Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ADAPTIVE TEMPORAL-SPATIAL MULTIMODAL ATTENTION NETWORK FOR EARLY BRAIN DISEASE PREDICTION USING MRI, FMRI, AND EEG

JEROMY R[1], Dr. JEBAMALAR TAMILSELVI J[2]

[1]Research Scholar, Department of Computer Science, School of Applied Science,
Faculty of Liberal Arts and Business Studies, SRM IST, Ramapuram Campus, Chennai, Tamilnadu, India,
0009-0003-4855-3413

[2]Associate Professor, Department of Cyber Security, School of Applied Science,
Faculty of Liberal Arts and Business Studies, SRM IST, Ramapuram Campus, Chennai, Tamilnadu, India,
0009-0004-1292-1293

[1]jerojerry13@gmail.com, [2]jebamalj@srmist.edu.in

ABSTRACT

Early and accurate prediction of brain diseases is essential for timely diagnosis, intervention, and effective clinical management. Neurological disorders are associated with complex changes in brain structure, functional connectivity, and electrophysiological activity, which can be captured using Magnetic Resonance Imaging (MRI), functional Magnetic Resonance Imaging (fMRI), and Electroencephalography (EEG), respectively. However, existing brain disease prediction approaches mainly depend on individual modalities, thereby limiting their ability to capture complementary disease-related information. Conventional multimodal approaches also frequently employ static feature fusion, which may not adequately adapt the contribution of structural, functional, and temporal information according to disease characteristics and progression. To address these challenges, this research proposes an Adaptive Temporal-Spatial Multimodal Attention Network (ATSMAN) for early brain disease prediction through adaptive integration of MRI, fMRI, and EEG information. Initially, the multimodal data are preprocessed to reduce noise and ensure data consistency. Subsequently, modality-specific Deep Learning (DL) architectures are employed to extract discriminative representations. A Three-Dimensional Convolutional Neural Network (3D-CNN) is utilized to learn spatial and structural features from MRI data, while a Graph Attention Network (GAT) constructs and analyzes functional connectivity representations from fMRI data. In parallel, a Bidirectional Long Short-Term Memory-Transformer (BiLSTM-Transformer) architecture captures temporal dependencies and electrophysiological patterns from EEG signals. The extracted structural, functional, and temporal representations are then integrated through an adaptive temporal-spatial attention mechanism, which dynamically assigns modality-specific relevance weights and generates a unified multimodal representation. The fused representation is subsequently processed by the disease prediction module to produce disease probabilities, predicted labels, and confidence scores. Finally, an Explainable Artificial Intelligence (XAI) module identifies important anatomical regions, functional connections, and temporal EEG patterns associated with the prediction. Thus, ATSMAN provides comprehensive multimodal integration, adaptive biomarker selection, and interpretable prediction, offering a promising framework for early brain disease prediction and precision neurological assessment.

Keywords: *Early Brain Disease Prediction, Multimodal Learning, MRI, fMRI, EEG, Adaptive Attention Network, DL, GAT, Explainable Artificial Intelligence, Neurological Disorders.*

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurological condition with impairment in memory, cognition and function that impacts daily living. Early and accurate diagnosis is important for effective disease management. Recently, MRI, fMRI, and EEG combined with DL has been applied to detect abnormalities in the brain structure, function, and timing. However, a single modality is not always sufficient to give information of disease changes. In recent years, DL and graph-based approaches have been used for analyzing MRI. Certain researchers adopted a graph neural network for AD classification Zhou, Z. et al, [1] and certain researchers designed a spatio-temporal graph neural network for identification of mild cognitive impairment An, X et al, [2]. Saoud et Al [3] applied an explainable 3D Vision Transformer to detect early Alzheimer's disease. Li, Y. et al. [4] combined structural MRI and diffusion imaging information for AD diagnosis. These works have proven the effectiveness of DL for MRI analysis, but they are primarily aimed at acquiring structural information. The use of Graph-based Learning has been also extended to brain analysis. Alharbi et al. [5] proposed a spectral graph convolutional neural network for the diagnosis of AD based on changes in the brain's function. Bapat et al. [6] predicted the onset of AD using an explainable deep CNN using MRI and cognitive information. Liu et al. [7] introduced a Transformer-based approach for the classification of cognitive impairment. These approaches have been shown to enhance feature learning but do not synergistically combine structural MRI, functional fMRI and temporal EEG.

Recently, the multimodal learning approach has been explored to overcome the limitations of single-modality learning approaches. To diagnose Alzheimer's disease, Abdelaziz et al. [8] have offered a multi-scale multimodal DL framework. Dardouri et al. [9] proposed an early detection approach for AD using CNN and MRI images. Liu et al. [10] introduced multimodal diagnosis of AD using resting-state EEG and structural MRI. Alruily et al. [11] adopted an ensemble of DL models for the classification of AD using MRI. The studies shown show that integration of learned features or enhancement of learned features is beneficial, but full integration of structural, functional, and electrophysiological information is not. There have been other research studies on advanced feature and graph-based learning. Jeyapaul et al. [12] proposed an ensemble of DL models for automated diagnosis of AD from MRI. Arunnehru et al. [13] used DL and histogram features and applied them for the classification of MRI images for mild cognitive impairment and Alzheimer's disease. Peng et al. [14] combined features extracted from the time- and frequency domain of fMRI with graph neural networks for AD classification. Early diagnosis of AD using a spatiotemporal graph neural network has been proposed by Zhang, S et al. [15]. As evidenced from these studies, spatial, temporal, and functional aspects are significant in brain disease analysis, usually with no joint use of all three types of information in one framework.

EEG-based methods offer helpful information regarding temporal and electrophysiological factors. Abadal et al. [16] studied graph neural networks to analyze EEG signals for diagnosing AD and epilepsy. Khan, W et al. [17] designed an explainable DL system for the diagnosis of AD and frontotemporal dementia based on EEG data. These are examples of how useful EEG can be in the field of brain disease analysis. EEG, however, only offers a very small amount of anatomical information compared to MRI, and MRI cannot capture the higher temporal resolution electrophysiological information that is offered by EEG. Thus, multimodal fusion is an important research trend. Zhang, R et al. [18] surveyed the multimodal fusion-based DL methods applied to AD and emphasized the need of fusing heterogeneous brain information. But there are different types of information in each modality and it may not be simple feature concatenation or static fusion that decides which modality is most relevant. So, adaptive fusion mechanisms are needed to assign different modalities varying degrees of importance. Clinically, interpretability is also of importance. MRI assisted White, T et al, [19] in the construction of an explainable DL model for staging AD. Wu, F et al. [20]. Attention-based fusion can leverage the relevant information among different modalities, but the structural MRI and clinical data do not include functional connectivity and electrophysiological temporal features of time series.

The major contributions of this study are summarized as follows:

- To propose a unified ATSMAN system to fuse MRI, fMRI and EEG data for early brain disease prediction.

- The 3D-CNN is used to develop a modality-specific representation learning strategy for structural MRI characteristics, GAT for fMRI functional connectivity and BiLSTM-Transformer for temporal EEG characteristics.
- An adaptive temporal-spatial attention fusion mechanism is proposed to dynamically select the relative contribution of MRI, fMRI and EEG representations instead of fixed modality weights.
- An XAI mechanism is built in to highlight key anatomical regions, functional connectivity patterns, and temporal EEG components that are involved in the prediction.
- The proposed framework offers an interpretable multimodal learning approach that overcomes the limitations of single-modality and conventional fusion-based approaches for brain disease prediction and empowers early diagnosis of neurological diseases.

2. LITERATURE REVIEW

Neha Prerna Tigga et al. [21] suggested the graph convolution neural network (GCNN) for Brain-region specific autism prediction from electroencephalogram signals. This research used EEG data from eight autistic and eight normally developing children diagnosed with the Childhood Autism Rating Scale at the Central Institute of Psychiatry, Ranchi. The HydroCel GSN with 257 channels and 71 10-10 international equivalent channels recorded EEGs. Twelve brain areas had electrodes. For ASD prediction, autoregressive and spectral feature extraction preceded GCNN. ASD was most accurately predicted by the anterior-frontal brain area, which controls emotion, memory, and social interaction, at 87.07%. GCNN is suitable for EEG-based ASD identification.

Shakhnoza Muksimova et al. [22] proposed an attention-based CNN with an MRI Segmentation Block (MSB) to classify Alzheimer's disease. The MSB executes skull stripping and the attention mechanism is centered on important brain regions to enhance feature extraction and classification. The segmentation quality was tested by comparing against the radiologist annotation using Dice Coefficient and Jaccard Index. However, since the method relies on MRI, it is devoted to structural information, and takes no complementary functional information derived from fMRI or temporal electrophysiological information from EEG.

The attention-based multimodal learning approach by Shivahare et al. [23] combined MRI with EEG, and showed the benefits of combining complementary biomarkers, but also used SNP information instead of the functional connectivity measured by fMRI. This study combines MRI and EEG in a cross-modality attention study, but does not include fMRI-based functional connectivity.

Although Dardouri et al. [24] obtained high performance by utilizing a multi-branch convolutional neural network (CNN) for MRI-based Alzheimer's disease classification, this method is constrained to the structural MRI data, and thus it does not integrate additional functional information from fMRI or temporal electrophysiological information from EEG. Furthermore, the method does not utilize adaptive multimodal attention to adaptively decide the weights of different neuroimaging modalities. To balance the dataset, data augmentation was used. The suggested model accurately identified Non-Dementia, Very Mild Dementia, Mild Dementia, and Moderate Dementia with 99.68% accuracy. The model's excellent accuracy makes it a viable tool for healthcare providers to analyze and diagnose AD in real time.

A recent multimodal approach by Gautum, R et al, [25] combined resting-state fMRI and EEG with modality-specific encoders, learnable attention-based weightings, and included an emphasis on explainability; however, they did not jointly incorporate structural MRI. This study already shows the benefit of fusion of multimodal information and XAI, but not of combining structural MRI, functional fMRI and EEG signals together. Instead of explicitly modeling the brain-region functional connectivity with a graph attention mechanism, its fMRI representation is processed through a 3D-CNN.

Ibrar, W et al. [26] discussed the inverted residual bottleneck with self-attention (IRBwSA) for Alzheimer's disease prediction. First, the dataset was balanced by data augmentation. After that, the dataset is used to construct and modify the vision model. A novel inverted bottleneck self-attention model is created. Train models on the supplemented dataset and fuse derived features using a unique search-based technique. The designed models are interpreted using LIME, an explainable AI approach. Final classification uses a shallow broad neural network and additional classifiers for fused features.

Experimental results on an expanded MRI dataset showed 96.1% accuracy and 96.05% precision. Comparing the suggested framework to current methods indicates its superiority. The method is mostly MRI-based and feature fusion is done inside these chosen deep models, but not jointly together with functional fMRI connectivity and temporal EEG information.

Thalapathiraj Sambandam et al. [27] deliberated on Mathematical Modeling, Hilbert Transforms, and Transformer-Based DL for Scalable Parkinson's Disease Prediction. The Hilbert-based embedding theorem encodes biologically inspired spatial correlations, ensuring brain and motor-function images remain stable and structured. Then, novel transformer designs like Swin Transformer and Vision Transformer (ViT) manage these altered aspects to capture local and global dependencies. The recommended method outperforms CNNs and baseline Machine Learning (ML) models in accuracy, precision, recall, and F1-score on spiral drawings, wave patterns, and fMRI images. Hilbert preprocessing enhances feature consistency, whereas transformer models allow system growth and diagnostic precision. This physiologically interpretable and computationally efficient technology may detect Parkinson's disease early, reliably, and scalable, advancing AI in precision healthcare. The study does not consider an integration of MRI structural information and EEG electrophysiological signals in a multimodal architecture, although it incorporates both MRI and fMRI and transformer-based learning.

