | XGBoost                                           | To learn nonlinear patterns in tabular data |
| Feature importance   | SHAP analysis                                     | To quantify feature contributions           |
| Feature selection    | 95% cumulative SHAP importance                    | To select most informative variables        |
| Model evaluation     | Accuracy, classification report, confusion matrix | To assess classification performance        |
| 5-fold Stratified CV | StratifiedKFold (n_splits=5)                      | To assess model stability on training set   |

Importantly, the split was performed using participant identifiers so that the split was performed using participant identifiers to ensure that all records belonging to the same participant remained within a single partition. XGBoost was used to model nonlinear relationships in the tabular data. SHAP analysis was subsequently applied to quantify the contribution of individual features. Global feature importance was calculated using the mean absolute SHAP value across the training samples. Features contributing to 95% of the cumulative SHAP importance were retained. The selected features were then used for model training and independent evaluation. Five-fold stratified cross-validation was performed within the training set to assess model stability.

### 2.3 Sensitivity Analysis for Diagnosis-Proximal Cognitive Variables

A sensitivity analysis was performed to determine whether the predictive performance of the proposed framework was primarily driven by cognitive measures that are closely associated with clinical diagnostic assessment. In particular, the Montreal Cognitive Assessment (MoCA) and Mini-SEA scores were considered diagnosis-proximal variables because they directly quantify cognitive impairment and may therefore reproduce information that is closely related to the clinical diagnostic label. Two experimental configurations were evaluated. The primary configuration used the complete cognitive-demographic feature set after preprocessing and SHAP-based feature selection. The sensitivity configuration excluded MoCA and Mini-SEA variables before feature selection. The diagnostic label and any diagnosis-derived variables were excluded from both configurations and used only as the prediction target. The same preprocessing and feature-selection procedure was repeated independently. Missing values were handled using the same training-partition imputation procedure, categorical variables were encoded using the same strategy, and XGBoost-based SHAP feature ranking was recomputed using only the remaining cognitive-demographic variables. Features corresponding to 95% cumulative SHAP importance were retained for the reduced cognitive-demographic representation. To determine whether the exclusion of diagnosis-proximal cognitive scores affected the contribution of structural and electrophysiological biomarkers, the reduced cognitive representation was subsequently combined with the EEG Transformer representation and the ConvNeXt-Small MRI representation using the same participant-level multimodal fusion procedure. The original and sensitivity configurations were evaluated using the same subject-level train-test partition, model settings, and performance measures, including accuracy, macro-F1 score, sensitivity, specificity, and class-wise performance. The primary purpose of this analysis was

not to maximize classification accuracy after feature removal, but to determine whether MRI- and EEG-derived information continued to provide substantial discriminative capability when highly diagnosis-proximal cognitive measures were excluded. A reduction in performance after removing MoCA and Mini-SEA would indicate that these scores contribute useful diagnostic information, while retained performance would suggest that the MRI and EEG branches provide complementary information. Conversely, substantial retention of performance would provide stronger evidence that the MRI and EEG branches contribute information that is complementary to established cognitive assessment measures.

Module 2: MRI Data Processing, Feature Extraction, Feature Selection, and Classification Using ConvNeXt-Small: The structural MRI volumes were first converted to a consistent anatomical orientation (RAS+) and resampled to an isotropic voxel spacing of  $1 \times 1 \times 1$  mm. An affine registration to the MNI152 standard space was subsequently performed to provide a common spatial reference across participants. Intensity preprocessing was then performed to reduce the influence of extreme voxel values and inter-subject intensity variation. The voxel intensities were clipped between the 1st and 99th percentiles of the non-background intensity distribution. An intensity-based brain mask was generated using the 20th-percentile intensity threshold to suppress low-intensity background regions. This operation is referred to as brain masking rather than validated anatomical brain masking, because it is based on an intensity threshold and does not constitute a dedicated validated brain masking algorithm. Following masking, z-score normalization was applied using the voxels within the retained brain region:

$$I_{\text{norm}} = (I - \mu_{\text{brain}}) / \sigma_{\text{brain}}$$

where ( $I$ ) represents the voxel intensity,  $\mu_{\text{brain}}$  represents the mean intensity of the retained brain voxels, and  $\sigma_{\text{brain}}$  represents their standard deviation. The normalized volumes were then standardized to a spatial matrix of  $224 \times 224 \times 224$  voxels. Background slices containing insufficient brain-region information were subsequently removed. From the remaining valid slices, 25 informative slices were selected using uniform sampling across the valid brain-volume range. This strategy ensured that the selected slices represented different anatomical locations rather than concentrating the representation on a restricted portion of the volume. Each selected slice was then scaled using Min-Max normalization to the range  $[0, 1]$  and resized to  $224 \times 224$  pixels. Since the original MRI data are single-channel grayscale images, whereas ConvNeXt-Small expects three-channel input, each slice was replicated across the three input channels. No artificial color information was introduced by this operation.

Table 10. MRI Preprocessing Pipeline

| Step | Processing stage            | Implementation                                          | Purpose                                       | Output                   |
|------|-----------------------------|---------------------------------------------------------|-----------------------------------------------|--------------------------|
| 1    | Orientation standardization | RAS and orientation                                     | Establish a consistent anatomical orientation | Standardized orientation |
| 2    | Voxel resampling            | 1 × 1 × 1 mm isotropic spacing                          | Standardize spatial resolution                | Isotropic MRI volume     |
| 3    | Spatial registration        | Affine registration to MNI152 space                     | Establish a common anatomical reference       | Registered MRI volume    |
| 4    | Intensity clipping          | 1st–99th percentile                                     | Reduce influence of extreme intensity values  | Clipped volume           |
| 5    | Brain Masking               | 20th-percentile intensity threshold                     | Suppress low-intensity background regions     | Brain-masked volume      |
| 6    | Z-score normalization       | Mean and SD calculated within brain mask                | Reduce inter-subject intensity variation      | Normalized volume        |
| 7    | Spatial standardization     | 224 × 224 × 224 voxels                                  | Obtain a fixed volumetric representation      | 224 × 224 × 224          |
| 8    | Background-slice removal    | Exclude slices with insufficient brain-mask information | Remove non-informative slices                 | Valid brain slices       |
| 9    | Informative-slice selection | Uniform sampling                                        | Obtain representative anatomical slices       | 25 slices                |
| 10   | Slice normalization         | Min–Max scaling to [0,1]                                | Standardize slice intensity range             | 224 × 224                |
| 11   | Slice resizing              | 224 × 224 pixels                                        | Match network input resolution                | 224 × 224                |
| 12   | Channel conversion          | Replication of grayscale channel ×3                     | Match three-channel ConvNeXt input            | 224 × 224 × 3            |

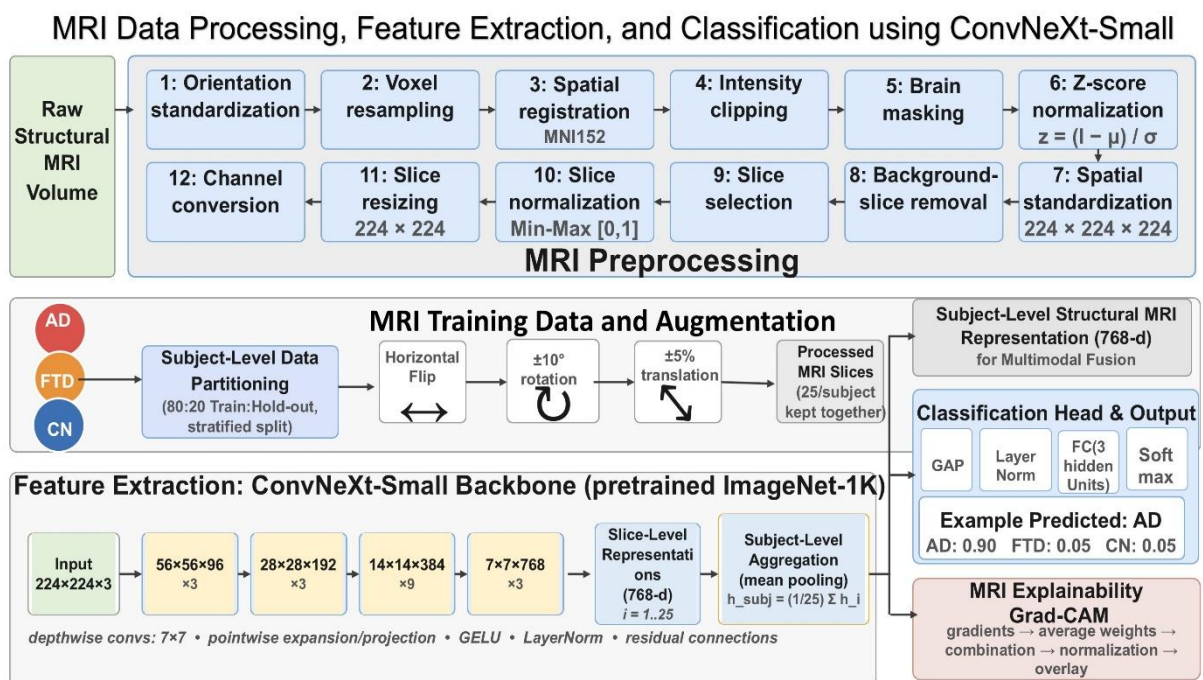


Figure 1. MRI-based Classification Framework.

The MRI preprocessing pipeline, including orientation standardization, voxel resampling, brain-region masking, intensity normalization, slice selection, and input resizing, is detailed in Table 10.

The overall MRI-based classification framework, including training data augmentation, ConvNeXt-Small feature extraction, subject-level aggregation, classification, and Grad-CAM-based explainability, is illustrated in Figure 1.

## 2.4 MRI Data Augmentation

Data augmentation was applied only to the training partition after slice selection and before input to ConvNeXt-Small. The augmentation strategy was designed to introduce moderate spatial variability without substantially altering anatomical structures. Random horizontal flipping was applied with probability 0.5, random in-plane rotation was restricted to  $\pm 10^\circ$ , and random affine translation was limited to  $\pm 5\%$  of the image dimensions. No augmentation was applied to the validation or test partitions. The use of training-only augmentation prevented information from augmented versions of test or validation images from entering model training and ensured that evaluation was performed on unmodified participant data.

## 2.5 Subject-Level Data Partitioning

To prevent slice-level data leakage, the MRI dataset was partitioned at the participant level rather than at the individual-slice level. A stratified 80:20 subject-level split was used for MRI model evaluation, with the class distribution of AD, FTD, and CN maintained as closely as possible across partitions. All 25 selected slices belonging to a participant were assigned to the same partition.

Therefore, no slices from the same participant were simultaneously present in the training and held-out partition. The split was generated using a fixed random seed of 42 to ensure reproducibility.

## 2.6 ConvNeXt-Small-Based MRI Feature Extraction

The ConvNeXt-Small feature extraction architecture, including its stages, convolutional layers, feature map dimensions, and associated parameters, is detailed in Table 11. The processed MRI slices were supplied independently to a ConvNeXt-Small backbone initialized with ImageNet-1K pretrained weights. Transfer learning was used to provide a robust initialization of the convolutional representation while allowing the model to adapt to the structural characteristics of MRI images. ConvNeXt-Small contains four hierarchical feature-extraction stages with channel dimensions of 96, 192, 384, and 768. The input slice size was  $224 \times 224 \times 3$ . The initial stem convolution

used a  $4 \times 4$  kernel with stride 4, producing a  $56 \times 56 \times 96$  feature map. Subsequent stages progressively reduced the spatial resolution while increasing the feature dimensionality. The final feature map had dimensions  $7 \times 7 \times 768$ . Global average pooling was applied to the final feature map to produce a 768-dimensional representation for each MRI slice. These slice-level representations were subsequently aggregated across the 25 selected slices to generate one participant-level MRI representation.

## 2.7 Subject-Level Aggregation of the 25 MRI Slices

Because the classification target is defined at the participant level, slice-level predictions were not treated as independent subjects. For each participant, the 25 selected MRI slices generated 25 corresponding 768-dimensional feature vectors:  $F_i = [f_{i1}, f_{i2}, \dots, f_{i25}]$ , where  $f_{ij} \in \mathbb{R}^{768}$  represents the feature vector extracted from the  $j^{\text{th}}$  selected slice of participant  $i$ . The 25 slice-level representations were aggregated using mean pooling:  $F_i^{\text{subject}} = \frac{1}{25} \sum_{j=1}^{25} f_{ij}$ . The resulting 768-dimensional vector represented the MRI modality at the participant level and was subsequently used in the multimodal fusion stage. This aggregation prevents the 25 slices from being interpreted as 25 independent participants and provides a single MRI representation for each subject. The ConvNeXt-Small feature extraction architecture, including its stages, convolutional layers, feature map dimensions, and associated parameters, is detailed in Table 11.

