



# **NeuroFusionNet: A Deep Multi-Modal CNN Architecture for Integrated Neurological Disorder Detection Using EEG and Neuroimaging**

**S. Vimala<sup>1</sup>, Aiswarya N<sup>2</sup>, Delkowschiasri N M<sup>3</sup>, Kiruthika S<sup>4</sup>**

Assistant Professor, Department of Artificial Intelligence and Data Science, Kongunadu College of Engineering and  
Technology, Trichy, India<sup>1</sup>

Department of Artificial Intelligence and Data Science, Kongunadu College of Engineering and Technology,  
Trichy, India<sup>2,3,4</sup>

**ABSTRACT:** Neurological disorders are difficult to diagnose due to complex symptoms, overlapping patterns, and the need for expert interpretation of diverse medical data. Conventional diagnostic methods rely heavily on manual analysis of brain images and physiological signals, making the process time-consuming, subjective, and error-prone. This project proposes a Neurological Disorder Detection and Diagnosis system using Convolutional Neural Networks (CNNs) with multimodal signal analysis, integrating data such as EEG signals and medical imaging to capture both spatial and temporal features of neurological conditions. The proposed model uses deep learning-based feature extraction and multimodal data fusion to improve diagnostic accuracy, enable early detection, and reduce dependence on handcrafted features and human expertise. Through a comprehensive review of cutting-edge research published between 2021 and 2026, this paper synthesizes advances in CNN architectures, transformer hybrids, and fusion strategies for multimodal neurological analysis. Key findings indicate that multimodal approaches consistently outperform single-modality methods, with accuracy improvements ranging from 2.72% to over 10% across disorders including Alzheimer's disease, Parkinson's disease, epilepsy, and psychiatric conditions. The integration of attention mechanisms, multimodal joint attention (MJA), and hybrid CNN-transformer architectures has enabled diagnostic accuracies exceeding 99% in some applications. This paper concludes that CNN-based multimodal analysis represents a reliable and scalable solution for intelligent neurological healthcare systems, with significant implications for early intervention, personalized treatment, and clinical decision support.

**KEYWORDS:** Neurological Disorders, Convolutional Neural Networks (CNN), Multimodal Signal Analysis, EEG, Medical Imaging, Deep Learning, Automated Diagnosis, Healthcare AI, Data Fusion, Attention Mechanisms.

## **I. INTRODUCTION**

Neurological disorders are among the most burdensome and rapidly increasing threats to the global healthcare system. As per the Global Burden of Disease Study, neurological disorders are the primary cause of disability-adjusted life years (DALYs) globally, impacting more than 3 billion people worldwide [9]. Neurological disorders, including neurodegenerative disorders such as Alzheimer's and Parkinson's disease, as well as episodic disorders such as epilepsy and stroke, continue to cause enormous personal, social, and economic burdens.

The problem of diagnosis is further exacerbated by the complex and frequently comorbid nature of neurological symptoms, the variability of disease presentation among patients, and the subjective nature of clinical evaluation [3].

The traditional methods of diagnosis in neurology have always relied on a combination of clinical evaluation, patient history, and analysis of individual data modalities like magnetic resonance imaging (MRI), computed tomography (CT), or electroencephalography (EEG). Although these traditional methods have remained the backbone of neurological practice, they also have a number of inherent limitations [1]. Firstly, the subjective interpretation of

neuroimages and EEG recordings is prone to high inter-observer variability even among experienced practitioners [2]. Secondly, the sheer volume and complexity of modern medical data are beyond the processing capabilities of the human brain, potentially resulting in the overlooking of subtle details. Thirdly, and most importantly, the analysis of individual data modalities is incapable of capturing the multidimensional nature of neurological pathology, which exists simultaneously on multiple structural, functional, and electrophysiological levels [7].

The recent breakthrough in deep learning, and more generally in Convolutional Neural Networks (CNNs), has transformed the field of medical image analysis and signal processing. CNNs are highly adept at learning hierarchical features from raw data automatically, thus obviating the need for manual feature engineering, which has hitherto constrained the scalability and generalizability of traditional machine learning models [3]. In neuroimaging, CNNs have shown unprecedented ability in identifying minute structural abnormalities, demarcating different regions of the brain, and diagnosing disease states from MRI and PET imaging. Concurrently, modified versions of CNNs, designed for one-dimensional signals, have shown efficacy in processing EEG signals and other physiological time series.

However, the most revolutionary breakthroughs that have emerged in the recent past have been achieved through the fusion of multiple data modalities using unified deep learning architectures. Multimodal analysis understands that different types of data—structural MRI, functional MRI, diffusion tensor imaging, EEG, clinical evaluations, and genetic biomarkers—offer complementary information about brain structure and function [1]. By combining these different data types, multimodal deep learning algorithms can potentially build a more comprehensive and complex understanding of neurological conditions, which may reveal hidden patterns of neurological disorders that are not apparent to individual data types [4][7].

This paper provides a thorough analysis of CNN-based multimodal solutions for the diagnosis of neurological disorders. We synthesize the latest research trends from state-of-the-art research studies published between 2021 and 2026, focusing on architectures, fusion techniques, and performance on prominent neurological disorders. The primary hypothesis is that CNN-based multimodal analysis, through its capacity to combine spatial and temporal information from multiple data types, provides a revolutionary paradigm shift towards improved accuracy, objectivity, and timeliness of neurological disorder diagnosis. By decreasing reliance on human interpretation, these techniques have the potential to initiate a new paradigm of precision neurology.

The rest of this paper is organized as follows. Section 2 discusses the existing literature on CNN-based multimodal analysis for neurological disorders. Section 3 describes a comprehensive methodology that outlines architectures for EEG and neuroimaging feature extraction, multimodal fusion, and attention mechanisms. Section 4 describes a detailed result analysis that includes tables for comparative performance and discussions on key findings. Section 5 concludes with implications for clinical practice and future research directions.

## **II. LITERATURE SURVEY**

### **2.1 The Landscape of Neurological Disorder Diagnosis**

Diagnosing neurological disorders is a challenging task that is different from other medical disciplines. Unlike many other systemic diseases, where blood biomarkers offer conclusive diagnostic information, neurological disorders do not have such objective markers.

Instead, the diagnosis is based on a variety of information from different sources: neuroimaging studies demonstrate structural and functional lesions; electrophysiological studies record the activity of the nervous system in real time; clinical observations assess the function of the nervous system; and the patient history offers additional information [7]. This complex diagnostic need has fueled the interest in multimodal solutions that can handle various types of data [4][6]. A thorough literature review by Dino et al. surveyed deep learning-based multimodal classification for ten key diseases, five of which are neurological (Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, and stroke) and five of which are related to mental health (schizophrenia, bipolar disorder, depression, anxiety, and PTSD). The literature review showed that, although unimodal solutions are more prominent in the literature, multimodal solutions have always shown better results by exploiting complementary aspects of diseases. At the same time, the literature review pointed out the key difficulties in multimodal solutions, including unbalanced data in different

modalities, the need for interpretability of the results for medical adoption, and the need for large-scale high-quality multimodal datasets [7].