Mehta et al. [28] also showcased the potential of explainable multimodal learning in the context of interpretable predictions for Parkinson's disease, highlighting the importance of clinically transparent AI systems. The above-mentioned recent studies suggest that the integration of multimodal information, adaptive attention and explainability are growing critical factors, but a common AdaBoost framework that integrates structural MRI, functional fMRI, and temporal EEG representations, with adaptive temporal-spatial fusion and modality-specific XAI has yet to be investigated in depth. While the study puts a strong emphasis on multimodal clinical prediction and interpretability, it does not offer a single deep-learning architecture that learns spatial features of MRI, functional connectivity of fMRI, and long-range temporal dependencies of EEG separately.

Sankari et al. [29] proposed the ViT framework that incorporates an innovative Hierarchical Multi-Scale Attention (HMSA) for accurate detection and classification of brain tumors in MRI-based medical imaging. This method includes innovative features such as multi-resolution patch embedding for feature extraction across various spatial scales (8, 16, and 32 patches), computationally optimized transformer architecture for 35% shorter training duration, and probabilistic calibration for decision-making applications. A large MRI dataset of 7023 T1-weighted contrast-enhanced images from the Brain Tumor MRI Dataset was used for experimental validation. The method outperformed conventional ML methods (Random Forest: 91.2%, Support Vector Machine: 89.8%, XGBoost: 92.5%), state-of-the-art CNN architectures (EfficientNet-B0: 96.5%, ResNet-50: 95.8%), standard transformers (ViT: 96.8%, Swin Transformer: 97.2%), and hybrid CNN-Transformer approaches (TransBTS: 96.9%, Swin-UNet: 96.6%). The model performs well with 0.986 accuracy, 0.988 recall, 0.987 F1-score, and 0.023 expected calibration error. The framework is limited to T1-weighted MRI and applies to spatial multi-scale representation learning for brain-tumor classification.

Ashif Mahmud Joy et al. [30] examined the ViT for Alzheimer's disease multi-stage classification from brain MRI scans. The suggested model improves AD stage identification by fine-tuning hyperparameters and adding layers to Google's ViT architecture. The ADNI dataset includes 1,296 brain MRI scans from five phases of AD: Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), Mild Cognitive Impairment (MCI), and Alzheimer's Disease. The convert grayscale to RGB, trim images, then apply a Laplacian sharpening filter to improve clarity. Data augmentation included horizontal/vertical flips, zoom, and rotation to guarantee model robustness. Training took up 85% of the dataset and testing 15%. After 20 epochs of training at 0.0001, ViTAD attained 99.98% accuracy, 100% precision, and an F1-score of 1.00. The ViTAD aims to classify Alzheimer's disease in a multi-stage method, based exclusively on structural MRI. Therefore, functional connectivity abnormalities and temporal electrophysiological changes which may yield complementary information for early disease prediction are not simultaneously modeled.

Although DL-based brain illness detection has advanced, some study limitations

remain. Most traditional techniques use single-modality data like MRI, fMRI, or EEG, missing out on additional anatomical, functional, and electrophysiological information needed for early-stage illness prediction. Most multimodal frameworks use static feature fusion algorithms, which give all modalities similar weight and ignore the varied contributions of structural, functional, and temporal indicators throughout illness phases. Additionally, existing approaches generally fail to detect complex spatial dependencies in MRI data, disturbed functional connectivity patterns in fMRI signals, and long-range temporal dynamics in EEG recordings. Lack of adaptive attention processes restricts models' capacity to flexibly focus information modalities throughout illness development. The lack of interpretability in most research reduces doctors' trust in computerized predictions and hinders practical adoption. The lack of explainable artificial intelligence methods for disease-specific anatomical areas, functional networks, and electrophysiological patterns that determine categorization judgments is another key restriction. Thus, an intelligent multimodal framework that learns spatial, functional, and temporal representations via adaptive attention-based fusion and predicts early brain illness is needed. The ATSMAN integrates MRI, fMRI, and EEG modalities via dynamic attention-guided fusion and explainable prediction processes to provide robust and accurate diagnosis.

2.1 Research gap

The existing approaches are mainly based on single-modality analysis, limited multimodal combinations, only static fusion, or low interpretability, based on the above studies. Hence, this study suggests the ATSMAN. The framework adopts a combination of 3D-CNN, GAT, and BiLSTM-Transformer for the MRI structural features, fMRI functional connectivity, and EEG temporal features, respectively. An adaptive temporal-spatial attention mechanism is then employed to fuse the three modalities adaptively, and XAI gives interpretable information on the importance of MRI regions, fMRI connections, and EEG segments. Therefore, ATSMAN intends to offer a more holistic and flexible multimodal brain disease prediction.

2.2 Problem statement

Brain disease prediction methods developed so far typically rely on one type of neuroimaging technique, such as MRI or fMRI or EEG, and cannot take advantage of multiple complementary neurostructural, functional and electrophysiological features. While multimodal approaches have been developed, traditional fusion methods typically rely on fixed or static modality contributions and may fail to capture individual or disease stage differences in information. Moreover, current methods cannot combine 3D anatomical structure information from MRI, functional connectivity information from fMRI, and temporal dependencies from EEG. The lack of integration of explainability in different modalities further diminishes the clinical interpretability and transparency of automatic predictions. Hence, a multimodal framework is needed that is adaptive and learns structural, functional and temporal representations of the brain, adapts dynamically to the relevance of the modalities, and is interpretable and suitable for the early assessment of brain disease. This study presents the ATSMAN, a model that combines 3D-CNN based MRI feature extraction, GAT based fMRI functional connectivity analysis, BiLSTM-Transformer based EEG temporal feature learning, adaptive temporal-spatial attention fusion, and XAI.

3. METHODOLOGY

This paper proposed an ATSMAN that jointly utilizes MRI, fMRI, and EEG for early prediction of brain diseases. 2 Preprocessing of the three modalities is first performed to eliminate noise and to enhance the consistencies of the data. Next, 3D-CNN is used to learn structural features of MRI, GAT to learn functional connectivity features of fMRI, and BiLSTM-Transformer to learn temporal features of EEG. These attributes are then merged with an adaptive temporal-spatial attention mechanism to weight the contribution of each modality. The combined representation is then employed for disease prediction, while XAI is applied to detect significant features that contribute to such prediction. The model is optimized with weighted cross-entropy loss, regularization and the Adam optimizer.

3.1 Proposed ATSMAN Framework

The proposed ATSMAN is designed to facilitate the early prediction of brain diseases by integrating information obtained from three complementary neuroimaging modalities: MRI, fMRI, and EEG. The rationale behind combining these modalities lies in their ability to capture

distinct aspects of neurological abnormalities associated with disease progression. Structural MRI provides detailed anatomical information, fMRI reveals alterations in functional brain connectivity, and EEG captures electrophysiological changes occurring at a high temporal resolution. Individually, each modality provides valuable insights; however, their integration offers a comprehensive understanding of brain pathology, thereby enhancing early disease prediction.

The ATSMAN framework consists of six major components: MRI structural feature extraction, fMRI functional connectivity analysis, EEG temporal feature learning, adaptive multimodal fusion, disease prediction, and explainable artificial intelligence. Initially, each modality undergoes preprocessing to eliminate noise and standardize the data. Subsequently, modality-specific DL architectures are employed to extract discriminative features. Structural characteristics are learned using a 3D-CNN from MRI data, functional interactions are modeled using GATs from fMRI connectivity matrices, and temporal dynamics are extracted from EEG signals using a BiLSTM-Transformer architecture. The learned features are then integrated through an adaptive attention mechanism that dynamically determines the importance of each modality. Finally, the fused representation is used to predict disease stages while explainability techniques provide insights into the factors contributing to the model's decisions.

The proposed framework addresses several limitations of existing approaches. First, it acknowledges that disease-related biomarkers evolve over time and therefore the contribution of each modality may vary across different stages of disease progression. Second, it integrates structural, functional, and electrophysiological information within a unified framework. Third, the incorporation of explainability mechanisms enhances clinical trust and facilitates the interpretation of predictions.

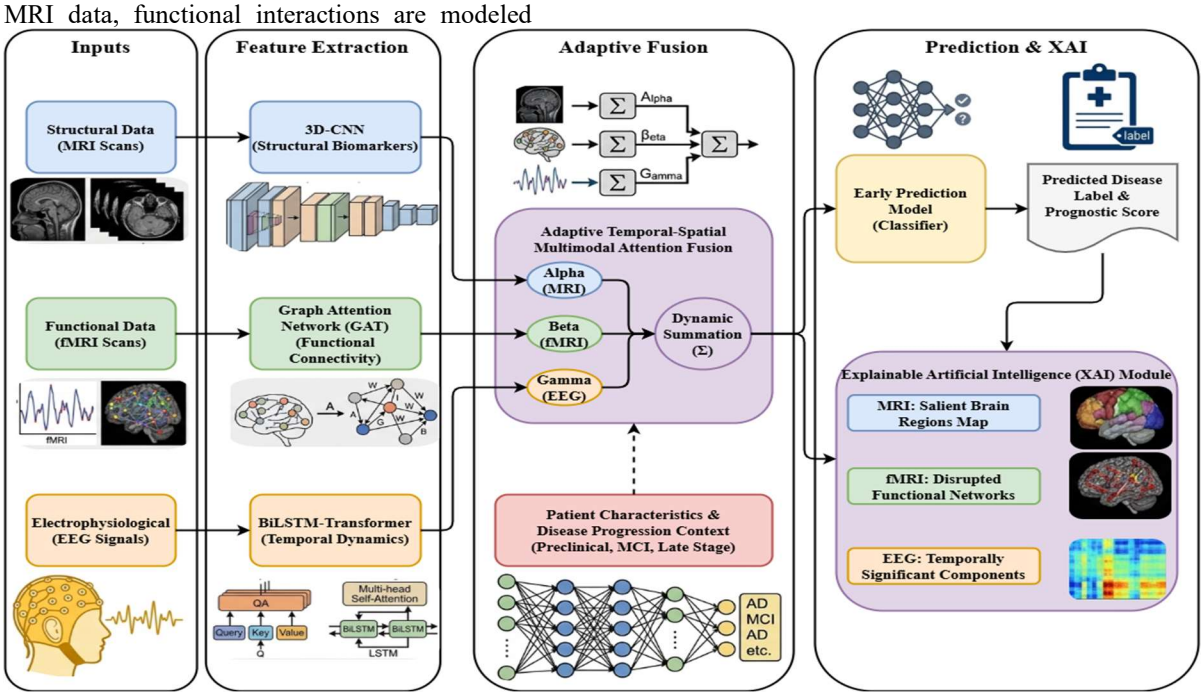

Figure 1: Proposed ATSMAN framework

Figure 1 shows the proposed ATSMAN Framework. In the first stage, heterogeneous neuroimaging data are collected from three complementary modalities. Structural MRI scans provide high-resolution anatomical information for identifying brain atrophy and tissue abnormalities. Functional MRI (fMRI) captures blood oxygen level-dependent (BOLD) signals and functional connectivity patterns among brain regions, reflecting alterations in neural communication. EEG signals provide high temporal resolution measurements of electrical brain activity, enabling the detection of subtle temporal abnormalities associated with early disease progression. The combination of these modalities allows the framework to exploit both

spatial and temporal characteristics of neurological disorders. During the feature extraction stage, modality-specific DL networks are employed to learn representative biomarkers from each data source. A 3D-CNN processes structural MRI volumes and extracts spatial features corresponding to anatomical biomarkers. The fMRI data are represented as connectivity graphs and analyzed using a GAT, which learns relationships among brain regions and emphasizes significant functional interactions through attention mechanisms. Simultaneously, EEG signals are fed into a BiLSTM-Transformer architecture, where BiLSTM captures long-term temporal dependencies in both forward and backward directions, while the Transformer with multi-head self-attention identifies important temporal patterns and frequency variations associated with abnormal neural activity. The extracted features are subsequently forwarded to the Adaptive Temporal-Spatial Multimodal Attention Fusion module, which forms the core of ATSMAN. Three modality-specific attention coefficients, namely α for MRI, β for fMRI, and γ for EEG, are assigned to emphasize the relative importance of each modality. The proposed architecture dynamically changes weights across patient features and disease progression phases, including preclinical, moderate cognitive impairment (MCI), Alzheimer's disease (AD), and advanced neurological disorders, unlike static fusion methods. This adaptive process makes the most informative modality more important to the final representation. A multimodal feature vector is created by integrating weighted features using a dynamic summation operation (Σ), preserving structural, functional, and temporal information. The early prediction model uses the fused representation to classify healthy people from neurological illness patients and estimate disease severity prognosis scores. The classifier uses multimodal representation to improve classification and early-stage diagnosis, when small defects may not be obvious. The approach uses Explainable Artificial Intelligence (XAI) to promote transparency and clinical dependability. XAI identifies salient anatomical regions in MRI scans responsible for prediction, disrupted functional connectivity networks in fMRI data, and temporally significant EEG components contributing to abnormal brain activity to produce interpretable outputs. These graphic explanations help physicians comprehend the model's conclusions and provide reliable diagnostic assistance.