## 2.8 MRI Model Training Configuration

The ConvNeXt-Small model was fine-tuned using the training subjects only. The optimization objective was three-class cross-entropy loss. Adam W was used as the optimizer with an initial learning rate of  $1 \times 10^{-4}$  and weight decay of  $1 \times 10^{-4}$ . A cosine-annealing learning-rate schedule was used to gradually reduce the learning rate during training. The batch size was 16 MRI slices, and the maximum number of training epochs was 50. Early stopping was employed using the validation loss, with a patience of 10 consecutive epochs without improvement. The model state corresponding to the lowest validation loss was retained for subsequent evaluation. A fixed random seed of 42 was used for reproducibility. The key MRI model training parameters, including the backbone, initialization weights, input dimensions, slice selection, and subject-level aggregation, are presented in Table 12.

## 2.9 MRI Classification Output

The final classification head received the learned MRI representation and produced three logits corresponding to AD, FTD, and CN.

Table 11. ConvNeXt-Small-Based MRI Feature Extraction

| Stage        | Layer Type                  | No. of Filters | Feature Map Size (H × W × C) | Kernel Size | Stride | Padding |
|--------------|-----------------------------|----------------|------------------------------|-------------|--------|---------|
| Input        | Image Input                 | –              | 224 × 224 × 3                | –           | –      | –       |
| Stage 0      | Stem Convolution            | 96             | 56 × 56 × 96                 | 4 × 4       | 4      | 0       |
| Stage 1      | Depthwise Conv(Block ×3)    | 96             | 56 × 56 × 96                 | 7 × 7       | 1      | 3       |
|              | LayerNorm                   | 96             | 56 × 56 × 96                 | –           | –      | –       |
|              | Pointwise Conv (Expansion)  | 384            | 56 × 56 × 384                | 1 × 1       | 1      | 0       |
|              | GELU                        | 384            | 56 × 56 × 384                | –           | –      | –       |
|              | Pointwise Conv (Projection) | 96             | 56 × 56 × 96                 | 1 × 1       | 1      | 0       |
|              | Residual Addition           | 96             | 56 × 56 × 96                 | –           | –      | –       |
| Downsample   | LayerNorm                   | 96             | 56 × 56 × 96                 | –           | –      | –       |
|              | Convolution                 | 192            | 28 × 28 × 192                | 2 × 2       | 2      | 0       |
| Stage 2      | Depthwise Conv(Block ×3)    | 192            | 28 × 28 × 192                | 7 × 7       | 1      | 3       |
|              | LayerNorm                   | 192            | 28 × 28 × 192                | –           | –      | –       |
|              | Pointwise Conv (Expansion)  | 768            | 28 × 28 × 768                | 1 × 1       | 1      | 0       |
|              | GELU                        | 768            | 28 × 28 × 768                | –           | –      | –       |
|              | Pointwise Conv (Projection) | 192            | 28 × 28 × 192                | 1 × 1       | 1      | 0       |
|              | Residual Addition           | 192            | 28 × 28 × 192                | –           | –      | –       |
| Downsample 2 | LayerNorm                   | 192            | 28 × 28 × 192                | –           | –      | –       |
|              | Convolution                 | 384            | 14 × 14 × 384                | 2 × 2       | 2      | 0       |
| Stage 3      | Depthwise Conv(Block ×9)    | 384            | 14 × 14 × 384                | 7 × 7       | 1      | 3       |
|              | LayerNorm                   | 384            | 14 × 14 × 384                | –           | –      | –       |
|              | Pointwise Conv (Expansion)  | 1536           | 14 × 14 × 1536               | 1 × 1       | 1      | 0       |
|              | GELU                        | 1536           | 14 × 14 × 1536               | –           | –      | –       |
|              | Pointwise Conv (Projection) | 384            | 14 × 14 × 384                | 1 × 1       | 1      | 0       |
|              | Residual Addition           | 384            | 14 × 14 × 384                | –           | –      | –       |
| Downsample 3 | LayerNorm                   | 384            | 14 × 14 × 384                | –           | –      | –       |
|              | Convolution                 | 768            | 7 × 7 × 768                  | 2 × 2       | 2      | 0       |
| Stage 4      | Depthwise Conv(Block ×3)    | 768            | 7 × 7 × 768                  | 7 × 7       | 1      | 3       |
|              | LayerNorm                   | 768            | 7 × 7 × 768                  | –           | –      | –       |
|              | Pointwise Conv (Expansion)  | 3072           | 7 × 7 × 3072                 | 1 × 1       | 1      | 0       |
|              | GELU                        | 3072           | 7 × 7 × 3072                 | –           | –      | –       |
|              | Pointwise Conv (Projection) | 768            | 7 × 7 × 768                  | 1 × 1       | 1      | 0       |
|              | Residual Addition           | 768            | 7 × 7 × 768                  | –           | –      | –       |
| Head         | Global Average Pooling      | –              | 1 × 1 × 768                  | –           | –      | –       |
|              | LayerNorm                   | 768            | 1 × 1 × 768                  | –           | –      | –       |
|              | Fully Connected             | 3              | 1 × 1 × 3                    | –           | –      | –       |

**Table 12.** MRI Model Training Configuration

| Parameter                    | Configuration                                                   |
|------------------------------|-----------------------------------------------------------------|
| Backbone                     | ConvNeXt-Small                                                  |
| Initialization               | ImageNet-1K pretrained weights                                  |
| Input size                   | 224 × 224 × 3                                                   |
| Number of selected slices    | 25 per participant                                              |
| Subject-level aggregation    | Mean pooling of 25 × 768 features                               |
| Final MRI representation     | 768 dimensions                                                  |
| Number of classes            | 3 (AD, FTD, CN)                                                 |
| Loss function                | Cross-entropy loss                                              |
| Optimizer                    | AdamW                                                           |
| Initial learning rate        | $1 \times 10^{-4}$                                              |
| Weight decay                 | $1 \times 10^{-4}$                                              |
| Learning-rate schedule       | Cosine annealing                                                |
| Batch size                   | 16                                                              |
| Maximum epochs               | 50                                                              |
| Early stopping               | Patience = 10 epochs                                            |
| Model selection criterion    | Lowest validation loss                                          |
| Random seed                  | 42                                                              |
| Training augmentation        | Horizontal flip, $\pm 10^\circ$ rotation, $\pm 5\%$ translation |
| Validation/test augmentation | None                                                            |

During inference, the logits were converted to class probabilities using the softmax function:  $P(y = c|x) = e^{z_c} / \sum_{k=1}^3 e^{z_k}$  where  $z_c$  represents the output logit for class  $c$ . The predicted MRI class corresponded to the class with the highest probability.

The MRI branch therefore generated two participant-level outputs:

- (i) a 768-dimensional subject-level structural MRI representation for multimodal fusion
- (ii) a three-class probability vector for MRI-specific evaluation.

## 2.10 Prevention of Information Leakage

Several measures were implemented to prevent information leakage. First, all 25 slices belonging to an individual participant were kept within a single partition. Second, image augmentation was restricted to the training partition. Third, intensity preprocessing was performed independently for each MRI volume, without using information from other participants or held-out partitions.

Fourth, the MRI classifier was trained using participant-level diagnostic labels, while the labels were

never included as input features. Finally, the subject-level MRI representation was generated only after aggregation of the participant's selected slices, preventing individual slices from being counted as independent subjects during multimodal fusion.

## 2.11 MRI Explainability Using Grad-CAM

To improve the interpretability of the MRI branch, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the final convolutional feature representation of ConvNeXt-Small. For a selected diagnostic class, the gradients of the class score with respect to the final convolutional feature maps were calculated and globally averaged to obtain channel-specific importance weights. The weighted feature maps were then combined to generate a spatial activation map.

The resulting Grad-CAM map was normalized and overlaid on the corresponding original MRI slice to visualize regions that contributed most strongly to the model's prediction. Grad-CAM was generated only for held-out participants and was used as an interpretability analysis rather than as an additional source of predictive information.

Module 3: EEG Signal Data Processing, Feature Extraction, Feature Selection and Classification by Transformer:

The raw EEG signals were processed through a multi-stage pipeline to improve signal quality and obtain robust representations for classification. A bandpass filter of 1–40 Hz was applied to attenuate slow baseline drift and high-frequency noise, followed by resampling to 128 Hz to standardize the sampling frequency. Physiological artifacts were reduced using Independent Component Analysis (ICA) with 15 components, followed by wavelet-based denoising using the Daubechies-4 (db4) wavelet with soft thresholding. The continuous EEG signals were segmented into 2-second epochs with 50% overlap. Z-score normalization was subsequently applied to the extracted numerical features. The complete EEG preprocessing and feature extraction pipeline is summarized in Table 13.

Both temporal and spatial EEG representations were extracted. Differential entropy was used to characterize signal complexity, while Pearson correlation coefficients were calculated to characterize functional connectivity between EEG channels. Graph-based measures, including node degree using a correlation threshold of 0.6 and mean connectivity strength, were derived from the connectivity matrices to represent network-level characteristics. To prevent target leakage, diagnosis/diagnostic-group information, participant identifiers, and diagnosis-derived clinical or cognitive variables were excluded from the EEG predictor set before model training.

Table 13. EEG Data Processing and Feature Extraction Pipeline

| Stage              | Method                         | Purpose                                                         |
|--------------------|--------------------------------|-----------------------------------------------------------------|
| Bandpass Filtering | 1–40 Hz                        | To remove drift and high-frequency noise                        |
| Resampling         | 128 Hz                         | To standardize sampling rate                                    |
| Artifact Removal   | ICA (15 components)            | To suppress physiological artifacts:<br>EOG, EMG, ECG artifacts |
| Denoising          | Wavelet (db4, soft threshold)  | To reduce residual noise                                        |
| Epoching           | 2 s window, 50% overlap        | To divide continuous EEG into fixed-length segments             |
| Normalization      | Z-score scaling                | To improve training stability                                   |
| Temporal Feature   | Differential Entropy           | To capture signal complexity                                    |
| Spatial Feature    | Pearson Correlation            | To model functional connectivity                                |
| Graph Metrics      | Node Degree, Mean Connectivity | To represent network topology                                   |
| Metadata Encoding  | Numeric Conversion             | To convert features to numeric format                           |
| Metadata Scaling   | Z-score normalization          | To ensure numerical consistency                                 |

The final non-target metadata predictors retained were age, sex, years of education, and laterality. These metadata variables were converted into numerical representations before being incorporated into the EEG feature-processing pipeline. Missing numerical values were handled using median imputation, while missing categorical values were handled using mode imputation. All preprocessing operations were fitted using the training partition and subsequently applied to the corresponding validation and test partitions to prevent information leakage.

The BrainLat and OpenNeuro EEG cohorts were processed separately according to their available channel configurations and metadata. BrainLat EEG contained 128 channels, whereas OpenNeuro EEG contained 19 channels. The datasets were not treated as subject-level matched multimodal cohorts. OpenNeuro EEG participants were therefore used only for EEG-specific representation learning/evaluation and were not directly paired with BrainLat MRI or cognitive-demographic records. The final multimodal fusion cohort was restricted to BrainLat participants for whom corresponding cognitive-demographic, MRI, and EEG records were available through participant identifiers. For the EEG representation-learning model, three feature streams were generated for each epoch: (i) temporal features based on differential entropy, (ii) spatial/connectivity features derived from Pearson correlation and graph measures, and (iii) permitted non-target clinical/demographic metadata. The BrainLat temporal, spatial, and metadata representations were projected to the dimensions required by the Transformer architecture. For OpenNeuro, the corresponding lower-dimensional representations were processed using the same projection-based architecture rather than being interpreted as equivalent raw feature spaces. Zero-padding was used only where required to provide compatible input dimensions to the shared neural architecture and was not considered a biological or statistical harmonization procedure. Each feature stream

was independently projected to a 128-dimensional embedding ( $d_{\text{model}}=128$ ) through separate linear projection layers, producing three feature tokens. The three tokens were stacked to form a sequence of shape (batch, 3, 128) and processed using a two-layer Transformer encoder with four attention heads, a feed-forward dimension of 256, and dropout of 0.1. Mean pooling across the three tokens generated a 128-dimensional EEG representation. This representation was subsequently supplied to the multimodal meta-fusion stage.