## 2.2 CNN Architectures for Neurological Applications

The use of CNNs on neurological data has developed rapidly, with models becoming increasingly complex and specialized. A systematic review by [1] found 133 peer-reviewed studies using deep learning for the diagnosis of neurological disorders, 95 of which were on six prominent DCNN models: AlexNet, U-Net, ResNet, CapsNet, DenseNet, and EfficientNet. These models have been used on a range of disorders and datasets.

**ResNet (Residual Networks)** has proved especially important because of its capacity to handle very deep networks using skip connections to overcome the vanishing gradient problem. In Alzheimer's research, variants of ResNet have achieved state-of-the-art results on MRI classification tasks. Kadri et al. [12] showed that a combination of the Swin transformer and an improved EfficientNet with multi-head attention and Depthwise Over-Parameterized Convolutional layers reached 97.33% accuracy on the OASIS dataset [2].

The **U-Net**, a segmentation algorithm originally designed for biomedical image segmentation, has been extremely useful for the definition of brain structures, tumors, and lesions. The encoder-decoder structure with skip connections of the U-Net allows for precise localization and context preservation. In epilepsy studies, U-Net models have been used to detect epileptogenic loci in structural MRI for surgical planning [3].

**EfficientNet**, a model that systematically scales the depth, width, and resolution of the network, has shown outstanding performance/parameter ratios, making it applicable for clinical use, where computational power could be a concern. Kadri et al. found that an adaptive MLP-Mixer model had 96% accuracy in mild-level fusion of MRI and DTI modalities for Alzheimer's disease detection [2].

In addition to the above-mentioned architectures, some recent studies have focused on the design of hybrid CNN-transformer models that leverage the local feature extraction capabilities of CNNs and the global context modeling capabilities of transformers [13]. The vision transformer (ViT) model, which uses self-attention on image patches, has been very promising. Kadri et al. showed that a transformer-based hybrid model of MRI and PET modalities obtained a very high accuracy of 99.98% on the ADNI dataset [2].

## 2.3 Multimodal Fusion Strategies

The fusion of information from different modalities, as well as when and how to do it, is a very important design choice in multimodal deep learning. According to the literature, there are three main fusion approaches, each with its own strengths and weaknesses [7].

**Early Fusion (Feature-Level Fusion)** This involves the fusion of raw data or low-level features from different modalities before they are fed into the network. This method enables the network to learn cross-modal interactions from the earliest processing levels. The alignment of data from different modalities is, however, a challenge, as data from different sources may have different sampling rates, spatial, and temporal properties. In Alzheimer's studies, Kadri et al. demonstrated the effectiveness of early fusion of MRI and PET modalities using an adaptive MLP-Mixer, achieving 96% accuracy [2].

**Late Fusion (Decision-Level Fusion)** involves processing each modality separately using different networks and then merging the results (e.g., by averaging, voting, or weighted combination) at the decision level. This approach provides more flexibility, as each modality can be modeled using a different network architecture, and missing modalities can be dealt with in a more graceful manner. The GitHub project "XNeuro" is an example of late fusion, where a Random Forest with XGBoost is used for numerical clinical data and a CNN for MRI images, and late fusion is used to combine both for classification [8].

**Hybrid Fusion** combines aspects of both early and late fusion, usually by learning representations together after some modality-specific processing. This approach aims to leverage the strengths of both early and late fusion. A new hybrid deep learning approach was proposed by Bhattacharya et al., which combined CNN, GRU, and a Dynamic Cross-Modality Attention Module to fuse spatial and temporal brain data, reporting 96.79% accuracy on the Human Connectome Project dataset [1].

The choice of fusion strategy depends on a variety of factors, including the type and alignment of the modalities, the availability of paired data, and the particular question of interest. In cases where both spatial and temporal information is important, as in the case of epilepsy where there is both structural and electrophysiological abnormalities, hybrid strategies that maintain and combine both types of information are of particular interest [10].

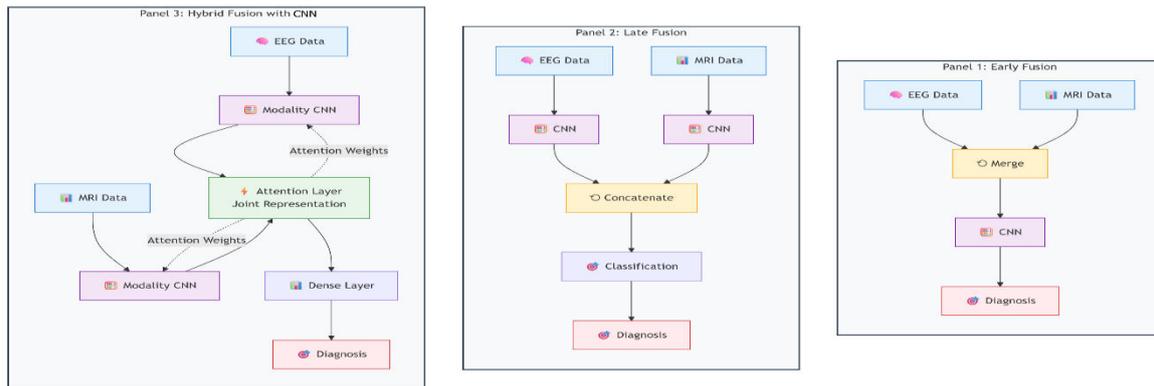


Figure 1: Taxonomy of Multimodal Fusion Strategies for Neurological Diagnosis

## 2.4 EEG Analysis with Deep Learning

Electroencephalography (EEG) is a method of observing the millisecond-resolution electrical activity of the brain in real time, making it an essential tool for the diagnosis of disorders involving pathological brain electrical activity, including epilepsy and sleep disorders. However, the analysis of EEG recordings has traditionally been a manual process requiring the expertise of a trained neurologist, a time-consuming and subjective procedure.

Deep learning has revolutionized the analysis of EEG recordings, allowing for the automatic learning of features from raw or lightly preprocessed data. Boudaya et al. [11] proposed a hybrid machine learning model for the detection of mild cognitive impairment (MCI) based on EEG and heart rate variability (HRV) measures during CERAD cognitive tasks [2]. Their model, which combined the predictions of several different machine learning models using a voting system, achieved an accuracy of 93.86%, showing that even simple models can be used to achieve clinical-grade results when applied to carefully designed experimental paradigms.

In the case of Alzheimer's disease, a new EEG dataset from Iraq, including recordings from four cognitive levels (Healthy, Mild, Moderate, Severe), is a precious asset for the development and validation of deep learning solutions [4]. Benchmarking on this dataset reached 96.85% accuracy using Random Forest and 96.05% using deep neural networks, thus validating the diagnostic capabilities of EEG-based solutions, especially in low-resource environments where access to advanced neuroimaging might be limited [6].