3.2 MRI Structural Feature Extraction

Brain structural alterations indicate neurological diseases. Progressive hippocampal shrinkage, cortical thinning, ventricular enlargement, and white matter tract degeneration are common in neurodegenerative disorders. As structural problems may develop years before severe clinical symptoms, MRI is important for early disease prediction. A variety of preprocessing steps increase picture quality and subject consistency in MRI data. Initially, skull stripping removes non-brain tissues like the scalp and skull to concentrate on brain regions. To eliminate scanner and acquisition technique differences, intensity normalization is employed. Spatial normalisation matches all images to a brain template for subject comparisons. Preprocessing methods like bias field correction reduce intensity inhomogeneities. Post-preprocessing, a 3D-CNN extracts structural characteristics from volumetric MRI data. Traditional methods use handmade features, whereas the 3D-CNN learns complicated spatial representations from data. Multiple convolutional layers detect local anatomical patterns, activation layers capture nonlinear interactions, and pooling layers decrease dimensionality while keeping critical information. Gradient flow and training efficiency are improved via residual connections. Structural feature representations show disease-related hippocampus volume, cortical thickness, and gray matter density variations. The multimodal framework relies on these variables to assess structural deterioration.

MRI and 3D-CNN are chosen for the stage of structural feature extraction, due to the fact that MRI enables the acquisition of volumetric anatomical data, while 3D-CNN can be applied to learning spatial relationship along three directions including depth, height and width. Hence, 3D-CNN is applicable in detecting 3D structural abnormality, e.g., hippocampal atrophy, cortical thickness, and gray matter density without considering individual slices as isolated 2D samples.

Spatial continuity between neighboring voxels and brain region-specific anatomical patterns are captured via three-dimensional convolution. Volumetric convolutions retain contextual linkages between nearby voxels using depth, height, and breadth information, unlike two-dimensional convolutions that evaluate slices independently. This method may reveal local structural abnormalities and disease-specific

anatomical representations to the model. Thus, convolutional formulation is necessary for MRI spatial biomarker extraction.

$$M_{u,v,w}^{(k)} = \sigma\left(\sum_{c=1}^{C^{(1)}}\sum_{i=0}^{K_d-1}\sum_{j=0}^{K_h-1}\sum_{m=0}^{K_w-1} W_{i,j,m,c}^{(k)} \cdot M_{u+i,v+j,w+m}^{(l-1,c)} + b_k^{(l)}\right) \quad (1)$$

In equation (1), where $M_{u,v,w}^{(k)}$ denotes the activation value of the k^{th} feature map at spatial location (w) in the l^{th} convolutional layer. $C^{(1)}$ represents the number of input channels inherited from the preceding layer. K_d, K_h, and K_w correspond to the kernel dimensions along depth, height, and width, respectively. $W_{i,j,m,c}^{(k)}$ denotes the trainable convolution weight connecting the c^{th} input channel to the k^{th}output feature map. $b_k^{(l)}$ represents the bias parameter associated with the k^{th} filter, while $\sigma(\cdot)$ denotes the nonlinear activation function employed to introduce representational complexity. The resulting tensor $M^{(k)}$ encapsulates localized anatomical characteristics extracted from MRI volumes and forms the basis for subsequent feature refinement. Although convolutional layers reliably capture local anatomical characteristics, their outputs change throughout training, which may affect convergence and model stability. Normalization and nonlinear activation enhance representation and control feature statistics. Networks learn more discriminative structural patterns by minimizing internal covariate changes. The design may replicate complex anatomical changes associated with sickness progression utilizing nonlinear transformations.

$$R_{u,v,w}^{(k)} = max\left(0, \frac{M_{u,v,w}^{(k)} - \mu_k^{(l)}}{\sqrt{\left(\sigma_k^{(l)}\right)^2 + \varepsilon}}\gamma_k^{(l)} + \beta_k^{(l)}\right) \quad (2)$$

As shown in equation (2), where $R_{u,v,w}^{(k)}$ denotes the normalized activation value corresponding to the k^{th} feature map. $\mu_k^{(l)}$ and $\sigma_k^{(l)}$ represent the mean and standard deviation computed across the mini-batch, respectively. The scalar ε is introduced to prevent numerical instability during division. Parameters $\gamma_k^{(l)}$ and $\beta_k^{(l)}$ denote learnable scaling and shifting coefficients associated with batch normalization. The operator $max\,(0,\cdot)$ corresponds to the Rectified Linear Unit (ReLU), which suppresses negative activations and preserves only meaningful structural responses. Redundant information and computational cost may rise with intermediate feature map dimensions. Thus, multi-scale pooling reduces feature dimensionality and aggregates discriminative responses. This method helps maintain dominant anatomical traits across receptive fields and increase resilience to slight spatial alterations. Feature aggregation strategies increase representation learning and model capture of disease-related structural anomalies.

$$P_k^{(l)} = \frac{1}{S}\sum_{s=1}^{S} max_{\,(u,v,w)\in\Omega_s}\left(R_{u,v,w}^{(k)}\right) + \lambda_p\sqrt{\frac{1}{N}\sum_{u,v,w}\left(R_k^{-(l)}\right)^2} \quad (3)$$

As computed in equation (3), where $P_k^{(l)}$ denotes the aggregated feature representation obtained from the k^{th} channel. S represents the total number of pooling regions, while Ω_s denotes the spatial neighborhood associated with the s^{th} pooling window. $R_k^{-(l)}$ corresponds to the average activation value of the feature map. The coefficient λ_p controls the contribution of statistical dispersion during aggregation, and N denotes the total number of voxels within the feature map. Hierarchical convolution and multi-scale feature aggregation provide a compact embedding for multimodal integration from derived structural representations. This embedding describes tissue shape and brain anatomy. The feature vector is the structural component of the proposed ATSMAN framework and engages in adaptive temporal-spatial multimodal fusion with functional and electrophysiological data.

$$F_{MRI} = \phi\left(b_f\right) + \lambda_f \parallel \oplus_{l=1}{}^{L} P^{(l)} \parallel_2^2 \quad (4)$$

As found in equation (4), where F_{MRI} denotes the final structural feature embedding generated from MRI data. The operator $\oplus$ represents feature concatenation across all convolutional layers, and $P^{(l)}$ denotes the pooled feature vector obtained from the l^{th} layer. W_f and b_f correspond to the learnable weight matrix and bias vector responsible for dimensionality reduction. The function $\phi(\cdot)$ denotes the nonlinear projection operation used to generate compact embeddings. The regularization coefficient λ_f controls the contribution of the L_2-norm term, which prevents overfitting and improves generalization. The resulting vector F_{MRI} encapsulates high-level anatomical biomarkers that are subsequently utilized during

adaptive multimodal fusion for early brain disease prediction. Figure 2 shows the Adaptive Attention-based Multimodal Fusion Mechanism.

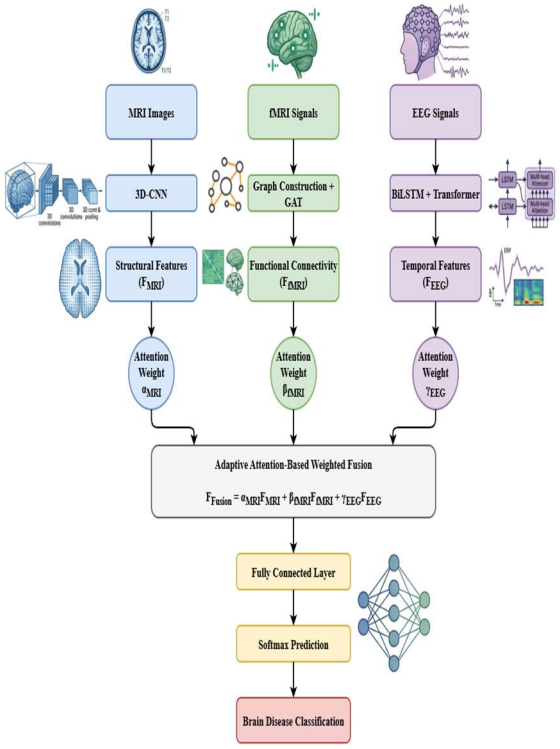

Figure 2: Adaptive Attention-based Multimodal Fusion Mechanism

GAT is chosen for fMRI data analysis since functional connectivity is about relationships between brain regions and not regional features in isolation. Considering brain regions as graph nodes and their functional interactions as edges, GAT can also learn the significance of neighbor links via attention. Hence, GAT is well suited to capture disease-related connectivity patterns and to filter important functional connections from less informative ones.

3.3 fMRI Functional Connectivity Analysis

While structural MRI provides information regarding anatomical abnormalities, functional MRI offers insights into brain activity and communication among different regions. Functional connectivity may be assessed using resting-state fMRI, which detects blood oxygenation variations caused by neural activity. Different connection patterns are connected to numerous neurological diseases and sometimes precede structural changes. Slice timing, motion correction, spatial normalization, smoothing, and temporal filtering are fMRI data preparation steps. Slice timing correction corrects slice-specific picture capture delays. Reduces scanning subject movement artifacts using motion correction. Smoothing improves signal-to-noise ratios, while spatial normalization aligns pictures to anatomical space. Temporal filtering reduces high-frequency noise and low-frequency drifts that might disrupt functional connectivity studies. A conventional atlas divides the brain into anatomically defined areas after preprocessing. Time-series data from each area of interest are extracted and pairwise interactions quantified to create functional connectivity matrices. These matrices show brain region communication strength and demonstrate large-scale brain network structure. GATs study connection patterns. This graph model treats brain areas as nodes and functional connections as edges. The attention mechanism lets the model prioritize the most revealing disease progression interactions by assigning various relationships different levels of relevance. Through multiple graph attention layers, the model learns discriminative functional representations that characterize abnormalities within important networks such as the Default Mode Network, Salience Network, and Executive Control Network. Functional connectivity features derived from fMRI complement structural MRI findings by identifying network disruptions that may emerge during the early stages of neurological disorders.