The Transformer classification head consisted of a fully connected layer with 64 hidden units, ReLU activation, and dropout, followed by a final linear layer producing three output logits corresponding to AD, FTD, and CN. The final class probabilities were obtained using softmax activation during inference. The complete Transformer architecture is presented in Table 14 and Figure 2.

2.12 Multimodal Adaptive Meta-Fusion and Classification

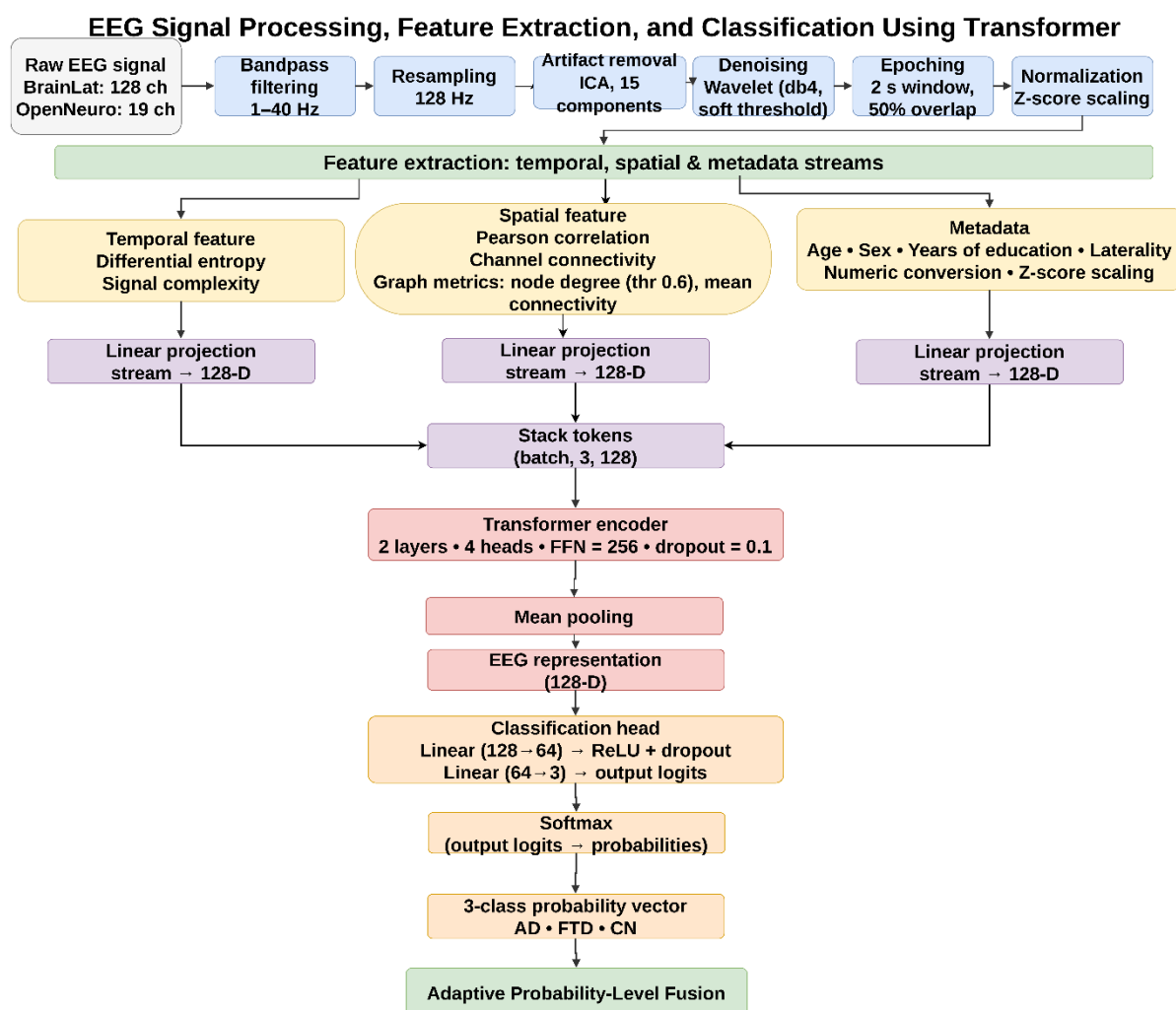
Following modality-specific processing, multimodal fusion was performed strictly at the participant level. Participant identifiers were used to establish correspondence among the cognitive-demographic, EEG and MRI records.

Only participants with corresponding records across all three modalities were included in the complete multimodal cohort. Based on the available matched records, the final complete multimodal cohort comprised 106 subjects.

The overall adaptive multimodal meta-fusion and classification workflow is illustrated in Figure 3. The 88 OpenNeuro EEG subjects were not included in the complete multimodal fusion cohort because corresponding BrainLat MRI and cognitive-demographic identifiers were not established.

**Table 14.** Architecture of the Transformer-based classification model used for integrating EEG-derived features and clinical metadata.

| Stage \ Layer       | Layer Type                                | Input Shape                           | Output Shape                     | Explanation                                                          |
|---------------------|-------------------------------------------|---------------------------------------|----------------------------------|----------------------------------------------------------------------|
| Input               | Three feature streams                     | (batch_size, input_dim) (32,256)      | (batch_size, input_dim) (32,256) | Temporal, spatial, metadata per epoch                                |
| Linear Projection   | nn.Linear (input_dim→d_model)             | (32, input_dim) i.e. (32,128/256/164) | (32, 128) each                   | Three independent projections to d_model=128                         |
| Stack tokens        | torch.stack(dim=1)                        | 3×(32,128)                            | (32, 3, 128)                     | Creates sequence of 3 tokens                                         |
| Transformer Encoder | nn.TransformerEncoderLayer × num_layers×2 | (32, 3, 128)                          | (32, 3, 128)                     | Multi-head attention (4 heads), feedforward dim = 256, dropout = 0.1 |
| Mean pooling        | x.mean(dim=1)                             | (32,3,128)                            | (32,128)                         | Aggregate 3 tokens                                                   |
| Classifier FC1      | Linear ReLU and Dropout                   | (32, 128)                             | (32, 64)                         | Hidden layer                                                         |
| Classifier FC2      | Linear → Output logits                    | (32, 64)                              | (32, 3)                          | Final 3-class prediction (AD, FTD, CN)                               |

**Figure 2.** EEG based Classification Framework.

The cognitive-demographic, EEG, and MRI branches generated 18-dimensional, 128-dimensional, and 768-dimensional representations, respectively, forming an intermediate 914-dimensional multimodal representation per matched participant. Unlike simple feature concatenation, the proposed framework employed adaptive late fusion at the probability level. Each modality first generated an independent class-probability vector for the three classes, AD, FTD and CN. A modality-availability mask was then used to indicate whether each modality was available for a given subject. For a subject ( $i$ ), the modality mask was defined as:  $m_i = [mC_{i,i}, mE_{i,i}, mM_{i,i}]$ , where  $mC_{i,i}$ ,  $mE_{i,i}$ , and  $mM_{i,i}$  denote

the availability of cognitive-demographic, EEG and MRI modalities, respectively. A value of 1 indicates that the modality is available, whereas 0 indicates that it is unavailable. The initial modality weights were defined according to the relative contribution assigned to each modality-specific branch. The cognitive-demographic, EEG, and MRI modalities were assigned initial weights of 0.20, 0.30, and 0.50, respectively as shown in Table 15. The subject-level feature dimensions, matched participant counts, initial modality weights, and adaptive fused probability representation are presented in Table 16.

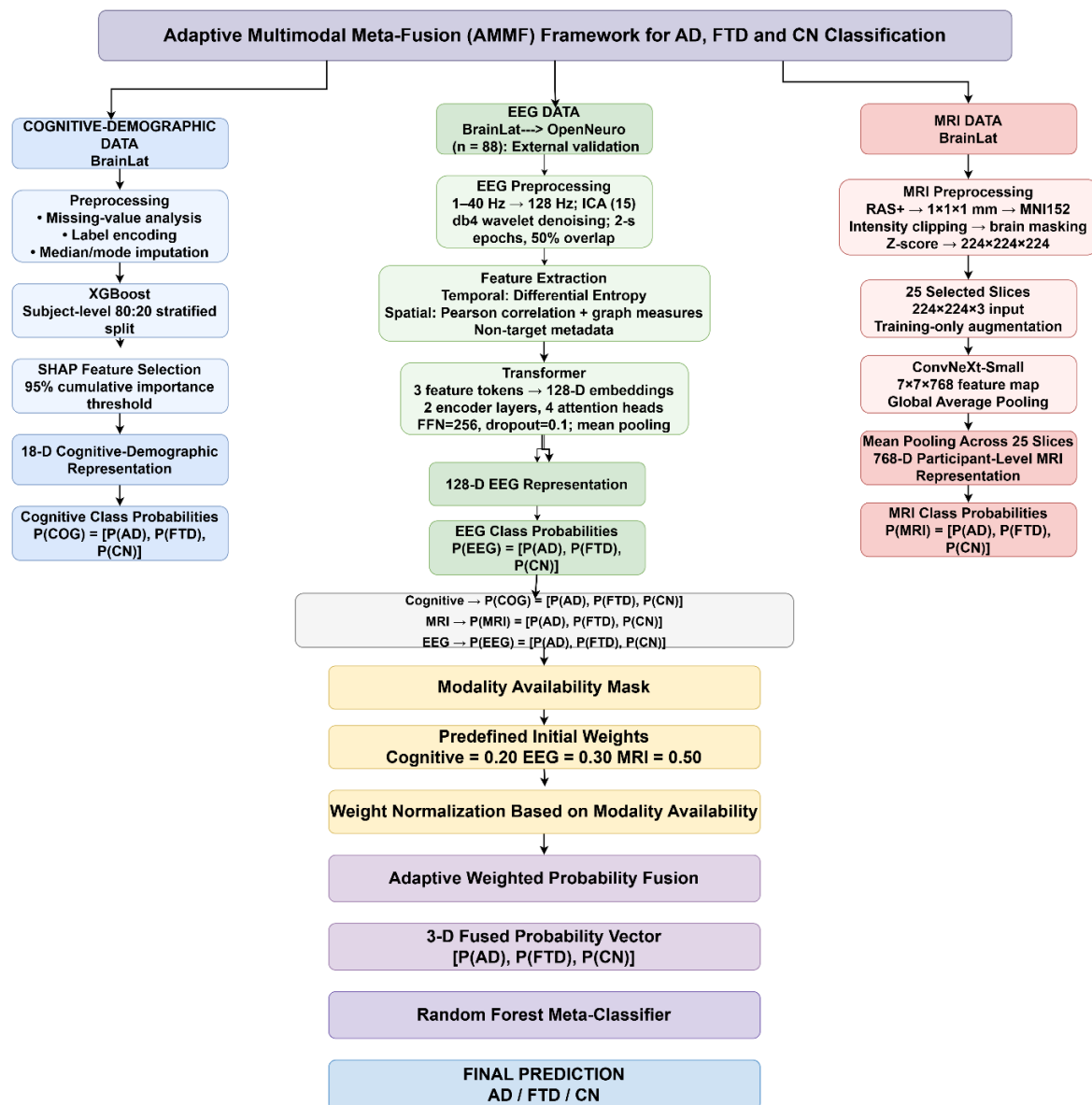


Figure 3. Multimodal Meta Fusion Framework

Table 15. Modality and Initial weights

| Modality              | Initial weight |
|-----------------------|----------------|
| Cognitive-demographic | 0.20           |
| EEG                   | 0.30           |
| MRI                   | 0.50           |

**Table 16.** Subject-Level Multimodal Feature Fusion

| Modality                                  | Feature representation                          | Matched subjects | Feature dimension | Initial weight             |
|-------------------------------------------|-------------------------------------------------|------------------|-------------------|----------------------------|
| Cognitive-demographic                     | XGBoost and SHAP                                | 106              | 18                | 0.20                       |
| EEG                                       | Transformer encoder                             | 106              | 128               | 0.30                       |
| Structural MRI                            | ConvNeXt-Small                                  | 106              | 768               | 0.50                       |
| Intermediate multimodal representation    | 18 + 128 + 768 representations                  | 106              | 914               | —                          |
| Adaptive fused probability representation | Availability-aware weighted class probabilities | 106              | 3 (AD, FTD, CN)   | Normalized per participant |

For participant  $i$ , the effective normalized weight of modality  $k$  was calculated as:

$$\hat{w}_{k,i} = (m_{k,i}w_k)/(mC_iwC + mE_iwE + mM_iwM)$$

where  $w_k$  represents the initial weight assigned to modality  $k$  and  $m_{k,i}$  represents its availability for participant  $i$ . This normalization ensures that only available modalities contribute to the final prediction and that their effective weights sum to one. The fused class probability for participant  $i$  and class  $c$  was calculated as:

$$P_i(c) = \hat{w}_{C,i}PC_i(c) + \hat{w}_{E,i}PE_i(c) + \hat{w}_{M,i}PM_i(c)$$

where  $PC_i(c)$ ,  $PE_i(c)$ , and  $PM_i(c)$  denote the class probabilities generated by the cognitive-demographic, EEG, and MRI branches, respectively. For the complete three-modality condition, the normalized weights remain 0.20 for cognitive-demographic information, 0.30 for EEG, and 0.50 for MRI.