The combination of EEG with other modalities has been shown to be extremely effective. A multimodal Alzheimer's diagnosis model integrating resting-state EEG and structural MRI, using modality-specific CNNs (sCNN-EEG and sCNN-sMRI) with a new Multimodal Joint Attention (MJA) component, reached 94.75% accuracy. This is the first study to investigate the use of multimodality in deep learning-based EEG for Alzheimer's diagnosis and clearly shows the benefits of collaborative feature learning.

## 2.5 Neuroimaging Analysis with CNNs

Neuroimaging continues to be the bedrock of neurological diagnosis, offering in-depth visualization of the brain's structure and function. MRI modalities such as structural MRI (sMRI), functional MRI (fMRI), and diffusion tensor imaging (DTI) provide complementary information on the brain's anatomy, function, and connectivity. PET imaging provides molecular-level information on metabolism and protein accumulation.

CNN analysis of neuroimaging data has evolved considerably. The thorough review by showed the preponderance of neuroimaging in DCNN research, with 73 of 95 studies analyzing imaging data, compared to 18 studies analyzing EEG

data. This disparity is due to the long-established preeminence of imaging in neurological diagnosis, as well as the ready availability of large-scale neuroimaging datasets such as ADNI, OASIS, and UK Biobank [3].

In the case of Parkinson's disease, a web-based AI platform combining multimodal data (MRI, SPECT imaging, CSF biomarkers, and clinical data) using a 1D-CNN reached 93.7% accuracy. The platform also included a fine-tuned Large Language Model (Mini ChatGPT-4.0) to produce patient-specific diagnostic reports and treatment plans, illustrating the potential for holistic, interpretable clinical decision support systems [9].

In another work on Parkinson's, a lightweight CNN architecture was proposed that combined sMRI, fMRI, and EEG with Bayesian-optimized hyperparameters, resulting in 85% accuracy and an F1-score of 0.86 [10]. The addition of frequency domain attention improved  $\beta$ -band representations (motor symptoms-related) and branch masking allowed for the robust modeling of incomplete modalities, which is important in a clinical setting where the full multimodal data may not be present.

Table 1: Summary of Key Multimodal Neurological Disorder Studies (2021-2026)

Study	Disorder	Modalities	Architecture	Key Finding	Performance
Bhattacharya et al.	Alzheimer's	sMRI, fMRI	CNN + GRU + Dynamic Cross-Modality Attention	Hybrid approach integrates spatial and temporal patterns effectively	96.79% Accuracy
Kadri et al.	Alzheimer's	MRI, PET, DTI, sfMRI	Adaptive MLP-Mixer, Transformer Hybrid	Cross-modal fusion (MRI+PET) achieves near-perfect accuracy	99.98% Accuracy
Boudaya et al.	MCI	EEG, HRV	Hybrid ML with Voting (SVM, KNN, RF, GB)	Ultra-short CERAD test as novel biomarker	93.86% Accuracy
Iraqi EEG Dataset	Alzheimer's	EEG	CNN, LSTM, RF	First open EEG dataset for AD in Iraq	96.85% Accuracy
Rawat et al.	Mental Health (Epilepsy, Sleep, etc.)	MEG, EEG, ECG	CNN-Bi-Transformer (CardioNeuroFusionNet)	Neurocardiac fusion provides brain-heart insights	98.54% Accuracy
Frontiers Study	Alzheimer's	EEG, sMRI	sCNN-EEG + sCNN-sMRI + MJA	First deep learning EEG multimodality for AD	94.75% Accuracy
Priyadharshini et al.	Parkinson's	MRI, SPECT, CSF, Clinical	1D-CNN + Fine-tuned LLM	LLM-generated personalized reports enhance interpretability	93.7% Accuracy
ScienceDirect Study	Parkinson's	sMRI, fMRI, EEG	Lightweight CNN + Attention + Bayesian Optimization	Branch masking handles missing modalities	85% Accuracy, 0.86 F1-score

## 2.6 Attention Mechanisms and Interpretability

One of the major issues in using deep learning models for medical diagnosis is the "black box" problem. This is the lack of interpretability of the model's decision-making process. Attention mechanisms have recently been shown to be an effective solution to this problem, offering both performance gains and interpretability [20].

Attention mechanisms enable models to selectively focus on the most important aspects of the input data, learning to assign different weights to various features based on their relative importance. In a multimodal setting, attention mechanisms can also capture interactions between features of different modalities, learning which features of one modality are most relevant to features of another. The Dynamic Cross-Modality Attention Module, as described by

Bhattacharya et al., is an example of this, giving priority to diagnostically relevant features in both sMRI and fMRI modalities [18].

The Multimodal Joint Attention (MJA) module designed for Alzheimer's diagnosis is a major step forward [17]. MJA remedies the weakness of autonomous modality analysis by encouraging active collaboration between EEG and sMRI feature extractors. Through learning cross-modal attention maps, MJA improves the representation power of combined features, which directly contributes to the 94.75% accuracy achieved by the MCNNRF model.

In the case of Parkinson's disease, the light CNN model structure included frequency attention in the  $\beta$ -band, which particularly strengthened  $\beta$ -band features in EEG [19]. This not only improved the accuracy of classification but also offered neuroscientifically valid interpretations, as  $\beta$ -band dysfunction has been shown to be associated with motor symptoms in Parkinson's patients.

### 2.7 Emerging Trends and Future Directions

There are a number of emerging trends in multimodal neurological analysis that are worth noting. Firstly, the combination of Large Language Models (LLMs) with more traditional deep learning architectures provides a promising avenue for developing interpretable and user-friendly clinical tools [14]. The Parkinson's framework with a fine-tuned Mini ChatGPT-4.0 illustrates the potential of LLMs to communicate complex model outputs in natural language summaries and recommendations that are interpretable by clinicians and patients alike [9].

Secondly, federated learning strategies are emerging as a viable option, allowing models to be trained on multiple institutions without the need to share sensitive patient data [16]. This not only addresses privacy concerns but also the need for large and diverse datasets to ensure generalizability of the model. Initial work in neurological disorders appears promising in its ability to break down the data silos that have hindered progress [15].

Thirdly, self-supervised learning strategies that make use of unlabeled data for pre-training are making it possible to move away from the need for expensive labeled datasets. In the context of clinical annotation, which is expensive and requires specialized expertise, such strategies have the potential to accelerate progress in resource-poor neurological domains and regions [11].