The temporal correlation between areas of interest (ROIs) must be quantified before graphing fMRI sequences. The system captures synchronized neural activity patterns and illness progression-related connectivity alterations via dynamic correlation estimation. Correlation-based connectivity analysis retains functional reliance across distant brain areas and supports graph learning better than static representations. Thus, the following algorithm calculates ROI signal pairwise connection strength and constructs the first functional connectivity matrix.

$$C_{ij} = \frac{\sum_{t=1}^{T}(X_i(t)-\mu_i)(X_j(t)-\mu_j)}{\sqrt{\sum_{t=1}^{T}(X_i(t)-\mu_i)^2}\sqrt{\sum_{t=1}^{T}(X_j(t)-\mu_j)^2}} + \lambda_c$$
$$exp\ exp\left(-\frac{\|X_i-X_j\|_2^2}{2\sigma_c^2}\right) \quad (5)$$

As discussed in equation (5), where C_{ij} denotes the functional connectivity strength between the i^{th} and j^{th} brain regions. $X_i(t)$ and $X_j(t)$ represent the BOLD signal intensities at time instant t, while μ_i and μ_j denote their corresponding mean values. T represents the total number of temporal observations. The coefficient

λ_c controls the contribution of the Gaussian similarity component, and σ_c denotes the bandwidth parameter governing neighborhood smoothness. The Euclidean norm $\|\cdot\|_2$ measures the distance between ROI signal vectors. The resulting matrix C_{ij} encapsulates both temporal correlation and similarity information required for graph construction. Correlation matrices give connection information, but weak associations have little diagnostic utility and add computational complexity. To maintain only relevant brain area interactions, an adjacency matrix is created. This transformation turns the functional connectivity matrix into a sparse graph for fast graph-based learning. The adjacency representation emphasizes disease-related brain network communication routes, improving interpretability.

$$A_{ij} = \frac{expexp\,(\eta C_{ij})}{\sum_{k=1}^{N} exp\,(\eta C_{ik})} \cdot \delta(\tau) + \lambda_a \left(\frac{C_{ij}}{max\,(C)}\right) \quad (6)$$

As derived in equation (6), where A_{ij} represents the weighted adjacency matrix corresponding to the connectivity between nodes i and j. C_{ij} denotes the previously computed correlation coefficient, while Nrepresents the total number of brain regions. The parameter ηcontrols the softmax scaling factor used for normalization. The function $\delta(\cdot)$ acts as a threshold operator governed by the threshold value τ, allowing only meaningful connections to be retained. The coefficient λ_a regulates the influence of normalized connectivity values. The matrix A_{ij} serves as the graph topology utilized in subsequent attention-based learning. Although not all nearby brain areas contribute equally to illness characterization, adaptive attention processes give graph nodes relevance weights. The framework may concentrate on useful functional interactions while eliminating irrelevant connections with this operation. The graph attention mechanism increases discriminative representation learning and sensitivity to neurological disorder-related aberrant connection patterns by dynamically evaluating neighborhood relevance.

$$\alpha_{ij} = \frac{expexp\left(LeakyReLU\left(a^T[W_h h_j]\right)\right)}{\sum_{k\in N(i)} exp\left(LeakyReLU\left(a^T[W_h h_k]\right)\right)} + \lambda_\alpha \frac{A_{ij}}{\sum_{k\in N(i)} A_{ik}} \quad (7)$$

As deliberated in equation (7), where α_{ij} denotes the normalized attention coefficient assigned to the connection between nodes $i$and j. h_i and h_j represent feature vectors corresponding to neighboring ROIs. W_h denotes the learnable weight matrix responsible for feature transformation, and a^T represents the attention vector used to estimate node importance. The operator $\|$ denotes vector concatenation, while $N(i)$represents the neighborhood set associated with node i. The coefficient λ_α controls the contribution of graph topology information during attention estimation. Consequently, α_{ij} quantifies the relative importance of neighboring regions and guides information propagation within the graph. After graph attention learning, high-dimensional connection representations become compact embeddings for multimodal fusion. Local neighborhood data and global network features are combined in this step. The feature vector describes brain functional communication patterns and complements structural MRI and EEG descriptions. A compact embedding increases processing efficiency and ATSMAN's adaptive temporal-spatial integration.

$$F_{fMRI} = \psi(b_g) + \lambda_g \left\|\sum_{j\in N(i)} \alpha_{ij} W_g h_j\right\|_2^2 \quad (8)$$

As shown in equation (8), where F_{fMRI} denotes the final functional connectivity embedding extracted from fMRI data. α_{ij} represents the attention coefficient obtained from graph attention learning, while W_g and b_g denote the trainable weight matrix and bias vector associated with graph feature transformation. The neighborhood set $N(i)$ includes all nodes connected to node i, and h_j represents the feature vector of neighboring regions. The nonlinear projection operator $\psi(\cdot)$ generates compact functional representations, whereas the regularization parameter λ_g controls the contribution of the L_2-norm term to prevent overfitting. The resulting embedding F_{fMRI} captures disease-specific functional interactions and serves as the functional component for adaptive multimodal fusion in the proposed ATSMAN framework.

3.4 EEG Temporal Feature Learning

EEG provides a non-invasive method for monitoring the electrical activity of the brain with millisecond temporal resolution. EEG detects fast neuronal dynamics and electrophysiological abnormalities better than MRI and fMRI. Numerous studies have shown that altered EEG rhythms may appear early in neurological illnesses, making EEG a useful early predictor. Several rounds of EEG preprocessing improve signal quality. Band-pass filtering reduces noise and extraneous frequency components.

Independent Component Analysis removes eye blinks, muscular movement, and electrical interference. For analysis, cleaned EEG data are split into fixed-duration epochs. Normalisation ensures recording uniformity. The proposed system learns temporal representations from EEG data using a hybrid BiLSTM-Transformer architecture. BiLSTM networks can describe sequential data relationships by processing information forward and backward. The model captures disease-related electrophysiological temporal patterns with bidirectional processing. BiLSTM networks can simulate sequential interactions, but complicated EEG recordings may make it difficult to discover long-range dependencies. Therefore, Transformer-based attention mechanisms are incorporated to enhance the learning of global temporal interactions. The self-attention mechanism enables the model to focus selectively on the most informative segments of the EEG signal while considering relationships across the entire sequence. The combined BiLSTM-Transformer architecture extracts features associated with alterations in brain rhythms, synchronization patterns, and oscillatory dynamics. Abnormal increases in theta activity, reductions in alpha power, and disruptions in neural synchronization serve as important biomarkers contributing to early disease prediction.

BiLSTM-Transformer is chosen for EEG analysis due to its sequential nature and the electrophysiological information of the EEG are reacting at both the short- and long-range temporal dependencies. BiLSTM captures temporal dependencies in forward and backward directions and the Transformer computes significant relations over distant signal parts via self-attention. Hence, the synthesis of the two methods can effectively represent both sequential and global temporal ECG patterns.

The proposed approach uses a bidirectional LSTM architecture to maintain past data and minimize extraneous EEG oscillations. Memory regulation is necessary to preserve relevant patterns and eliminate duplicate information as disease-related temporal features span several time periods. This system governs the contribution of prior memory states and dynamically decides the quantity of historical knowledge needed for future prediction. Therefore, the following formulation determines adaptive memory retention in the recurrent architecture.

$$f_t = \sigma\left(W_f^{(h)} h_{t-1} + W_f^{(x)} x_t + b_f + \lambda_f tanh\left(\frac{\|h_{t-1}\|_2^2 + \|x_t\|_2^2}{1+exp\,(-t/T)}\right)\right) \quad (9)$$

In equation (9), where f_t denotes the forget gate activation at time instant t, which determines the proportion of information retained from the previous memory state. The vectors h_{t-1} and x_t represent the hidden state and EEG input sequence, respectively. The matrices $W_f^{(h)}$ and $W_f^{(x)}$ denote trainable weights corresponding to hidden and input components, whereas $b_f$represents the bias vector. The coefficient $\lambda_f$controls the nonlinear memory modulation term, and T denotes the sequence length. The sigmoid function $\sigma(\cdot)$ maps the output into the interval [1], thereby regulating memory preservation. Despite memory preservation, the network must update its internal state with fresh electrophysiological data. Thus, historical relationships and EEG data are combined via adaptive memory evolution. The system can capture disease-related alterations across diverse time scales with continuous temporal learning. Thus, cell status updating is essential for maintaining long-range EEG dependencies.

$$C_t = f_t \odot C_{t-1} + i_t \odot tanh\left(\lambda_c \frac{\|x_t - h_{t-1}\|_2^2}{1+\|C_{t-1}\|_2}\right) \quad (10)$$

As inferred from equation (10), where C_t denotes the memory cell state at time instant t, while C_{t-1} represents the previous memory state. The operator $\odot$ indicates element-wise multiplication. The variable i_t denotes the input gate responsible for regulating new information updates. The matrices $W_c^{(h)}$and $W_c^{(x)}$correspond to trainable weight parameters associated with hidden and input vectors, respectively, whereas b_c denotes the bias term. The coefficient λ_c controls the influence of the adaptive regularization component. The hyperbolic tangent function provides nonlinear transformation for memory updates. BiLSTM retains sequential dependencies but cannot describe long-range contextual interactions in longer EEG sequences. The Transformer encoder uses multi-head self-attention to circumvent this constraint. The framework may concentrate on informative temporal segments and build links between distant signal components using this technique. EEG representations become more discriminative and sensitive to abnormal neuronal oscillations associated with illness progression with attention-driven feature learning.

$$A^{(m)} = Softmax\left(\lambda_a \frac{Q^{(m)}Q^{(m)^T} + K^{(m)}K^{(m)^T}}{\|Q^{(m)}\|_F + \|K^{(m)}\|_F + \varepsilon}\right)V^{(m)} \quad (11)$$

As discussed in equation (11), where $A^{(m)}$ denotes the attention output corresponding to the m^{th} attention head. The matrices $Q^{(m)}$, $K^{(m)}$, and $V^{(m)}$ represent query, key, and value matrices generated from EEG feature sequences. The scalar d_k denotes the dimensionality of key vectors and is used for scaling the attention scores. The coefficient λ_a controls the contribution of similarity regularization, while $\|\cdot\|_F$ represents the Frobenius norm. The constant εprevents numerical instability. The Softmax operator normalizes attention scores and assigns relative importance to different temporal segments. The retrieved information is compressed for multimodal fusion after BiLSTM and self-attention learn temporal relationships and contextual interactions. Sequential consistency and attention-based contextual linkages are preserved in the storage of discriminative temporal properties. This compact embedding is ATSMAN's electrophysiological component and complements MRI and fMRI structural and functional representations. The following formulation provides the final EEG feature vector for adaptive temporal-spatial fusion.

$$F_{EEG} = \phi(b_e) + \lambda_e \| W_e[h_T^b] \|_2^2 \quad (12)$$

As shown in equation (12), where F_{EEG} denotes the final temporal feature embedding extracted from EEG signals. The operator $\oplus$ represents concatenation across all attention heads, while $A^{(m)}$denotes the output corresponding to the m^{th}self-attention head. The vectors h_T^f and h_T^b represent forward and backward hidden states generated by the BiLSTM network. The trainable matrix W_e and bias vector b_e perform feature transformation and dimensionality reduction. The nonlinear projection function $\phi(\cdot)$ generates compact temporal representations, whereas the regularization coefficient λ_e controls the contribution of the L_2-norm penalty.

3.5 Adaptive Attention-Based Multimodal Fusion

One of the major innovations of the proposed ATSMAN framework is the adaptive attention-based multimodal fusion strategy. Traditional multimodal techniques use simple feature concatenation or set weighting algorithms that presume equal contributions from all modalities. Neurological illnesses are dynamic, therefore biomarkers may change over time. MRI structural alterations may be less noticeable than EEG electrophysiological abnormalities in early illness development. In intermediate phases, fMRI may show functional connectivity changes, whereas MRI shows morphological deterioration in later stages. Thus, giving all modalities equal weight may reduce prediction performance. An adaptive attention mechanism in ATSMAN dynamically selects the contribution of MRI, fMRI, and EEG characteristics for each person to overcome this constraint. The attention module weighs modality-specific representations based on their prediction task relevance. The framework prioritizes informative biomarkers and reduces less relevant modalities using this adaptive technique. The fused multimodal representation generated through this process captures complementary information from structural, functional, and electrophysiological domains. Consequently, the integrated feature space provides a more comprehensive characterization of disease progression compared with unimodal approaches.

Adaptive temporal-spatial attention is chosen as the relevance of MRI, fMRI and EEG data may vary between subject and disease specific aberrancies. This fixed weighting or simple concatenation does not allow for dynamic modulation of these weights of contributions. Hence, the proposed attention mechanism learns modality-specific relevance weights and highlights the most informative representations in fusion, which fits well for fusing heterogeneous structural, functional as well as temporal information.