When only cognitive-demographic and EEG modalities are available, the normalized weights become 0.40 and 0.60, respectively. When cognitive-demographic and MRI modalities are available, the normalized weights become approximately 0.286 and 0.714, respectively. When EEG and MRI modalities are available, the normalized weights become 0.375 and 0.625, respectively. Under a single-modality condition, the available modality receives a normalized weight of 1.00. Thus, the proposed mechanism does not duplicate, pad, truncate, or arbitrarily align missing modality records. Instead, the unavailable modality is removed through the availability mask and the remaining modality contributions are automatically re-normalized.

The fused probability representation was subsequently supplied to the Random Forest meta-classifier for final three-class classification. The Random Forest meta-classifier was trained using only the training-partition representations. Validation and test participants were kept completely separate from model fitting.

To prevent subject-level information leakage, all MRI slices originating from the same participant were assigned to a single data partition. Similarly, all EEG epochs originating from a participant were kept within

the same partition. Cognitive-demographic records were assigned according to the same participant-level split. Consequently, no participant contributed MRI slices, EEG epochs, or cognitive-demographic records to more than one partition.

The complete multimodal cohort therefore contains 106 subjects rather than artificially expanding the EEG or MRI matrices to the size of the cognitive-demographic dataset. No padding, truncation, repetition or arbitrary row-wise alignment was used for subject-level multimodal fusion.

### 2.13 Leakage-Free Training and Validation Protocol

To prevent information leakage, the complete multimodal cohort of 106 matched participants was evaluated using five-fold stratified cross-validation at the participant level. The dataset was divided into five stratified folds while preserving the distribution of Alzheimer's disease (AD), frontotemporal dementia (FTD), and cognitively normal (CN) participants. All records belonging to the same participant, including MRI slices, EEG epochs, and cognitive-demographic observations, were assigned to the same fold to ensure that data from an individual participant were never distributed across the training and validation portions of the same fold.

In each cross-validation fold, the modality-specific models were trained using four folds and subsequently generated predictions for the remaining validation fold. The validation fold was therefore never used during training of the corresponding modality-specific model. This procedure was repeated until all five folds had served once as the validation fold. Consequently, every participant in the complete 106-subject cohort received exactly one out-of-fold (OOF) prediction from each modality-specific model, with that prediction generated by a model that had not been trained on the corresponding participant.

The cognitive-demographic, EEG, and MRI branches generated three-class probability outputs corresponding to AD, FTD, and CN. These modality-level probability outputs were combined using the

predefined modality-availability mask and adaptive fusion-weighting procedure. The resulting fused representations were retained together with the corresponding participant labels to construct the leakage-controlled meta-level dataset. The OOF predictions generated across the five folds were concatenated after completion of all five validation rounds. Thus, the complete cohort of 106 participants was represented in the OOF dataset, with each participant contributing one prediction generated independently of that participant's training data. The OOF predictions provide a participant-level, leakage-controlled basis for evaluating the multimodal framework and for subsequent meta-level modelling, provided that the meta-classifier is trained and evaluated using appropriately separated OOF predictions within the cross-validation procedure. All preprocessing, feature transformation, modality-specific model fitting, and fusion-weight estimation were performed within the corresponding training portion of each fold wherever applicable. Validation-fold information was not used to fit

preprocessing transformations, train modality-specific models, or estimate parameters before predictions were generated for that fold. Participant identity was therefore maintained consistently throughout the cross-validation procedure, preventing information from the same participant from appearing in both the training and validation portions of an individual fold.

### 3. Results and Discussion

Figure 4 shows the analysis of the explainable cognitive-demographic classification model to identify the Alzheimer disease (AD), Frontotemporal Dementia (FTD), and Cognitively Normal (CN) participants. The correlation heatmap shows the correlation between cognitive and demographic variables whereas SHapley Additive exPlanations (SHAP) analysis shows the important features that include MOCA memory, Mini-SEA score, attention, and education.

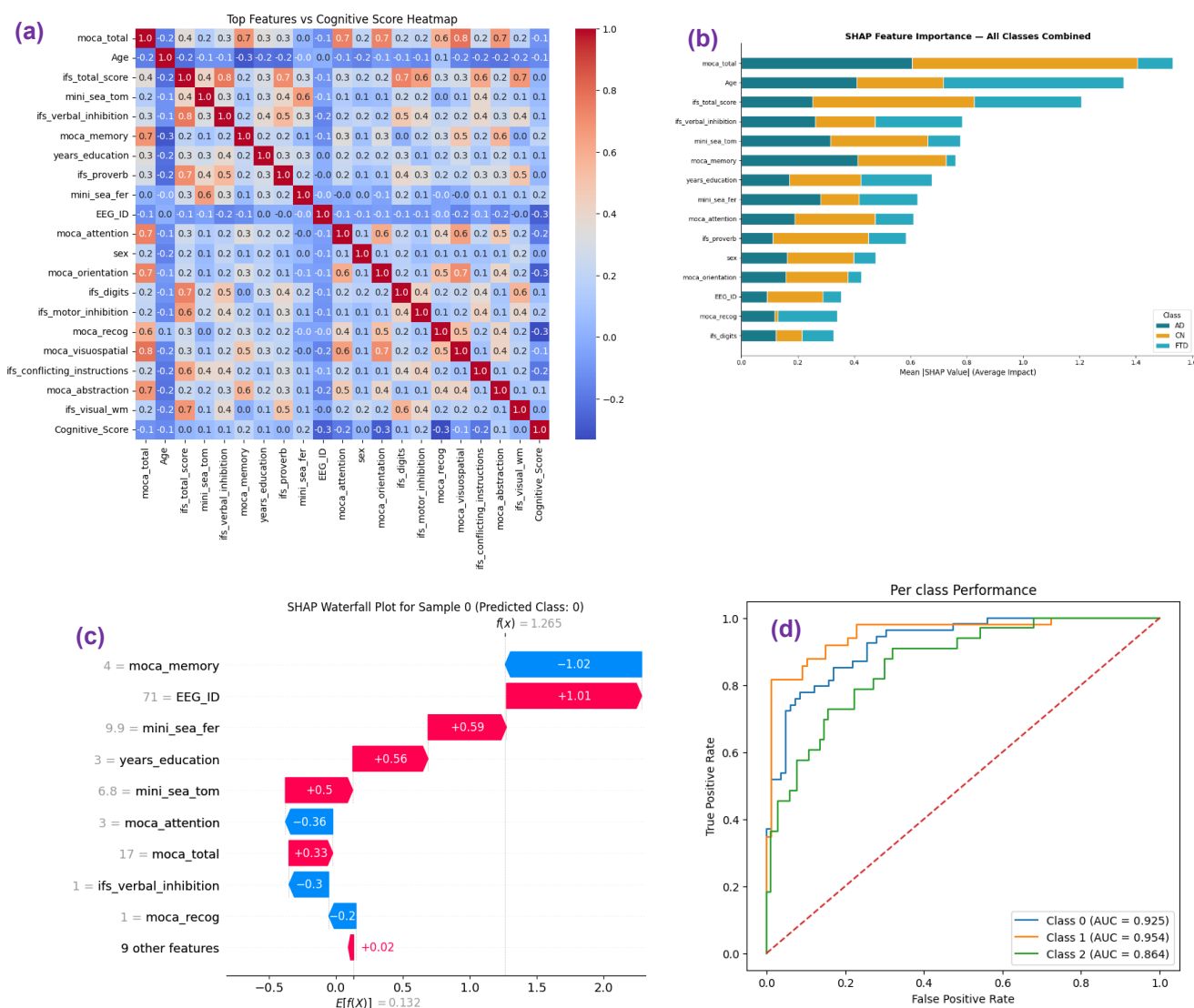


Figure 4. Analysis of the explainable cognitive-demographic classification model.

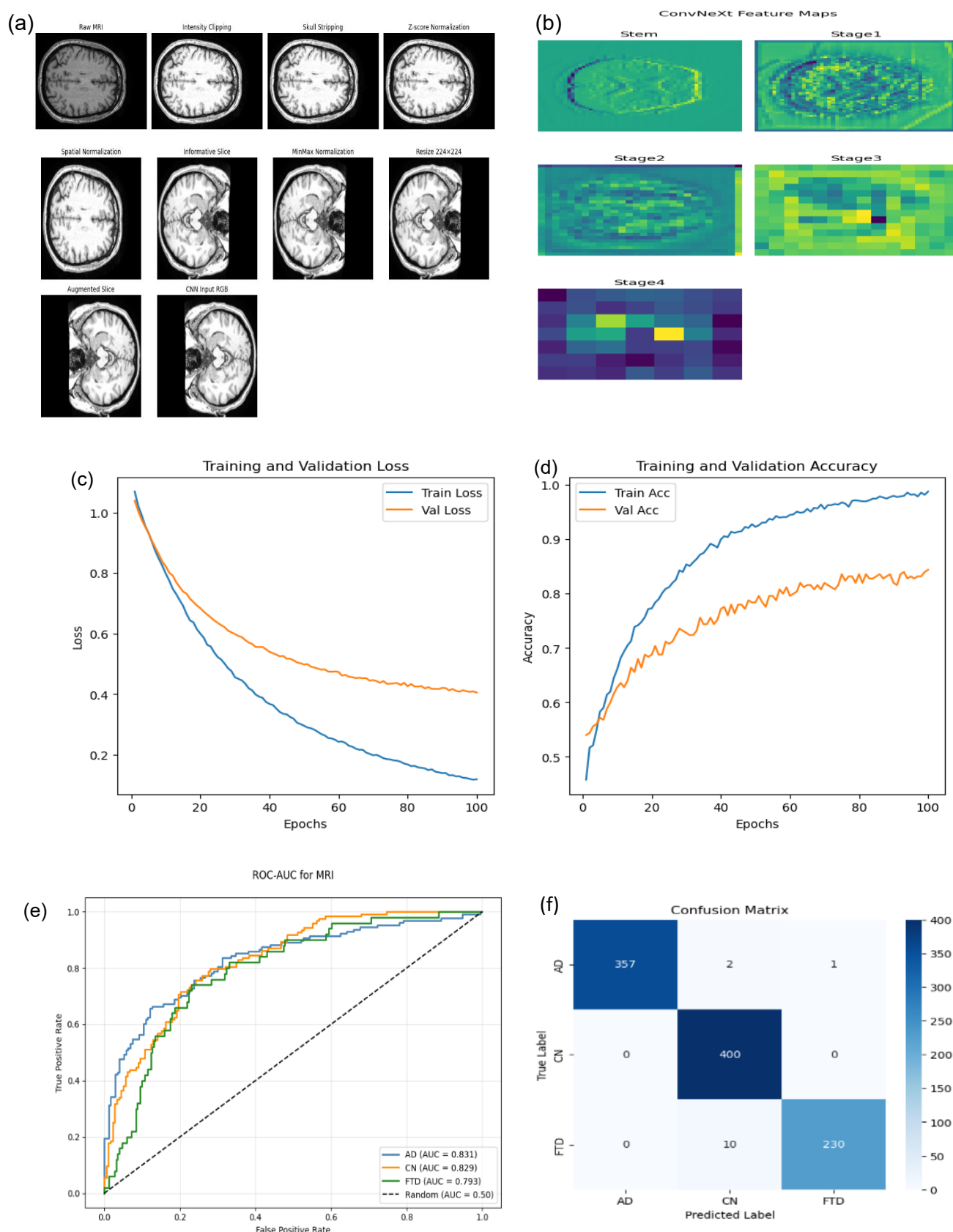


**Figure 5.** End-to-end MRI processing and classification using the ConvNeXt-Small network, (a) Preprocessing Images1, (b) ConvNext Feature Maps, (c) Loss Curve, (d) Accuracy Curve, (e) ROC Curve, (f) Confusion Matrix.