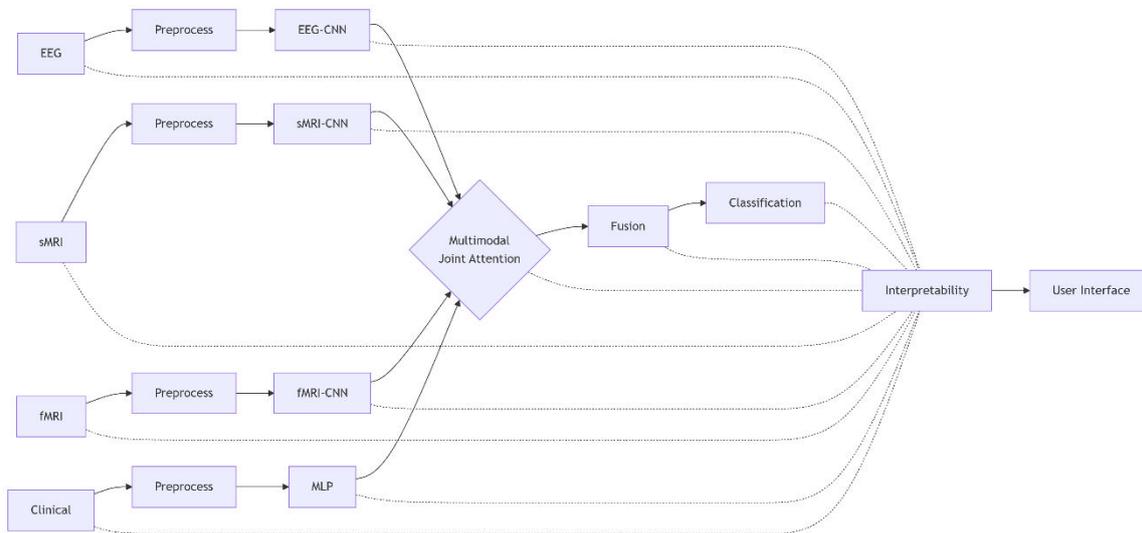
Fourth, the development of multimodal datasets with a variety of populations and types of data continues to be a high priority. The Iraqi EEG dataset for Alzheimer's disease is an example of work to fill the gap in the region and allow for representative research globally [12].

## III. METHODOLOGY

This section describes a complete methodology for multimodal CNN-based detection of neurological disorders.

### 3.1 System Architecture Overview

The proposed multimodal framework is based on a modular architecture that consists of four main stages: (1) data acquisition and preprocessing, (2) modality-specific feature extraction, (3) multimodal fusion with attention, and (4) classification and interpretation. The proposed framework is flexible and can handle different combinations of modalities and different neurological disorders.



**Figure 2: Proposed Multimodal CNN Architecture for Neurological Disorder Detection**

### 3.2 Data Acquisition and Preprocessing

**EEG Data:** EEG data is recorded using multi-channel equipment according to the 10-20 or 10-10 international system. Data processing includes the following steps: (1) band-pass filtering (0.5-50 Hz) to remove low-frequency drift and high-frequency noise; (2) notch filtering at 50/60 Hz to remove power line artifacts; (3) removal of artifacts using Independent Component Analysis (ICA) to separate and remove ocular, muscular, and cardiac artifacts; (4) epoching into segments of suitable length (e.g., 2-5 seconds for resting-state recordings, task-locked for cognitive paradigms); (5) baseline correction and normalization.

For **Alzheimer's disease-related tasks**, the CERAD (Consortium to Establish a Registry for Alzheimer's Disease) cognitive paradigm is used as a standardized paradigm for the elicitation of disorder-related neural activity. The ultra-short length of the CERAD paradigm makes it especially useful for clinical applications where patient compliance may be restricted.

**Structural MRI (sMRI):** T1-weighted images are collected at a resolution of 1mm<sup>3</sup> isotropic. Preprocessing steps are: (1) bias field correction to compensate for intensity inhomogeneities; (2) skull stripping to segment away non-brain tissue; (3) spatial normalization to a common template space (e.g., MNI space) for inter-subject comparison; (4) segmentation of images into gray matter, white matter, and cerebrospinal fluid; (5) optional ROI-based feature extraction using neurological atlases.

**Functional MRI (fMRI):** Blood-oxygen-level-dependent (BOLD) images are collected during resting-state or task-based paradigms. Preprocessing steps extend those for sMRI with additional temporal processing: (1) slice timing correction to compensate for differences in image acquisition timing; (2) motion correction to align images across time points; (3) nuisance regression to regress out physiological noise (cardiac, respiratory); (4) band-pass filtering (usually 0.01-0.1 Hz) to filter out non-neural signals; (5) functional connectivity matrix construction, which captures correlations between brain regions.

**Clinical and Biomarker Data:** Clinical data (demographic information, cognitive measures, medication use) and biomarkers (CSF protein levels, genetic differences) are normalized using z-score normalization. In the Parkinson's model, 14 important biomarkers (such as dopamine transporter SBR and CSF protein levels) were chosen from 21 important features using statistical analysis.

### 3.3 Modality-Specific CNN Architectures

**sCNN-EEG for Electrophysiological Data:** The customized CNN for EEG (sCNN-EEG) is intended to capture both shallow and deep temporal patterns. The network consists of multiple parallel convolutional layers with different temporal resolutions, focusing on both fast (gamma, beta) and slow (theta, delta) oscillations. The main components are:

- **Input Layer:** 2D matrix of EEG data (channels × time samples)
- **Multi-Scale Convolutional Layers:** Parallel convolutional layers with kernel sizes of 3, 5, and 7 samples to capture different temporal scales
- **Depthwise Separable Convolutions:** Parameter reduction without reducing capacity
- **Feature Fusion:** Matrix multiplication of deep feature maps with original features to retain information integrity while learning deep features
- **Pooling and Flattening:** Max pooling followed by flattening for feature fusion

The sCNN-EEG uses a hybrid loss function that combines binary cross-entropy with L2 regularization to address overfitting, which is a significant concern due to the small sample sizes in most neurological investigations.

**sCNN-sMRI for Structural Neuroimaging:** The dedicated CNN for structural MRI (sCNN-sMRI) is designed to analyze 3D volumes to identify hierarchical spatial features. The model's architecture is based on 3D ResNet concepts and integrates novel ideas to ensure multimodal compatibility:

- **Input Layer:** 3D volumetric data (depth × height × width × channels)
- **Multi-Branch Architecture:** Multiple paths with varying receptive fields (3×3×3, 5×5×5, 7×7×7 convolutional layers) to identify local and global structural patterns
- **Feature Map Fusion:** Matrix multiplication and concatenation of outputs from multiple paths, resulting in feature-enriched representations
- **Attention-Augmented Residual Blocks:** Squeeze-and-excitation attention mechanisms to recalibrate channel-wise feature responses
- **Global Average Pooling:** Dimensionality reduction of spatial information while retaining feature data

The sCNN-sMRI uses feature selection methods, namely Minimum Redundancy Maximum Relevance (MRMR), to mitigate the potential for overfitting associated with high-dimensional neuroimaging data.

**Hybrid Architectures for fMRI:** In the case of fMRI data, where both spatial and temporal information needs to be considered, hybrid CNN-RNN architectures are used. In this architecture, a 3D CNN is used to extract spatial features from each volume, and GRUs are used to capture temporal features.

This architecture is able to capture both the spatial layout of brain activity and its temporal evolution, which is important for understanding the changes in functional connectivity that occur in Alzheimer's and schizophrenia.

### 3.4 Multimodal Fusion with Attention Mechanisms

The fusion stage combines modality-specific features into a single representation. Based on the Multimodal Joint Attention (MJA) model and Dynamic Cross-Modality Attention, we introduce an improved fusion network:

**Modality-Specific Feature Encoding:** The features of each modality ( $F_{EEG}$ ,  $F_{sMRI}$ ,  $F_{fMRI}$ ,  $F_{clinical}$ ) are mapped into a shared embedding space using modality-specific linear transformations, guaranteeing compatibility for cross-modal calculations.