The proposed methodology initially computes a modality-specific attention coefficient for structural characteristics to measure the significance of MRI-extracted anatomical information. Anatomical biomarkers vary in relevance throughout illness stages, therefore assigning a set contribution to MRI characteristics may lose or duplicate information. Thus, nonlinear interactions among multimodal representations are used to evaluate structural contribution via adaptive weighting. The framework preserves disease-relevant anatomical traits and balances functional and electrophysiological properties with this operation.

$$\alpha_{MRI} = \frac{expexp\left((b_m)\right)}{exp\left((b_m)\right) + \sum_{q\in\{fMRI,EEG\}} exp\left(S_q\right)} + \lambda_m \frac{\|F_{MRI}\|_2}{\|F_{MRI}\|_2 + \|F_{fMRI}\|_2 + \|F_{EEG}\|_2} \quad (13)$$

As shown in equation (13), where α_{MRI} denotes the adaptive attention coefficient associated with MRI structural features. The vectors F_{MRI}, F^{-}_{fMRI}, and F^{-}_{EEG} represent structural, functional, and electrophysiological feature embeddings, respectively. The matrices W_m, U_m, and V_m denote trainable transformation parameters, whereas b_m represents the bias vector. The parameter ω_m corresponds to the attention projection vector. The quantity S_q represents attention scores associated with the remaining modalities. The coefficient λ_m regulates the contribution of feature-energy normalization. Consequently, α_{MRI} determines the relative importance of anatomical biomarkers during multimodal integration. Functional interactions from fMRI give supplementary information on brain connectivity problems, whereas structural aspects provide significant anatomical information. Since illness progression often impacts functional communication routes before structural deterioration, fMRI representations must be evaluated adaptively. To evaluate the relevance of functional connectivity traits and dynamically manage their effect during multimodal aggregation, a specialized attention method is used.

$$\beta_{fMRI} = \frac{expexp\left((b_f)\right)}{exp\left((b_f)\right)+\sum_{q\in\{MRI,EEG\}} exp\,(S_q)} + \lambda_f \frac{\|F_{fMRI}\|_2}{\|F_{MRI}\|_2+\|F_{fMRI}\|_2+\|F_{EEG}\|_2} \quad (14)$$

In equation (14), where β_{fMRI} denotes the adaptive attention weight assigned to functional connectivity representations. The vector F_{fMRI} represents graph-based functional embeddings, whereas F^{-}_{MRI} and F^{-}_{EEG} denote complementary feature representations obtained from MRI and EEG modalities. The matrices W_f, U_f, and V_f correspond to learnable transformation parameters, and $b_f$denotes the bias vector. The attention projection vector is represented by ω_f, while S_q indicates attention scores of competing modalities. The coefficient λ_f regulates feature-energy normalization. The resulting coefficient β_{fMRI} quantifies the significance of functional connectivity patterns for disease characterization. EEG data has excellent temporal resolution and may identify dynamic brain disorders that structural or functional imaging cannot. Since temporal patterns vary widely between individuals and illness stages, fixed weights for EEG representations may limit prediction accuracy. To assess EEG data' importance and dynamically decide their contribution to the final multimodal representation, an adaptive attention mechanism is proposed.

$$\gamma_{EEG} = \frac{expexp\,((b_e))}{exp\,((b_e))+\sum_{q\in\{fMRI,fMRI\}} exp\,(S_q)} + \lambda_e \frac{\|F_{EEG}\|_2}{\|F_{MRI}\|_2+\|F_{fMRI}\|_2+\|F_{EEG}\|_2} \quad (15)$$

As derived in equation (15), where γ_{EEG} represents the adaptive attention coefficient associated with temporal EEG features. The vector F_{EEG} denotes the electrophysiological embedding extracted by the BiLSTM-Transformer architecture. The matrices W_e, U_e, and V_e correspond to trainable weight parameters responsible for nonlinear transformations, while b_e represents the bias vector. The projection vector ω_e estimates the importance of EEG features, and S_q denotes competing modality scores. The coefficient λ_e controls the energy-based regularization term. Consequently, γ_{EEG} dynamically determines the contribution of electrophysiological information during multimodal fusion. ATSMAN combines structural, functional, and temporal representations into a latent space after calculating modality-specific attention coefficients. The adaptive fusion technique prioritizes modality based on its discriminative capacity, unlike concatenation. This technique creates a compact multimodal representation for illness prediction by preserving complimentary information and minimizing duplicate characteristics. Fused embedding is the classification module's fundamental representation.

$$F_{Fusion} = \phi\left(b_{fus}\right) + \lambda_{fus} \| \alpha_{MRI}F_{MRI} + \beta_{fMRI}F_{fMRI} + \gamma_{EEG}F_{EEG} \|_2^2 \quad (16)$$

As computed in equation (16), where F_{Fusion} denotes the final multimodal feature representation generated by the adaptive temporal-spatial attention mechanism. The coefficients α_{MRI}, β_{fMRI}, and γ_{EEG} represent attention weights corresponding to MRI, fMRI, and EEG modalities, respectively. The operator ‖denotes feature concatenation, while W_{fus} and b_{fus} represent trainable fusion parameters. The nonlinear mapping function $\phi(\cdot)$ transforms heterogeneous representations into a common latent space. The coefficient λ_{fus} regulates the regularization term to improve generalization capability. The resulting vector F_{Fusion} preserves complementary anatomical, functional, and electrophysiological information and serves as the input to the disease prediction module of the proposed ATSMAN framework.

3.6 Disease Prediction Module

Following multimodal fusion, the integrated feature representation is provided to the disease prediction module. This component consists of fully connected neural network layers responsible for learning complex relationships between the extracted features and disease outcomes. The prediction module is designed to classify individuals into different disease categories or stages. Depending on the study objectives, these categories may include healthy controls, individuals at risk, patients with mild cognitive impairment, early-stage disease cases, and advanced disease cases. The network progressively transforms the fused feature representation into high-level discriminative patterns associated with these diagnostic categories. Dropout regularization techniques are employed to reduce overfitting and improve model generalization. Batch normalization is incorporated to stabilize training and accelerate convergence. The final output layer generates probability distributions corresponding to each disease class, enabling the identification of the most likely diagnostic outcome. By leveraging multimodal representations generated through ATSMAN, the prediction module facilitates accurate identification of individuals experiencing early pathological changes, thereby supporting timely intervention strategies.

First, the fused feature representation must be translated into a high-dimensional discriminative space where class-specific properties become separable to identify healthy people from diseased circumstances. Using latent interactions among multimodal embeddings and class-wise activation scores, the nonlinear fully connected prediction layer transforms. A formulation that improves feature separability allows the framework to capture complicated disease progression decision boundaries. The following equation calculates the discriminative score vector for probability estimation.

$$Z_k = \phi\left(\sum_{i=1}^{D}\sum_{j=1}^{H} W_{ij}^{(2)} \cdot \sigma\left(b_j^{(1)}\right) + b_k^{(2)}\right) + \lambda_z \parallel W^{(1)}W^{(2)} \parallel_F^2 \quad (17)$$

As shown in equation (17), where Z_k denotes the discriminative activation score corresponding to the k^{th}disease category. The vector $F_{Fusion}^{(i)}$ represents the i^{th} component of the multimodal feature representation obtained after adaptive attention fusion. The matrices $W^{(1)}$ and $W^{(2)}$ denote trainable weight parameters associated with the hidden and output layers, respectively, whereas $b_j^{(1)}$ and $b_k^{(2)}$ represent bias terms. The functions $\sigma(\cdot)$ and $\phi(\cdot)$ denote nonlinear activation operators responsible for introducing representational flexibility. The dimensions D and H correspond to the input and hidden layer sizes. The parameter $\lambda_z$controls the Frobenius norm regularization term, which prevents excessive parameter growth and improves generalization. The resulting score vector Z_k provides discriminative information required for probabilistic disease prediction. Discriminative scores measure class separability, but clinical decision-making needs normalized probability estimates of illness category likelihood. Therefore, Softmax-based probabilistic inference is used to convert activation ratings into posterior probabilities. This approach lets the network assess uncertainty and distinguish numerous neurological diseases. Probabilistic outputs aid interpretation and confidence-based diagnosis. Thus, the following equation gives illness class probability distribution.

$$P(y = k \mid F_{Fusion}) = \frac{exp\,exp\left(\lambda_p \frac{\|F_{Fusion}\|_2^2}{1+\sum_{c=1}^{C}\|Z_c\|_2}\right)}{\sum_{c=1}^{C} exp\left(\lambda_p \frac{\|F_{Fusion}\|_2^2}{1+\sum_{c=1}^{C}\|Z_c\|_2}\right)} \quad (18)$$

As calculated in equation (18), where $P(y = k \mid F_{Fusion})$ denotes the posterior probability that the fused representation belongs to the k^{th} disease category. The quantity Z_k represents the discriminative score corresponding to class k, while C denotes the total number of disease categories considered during classification. The parameter λ_p controls the contribution of feature-energy normalization, and $\| F_{Fusion} \|_2$ denotes the Euclidean norm of the fused feature vector. The exponential operator ensures positive probability values, whereas the denominator performs normalization across all classes. While posterior probabilities give class membership information, early illness detection needs confidence estimate to measure prediction reliability and disease severity. Thus, the proposed ATSMAN system generates a disease confidence score using probability distributions and feature-energy characteristics. This process improves interpretability and lets doctors assess forecast confidence before making decisions. Confidence estimate also enhances model transparency and risk assessment in diagnostic settings.

$$S_{conf} = \eta_1 max_{\,k}(P(y = k \mid F_{Fusion})) + \eta_2\left(\frac{-\sum_{k=1}^{C} P(y=k|F_{Fusion}) log\ P(y=k|F_{Fusion})}{loglog\ C}\right) + \eta_3 \frac{\|F_{Fusion}\|_2}{\|F_{Fusion}\|_2+\|Z\|_2} \quad (19)$$

As shown in equation (19), where S_{conf} denotes the overall disease confidence score generated by the prediction module. The term $P(y = k \mid F_{Fusion})$ represents the posterior probability corresponding to class k, whereas C denotes the total number of disease categories. The coefficients η_1, η_2, and η_3 regulate the contributions of maximum probability, entropy-based uncertainty reduction, and feature-energy consistency, respectively. The logarithmic term measures prediction entropy and quantifies uncertainty associated with the classification process. The Euclidean norms $\| F_{Fusion} \|_2$ and $\| Z \|_2$ characterize the magnitudes of fused features and score vectors. Consequently, the confidence score S_{conf} provides an additional measure of diagnostic reliability and serves as an important indicator for prognosis estimation and clinical interpretation within the proposed ATSMAN framework.

3.7 Model Training and Optimization

Training the ATSMAN framework uses supervised learning. The dataset is split into training, validation, and testing subsets for accurate model assessment. Class distributions across subsets are preserved via stratified sampling. Optimization methods reduce training classification mistakes. Adaptive optimization approaches dynamically alter learning rates and update parameters efficiently. Overfitting is prevented by monitoring validation performance using early stopping tactics. When gains plateau, learning rate scheduling reduces learning rates to improve convergence. Additionally, data augmentation may increase generalization and alleviate class imbalance. Rotations and translations boost dataset variety for MRI and fMRI data. Signal perturbations and windowing may be used for EEG recordings. Learning rates, batch sizes, attention dimensions, and hidden units are optimized by hyperparameter adjustment. Together, these optimization procedures provide model stability.