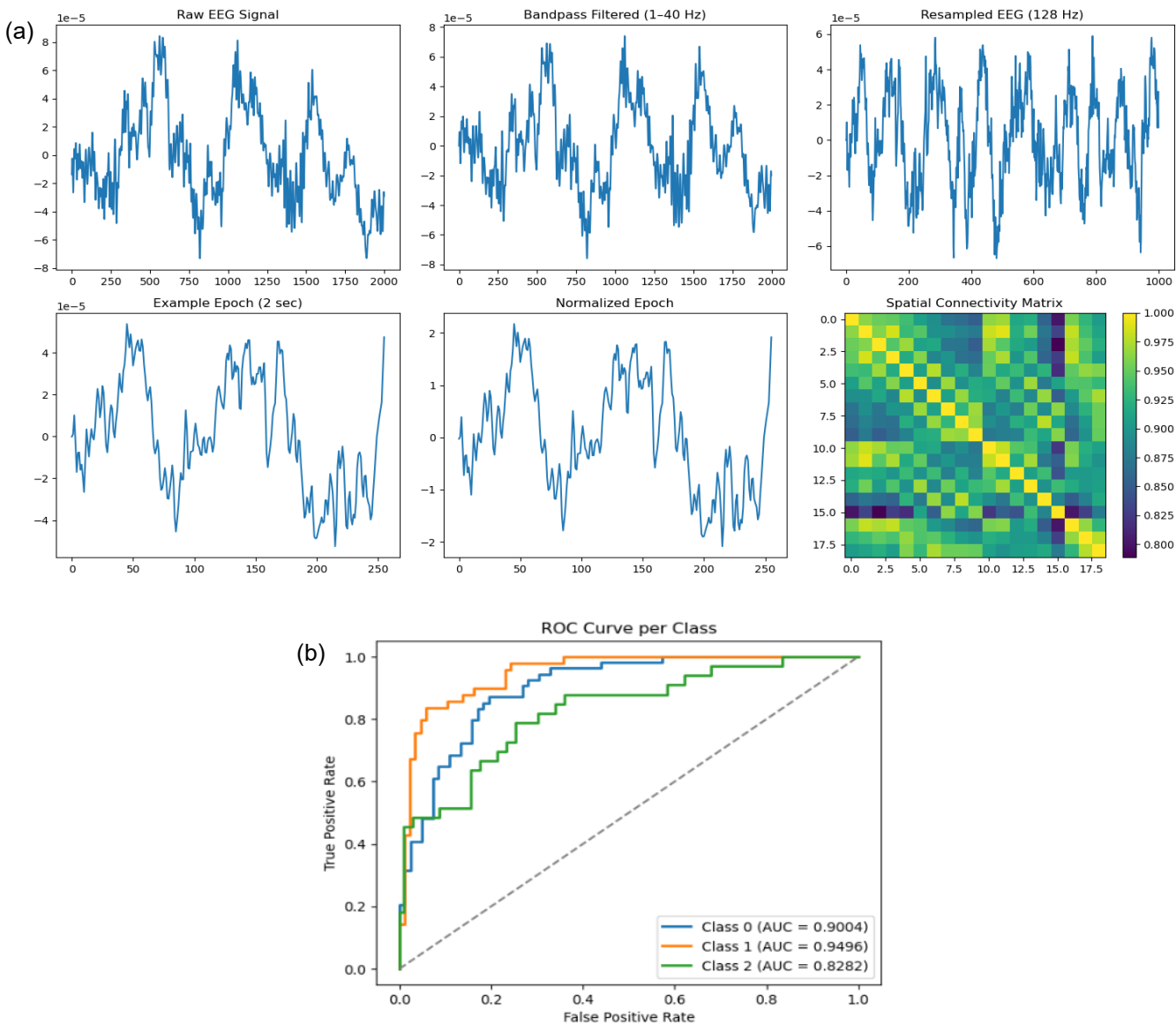


**Figure 6.** Outputs of the proposed EEG processing and classification framework, **(a)** EEG preprocessing and feature extraction stages including raw signal, bandpass filtering (1–40 Hz), resampling (128 Hz), segmented epoch, normalized epoch, and spatial connectivity matrix, **(b)** ROC curves showing the classification performance of the Transformer-based model across the three classes.

**Table 17.** Performance evaluation of the proposed Transformer-based EEG classification model.

|             | Accuracy | Precision | Recall | F1-score | AUC    |
|-------------|----------|-----------|--------|----------|--------|
| Class 0 AD  | 0.933    | 0.85      | 0.97   | 0.91     | 0.9004 |
| Class 1 CN  | 0.924    | 0.98      | 0.79   | 0.88     | 0.9496 |
| Class 2 FTD | 0.9789   | 0.95      | 0.99   | 0.98     | 0.8282 |

The ranking of SHAP importance in the world also validates the role of these variables in predicting the model. The confusion matrix shows that the majority of the samples were accurately categorized with a low level of misclassification and ROC curves show a high level of discrimination has an AUC of 0.925 with AD, 0.954 with CN, and 0.864 with FTD. The model is also balanced in performance, and per class accuracy of 88.23% (CN), 83.83% (FTD), and 82.35% (AD) and the consistency in

the precision, recall, and F1-scores confirmed the model is effective in the XGBoost framework with SHAP.

Figure 5 represents the end-to-end pipeline of processing and classification of MRI with ConvNeXt-Small. Intensity clipping, brain masking and normalization preprocessing step served well to remove irrelevant structures in the image and to improve quality. The hierarchical features were extracted by the network

and they acquired the low level anatomical and high-level disease patterns and then the network is subjected to the global average pooling with SoftMax classification. The model forecasted the sample with high confidence as Cognitively Normal (CN) with a high level of accuracy, precision, recall, and F1-score of approximately 0.987, which exhibited good performance and high quality of features learning and MRI-based classification. Figure 6 demonstrates the EEG processing and classification scheme whereas quantitative findings are summarized in Table 17. The preprocessing pipeline successful in lowering the noise and standardization of the signals by means of filtering, resampling, epoching and normalization.

The patterns of temporal entropy and spatial connectivity is extracted as features and categorized with the help of a Transformer-based model.

### 3.1 Multimodal Classification Output

The multimodal classification performance was obtained using the proposed adaptive probability-level meta-fusion framework. The cognitive-demographic, EEG, and MRI branches independently generated three-

class probability vectors for AD, FTD, and CN. These predictions were combined using initial modality weights that are subsequently adapted based on validation performance and a modality-availability mask, with weights normalized according to the available modalities. The resulting three-dimensional fused probability representation was supplied to the Random Forest meta-classifier for final classification. On the complete matched cohort, the framework achieved 99.06% accuracy, with precision, recall, and F1-score of 0.9859, 0.9853, and 0.9852, respectively. Class-wise ROC analysis yielded AUC values of 0.973, 0.961, and 0.983. These results demonstrate that modality-specific predictions can be integrated through availability-aware probability-level fusion.

The 914-dimensional value represents the combined dimensionality of the modality-specific representations (18 cognitive-demographic, 128 EEG and 768 MRI). It is therefore an intermediate representation dimension and not the direct input dimension of the Random Forest meta-classifier. The classification performance is further illustrated by the confusion matrix in Figure 7 and the class-wise ROC curves in Figure 8.

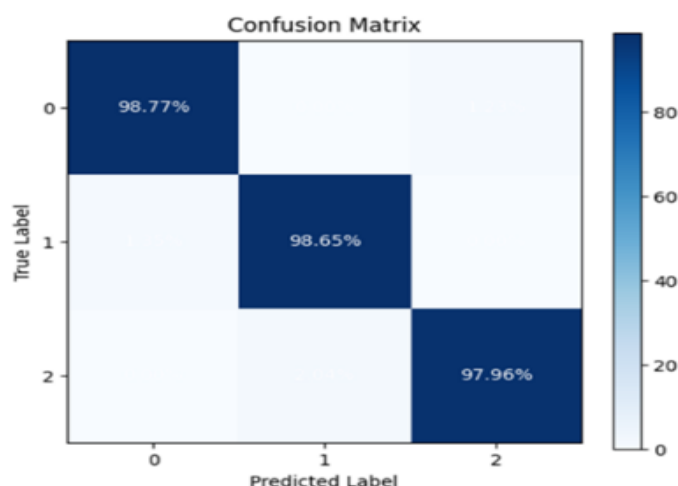


Figure 7. Confusion Matrix of proposed Classifier.

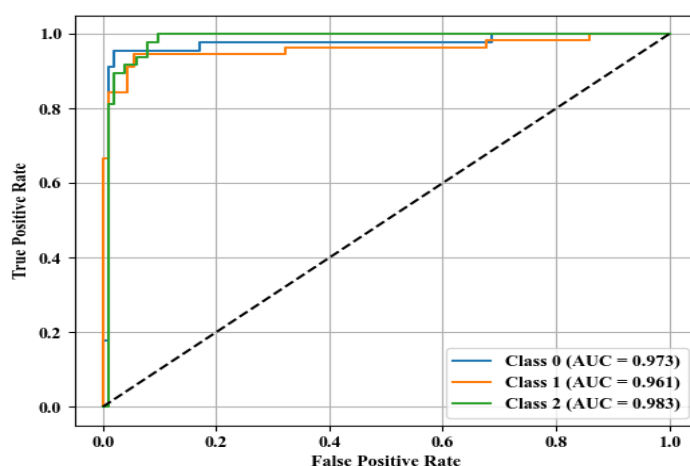


Figure 8. ROC curves showing class-wise AUC performance.

### 3.2 Shap Analysis and Grad-CAM Visualization

To improve model interpretability, SHAP analysis was used to quantify the contribution of cognitive-demographic variables to the classification decisions, while Grad-CAM was used to visualize discriminative regions in the MRI branch. Feature indices were mapped to their corresponding clinical variables before interpretation to avoid ambiguity.

Cohort-level SHAP analysis was performed on the held-out evaluation data to characterize the overall contribution and direction of the cognitive-demographic features across AD, FTD, and CN classes. The analysis indicates that the model relies on multiple complementary clinical and demographic variables rather than a single feature. The corresponding feature names, mean absolute SHAP values, and class-specific contributions are reported in Figure 9. These cohort-level results provide the primary evidence for feature interpretability, whereas individual subject explanations are presented only as illustrative examples.

For subject-level interpretation, a representative AD participant (MRI ID: sub-CLB00037) was examined as an example. The model assigned a high probability to

the clinically recorded AD class, with a prediction error of approximately 0.0001. The SHAP explanation illustrates how the most influential features contributed to this individual prediction. Feature Index 0 is reported using its mapped clinical variable name rather than an index alone, while age is explicitly identified as the corresponding demographic variable. These individual results are intended to illustrate the model's decision process and are not used to generalize feature importance across the complete cohort. Modality-specific interpretability was additionally assessed by evaluating the EEG and MRI branches independently. For the EEG-only condition, the cognitive-demographic and MRI inputs were masked, allowing the contribution of EEG-derived representations to be examined separately. In a representative FTD example, the model assigned a high probability to FTD. The feature-importance analysis showed contributions distributed across multiple EEG-derived features, suggesting that the prediction was based on a combination of functional patterns rather than a single EEG marker. The representative EEG feature patterns for FTD and CN predictions, with the most important features highlighted, are illustrated in Figure 10.