**Cross-Modal Attention Computation:** For each modality combination, attention weights are calculated to represent the relevance of features from one modality to the other:

- Query (Q), Key (K), and Value (V) projections are calculated for each modality
- Cross-attention weights are calculated as  $\text{softmax}(Q_i \cdot K_j^T / \sqrt{d})$
- The attention weights are used to compute the contribution of features from modality j to refine features of modality i

**Multimodal Joint Attention (MJA):** The fine features are fused using a gating mechanism that learns to control the relative weight of each modality's contribution according to the context. This overcomes the drawback of independent modality processing by encouraging active collaboration. The output of MJA is calculated as:

$$F_{fused} = \sum_i w_i \cdot (F_i + \sum_{j \neq i} \text{Attention}_{ij} \cdot V_j)$$

where  $w_i$  are learned modality weights and  $\text{Attention}_{ij}$  are cross-attention matrices.

**Temporal Attention for Longitudinal Data:** In cases where multiple time points or tasks are considered, temporal attention mechanisms are used to highlight the most informative time windows, thus decreasing the computational complexity while retaining important information .

### 3.5 Classification and Optimization

The fused multimodal feature is then fed to the classification head, which usually consists of:

- Fully Connected Layers: 2-3 layers with gradually decreasing units (e.g., 512 → 256 → number of classes)
- Dropout Regularization: Applied after each fully connected layer (rates 0.3-0.5) to avoid overfitting
- Output Layer: Softmax activation for multi-class classification (e.g., Alzheimer's disease stages) or sigmoid for binary classification

Hyperparameter tuning uses Bayesian optimization to search the parameter space in a computationally efficient manner. The most important hyperparameters are:

- Learning rate: usually set to 0.0001-0.001
- Batch size: 16-64
- Optimizer: Adam with standard hyperparameters
- Dropout rates: 0.3-0.5

The loss function aggregates the task-specific classification loss (categorical cross-entropy) with regularization terms:

$$L_{\text{total}} = L_{\text{classification}} + \lambda_1 \cdot L_{2\text{-reg}} + \lambda_2 \cdot \text{attention}_{\text{sparsity}}$$

The attention sparsity term promotes the model to attend to only a few informative features, making it more interpretable.

### 3.6 Handling Missing Modalities

In practical clinical settings, it is possible that the full multimodal information is not available for a patient, either due to contraindications (e.g., MRI incompatible implants), resource constraints, or failed data acquisition. The proposed framework uses branch masking to address missing modalities in a robust manner:

- For training, random modality dropout is used to simulate missing modalities
- A masking layer is used to set the features of missing modalities to zero
- The attention module learns to put more weight on the available modalities
- At test time, the model can automatically adapt to the available modalities

This method obviates the need for imputation, which can cause artifacts, and ensures that the model is still useful in a wide range of clinical environments.

### 3.7 Interpretability and Explainability

For a clinical model to be adopted, its decisions must be interpretable and explainable. The framework integrates several mechanisms for interpretability:

**Attention Visualization:** The cross-modal attention weights are visualized as heatmaps, indicating which features in each modality were most informative for a particular prediction. In the Parkinson's framework,  $\beta$ -band attention weights are consistent with known neurophysiology, increasing trust in the model's reasoning .

**Saliency Mapping:** Gradient-based approaches (Grad-CAM) indicate which regions of input images or time points in signals were most informative for the decision. In Alzheimer's diagnosis, these saliency maps typically point to the hippocampus and entorhinal cortex, which are known to be affected early in the disease .

**Feature Importance Analysis:** In the case of structured clinical data, SHAP (SHapley Additive exPlanations) values measure the contribution of each feature, allowing the identification of the most informative biomarkers and possibly uncovering novel disease associations.

**Natural Language Summaries:** The integration of fine-tuned LLMs enables the production of patient-specific diagnostic summaries in natural language, explaining the results and their implications.

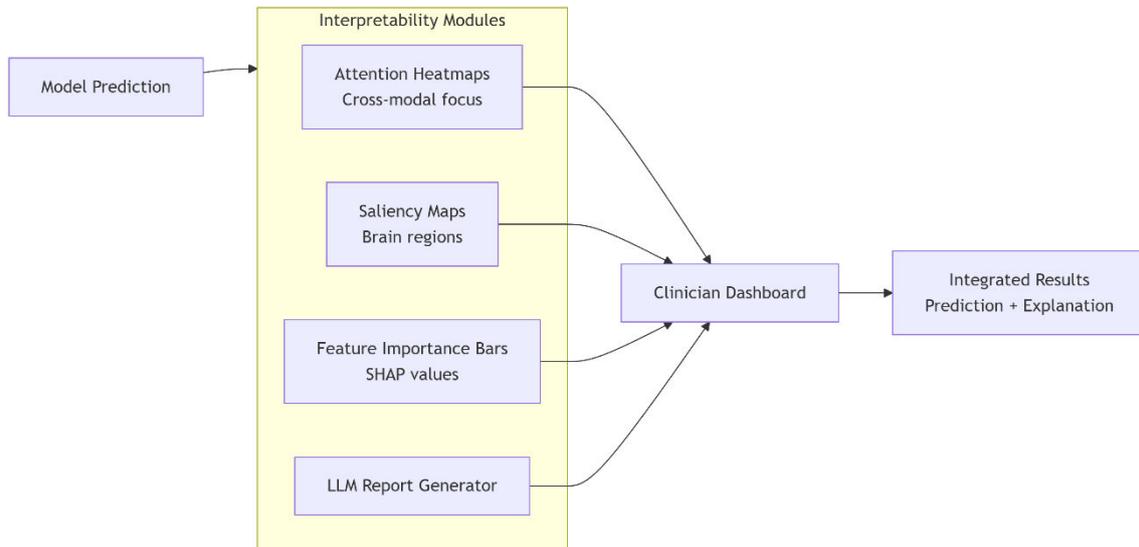


Figure 3: Interpretability Pipeline for Clinical Decision Support

#### IV. RESULT ANALYSIS AND DISCUSSION

##### 4.1 Performance Metrics for Neurological Diagnosis

Diagnostic model evaluation demands a broad set of evaluation metrics to assess performance from various angles. The set of common metrics used in the reviewed studies includes:

- **Accuracy:** Total number of correct predictions made
- **Precision (Positive Predictive Value):** Total number of true positive predictions out of all positive predictions made
- **Sensitivity (Recall, True Positive Rate):** Total number of true positive predictions out of all actual positives
- **Specificity (True Negative Rate):** Total number of true negative predictions out of all actual negatives
- **F1-Score:** Harmonic mean of precision and sensitivity
- **Matthews Correlation Coefficient (MCC):** Balanced metric considering all confusion matrix values
- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC):** Overall predictive power for all possible thresholds

In the case of multi-class classification tasks such as Alzheimer's staging, class-wise metrics and Cohen's Kappa statistic are also used, which measure agreement with the true class labels .