To enhance sensitivity toward minority disease classes and mitigate the influence of imbalanced training samples, the proposed framework employs a weighted categorical cross-entropy objective function. Unlike conventional loss functions that assign equal importance to all classes, the weighted formulation adaptively penalizes misclassifications according to class distributions. This mechanism improves the ability of the network to learn subtle disease characteristics and prevents the dominance of majority classes during optimization. Consequently, the following formulation computes the primary classification loss minimized during training.

$$L_{WCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\omega_c y_{ic} log\ (\varepsilon) + \lambda_w \left(\frac{\|F_{Fusion}^{(i)}\|_2^2}{1+\sum_{c=1}^{C}\hat{y}_{ic}}\right) \quad (20)$$

As discussed in equation (20), where L_{WCE} denotes the weighted categorical cross-entropy loss associated with multimodal classification. The quantity N represents the total number of training samples, whereas C denotes the number of disease categories. The variables y_{ic} and $\hat{y}_{ic}$ correspond to the ground-truth label and predicted probability of the c^{th} class for the i^{th}sample, respectively. The coefficient ω_c represents the class-specific weighting factor introduced to compensate for class imbalance. The constant εprevents numerical instability caused by logarithmic operations. The parameter λ_w regulates the contribution of feature-energy normalization, and $F_{Fusion}^{(i)}$ denotes the fused multimodal representation corresponding to the i^{th} sample. Due of its many trainable parameters, deep neural networks may overfit, even when weighted cross-entropy minimizes classification mistakes. Large parameter magnitudes can cause poor generalization and inconsistent performance on unseen examples. Therefore, a regularization approach is used to limit model complexity and smoother parameter distributions. Regularization enhances resilience and optimization convergence. Thus, the following statement regularizes training parameters.

$$L_{Reg} = \lambda_1 \sum_{l=1}^{L} \| W^{(l)} \|_F^2 + \lambda_2 \sum_{l=1}^{L} \| b^{(l)} \|_2^2 + \lambda_3 \sum_{m=1}^{M} \| \alpha_m F_m \|_2^2 \quad (21)$$

In equation (21), where L_{Reg} denotes the regularization loss imposed on network parameters. The matrices $W^{(l)}$ and vectors $b^{(l)}$ represent trainable weights and bias parameters associated with the l^{th} layer, respectively. The total number of layers is denoted by L, whereas M corresponds to the number of modalities participating in multimodal fusion. The coefficient α_m represents the attention weight assigned to the m^{th}modality, and F_m denotes the corresponding feature embedding. The parameters λ_1, λ_2, and λ_3 control the contributions of weight regularization, bias regularization, and feature regularization, respectively. The Frobenius norm $\|\cdot\|_F$ and Euclidean norm $\|\cdot\|_2$ constrain parameter magnitudes and prevent overfitting. The Adam

optimizer is used to optimize parameters after loss computation and regularization because it can adapt learning rates to first- and second-order gradient statistics. Adaptive optimization improves convergence and numerical stability during training. Adam supports sparse gradients and nonstationary objective functions in multimodal DL frameworks well. The following equation repeatedly changes network parameters using adaptive moment estimation.

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} + \lambda_\theta \left(\frac{\nabla_\theta L_{WCE} + \nabla_\theta L_{Reg}}{1+\|\nabla_\theta L_{WCE} + \nabla_\theta L_{Reg}\|_2} \right) \quad (22)$$

As shown in equation (22), where θ_t and θ_{t+1} denote parameter vectors at iterations t and $t+1$, respectively. The scalar η represents the learning rate. The variables $\hat{m}_t$ and $\hat{v}_t$ correspond to bias-corrected first-order and second-order moment estimates obtained from gradient information. The constant ϵprevents division by zero during parameter updates. The coefficient λ_θ controls the contribution of gradient normalization, while ∇_θ denotes the gradient operator with respect to model parameters. The terms L_{WCE} and L_{Reg} represent classification and regularization losses, respectively. The ATSMAN framework aims to reduce classification errors and parameter complexity while increasing multimodal representation discrimination. Thus, the overall optimization goal consolidates weighted cross-entropy loss and regularization requirements. This goal function directs the training process and guarantees the network converges to an optimum parameter configuration for illness prediction. The following formulation reflects the ultimate optimization target reduced during training.

$$L_{Total} = L_{WCE} + L_{Reg} + \lambda_T \left(\frac{\sum_{i=1}^{N} F_{Fusion}^{(i)} \cdot \bar{F}_{Fusion}}{\|F_{Fusion}^{(i)}\|_2 \|\bar{F}_{Fusion}\|_2} \right) \quad (23)$$

As inferred from equation (23), L_{Total} denotes the overall objective function optimized during network training. The terms L_{WCE} and L_{Reg} represent weighted classification loss and parameter regularization loss, respectively. The coefficient λ_T regulates the contribution of feature consistency constraints. The vector $F_{Fusion}^{(i)}$ denotes the multimodal feature representation corresponding to the i^{th} training sample, whereas $\bar{F}_{Fusion}$ represents the average fused representation across the training set. The inner product operation quantifies feature similarity, and the Euclidean norms ensure normalization. Consequently, minimizing L_{Total} enables ATSMAN to achieve stable convergence, improved generalization capability, and reliable early brain disease prediction.

3.8 Explainable Artificial Intelligence Module

XAI is selected because multimodal DL models can produce accurate predictions while providing limited information about the factors responsible for those predictions. In a clinical context, identifying the evidence supporting a prediction is important for interpretation and decision support. Therefore, modality-specific XAI is incorporated to identify relevant MRI anatomical regions, fMRI functional connections, and informative EEG temporal patterns, providing interpretable evidence for the model's decisions.

Deep learning models offer tremendous predictive power, but their black-box nature is a major problem in clinical applications. Before using artificial intelligence systems, doctors need clear explanations of diagnostic variables. ATSMAN uses modality-specific Explainable Artificial Intelligence to improve interpretability. Visualization approaches help doctors discover disease-related structural anomalies in MRIs by highlighting anatomical areas that predict. For fMRI, graph attention mechanisms reveal important functional connections and disrupted brain networks associated with disease progression. For EEG, temporal attention mechanisms identify critical signal segments and frequency components influencing classification outcomes. The integration of explainability mechanisms improves clinician trust, supports decision-making processes, and facilitates the translation of artificial intelligence models into real-world healthcare environments.

A modality-specific relevance score is computed by integrating attention coefficients and feature activation intensities to identify each modality's contribution to the final judgment. The framework may assess structural MRI, functional fMRI, and temporal EEG importance using this approach. A technique that shows the multimodal network's internal dynamics enables physicians to determine which modality dominates a prediction.

$$I_m = \frac{\alpha_m \|F_m\|_2 exp\,(\omega_m^T F_m)}{\sum_{r=1}^{M} \alpha_r \|F_r\|_2 exp\,(\omega_r^T F_r)} + \lambda_I \frac{\|F_m - \bar{F}\|_2^2}{1+\sum_{r=1}^{M} \|F_r\|_2} \quad (24)$$

In equation (24), where I_m denotes the interpretability score associated with the m^{th} modality. The feature vector F_m represents the

embedding extracted from MRI, fMRI, or EEG data, while α_m corresponds to the attention coefficient assigned during multimodal fusion. The vector ω_m denotes the learnable importance projection parameter, and Mrepresents the total number of modalities. The average feature representation is represented by $\bar{F}$, whereas λ_I controls the regularization contribution. The Euclidean norm $\|\cdot\|_2$ measures feature magnitude and similarity. Consequently, the score I_m quantifies the diagnostic contribution of each modality toward disease prediction. Although modality-level interpretation gives broad explanations, local anatomical and functional structures responsible for prediction must be identified. To estimate discriminative feature saliency distributions, gradient-weighted activation analysis is used. This procedure emphasizes disease-related areas and helps physicians see prediction rationale. Explainability boosts credibility and clinical validity of the ATSMAN system..

$$H_{sal}(x,y,z) = ReLU\left(\sum_{k=1}^{K}\frac{1}{Z_k}\sum_{u,v,w}\frac{\partial S_{conf}}{\partial A_{uvw}^{(k)}}A_{xyz}^{(k)}\right) + \lambda_H\frac{\|A^{(k)}\|_F}{1+\sum_{k=1}^{K}\|A^{(k)}\|_F} \quad (25)$$

As found in equation (25), where $H_{sal}(x,y,z)$ denotes the saliency heat map corresponding to spatial location (z). The quantity $A_{uvw}^{(k)}$ represents the activation of the k^{th} feature map, while S_{conf} denotes the disease confidence score generated by the prediction module. The parameter Z_k corresponds to the normalization coefficient associated with the k^{th} feature channel. The ReLU operator suppresses negative responses and preserves informative activations. The coefficient λ_H regulates feature-map regularization, and $\|\cdot\|_F$ denotes the Frobenius norm. The approach produces an explanation confidence score by combining prediction certainty, modality significance, and saliency consistency to assess explanation dependability. This framework assesses produced explanations' credibility and helps doctors evaluate automated judgments. Thus, the explanation confidence metric aids clinical interpretation and decision-making.

$$E_{conf} = \mu_1 S_{conf} + \mu_2\left(\frac{1}{M}\sum_{m=1}^{M} I_m\right) + \mu_3\left(\frac{-\sum_{k=1}^{K} p_k log\ p_k}{loglog\ K}\right) + \mu_4\frac{\|H_{sal}\|_F}{\|H_{sal}\|_F+\|F_{Fusion}\|_2} \quad (26)$$

In equation (26), where E_{conf} denotes the overall explanation confidence score. The variable S_{conf} represents the prediction confidence obtained from the disease prediction module, while I_m denotes modality-specific importance scores. The probability p_k corresponds to normalized saliency responses associated with the k^{th} feature channel, and K represents the total number of channels. The coefficients μ_1, μ_2, μ_3, and μ_4 control the relative contributions of prediction confidence, modality importance, entropy reduction, and saliency consistency, respectively. The norms $\| H_{sal} \|_F$ and $\| F_{Fusion} \|_2$ measure the magnitudes of saliency maps and fused feature representations. Therefore, E_{conf} provides a quantitative measure of explanation reliability and enhances the interpretability of the proposed ATSMAN framework.

3.9 Algorithmic Workflow of ATSMAN

The overall workflow of the proposed ATSMAN framework begins with the acquisition of MRI, fMRI, and EEG data from each participant. After preprocessing, modality-specific feature extraction generates structural, functional, and temporal representations. Later, adaptive attention systems judge modality relevance and merge retrieved information into a single representation. Fused representation is analyzed by disease prediction network to produce diagnostic results. Finally, explainability modules illuminate decision-making. Comprehensive multimodal integration, adaptive biomarker selection, better sensitivity to early illness changes, and improved interpretability are among the suggested methods' benefits. ATSMAN is a promising platform for early brain illness prediction and precision neurology due to these properties.

Algorithm 1: Pseudocode of Adaptive Temporal-Spatial Multimodal Attention Network (ATSMAN)

Input: MRI data X_{MRI}, fMRI data X_{fMRI}, EEG signals X_{EEG}, maximum epochs $MaxEpoch$
Output: Disease label $\hat{y}$, probability $P(y)$, confidence score S_{conf}

1. Initialize network parameters θ, attention weights α_{MRI}, β_{fMRI}, and γ_{EEG}.
2. WHILE $Epoch \leq MaxEpoch$DO
3. FOR each mini-batch BDO
4. Preprocess X_{MRI}, X_{fMRI}, and X_{EEG}.
5. IF MRI modality exists THEN
6. Extract structural features F_{MRI}using 3D-CNN.
7. END IF
8. IF fMRI modality exists THEN
9. Construct connectivity graph G.
10. Obtain functional features F_{fMRI}using GAT.

11. END IF
12. IF EEG modality exists THEN
13. Generate temporal features F_{EEG}using BiLSTM-Transformer.
14. END IF
15. Compute adaptive attention coefficients α_{MRI}, β_{fMRI}, and γ_{EEG}.
16. Fuse multimodal features to obtain F_{Fusion}.
17. Predict probability vector $P(y)$using Softmax.
18. Assign disease label $\hat{y} = argmax(P(y))$.
19. IF Training = TRUE THEN
20. Compute weighted cross-entropy loss L_{WCE}.
21. Compute regularization loss L_{Reg}.
22. Evaluate total loss L_{Total}.
23. Update parameters θusing Adam optimizer.
24. END IF
25. END FOR
26. IF convergence criterion is satisfied THEN
27. BREAK
28. END IF
29. $Epoch \leftarrow Epoch + 1$
30. END WHILE
31. Compute confidence score S_{conf}.
32. Return $\hat{y}$, $P(y)$, and S_{conf}.

Algorithm 1 shows the Pseudocode of ATSMAN. For each iteration, MRI, fMRI, and EEG data are pre-processed to remove noise and maintain uniformity. A 3D-CNN recovers anatomical features from structural MRI data, whereas a GAT processes functional connection graphs from fMRI signals to produce functional representations. Furthermore, used a BiLSTM-Transformer architecture to analyze EEG data for temporal features. After that, adaptive attention coefficients are calculated to identify each modality's relevance, and the retrieved features are combined into a multimodal representation. Fused characteristics are processed via the prediction module to estimate illness probability and apply diagnostic labels. Weighted cross-entropy and regularization losses are assessed and model parameters modified using the Adam optimizer during training. After convergence conditions are met, the framework outputs the predicted illness class, probability distribution, and confidence score, allowing reliable and interpretable early brain disease prediction.