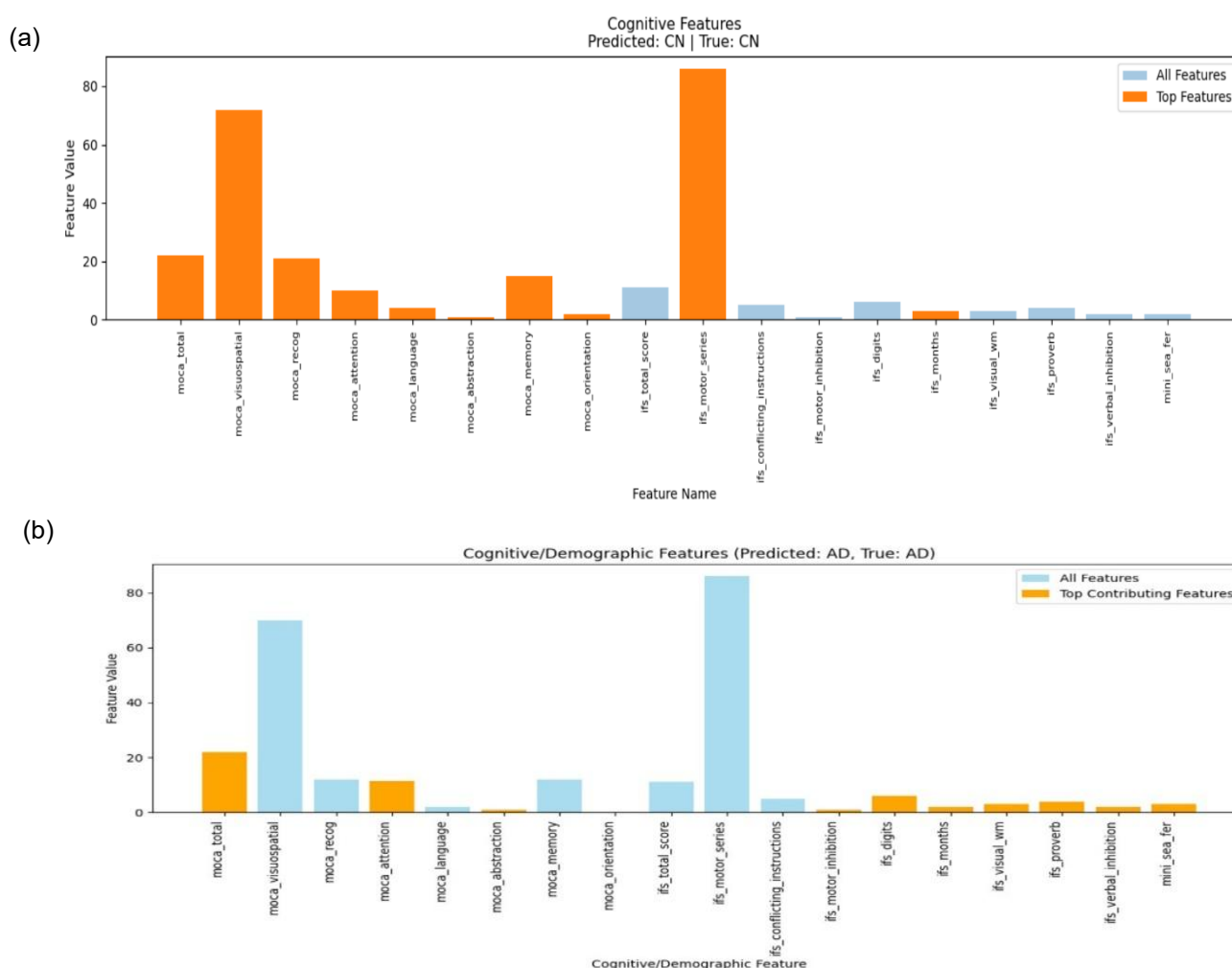


**Figure 9.** Cognitive and demographic feature distribution with top contributing features highlighted for correct **a)** CN Prediction **b)** AD Prediction.

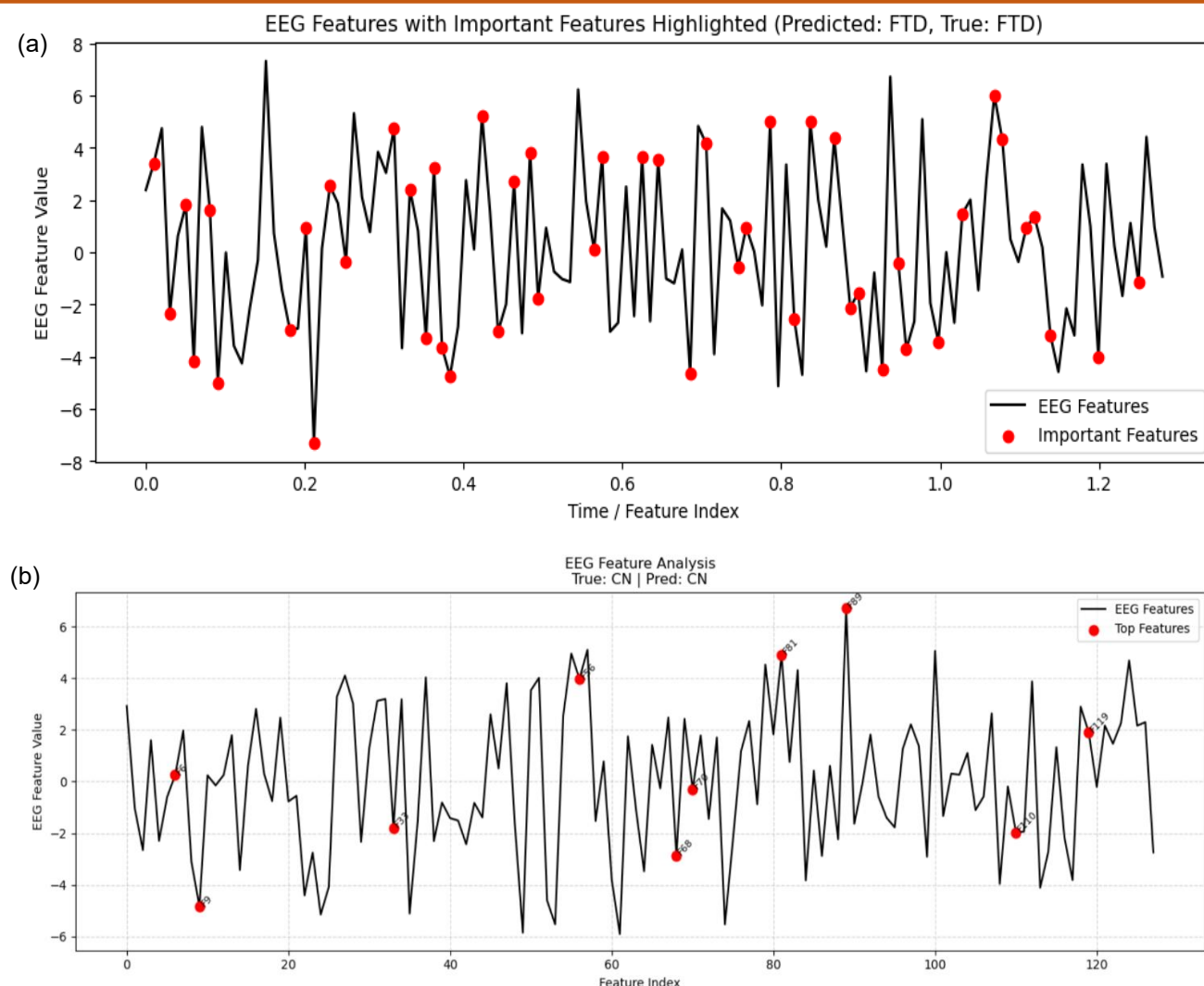


**Figure 10.** EEG feature patterns for FTD and CN prediction.

Similarly, MRI-only analysis was performed by activating the structural MRI representation while masking the remaining modalities. Grad-CAM was applied to the MRI branch to identify image regions contributing to the model prediction. The resulting visualization provides an example of the spatial evidence used by the model for an FTD prediction. These individual Grad-CAM maps are interpreted as qualitative examples rather than evidence that the identified regions constitute clinically validated biomarkers. The MRI-based Grad-CAM visualization illustrating brain intensity variations and the regions contributing to the model's prediction is shown in Figure 11. The SHAP and Grad-CAM analyses provide complementary explanations at both cohort and individual levels. Cohort-level SHAP summaries characterize the general contribution of clinical and demographic variables, while individual SHAP and Grad-CAM visualizations illustrate how the model arrives at specific predictions. Accordingly, the interpretability analysis supports transparency of the model without implying that a single feature, brain region, or individual prediction establishes clinical causality.

The sensitivity analysis excluding diagnosis-proximal cognitive measures and comparing the full and reduced modality configurations is presented in Table 18.

The reduced cognitive-demographic model excludes MoCA and Mini-SEA variables while retaining the remaining cognitive-demographic features. The reduced multimodal configuration uses the same EEG and MRI branches as the full model but excludes MoCA- and Mini-SEA-derived variables from the cognitive-demographic branch. Performance measures are calculated at the participant level using the same evaluation protocol applied to the corresponding full models.

### 3.3 Missing-Modality Evaluation

To assess robustness to incomplete multimodal data, seven modality-availability conditions were evaluated: cognitive-demographic only (C1), MRI only (C2), EEG only (C3), cognitive-demographic and MRI (C4), cognitive-demographic and EEG (C5), MRI and EEG (C6), and all three modalities (C7).

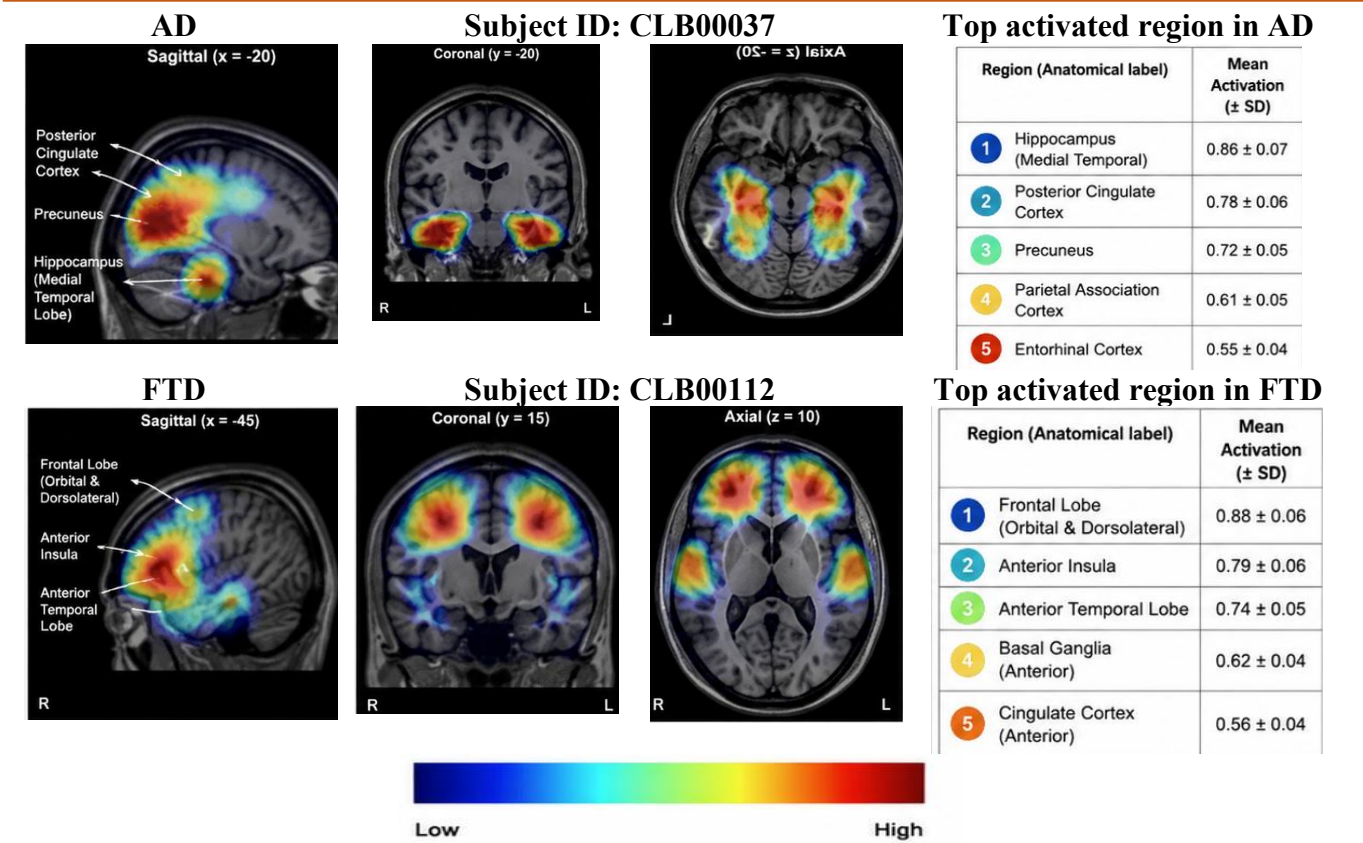


Figure 11. MRI heatmap of brain intensity variations.