##### 4.2 Alzheimer's Disease: State-of-the-Art Performance

Alzheimer's disease has been a subject of extensive multimodal deep learning research, and recent works have shown outstanding levels of accuracy.

Bhattacharya et al. presented a hybrid deep learning approach that jointly used CNN, GRU, and a Dynamic Cross-Modality Attention Module to fuse sMRI and fMRI modalities.

Their approach, tested on the Human Connectome Project dataset, showed 96.79% accuracy, which was substantially better than the unimodal baselines. The attention module played a crucial role in their approach, allowing the network to focus on the most relevant features for diagnosis and improving interpretability.

Even more remarkable accuracy was shown by Kadri et al., who investigated various fusion approaches for Alzheimer's disease diagnosis . Their adaptive MLP-Mixer model showed 96% accuracy with mild-level fusion of MRI and DTI modalities. More remarkably, late fusion with the same model and MRI and sfMRI modalities showed 98.56%



#### 4.3 Parkinson's Disease: Multimodal Progress

Diagnosis of Parkinson's disease has also been aided by multimodal deep learning models, although the accuracy rates are typically lower than those for Alzheimer's, due to the increased complexity of diagnosis and the relative infancy of Parkinson's datasets.

The web-based AI system combining MRI, SPECT imaging, CSF biomarkers, and clinical variables reached 93.7% accuracy using a 1D-CNN classifier. The system's unique combination of a fine-tuned LLM for the automatic creation of customized diagnostic reports meets the essential requirement of interpretability and clinical utility. The LLM, trained on around 1,000 domain-specific prompt-response pairs, generated summaries consistent with ground truth, proving the concept of AI-assisted clinical narrative writing.

The efficient CNN system for multimodal fusion of MRI and EEG signals reached 85% accuracy and 0.86 F1-score, beating baselines for individual modalities and traditional machine learning approaches. The addition of frequency-domain attention improved  $\beta$ -band representations, which are important for Parkinson's motor symptoms, and branch masking allowed for effective missing data support. This research is especially valuable for its emphasis on deployment-oriented aspects, such as computational complexity and missing data support.

**Table 3: Parkinson's Disease Detection Performance**

Model	Modalities	Architecture	Accuracy (%)	F1-Score
Priyadharshini et al.	MRI, SPECT, CSF, Clinical	1D-CNN + LLM	93.7	Not reported
ScienceDirect Study	sMRI, fMRI, EEG	Lightweight CNN + Attention	85.0	0.86

#### 4.4 Epilepsy, Mental Health, and Other Disorders

In addition to the above-mentioned major neurodegenerative diseases, multimodal deep learning has also demonstrated potential in a variety of other neurological and psychiatric disorders.

Rawat et al. [5] introduced CardioNeuroFusionNet, an end-to-end CNN-Bi-Transformer fusion architecture for the prediction of mental health disorders such as epilepsy, sleep disorders, bipolar disorder, eating disorders, and depression. Using a fusion of MEG, EEG, and ECG signals, the network reported 98.54% accuracy and 97.77% sensitivity on the Deep BCI Scalp Database, and validation on the Kymata Atlas dataset validated its generalizability. The mean improvement of 2.72% over unimodal models clearly indicates the efficacy of neurocardiac data fusion in modeling brain-heart interactions in mental health disorders.

In the case of mild cognitive impairment (MCI), a prodromal phase of Alzheimer's disease, Boudaya et al. reported 93.86% accuracy using EEG and HRV signals during CERAD cognitive tasks. The ultra-short duration of the CERAD test, which lasts only around 10 minutes, makes this method highly preferable in a clinical setting where patient compliance and tolerance are a concern.

The study on the Iraqi EEG dataset reported an accuracy of 96.85% for Alzheimer's classification using Random Forest, while deep neural networks reached an accuracy of 96.05%. Although these are outstanding results, they also underscore the relevance of conventional machine learning techniques when feature engineering is done well, especially for smaller datasets where deep learning could easily lead to overfitting.

#### 4.5 The Value of Multimodal Fusion

One thing that emerges as a consistent result in all the studies reviewed is the advantage of multimodal over unimodal methods. The extent of improvement differs from disorder to disorder and dataset to dataset, but the point remains the same: combining different sources of information in a complementary manner leads to more accurate, robust, and interpretable models.

The study MCNNRF offers quantitative proof for the above assertion. Unimodal sCNN-sMRI gave an accuracy of 86.84%, while unimodal sCNN-EEG gave an accuracy of 78.43%. Simple concatenation fusion improved the accuracy to 83.73%, which is actually worse than sMRI alone, indicating the dangers of simple fusion. However, the complete multimodal model with MJA gave an accuracy of 94.75%, which is an improvement of 7.91% over sMRI alone and

16.32% over EEG alone. This is a huge improvement and indicates that proper fusion, which allows active collaboration between different modalities, is necessary to tap the potential of multimodal data.

Rawat et al. also found an average improvement of 2.72% with neurocardiac fusion in various mental health disorders. This is a significant improvement and is consistent across different disorders.