4. RESULTS AND DISCUSSION

ATSMAN was tested using Kaggle's Simultaneous EEG-fMRI Dataset [21]. Dataset comes from an open-access neuroimaging archive, including data from 200 healthy young male individuals aged 26 ± 3.8 years. Both resting-state and task-based EEG, fMRI, and MRI data are included. EEG recordings, including gradient artifact-corrected versions, were taken inside and outside the MRI environment. The fMRI data includes BOLD functional volumes and structural T1-weighted MRI scans organized according to the Brain Imaging Data Structure (BIDS) standard. Multimodal fusion and attention-based learning are ideal for the 13 GB dataset, which includes complementary temporal and spatial information from EEG signals and MRI/fMRI images.

Table 1. Simulation Parameter

Parameter	Specification
Dataset	Simultaneous EEG-fMRI Dataset
Dataset name	Simultaneous EEG-fMRI Dataset
Total	200
Explainability	XAI
Optimizer	Adam

Simultaneous EEG-fMRI Dataset is complete available for all 200 subjects, allowing for multimodal brain analysis, as shown in table 1. The complementary brain content in this dataset enables analysis of the structural, functional and electrophysiological features. Explainable Artificial Intelligence (XAI) is integrated in order to enhance the interpretability of the model, by revealing which features have the biggest impact on the predictions. The Adam optimization algorithm is employed during model training, which simultaneously performs efficient stochastic gradient descent updates.

4.1 Performance Evaluation of the Proposed ATSMAN Framework

The effectiveness of the proposed ATSMAN was evaluated for the early prediction of brain diseases by integrating structural MRI, functional MRI, fMRI, and EEG data. The primary objective of the evaluation was to determine whether multimodal fusion could improve predictive performance compared with individual modalities and conventional fusion strategies. The experimental analysis demonstrated that ATSMAN successfully identified subtle neurological abnormalities associated with early disease stages. Anatomical, functional, and electrophysiological indicators allowed the model to discover complementary information that single-modality techniques would miss. According to the findings, the suggested framework was better at identifying healthy people from early pathological

alterations. This improvement can be attributed to the adaptive attention mechanism, which dynamically prioritizes the most informative modality based on disease progression.

The proposed ATSMAN system was evaluated using five performance metrics: BA, SP, MCC, κ, and JSI. Balanced Accuracy controls for class imbalance by balancing sensitivity and specificity, while Specificity reflects the framework's capacity to identify healthy people. True positives, true negatives, false positives, and false negatives make Matthews Correlation Coefficient a fair evaluation. Cohen's Kappa Coefficient measures the agreement between anticipated and real illness labels beyond chance, whereas the Jaccard Similarity Index measures predicted and ground-truth classifications.

4.2 Balanced Accuracy

Balanced Accuracy was used to quantify the proposed ATSMAN classification capacity under class-imbalanced neurological dataset settings. In contrast to majority-dominated accuracy metrics, Balanced Accuracy computes the arithmetic mean of sensitivity and specificity, giving sick and healthy patients equal weight. The combined MRI, fMRI, and EEG representations' capacity to decrease false positives and negatives is measured to assess their discriminative strength. ATSMAN's adaptive temporal-spatial attention mechanism stores complementary anatomical, functional, and electrophysiological information, improving class separability throughout illness stages. Balanced Accuracy shows the multimodal fusion strategy's tolerance to skewed data distributions and is a better indicator of generalization performance than conventional accuracy. Higher Balanced Accuracy values show that the proposed framework predicts consistently across all categories and captures subtle neurological abnormalities associated with early disease progression, validating the adaptive attention-based multimodal learning architecture. Figure 3 shows the balanced accuracy.

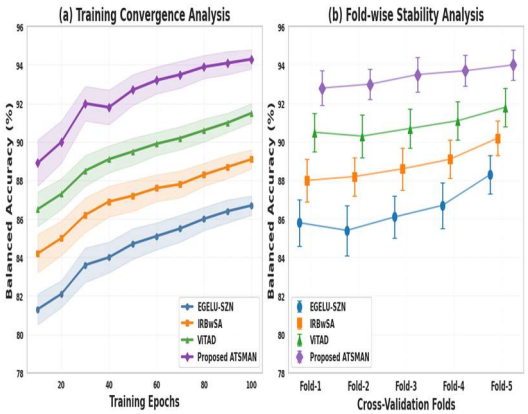

Figure 3: Balanced Accuracy

4.3 Specificity

Specificity was used to test the proposed ATSMAN framework's capacity to identify healthy individuals with few false positives. Too many false alarms in early brain illness detection might lead to needless clinical investigations and diagnostic ambiguity. Thus, specificity is a key sign of the model's capacity to discriminate normal neurological situations from pathologies. The system suppresses irrelevant variables and retains discriminative biomarkers by merging structural MRI, functional connectivity information from fMRI, and temporal EEG characteristics via adaptive attention-based fusion. Thus, greater specificity values suggest that ATSMAN minimizes false positives while preserving decision limits. This shows that the suggested multimodal representation preserves healthy brain patterns and improves framework diagnostic accuracy. Figure 4 shows the specificity ratio.

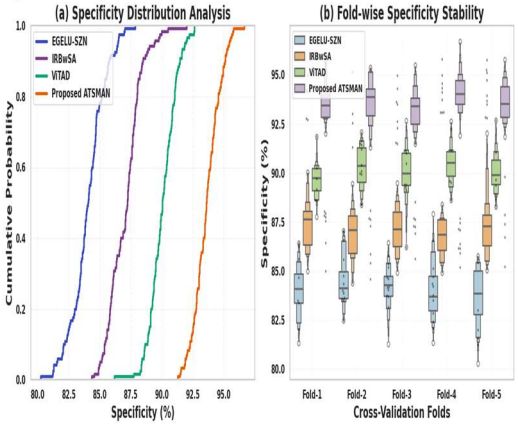

Figure 4: Specificity Ratio

4.4 Matthews Correlation Coefficient (MCC)

Matthews Correlation Coefficient estimated classification performance by

integrating true positives, true negatives, false positives, and false negatives. MCC assesses prediction consistency and correlation between actual and predicted labels more fully than conventional metrics, which may be affected by imbalanced datasets. The proposed ATSMAN system adaptively fuses MRI, fMRI, and EEG modalities to strengthen feature discrimination and remove classification ambiguity to improve sickness category correlation. MCC correctly measures model stability under varied data distributions and learned multimodal representation robustness. The suggested framework shows higher predictive consistency and balanced classification performance across all sickness classes with high MCC values, confirming the temporal-spatial attention mechanism. Figure 5 shows the Matthews Correlation Coefficient.

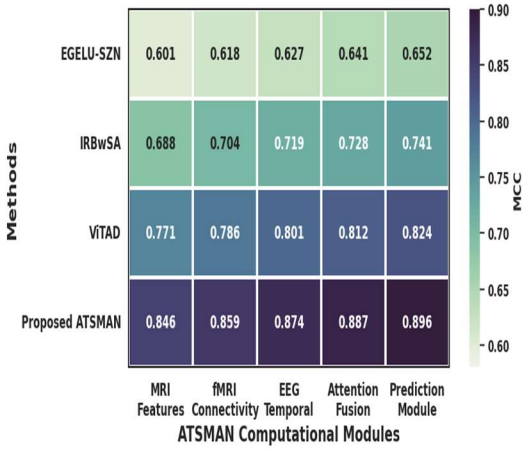

Figure 5: Matthews Correlation Coefficient

4.5 Cohen's Kappa Coefficient (κ)

While accounting for chance, Cohen's Kappa coefficient measured ATSMAN's predicted labels' agreement with ground-truth sickness categories. Kappa statistics measure classification reliability better than accuracy measures because neurological datasets have overlapping properties and unequal class distributions. The proposed adaptive attention mechanism effectively combines anatomical, functional, and electrophysiological information to improve inter-class discrimination and prediction confidence. Model predictions and sickness classifications match higher Kappa values. This suggests that the proposed framework improves diagnostic consistency and decision-making, making it perfect for early brain ailment detection clinical applications. Figure 6 shows the Cohen's Kappa Coefficient.

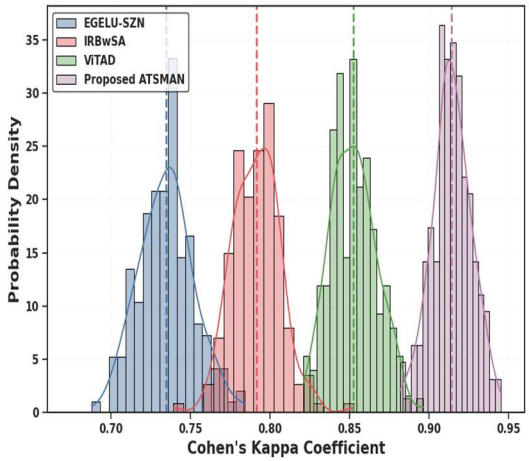

Figure 6: Cohen's Kappa Coefficient

4.6 Jaccard Similarity Index (JSI)

Jaccard Similarity Index was used to compare predicted disease categories to ground-truth labels. This metric directly evaluates classification similarity and is useful for medical diagnostics, multimodal feature representations. In the proposed ATSMAN system, adaptive temporal-spatial fusion merges complementary MRI, fMRI, and EEG data to improve class separability and prediction differences. The Jaccard index assesses how well projected labels match sickness categories in learned multimodal embeddings. As Jaccard similarity values increase, prediction results and reference labels match more, indicating that the proposed architecture captures subtle neurological features and classifies illnesses. ATSMAN's enhanced overlap suggests attention-guided fusion may preserve clinically relevant biomarkers and improve diagnostics. Figure 7 shows the Jaccard Similarity Index.

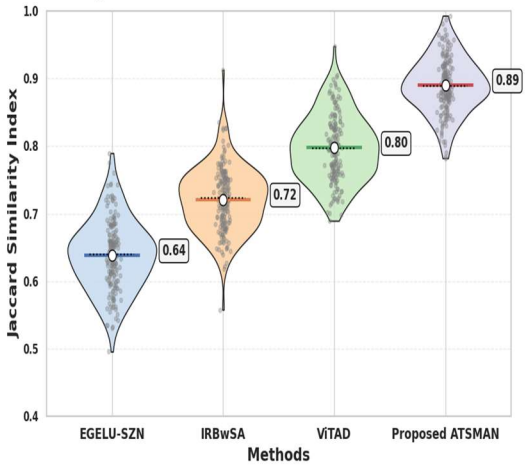

Figure 7: Jaccard Similarity Index

4.7 Discussion

This research shows that MRI, fMRI, and EEG biomarkers improve early brain illness prediction. The proposed ATSMAN system uses adaptive attention processes to integrate structural, functional, and electrophysiological data to characterize illness development. Experimental findings show that disease-related aberrations occur progressively across modalities. EEG abnormalities precede large-scale structural deterioration, but fMRI functional connectivity disturbances indicate disease that is emerging. Later structural MRI shows neurodegeneration. Neurological illnesses are complicated and need extensive examination across biological areas. ATSMAN's adaptive fusion technique dynamically adjusts modality contributions by illness stage to handle this complexity. Overall, the ATSMAN framework is a promising intelligent multimodal system for early brain illness prediction. ATSMAN may improve diagnosis accuracy, intervention time, and precision neurology by combining complementary neuroimaging modalities into an interpretable and adaptable framework.

4.8 Difference from Prior Work, Strengths and Limitations

Comparing to the latest researches, ATSMAN combines MRI, fMRI, and EEG in one approach. For MRI, it employs 3D-CNN, while it employs GAT for the connectivity of fMRI brain function. EEG is processed through BiLSTM-Transformer for capturing temporal details. The process of adaptive attention acts as a system to evaluate importance of each modality based on developments of the system. Meanwhile, XAI allows to gather interpretable data from three modalities.