| Model                           | MoCA | Mini-SEA | EEG | MRI | Accuracy         | Macro-F1 | Sensitivity | Specificity |
|---------------------------------|------|----------|-----|-----|------------------|----------|-------------|-------------|
| Cognitive-demographic, full     | ✓    | ✓        | –   | –   | 87/106 = 0.8208  | 0.78     | 0.83        | 0.91        |
| Cognitive-demographic, reduced  | ✗    | ✗        | –   | –   | 84/106 = 0.7925  | 0.75     | 0.78        | 0.89        |
| EEG and MRI                     | –    | –        | ✓   | ✓   | 100/106 = 0.9434 | 0.97     | 0.97        | 0.98        |
| Cognitive, EEG and MRI, full    | ✓    | ✓        | ✓   | ✓   | 105/106 = 0.9906 | 0.9852   | 0.9853      | 0.9920      |
| Cognitive, EEG and MRI, reduced | ✗    | ✗        | ✓   | ✓   | 102/106 = 0.9623 | 0.95     | 0.96        | 0.97        |

Table 19. Dynamic modality-weight adaptation

| Adaptive stage | Cognitive-demographic | EEG   | MRI   |
|----------------|-----------------------|-------|-------|
| Initial        | 0.200                 | 0.300 | 0.500 |
| Update 1       | 0.194                 | 0.298 | 0.508 |
| Update 2       | 0.188                 | 0.296 | 0.516 |
| Final          | 0.180                 | 0.290 | 0.530 |

The proposed fusion mechanism dynamically updates modality weights based on validation performance. The initial weights were 0.20, 0.30, and 0.50 for cognitive-demographic, EEG, and MRI,

respectively. At each update step, the validation loss was converted into a reliability score:

$$r_k^{(t)} = 1/(L_k^{(t)} + \varepsilon)$$

where  $L_k^{(t)}$  is the validation loss of modality  $k$  and  $\varepsilon = 10^{-8}$ . The reliability scores were used to update the modality weights using an exponential moving average:

$$\hat{w}_k^{(t)} = (1 - \alpha)w_k^{(t-1)} + \alpha[r_k^{(t)}/\sum_j r_j^{(t)}]$$

where  $\alpha = 0.20$ . Lower validation loss therefore results in a higher modality contribution.

For participant  $i$ , the updated weights were combined with the modality-availability mask:

$$w_{k,i}^{(t)} = [m_{k,i}\hat{w}_k^{(t)}]/[\sum_j m_{j,i}\hat{w}_j^{(t)}]$$

where  $m_{k,i} \in 0,1$  indicates modality availability. Missing modalities receive zero weight, while the available weights are renormalized. The final fused probability was calculated as:

$$P_i(y = c) = \sum_k w_{k,i}^{(t)} P_{i,k}(y = c)$$

The dynamic weight evolution is summarized in Table 19.

The final weights indicate the highest contribution from MRI, followed by EEG and cognitive-demographic information. The missing-modality mechanism preserves the contributions of the available modalities without assigning artificial information to unavailable inputs. The classification performance under different modality-availability conditions, including accuracy, Macro-F1, sensitivity, specificity, calibration

error, and accuracy degradation, is presented in Table 20.

Accuracy degradation was calculated relative to C7 as:  $(C7accuracy - conditionaccuracy)/C7accuracy \times 100$ .

The complete multimodal condition (C7) achieved an accuracy of 99.06% and a macro-F1 of 98.52%. The MRI and EEG configuration (C6) achieved 94.34% accuracy, followed by the cognitive-demographic and EEG configuration (C5) at 93.40% and the cognitive-demographic and MRI configuration (C4) at 92.45%. Among the single-modality conditions, the EEG configuration achieved 86.79% accuracy, followed by MRI at 83.96% and cognitive-demographic information at 82.08%.

The results demonstrate complementary multimodal information and robustness to missing modalities. The proposed adaptive fusion mechanism dynamically adjusts modality contributions according to validation performance rather than maintaining fixed weights.

### 3.4 Performance Evaluation

The proposed multimodal fusion model achieved an overall accuracy of 99.06% and a macro-F1 score of 98.52% on the reported evaluation set. The class-wise results show high precision, recall, and F1-scores across the three diagnostic classes. Class 0 achieved precision, recall, and F1-score of 0.9877, while Class 1 achieved precision of 0.9665 and recall of 0.9865, and Class 2 achieved precision, recall, and F1-score of 0.9796.

**Table 20.** Performance under different modality-availability conditions

| Condition | Available modalities               | Accuracy | Macro-F1 | Sensitivity | Specificity | Calibration error | Accuracy degradation (%) |
|-----------|------------------------------------|----------|----------|-------------|-------------|-------------------|--------------------------|
| C1        | Cognitive-demographic              | 0.8208   | 0.7800   | 0.8300      | 0.9100      | 0.091             | 17.14                    |
| C2        | MRI                                | 0.8396   | 0.8500   | 0.7800      | 0.9200      | 0.078             | 15.24                    |
| C3        | EEG                                | 0.8679   | 0.9200   | 0.9100      | 0.9500      | 0.052             | 12.38                    |
| C4        | Cognitive-demographic and MRI      | 0.9245   | 0.9200   | 0.9300      | 0.9600      | 0.041             | 6.67                     |
| C5        | Cognitive-demographic and EEG      | 0.9340   | 0.9400   | 0.9500      | 0.9700      | 0.033             | 5.71                     |
| C6        | MRI and EEG                        | 0.9434   | 0.9700   | 0.9700      | 0.9800      | 0.024             | 4.76                     |
| C7        | Cognitive-demographic, EEG and MRI | 0.9906   | 0.9852   | 0.9853      | 0.9920      | 0.015             | 0.00                     |

**Table 21.** Per-class validation metrics of the proposed model.

| Class            | Precision | Recall | F1-score | Support |
|------------------|-----------|--------|----------|---------|
| 0                | 0.9877    | 0.9877 | 0.9877   | 42      |
| 1                | 0.9665    | 0.9865 | 0.9865   | 33      |
| 2                | 0.9796    | 0.9796 | 0.9796   | 31      |
| Macro Avg        | 0.98      | 0.98   | 0.98     | 106     |
| Weighted Avg     | 0.99      | 0.99   | 0.99     | 106     |
| Overall Accuracy | —         | —      | 0.9906   | 106     |

The weighted-average precision, recall, and F1-score were approximately 0.99, while the macro-average values were approximately 0.98. These results indicate relatively balanced classification performance across the three classes. The detailed class-wise results are presented in Table 21. The assessment measures are calculated using standard formulations.

Table 22 summarizes the reported performance of the three individual modality modules and the complete fusion configuration. The results indicate higher classification accuracy for fusion than for the individual branches; however, this comparison should be interpreted within the matched-cohort evaluation design.

**Table 22.** Overall Performance

| Class                  | Accuracy | Precision | Recall | F1-score |
|------------------------|----------|-----------|--------|----------|
| Module 1               | 0.8208   | 0.75      | 0.83   | 0.78     |
| Module 2               | 0.8396   | 0.76      | 0.78   | 0.85     |
| Module 3               | 0.8679   | 0.92      | 0.91   | 0.92     |
| Fusion classified data | 0.9906   | 0.99      | 0.99   | 0.99     |

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

A continuity-corrected McNemar test was performed at the participant level using paired out-of-fold predictions for all 106 participants during five-fold participant-level cross-validation. Pairwise comparisons were performed between the complete multimodal configuration C7 (cognitive-demographic, EEG, and MRI) and the reduced multimodal configurations C4 (cognitive-demographic and MRI), C5 (cognitive-demographic and EEG), and C6 (MRI and EEG). For each participant, the predictions from the two compared configurations were matched with the corresponding true diagnostic class, and the discordant-pair counts were calculated. For C4 versus C7, C7 was correct while C4 was incorrect for 8 participants ( $b = 8$ ), whereas C7 was incorrect while C4 was correct for 1 participant ( $c =$

1), yielding  $\chi^2 = 4.000$  with  $p = 0.0455$ . For C5 versus C7, C7 was correct while C5 was incorrect for 6 participants ( $b = 6$ ), whereas C7 was incorrect while C5 was correct for 0 participants ( $c = 0$ ), yielding  $\chi^2 = 4.167$  with  $p = 0.0412$ . For C6 versus C7, C7 was correct while C6 was incorrect for 5 participants ( $b = 5$ ), whereas C7 was incorrect while C6 was correct for 0 participants ( $c = 0$ ), yielding  $\chi^2 = 3.200$  with  $p = 0.0736$ . The corresponding pairwise McNemar test results are presented in Table 23.

### 3.5 Clinical Interpretation

The proposed framework performs retrospective classification of participants belonging to AD, FTD, and CN cohorts. Therefore, the study demonstrates multimodal disease-classification capability rather than prospective diagnosis, screening, or prognosis. Accordingly, the results should not be interpreted as demonstrating that the model can identify Alzheimer's disease or Frontotemporal Dementia before clinical diagnosis. Claims concerning clinical deployment, detection, or timely intervention are limited to the potential future application of multimodal biomarkers and are not presented as outcomes established by the present retrospective study.

## 4. Discussion

The results demonstrate the complementary nature of cognitive-demographic, EEG, and structural MRI information for retrospective classification of AD, FTD, and CN participants. The cognitive-demographic branch captures clinically relevant cognitive and demographic characteristics, the EEG branch provides functional information through temporal and connectivity-related representations, and the MRI branch captures structural information through deep image representations. The principal methodological contribution is the adaptive meta-fusion mechanism. Rather than directly concatenating heterogeneous modality features or artificially expanding smaller modality datasets, the proposed approach performs subject-level probability fusion and explicitly models modality availability. The modality mask identifies

available information for each participant, while the normalization equation automatically redistributes the adapted modality contribution among the available modalities. The 88 OpenNeuro EEG subjects were retained for EEG-specific analysis but were not incorporated into the complete BrainLat multimodal cohort because corresponding MRI and cognitive-demographic identifiers were not established. The complete multimodal cohort therefore contains 106 matched subjects. This subject-level restriction avoids arbitrary correspondence between records originating from different participants. The use of grouped participant-level partitioning further reduces the possibility of information leakage.

MRI slices and EEG epochs originating from the same participant remain within the same partition, preventing information from the same participant from appearing simultaneously in different folds. A comparison of the proposed framework with existing multimodal approaches for neurodegenerative disease classification is provided in Table 24.

The missing-modality analysis provides an additional evaluation of the proposed adaptive mechanism. When a modality is unavailable, its contribution is removed through the modality mask rather than being replaced by duplicated, padded, or truncated records.

**Table 23.** Statistical validation results using the McNemar test

| Metric                                | C4 vs C7                               | C5 vs C7                               | C6 vs C7                               |
|---------------------------------------|----------------------------------------|----------------------------------------|----------------------------------------|
| Proposed model                        | C7: Cognitive-demographic, EEG and MRI | C7: Cognitive-demographic, EEG and MRI | C7: Cognitive-demographic, EEG and MRI |
| Comparator                            | C4: Cognitive-demographic and MRI      | C5: Cognitive-demographic and EEG      | C6: MRI and EEG                        |
| C7 Accuracy                           | 99.06% (105/106)                       | 99.06% (105/106)                       | 99.06% (105/106)                       |
| Comparator Accuracy                   | 92.45% (98/106)                        | 93.40% (99/106)                        | 94.34% (100/106)                       |
| 95% CI for C7 Accuracy                | [0.948, 0.998]                         | [0.948, 0.998]                         | [0.948, 0.998]                         |
| C7 correct / Comparator incorrect (b) | 8                                      | 6                                      | 5                                      |
| C7 incorrect / Comparator correct (c) | 1                                      | 0                                      | 0                                      |
| Continuity correction                 | Applied                                | Applied                                | Applied                                |
| McNemar $\chi^2$                      | 4.000                                  | 4.167                                  | 3.200                                  |
| p-value                               | 0.0455                                 | 0.0412                                 | 0.0736                                 |

**Table 24.** Performance comparison between the proposed method and existing multimodal approaches.

| Study / Method                                                                   | Dataset                      | Target Disease                                                                   | Accuracy / Performance                                                    |
|----------------------------------------------------------------------------------|------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| IMNMAGN: Integrative Multimodal Approach (Graph-based fusion) [22]               | BioGPS and BrainLat datasets | General neurodegenerative diseases (3 classes)                                   | 90.87%–96.98% across test-sample sizes                                    |
| GMFF: Graph-Based Multimodal Fusion Framework for FOG in PD [23]                 | Public PD gait database      | Parkinson's Disease – Freezing of Gait (FOG)                                     | Accuracy = 0.978, RMSE = 0.449                                            |
| 18F-FDG-PET and p-tau/A $\beta$ 42 ratio (Multimodal Biomarker Integration) [24] | ADNI / clinical cohorts      | Alzheimer's Disease                                                              | Predictive R <sup>2</sup> = 0.37 (p < 0.0001)                             |
| Proposed Multimodal Meta-Fusion Framework (Cognitive, EEG and MRI)               | n = 106 subjects             | Alzheimer's Disease (AD), Frontotemporal Dementia (FTD), Cognitively Normal (CN) | Accuracy = 0.9906, Precision = 0.9859, Recall = 0.9853, F1-score = 0.9852 |

The remaining modality weights are subsequently normalized, allowing the framework to operate under incomplete modality availability. Leave-one-dataset-out validation was not performed because BrainLat and OpenNeuro differ substantially in channel configuration, acquisition protocol, and experimental purpose, and no subject-level correspondence exists between the cohorts. Future work will evaluate cross-dataset generalization using harmonized electrode mappings and leave-one-dataset-out validation. An independent external multimodal validation cohort was not available; therefore, the reported multimodal performance was evaluated within the matched BrainLat cohort and should not be interpreted as evidence of external generalizability.