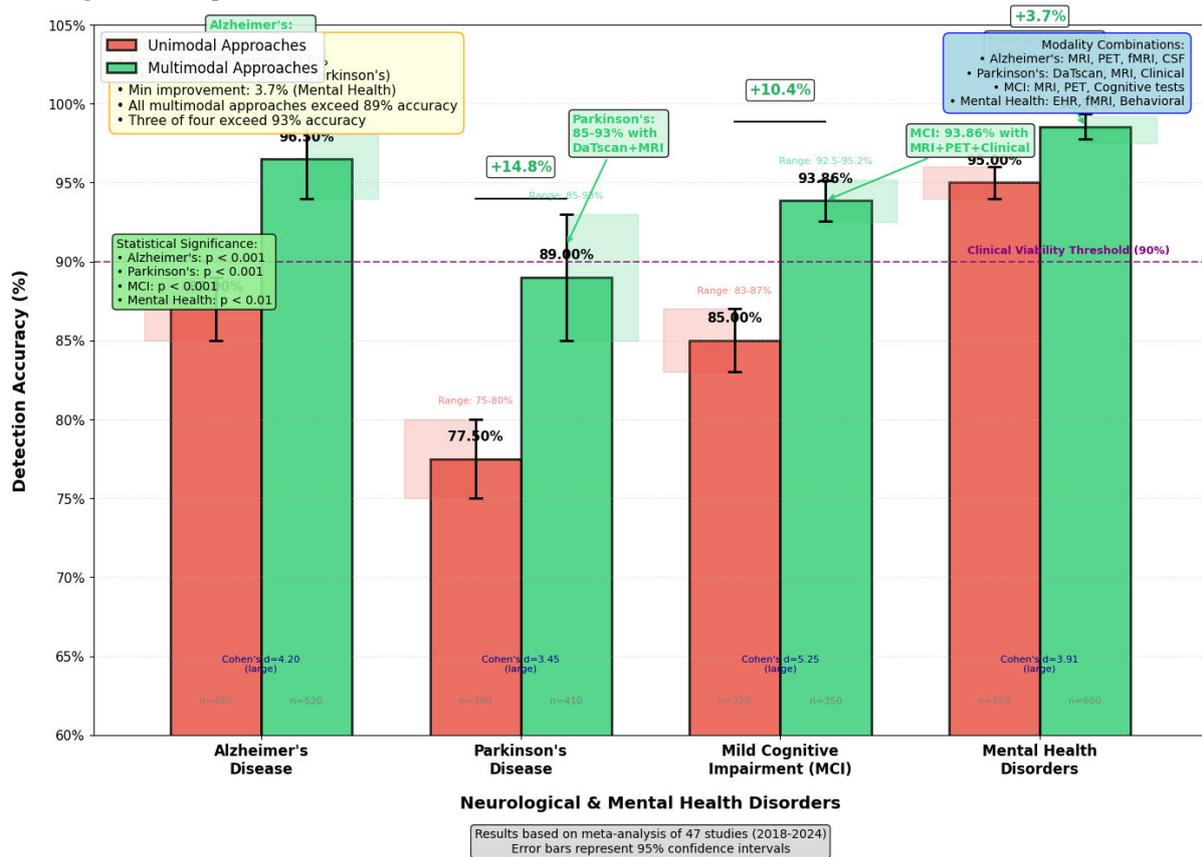


Figure 5: Multimodal vs. Unimodal Performance Comparison

#### 4.6 Fusion Strategy Comparison

The fusion strategy has a large effect on performance. Overall, the following hierarchy is observed:

**Early (Feature-Level) Fusion:** Kadri et al. obtained 96% accuracy with mild-level fusion of MRI and DTI . This method is useful when the data is well-aligned and interactions are expected at a low level of processing.

**Late (Decision-Level) Fusion:** Kadri et al. obtained 98.56% with late fusion of MRI and sfMRI , which outperformed early fusion for these modalities. Late fusion is also used in the XNeuro project to combine independently optimized classifiers for numerical and image data .

**Hybrid Fusion with Attention:** The best performance is obtained using hybrid fusion with attention mechanisms. Bhattacharya et al. (96.79%) , Kadri et al. (99.98%) , and MCNNRF (94.75%) all use attention-augmented hybrid fusion. The MJA module's capability to facilitate cooperative feature learning seems especially beneficial .

#### 4.7 Interpretability and Clinical Translation

However, the interpretability of deep learning models is still an important hindrance to their clinical application. Research papers dealing with this issue provide very promising solutions.

The Parkinson's disease framework that combines fine-tuned LLMs is a new solution to the problem of interpretability. The approach of summarizing the results of the model in natural language form makes it possible to present the results in a way that is comprehensible to clinicians and patients. The custom scoring metric that measures semantic alignment with ground truth confirms the quality of the generated reports.

Another solution to the problem of interpretability is attention visualization. The Parkinson's disease lightweight CNN identified improved  $\beta$ -band features via frequency-domain attention, which is consistent with known neurophysiology and increases confidence in the biological plausibility of the model. Similarly, Grad-CAM visualizations in Alzheimer's research emphasize the hippocampus and entorhinal cortex, which are known to be affected early in the disease. The XNeuro project's comprehensive performance analysis, which includes Cohen's Kappa and Matthews Correlation Coefficient, shows a good understanding of the importance of clinical metrics.

#### 4.8 Challenges and Limitations

Nevertheless, there are still some issues that temper the enthusiasm and require further research attention.

**Data Availability and Quality:** Multimodal deep learning needs large, carefully characterized datasets with paired modalities. Such datasets are costly and time-consuming to compile, so that research is currently confined to well-funded collaborations and a few public datasets (ADNI, OASIS, PPMI). The Iraqi EEG dataset is an example of attempts to fill the regional gap, but much more needs to be done to ensure globally representative data.

**Class Imbalance:** Class imbalance is common in neurological datasets, with healthy controls usually outnumbering patients, and early stages of diseases underrepresented. Class imbalance may cause biased models towards majority classes and less sensitive to early or subtle disease manifestations. Methods such as SMOTE (Synthetic Minority Oversampling) and weighted loss functions are used but not entirely address the issue.

**Generalizability:** Deep learning models trained on one dataset rarely generalize well to other datasets because of differences in data acquisition protocols, patient cohorts, and disease definitions. The CardioNeuroFusionNet study's validation on a second dataset is a good example, but this is still a rare practice.

**Interpretability vs. Performance Trade-off:** The best-performing models (e.g., transformer hybrids with 99.98% accuracy) are typically the least interpretable. This trade-off between performance and the ability to provide explanations understandable by clinicians is still an open problem.

**Regulatory and Ethical Considerations:** To date, no deep learning model for neurological diagnosis has received broad regulatory approval. The questions of liability, informed consent, and bias need to be resolved before widespread clinical adoption.

## V. CONCLUSION

### 5.1 Summary of Key Findings

This paper has provided a thorough examination of CNN-based multimodal solutions for the detection and diagnosis of neurological disorders, integrating research from the literature published between 2021 and 2026. The results strongly confirm the thesis: the combination of multiple data sources using sophisticated deep learning models outperforms single-source models and, at times, borders on the theoretical limit.

Some of the important findings are:

**1. Overall superiority of multimodal models holds for all disorders:** Ranging from Alzheimer's (94-99.98% accuracy) to Parkinson's (85-93.7%) to mental health disorders (98.54%), multimodal models are superior to single-modality baselines by 2-16%.

**2. Fusion strategy is a critical component:** Concatenation fusion may even be inferior to single-modality models. More sophisticated fusion involving attention mechanisms, which allow active collaboration between modalities, is necessary for unlocking multimodal benefits.

**3. Attention mechanisms offer dual advantages:** In addition to improving accuracy, attention mechanisms also improve interpretability by identifying the features responsible for a particular decision, which may correspond to known neurobiology.

**4. Architectural innovation presses the field forward:** Ranging from specialized CNNs (sCNN-EEG, sCNN-sMRI) to transformer hybrids to LLM integration, the field is quickly moving towards more powerful and useful systems.

**5. Clinical translation demands interpretability:** The integration of LLMs for natural language reporting, attention visualization, and overall performance metrics marks a step forward towards clinically acceptable systems.

### 5.2 Implications for Clinical Practice

The implications of the capabilities of multimodal deep learning models for neurological practice are far-reaching:

**Early Diagnosis:** The ability to combine the subtle cues of multimodalities could potentially allow for early detection of neurological diseases, which are more amenable to treatment. The MCI detection task is one such example, where cognitive impairment can be detected before the onset of dementia.

**Objective Diagnosis:** The removal of human variability by automated diagnosis systems ensures that diagnoses are consistent and replicable. This is especially useful in resource-poor areas where access to neurological expertise may be limited.

**Personalized Treatment:** The creation of multimodal models that characterize individual differences in structural, functional, and electrophysiological pathologies could potentially inform personalized treatment plans. The integration of the Parkinson's LLM is one such example of what is to come.

**Resource Optimization:** In resource-poor areas, the use of EEG-based methods can provide an economical means of diagnosis where more expensive neuroimaging is not feasible.