The key advantages of ATSMAN are that it employs multimodal learning, enables adaptive fusion, and interpretability. The frame work fuses structural, functional, and temporal information by integrating MRI, fMRI and EEG. The obtained results of 94.3% Balanced Accuracy, 93.6% Specificity, 0.896 MCC, and 89.0% Jaccard Similarity Index signify the efficacy of the proposed method in the reported evaluation.

Nevertheless, ATSMAN has some constraints. The dataset draws on limited population, which may limit generalizability to other clinical populations. The protocol demands all the three modalities, which may be not always achievable in practice. In addition, integrating multiple DL and XAI modules enhance computational complexity. Future works should consider larger multi-center datasets as well as missing-modality handling and lighweight models.

5. CONCLUSION AND FUTURE DIRECTIONS

This study aimed to address the main issue of insufficient capacity of current brain disease prediction methods to simultaneously capture complementary structural, functional, and electrophysiological abnormalities at early stages. Conventional multimodal approaches may be static fusion and lack interpretability, while existing single-modality approaches cannot fully utilize the information from MRI, fMRI, and EEG. These methods are unable to fully encompass complementary structural, functional, and temporal information. Existing multimodal methods also often use static fusion without dynamically adjusting the importance of each modality. To tackle these limitations, the proposed ATSMAN is designed as a three-part model comprising three deep neural networks: 3D-CNN, GAT, and BiLSTM-Transformer, for MRI, fMRI, and EEG, respectively. A temporal-spatial attention mechanism, which dynamically selects the modality importance for fusion, is utilized. Furthermore, XAI is capable of highlighting MRI regions, fMRI connections, and EEG patterns. Hence, ATSMAN is a framework for multimodal learning, adaptive fusion and interpretable prediction. The major advantage of ATSMAN is the combination of MRI, fMRI and EEG, which gives complementary information at the structural, functional and temporal levels. The adaptive temporal-spatial attention mechanism dynamically allocates attention across each modality, so that the model can attend to the most informative features. Moreover, XAI can also facilitate interpretability by highlighting relevant MRI regions, fMRI connections, and EEG patterns. The experimental results of ATSMAN in multimodal brain disease prediction showed 94.3% Balanced Accuracy, 93.6% Specificity, 0.896 MCC, and 89.0% Jaccard Similarity Index, which proved that the ATSMAN is effective. ATSMAN, although promising, has some drawbacks. First, the population of the dataset is only 200 healthy young males, reducing the diversity and generalizability of the population. Second, the framework relies on MRI, fMRI and

EEG which are not always available in a clinical setting. Third, the model is relatively complex and computationally expensive from the combination of 3D-CNN, GAT, BiLSTM-Transformer, attention and XAI. For future research, more large-scale multi-center datasets and missing-modality handling as well as lightweight architectures are desirable for more robust, efficient and clinically applicable systems.

REFERENCES

[1] Zhou, Z., Wang, Q., An, X., Chen, S., Sun, Y., Wang, G., & Yan, G. "A novel graph neural network method for Alzheimer's disease classification." Computers in Biology and Medicine, 180, 108869, 2024. DOI: 10.1016/j.compbiomed.2024.108869.

[2] An, X., Zhou, Y., Di, Y., Han, Y., & Ming, D. "A Novel Method to Identify Mild Cognitive Impairment Using Dynamic Spatio-Temporal Graph Neural Network." IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 3328–3337, 2024. DOI: 10.1109/TNSRE.2024.3450443.

[3] Saoud, L. S., & AlMarzouqi, H. "Explainable early detection of Alzheimer's disease using ROIs and an ensemble of 138 3D vision transformers." Scientific Reports, 14, 27756, 2024. DOI: 10.1038/s41598-024-76313-0.

[4] Li, Y., Chen, G., Wang, G., Zhou, Z., An, S., Dai, S., Jin, Y., Zhang, C., Zhang, M., & Yu, F. "Dominating Alzheimer's disease diagnosis with deep learning on sMRI and DTI-MD." Frontiers in Neurology, 15, 1444795, 2024. DOI: 10.3389/fneur.2024.1444795.

[5] Alharbi, H., Juanatas, R. A., Al Hejaili, A., & Lim, S.-J. "Spectral graph convolutional neural network for Alzheimer's disease diagnosis and multi-disease categorization from functional brain changes in magnetic resonance images." Frontiers in Neuroinformatics, 18, 1495571, 2024. DOI: 10.3389/fninf.2024.1495571.

[6] Bapat, R., Ma, D., & Duong, T. Q. "Predicting Four-Year's Alzheimer's Disease Onset Using Longitudinal Neurocognitive Tests and MRI Data Using Explainable Deep Convolutional Neural Networks." Journal of Alzheimer's Disease, 97(1), 459–469, 2024. DOI: 10.3233/JAD-230893.

[7] Liu, H., Weakley, A. M., Zhang, J., & Liu, X. "A Transformer Approach for Cognitive Impairment Classification and Prediction." Alzheimer Disease & Associated Disorders, 38(2), 189–194, 2024. DOI: 10.1097/WAD.0000000000000619.

[8] Abdelaziz, M., Wang, T., Anwaar, W., & Elazab, A. "Multi-scale multimodal deep learning framework for Alzheimer's disease diagnosis." Computers in Biology and Medicine, 184, 109438, 2025. DOI: 10.1016/j.compbiomed.2024.109438.

[9] Dardouri, S. "An efficient method for early Alzheimer's disease detection based on MRI images using deep convolutional neural networks." Frontiers in Artificial Intelligence, 8, 1563016, 2025. DOI: 10.3389/frai.2025.1563016.

[10] Liu, J., Wu, S., Fu, Q., Luo, X., Luo, Y., Qin, S., Huang, Y., & Chen, Z. "Multimodal diagnosis of Alzheimer's disease based on resting-state electroencephalography and structural magnetic resonance imaging." Frontiers in Physiology, 16, 1515881, 2025. DOI: 10.3389/fphys.2025.1515881.

[11] Alruily, M., Abd El-Aziz, A. A., Mostafa, A. M., Ezz, M., Mostafa, E., Alsayat, A., & Abd El-Ghany, S. "Ensemble deep learning for Alzheimer's disease diagnosis using MRI: Integrating features from VGG16, MobileNet, and InceptionResNetV2 models." PLOS ONE, 20(4), e0318620, 2025. DOI: 10.1371/journal.pone.0318620.

[12] Jeyapaul, S. R., et al. "Ensemble deep learning model for early diagnosis and classification of Alzheimer's disease using MRI scans." Journal of Alzheimer's Disease, 107(1), 57–74, 2025. DOI: 10.1177/13872877251359950.

[13] Arunnehru, J., Khan, S., Basha, H. A., Albarrak, A. M., & Ali, A. "An MRI based histogram-oriented gradient and deep learning approach for accurate classification of mild cognitive impairment and Alzheimer's disease." Frontiers in Medicine, 12, 1529761, 2025. DOI: 10.3389/fmed.2025.1529761.

[14] Peng, W., Li, C., Ma, Y., Dai, W., Fu, D., Liu, L., Liu, L., Yu, N., & Liu, J. "Integrating Time and Frequency Domain Features of fMRI Time Series for Alzheimer's Disease Classification Using Graph Neural Networks." Interdisciplinary Sciences: Computational Life Sciences, 2025. DOI: 10.1007/s12539-025-00759-7.

[15] Zhang, S., Wang, Q., Wei, M., Zhong, J., Zhang, Y., Song, Z., Li, C., Zhang, X., Han, Y., Li, Y., Lv, H., & Jiang, J. "FROG: A Fine-Grained Spatiotemporal Graph Neural Network With Self-Supervised Guidance for Early Diagnosis of Alzheimer's Disease." IEEE Journal of Biomedical and Health Informatics, 2025. DOI: 10.1109/JBHI.2025.3552638.

[16] Abadal, S., et al. "Graph neural networks for electroencephalogram analysis: Alzheimer's disease and epilepsy use cases." Neural Networks, 2025, 106792. DOI: 10.1016/j.neunet.2024.106792.

[17] Khan, W., Khan, M. S., Qasem, S. N., et al. "An explainable and efficient deep learning framework for EEG-based diagnosis of Alzheimer's disease and frontotemporal dementia." Frontiers in Medicine, 12, 1590201, 2025. DOI: 10.3389/fmed.2025.1590201.

[18] Zhang, R., Sheng, J., Zhang, Q., Wang, J., & Wang, B. "A review of multimodal fusion-based deep learning for Alzheimer's disease." Neuroscience, 576, 80–95, 2025. DOI: 10.1016/j.neuroscience.2025.04.035.

[19] White, T., & Rowlands, S. "Cogninet: an explainable deep learning model for multi-class MRI-based Alzheimer's disease staging." International Journal of Computer Assisted Radiology and Surgery, 2026. DOI: 10.1007/s11548-026-03617-z.

[20] Wu, F., Ai, F., Zhao, X., Wang, M., Yang, J., Hao, T., et al. "A multi-modal agent attention model for Alzheimer's disease diagnosis with structural MRI and clinical data." Scientific Reports, 2026. DOI: 10.1038/s41598-026-63773-9.

[21] Tigga, N. P., Garg, S., Goyal, N., Raj, J., & Das, B. (2025). Brain-region specific autism prediction from electroencephalogram signals using graph convolution neural network. *Technology and Health Care*, *33*(1), 77-101.

[22] Muksimova, S., Umirzakova, S., Iskhakova, N., Khaitov, A., & Im Cho, Y. (2025). Advanced convolutional neural network with attention mechanism for Alzheimer's disease classification using MRI. *Computers in Biology and Medicine*, *190*, 110095.

[23] B. D. Shivahare, H. Rajadurai, D. K. Deeba, S. Kansal, S. K. Mathivanan, et al., "An attention-augmented multimodal classification of Alzheimer's disease and Parkinson's disease vs healthy controls using MRI, EEG, and SNP data," Scientific Reports, vol. 16, Art. no. 2422, 2026. DOI: 10.1038/s41598-025-32274-6.

[24] Dardouri, S. (2025). An efficient method for early Alzheimer's disease detection based on MRI images using deep convolutional neural networks. *Frontiers in Artificial Intelligence*, *8*, 1563016.

[25] Gautum, R., Tanwar, S., & Singh, R. (2025). Explainable AI for Early Detection of Neurological Disorders Using Multimodal Biomedical Data. Procedia Computer Science, 283, 5221-5232. https://doi.org/10.1016/j.procs.2026.06.574.

[26] Ibrar, W., Khan, M. A., Hamza, A., Rubab, S., Alqahtani, O., Alouane, M. T. H., ... & Nam, Y. (2025). A novel interpreted deep network for Alzheimer's disease prediction based on inverted self attention and vision transformer. *Scientific Reports*, *15*(1), 29974.

[27] SAMBANDAM, T., Vivekanandam, B., & kumar Tripathi, R. (2025). Scalable Parkinson's Disease Prediction Using Mathematical Modeling, Hilbert Transforms, and Transformer-Based Deep Learning. *SGS-Engineering & Sciences*, *1*(4).

[28] Mehta, V., Maram, B., Garg, J. et al. A multimodal explainable artificial intelligence framework for interpretable Parkinson's disease prediction. Sci Rep 16, 18143 (2026). https://doi.org/10.1038/s41598-026-47769-z.

[29] Sankari, C., Jamuna, V., & Kavitha, A. R. (2025). Hierarchical multi-scale vision transformer model for accurate detection and classification of brain tumors in MRI-based medical imaging. *Scientific Reports*, *15*(1), 38275.

[30] Joy, M. A. M., Nasrin, S., Siddiqua, A., & Farid, D. M. (2025). ViTAD: Leveraging modified vision transformer for Alzheimer's disease multi-stage classification from brain MRI scans. *Brain Research*, *1847*, 149302. https://www.kaggle.com/datasets/showrinrahman/simultaneous-eeg-fmri-dataset