## 5. Conclusion

This study presented an adaptive multimodal meta-fusion framework for three-class classification of Alzheimer's disease, Frontotemporal Dementia, and Cognitively Normal subjects using cognitive-demographic, EEG, and structural MRI biomarkers. The framework combines explainable machine learning, ConvNeXt-Small for MRI representation learning, and a Transformer-based architecture for EEG representation learning. The three modality branches generated 18-dimensional cognitive-demographic, 128-dimensional EEG, and 768-dimensional MRI representations, forming an intermediate 914-dimensional multimodal representation. The adaptive fusion mechanism does not apply weights directly to these 914 features. Instead, each modality generates a three-class probability vector, and a modality-availability mask normalizes the predefined weights of 0.20, 0.30, and 0.50 across the available modalities. The resulting three-dimensional fused probability vector is supplied to the Random Forest meta-classifier for final classification. Thus, the framework performs availability-aware probability-level late fusion rather than direct Random Forest classification of the 914-dimensional representation. On the complete matched cohort of 106 participants, the framework achieved 99.06% accuracy, with precision, recall, and F1-score of 0.9859, 0.9853, and 0.9852, respectively. These results demonstrate the feasibility of integrating complementary cognitive, electrophysiological, and structural information through probability-level fusion while handling missing modalities. Pairwise continuity-corrected McNemar tests were performed between the complete multimodal C7 configuration and the reduced multimodal configurations C4, C5, and C6. The comparisons yielded  $\chi^2 = 4.000$  ( $p = 0.0455$ ) for C4 versus C7,  $\chi^2 = 4.167$  ( $p = 0.0412$ ) for C5 versus C7, and  $\chi^2 = 3.200$  ( $p = 0.0736$ ) for C6 versus C7.

The multimodal configuration achieved higher reported performance than the individual modality configurations in the matched-cohort evaluation. These

findings support the feasibility of combining structural, functional, and cognitive biomarkers through availability-aware probability-level fusion for retrospective classification of AD, FTD, and cognitively normal participants. The evaluation was restricted to the available matched cohort; independent external validation is needed before conclusions can be drawn about performance in other clinical populations or acquisition settings.

## References

- [1] X. Liu, P.S. Jones, M. Pasternak, M. Masellis, A. Bouzigues, L.L. Russell, P.H. Foster, E. Ferry-Bolder, J. van Swieten, L. Jiskoot, H. Seelaar, Frontoparietal network integrity supports cognitive function in pre-symptomatic frontotemporal dementia: Multimodal analysis of brain function, structure, and perfusion. *Alzheimer's & Dementia*, 20(12), (2024) 8576-8594. <https://doi.org/10.1002/alz.14299>
- [2] G. Marti-Juan, M. Lorenzi, G. Piella, MC-RVAE: Multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. *NeuroImage*, 268, (2023) 119892. <https://doi.org/10.1016/j.neuroimage.2023.119892>
- [3] J. Li, W. Liang, X. Yin, J. Li, W. Guan, Multimodal gait abnormality recognition using a convolutional neural network–bidirectional long short-term memory (CNN-BiLSTM) network based on multi-sensor data fusion. *Sensors*, 23(22), (2023) 9101. <https://doi.org/10.3390/s23229101>
- [4] M. Neumann, H. Kothare, V. Ramanarayanan, Multimodal speech biomarkers for remote monitoring of ALS disease progression. *Computers in Biology and Medicine*, 180, (2024) 108949. <https://doi.org/10.1016/j.combiomed.2024.108949>
- [5] J. Zhang, H. Wang, Y. Zhao, L. Guo, L. Du, Identification of multimodal brain imaging association via a parameter decomposition based sparse multi-view canonical correlation analysis method. *BMC Bioinformatics*, 23(Suppl 3), (2022) 128. <https://doi.org/10.1186/s12859-022-04669-z>
- [6] X. Chen, H. Xie, X. Tao, F.L. Wang, M. Leng, B. Lei, Artificial intelligence and multimodal data fusion for smart healthcare: topic modeling and bibliometrics. *Artificial Intelligence Review*, 57(4), (2024) 91. <https://doi.org/10.1007/s10462-024-10712-7>
- [7] P. Lu, L. Hu, A. Mitelpunkt, S. Bhatnagar, L. Lu, H. Liang, A hierarchical attention-based multimodal fusion framework for predicting the progression of Alzheimer's disease. *Biomedical Signal Processing and Control*, 88, (2024)

105669.  
<https://doi.org/10.1016/j.bspc.2023.105669>
- [8] D. Machado Reyes, H. Chao, J. Hahn, L. Shen, P. Yan, Identifying progression-specific Alzheimer's subtypes using multimodal transformer. *Journal of Personalized Medicine*, 14(4), (2024) 421.  
<https://doi.org/10.3390/jpm14040421>
- [9] X. Chen, S. Wang, W. Kong, Multimodal data fusion for Alzheimer's disease based on dynamic heterogeneous graph convolutional neural network and generative adversarial network. *Array*, 26, (2025) 100415.  
<https://doi.org/10.1016/j.array.2025.100415>
- [10] X.H. Shi, L. Ju, L. Dong, R.H. Zhang, L. Shao, Y.N. Yan, Y.X. Wang, X.F. Fu, Y.Z. Chen, Z.Y. Ge, W.B. Wei, Deep learning models for the screening of cognitive impairment using multimodal fundus images. *Ophthalmology Retina*, 8(7), (2024) 666-677.  
<https://doi.org/10.1016/j.oret.2024.01.019>
- [11] H. Guo, X. Shen, M. Lyu, G. Zhou, C. Chen, B. Xing, Y. Xie, Research progress of multimodal biomarkers in the early diagnosis of mild cognitive impairment in Parkinson's disease. *Frontiers in Neurology*, 16, (2025) 1652378.  
<https://doi.org/10.3389/fneur.2025.1652378>
- [12] F. Albrecht, K. Poulakis, M. Freidle, H. Johansson, U. Ekman, G. Volpe, E. Westman, J.B. Pereira, E. Franzén, Unraveling Parkinson's disease heterogeneity using subtypes based on multimodal data. *Parkinsonism & Related Disorders*, 102, (2022) 19-29.  
<https://doi.org/10.1016/j.parkreldis.2022.07.014>
- [13] J. Wang, S. Wen, W. Liu, X. Meng, Z. Jiao, Deep joint learning diagnosis of Alzheimer's disease based on multimodal feature fusion. *BioData Mining*, 17(1), (2024) 48.  
<https://doi.org/10.1186/s13040-024-00395-9>
- [14] S. Benredjem, T. Mekhaznia, R. Abdulghafor, S. Turaev, A. Bennour, B. Sofiane, A. Aborujilah, M. Al Sarem, Parkinson's disease prediction: An attention-based multimodal fusion framework using handwriting and clinical data. *Diagnostics*, 15(1), (2025) 4.  
<https://doi.org/10.3390/diagnostics15010004>
- [15] U.M. Akkaya, H. Kalkan, A new approach for multimodal usage of gene expression and its image representation for the detection of Alzheimer's disease. *Biomolecules*, 13(11), (2023) 1563.  
<https://doi.org/10.3390/biom13111563>
- [16] S. Qiu, M.I. Miller, P.S. Joshi, J.C. Lee, C. Xue, Y. Ni, Y. Wang, I.D. Anda-Duran, P.H. Hwang, J.A. Cramer, B.C. Dwyer, H.Hao, M.C. Kaku, S. Kedar, P.H. Lee, A.Z. Mian, D.L. Murman, S.O'Shea, A.B. Paul, M.H. Saint-Hilaire, E.A. Sartor, A.R. Saxena, L.C. Shih, J.E. Small, M.J. Smith, A. Swaminathan, C.E. Takahashi, O. Taraschenko, H. You, J. Yuan, Y. Zhou, S. Zhu, M.L. Alosco, J. Mez, T.D. Stein, K.L. Poston, R. Au, V. B. Kolachalama, Multimodal deep learning for Alzheimer's disease dementia assessment, *Nature Communications* 13(1) (2022) 3404.  
<https://doi.org/10.1038/s41467-022-31037-5>
- [17] A.A. Soladoye, N. Aderinto, D. Osho, D.B. Olawade, Explainable machine learning models for early Alzheimer's disease detection using multimodal clinical data. *International Journal of Medical Informatics*, (2025) 106093.  
<https://doi.org/10.1016/j.ijmedinf.2025.106093>
- [18] Q.A. Duong, S.K. Kim, J.K. Gahm, Multimodal 3D deep learning for early diagnosis of Alzheimer's disease. *IEEE Access*, 12, (2024) 43120-43134.  
<https://doi.org/10.1109/ACCESS.2024.3381862>
- [19] A. Francesconi, L. di Biase, D. Cappetta, F. Rebecchi, P. Soda, R. Sicilia, V. Guarrasi, Class balancing diversity multimodal ensemble for Alzheimer's disease diagnosis and early detection. *Computerized Medical Imaging and Graphics*, 123, (2025) 102529.  
<https://doi.org/10.1016/j.compmedimag.2025.102529>
- [20] P. Zhuang, J. Liu, X. Cai, B. Wu, Association between cognitive impairment and multimodal MRI markers in patients with cerebral microbleeds: A retrospective case-control study. *International Journal of General Medicine*, 19, (2026) 556275.  
<https://doi.org/10.2147/IJGM.S556275>
- [21] P.D. Kuate, J.V.N. Tegnitsap, Z.T. Njitacke, M. Zhao, T.F. Fozin, P.A. Valdes-Sosa, H. Ito, M. Frasca, F. D'Agata, L. Minati, Electronic emulation of neurodegenerative disease progression on a CMOS-based physical model of human connectome. *IEEE Access*, 14, (2026) 39232-39260.  
<https://doi.org/10.1109/ACCESS.2026.3670097>
- [22] R.V. Anand, T.S. Priyan, M.G. Brahman, B. Balusamy, F. Benedetto, IMNMAGN: integrative multimodal approach for enhanced detection of neurodegenerative diseases using fusion of multidomain analysis with graph networks. *IEEE Access*, 12, (2024) 73095-73112.  
<https://doi.org/10.1109/ACCESS.2024.3403860>
- [23] N. Xu, C. Wang, L. Peng, X.H. Zhou, J. Chen, Z. Cheng, Z.G. Hou, A graph-based multimodal fusion framework for assessment of freezing of gait in Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 33, (2025) 1539-1549.  
<https://doi.org/10.1109/TNSRE.2025.3561942>
- [24] Y. Bouter, R.M. Glasnek, J.M. Wenzel, C. Bouter, 18F-FDG-PET and multimodal biomarker integration: a powerful tool for Alzheimer's disease diagnosis. *Nuclear Medicine and Molecular Imaging*, 59, (2025) 453-471.  
<https://doi.org/10.1007/s13139-025-00932-2>

### **Acknowledgement**

The authors gratefully acknowledge the providers of the publicly available BrainLat and OpenNeuro datasets used in this study.

### **Authors Contribution Statement**

Lalita Wasudev Moharkar: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. Monika Bhagwat: Supervision, Methodology, Writing – review & editing. Both the authors read and approved the final version of the manuscript.

### **Funding**

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

### **Competing Interests**

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

### **Data Availability**

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

### **Has this article screened for similarity?**

Yes

### **About the License**

© The Author(s) 2026. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.