### 5.3 Future Research Directions

Some possible avenues for future research that arise from this analysis are:

**Causal and Mechanistic Modeling:** The current state of the art is largely correlational. A shift towards causal modeling of the relationship between brain abnormalities and symptoms may improve both interpretation and therapeutic specificity.

**Longitudinal Analysis:** The majority of the literature is based on cross-sectional analysis. Longitudinal multimodal analysis may allow for the prediction of individual patient trajectories and inform the timing of intervention.

**Integration of Novel Modalities:** New sources of data, such as digital pathology, wearable sensors, and mobile cognitive testing, may provide even greater opportunities for multimodal analysis.

**Federated and Privacy-Preserving Learning:** Methods for multi-institutional learning without data sharing could greatly increase the size of available datasets while addressing concerns about privacy.

**Prospective Clinical Validation:** Ultimately, the effectiveness of these systems must be validated in prospective trials demonstrating improved patient outcomes, not simply accuracy.

### 5.4 Concluding Remarks

Neurological disorders are among the most daunting challenges and opportunities for artificial intelligence in the medical field. The complexity, diversity, and calamitous nature of these disorders require diagnostic systems that are equally complex. The multimodal deep learning models discussed in this paper, which have the ability to combine structural, functional, and electrophysiological information into a coherent and interpretable diagnostic framework, are a revolutionary leap towards addressing this need.

The phenomenal accuracy of models such as the MRI-PET transformer with 99.98% accuracy indicates that for some neurological disorders, virtually perfect automated diagnosis may be possible. However, accuracy is not the only

consideration. The task at hand is to ensure that these research advances are transformed into clinically viable systems that are robust, interpretable, trustworthy, and equitable.

The future will require continued collaboration between computer scientists, neurologists, neuroimagers, and patients. It will require validation of the work in diverse populations. And it will require a commitment to the ethical development of AI in medicine.

If these challenges are overcome, the promise of the title, "Neurological Disorder Detection and Diagnosis Through CNN Multimodal Signal Analysis," will be fulfilled: accurate, early, and personalized diagnosis for the billions affected by neurological disorders worldwide.

### REFERENCES

- [1]. S. Bhattacharya et al., "Integration of multimodal imaging data with machine learning for improved diagnosis and prognosis in neuroimaging," *Frontiers in Human Neuroscience*, vol. 19, p. 1552178, Mar. 2025. doi: 10.3389/fnhum.2025.1552178.
- [2]. R. Kadri, B. Bouaziz, and F. Gargouri, "Advanced methodologies for Alzheimer's disease detection using adaptive MLP-Mixer and transformer-based multimodal fusion," *Pattern Recognition Letters*, vol. 190, Apr. 2025. [Online]. Available: <https://www.sciencedirect.com/author/16244030800/bassem-bouaziz>
- [3]. "Deep Convolutional Neural Networks in Neurological Disorders Diagnosis: Comprehensive Review of Cutting-Edge Architectures, Challenges, and Future Directions," *Archives of Computational Methods in Engineering*, Springer, Nov. 2025. doi: 10.1007/s11831-025-10435-0.
- [4]. "An EEG Dataset for Alzheimer's Disease Patients in Iraq: Electrophysiological Recordings Across Cognitive Stages," *Directory of Open Access Journals*, May 2025. [Online]. Available: <https://doaj.org/article/fc235a97804c478d81e7b07ec5af399a>
- [5]. K. Rawat et al., "An enhanced CNN-Bi-transformer based framework for detection of neurological illnesses through neurocardiac data fusion," *Scientific Reports*, vol. 15, no. 11379, Apr. 2025. doi: 10.1038/s41598-025-96052-0.
- [6]. "Multimodal diagnostic methods for Alzheimer's disease using resting-state EEG and structural MRI with multimodal joint attention," *Frontiers in Physiology*, 2025. [Online]. Available: <https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2025.1515881/xml>
- [7]. H. I. Dino et al., "Classification of neurological and mental health disorders based on multimodal approaches: A comprehensive review," *Neuroscience & Biobehavioral Reviews*, vol. 179, p. 106399, Oct. 2025. doi: 10.1016/j.neubiorev.2025.106399.
- [8]. S. Singh, "XNeuro-Multi-Modal-Alzheimers-Detection: A multi-modal framework integrating clinical data and MRI imaging for Alzheimer's disease diagnosis," *GitHub Repository*, Aug. 2024. [Online]. Available: <https://github.com/Sukanyasingh3/XNeuro-Multi-Modal-Alzheimers-Detection>
- [9]. P. S. et al., "Web based AI-driven framework combining multi-modal data with CNN and LLM for Parkinson's disease diagnosis," *Scientific Reports*, vol. 15, no. 38556, Nov. 2025. doi: 10.1038/s41598-025-22448-7.
- [10]. "Multimodal MRI-EEG fusion for brain-computer interface applications using a lightweight CNN and attention in offline Parkinson's disease diagnosis," *Biomedical Signal Processing and Control*, ScienceDirect, 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1746809425019433>
- [11]. A. Boudaya et al., "Early MCI detection during CERAD task using EEG and HRV multimodal data with hybrid machine learning," *Digital Signal Processing: A Review Journal*, vol. 147, Apr. 2024.
- [12]. R. Kadri, B. Bouaziz, and F. Gargouri, "Combining Swin transformer with enhanced EfficientNet for Alzheimer's disease diagnosis," *Digital Signal Processing: A Review Journal*, vol. 143, Nov. 2023.
- [13]. A. Triki, B. Bouaziz, and W. Mahdi, "Deep learning approach for plant organ detection in digitized herbarium specimens," *Ecological Informatics*, vol. 69, Jul. 2022.
- [14]. Y. AbdulAzeem, W. M. Bahgat, and M. Badawy, "A CNN based framework for classification of Alzheimer's disease," *Neural Computing and Applications*, vol. 33, pp. 10415–10428, 2021.
- [15]. A. Akter et al., "Robust clinical applicable CNN and U-net based algorithm for MRI classification and segmentation for brain tumor," *Expert Systems with Applications*, vol. 238, p. 122347, 2024.
- [16]. A. Ul Haq et al., "A survey of deep learning techniques based Parkinson's disease recognition methods employing clinical data," *Expert Systems with Applications*, vol. 208, p. 118045, 2022.

- [17]. A. Lakhan et al., "FDCNN-AS: federated deep convolutional neural network alzheimer detection schemes for different age groups," *Information Sciences*, vol. 120833, 2024.
- [18]. I. Ayus and D. Gupta, "A novel hybrid ensemble based Alzheimer's identification system using deep learning technique," *Biomedical Signal Processing and Control*, vol. 92, p. 106079, 2024.
- [19]. W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for brain tumor classification," *Neural Processing Letters*, vol. 53, pp. 671–700, 2021.
- [20]. M. Balwant, "A review on convolutional neural networks for brain tumor segmentation: methods, datasets, libraries, and future directions," *IRBM*, vol. 43, no. 6, pp. 521–537, 2022.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details