International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 286-309
https://doi.org/10.70917/ijcisim-2026-2078
Article

# A SELF-SUPERVISED HYBRID 3D CNN–VISION TRANSFORMER FRAMEWORK FOR MULTI-STAGE ALZHEIMER'S DISEASE CLASSIFICATION FROM MRI

**Anusha Jain[1], Swati Pandey[2]**

[1] Department of Computer Science & Engineering, Oriental University, Indore, Madhya Pradesh, India. anusha.jain9@gmail.com
[2] Department of Computer Science & Engineering, Oriental University, Indore, Madhya Pradesh, India.swatipandey@oriental.ac.in
**Corresponding Author:** Anusha Jain (anusha.jain9@gmail.com)

**Abstract:** Alzheimer's disease is an irreversible, progressive brain disorder that slowly destroys memory and thinking skills and, eventually, the ability to carry out the simplest tasks. Conventional deep learning methods usually find it difficult to simultaneously extract local anatomical features and long-range contextual relationships from brain MRI images. To this end, we introduce a hybrid3D CNN–Vision Transformer architecture for multi-stage Alzheimer's disease classification in this work. The proposed model integrates volumetric feature extraction with a 3D CNN and global attention modeling via the Vision Transformer, together with a self-supervised pretraining approach. Experiments show that the proposed model can achieve an overall classification accuracy of 98.23%, which outperforms other ensemble CNN and multimodal fusion methods. The results suggest better feature representation, higher class-wise performance and more effective generalization among disease stages.

**Keywords:** Alzheimer's disease, 3D CNN, Vision Transformer, hybrid deep learning, MRI classification, self-supervised learning.....

## 1. INTRODUCTION

Alzheimer's disease (AD) is a degenerative brain disease associated with gradual and irreversible decline of memory, cognition, and overall daily functions [1]. It is the most prevalent cause of dementia globally, and it is a major public health challenge given the growing aging population. Identifying individuals at risk for Alzheimer's disease is important, as interventions are more likely to be successful earlier during cognitive decline. Magnetic resonance imaging (MRI) is the method of choice for detecting structural brain changes linked to various stages of AD [2]. But manual diagnosis is a long ordeal that human beings may not always be patient to perform, and it is also prone to error which has prompted interest in AI-powered diagnostic systems.

In the last few years, deep learning approaches in general and CNNs have shown excellent results for medical image classification. Due to the capability of automatically learning discriminative features from MRI, CNN-based methods are very popular in AD detection. However, traditional CNN architecture typically only considers local spatial patterns and may hardly integrate long-range interactions between brain regions [4]. Meanwhile, transformer-based architectures have achieved great success on modeling global contextual relationships but heavily rely on large scale datasets and may not perform well in terms of strong local feature extraction if they are used alone [5].

The purpose of the present work is to alleviate the drawbacks of single deep learning architecture by attempting to model features from CNNs and transformers within the same hybrid framework. And if we combine volumetric feature extraction with global attention mechanisms [6], it would be able to have better and more robust classification results for the multi-stage of Alzheimer's disease.

**Copyright:** © 2026 by the authors

The goal of this study is to implement a novel hybrid 3D CNN-ViT model that can accurately classify MCI stages to AD as well as the baseline accuracy and generalization to different subjects. The paper also presents a self-supervised pretraining method to improve feature learning quality, by leveraging the small amount of labeled medical data.

The new key points of this Work are:

A hybrid architecture that combines a 3D CNN for local volumetric feature extraction and a Vision Transformer for global context modeling.

The introduction of self-supervised pretraining for better feature representation learning and over-fitting reduction.

Application of a universal algorithm that is capable of multi-stage Alzheimer's diagnosis.

Improved classification performance, with accuracy of about 98.23%, achieving the better CNN-based and multimodal fusion methods previously reported.

The rest of this paper is structured as follows. Section 2 provides an extensive survey of related works in Alzheimer's disease diagnosis and deep learning methods. In section 3, we detail our experimental configuration of the proposed hybrid method regarding architecture, algorithm and training policy. Implementation environment, dataset properties and experimental setting are described in Section 4. Results analysis and comparisons with previous works are reported in section 5. Finally, Section -6 discusses the conclusion of the paper and future possible research directions.

## 2. REVIEW OF LITERATURE

Alzheimer's disease (AD) is a chronic and progressive neurodegenerative disorder featuring cognitive, behavioral and memory dysfunction. It is the most prevalent among dementia disorders in addition to still remaining as a critical worldwide public health threat. According to more recent estimates, millions of people are affected world-wide and there is extraordinary suggestion that this prevalence will increase significantly due to aging [11], [24]. Epidemiological studies indicate that incidence and prevalence rates are steadily rising in all regions, albeit with very high and rapidly increasing rates reported from developed countries and East Asia, respectively [4], [19]. Health system data also indicate a significant economic and clinical burden, such as high hospitalization rates, mortality and cost of care, particularly among older individuals and women [3]. These results underline the importance of having early diagnostic strategies and disease modifying interventions.

### 1. Pathophysiology and Disease Progression

Alzheimer's disease is associated with a complicated pathological process, which includes amyloid-beta (Aβ) deposition of plaque, tau protein hyperphosphorylation, synaptic impairment and so on [5]. Although the amyloid cascade hypothesis is a prevailing model in AD research, recent studies have highlighted other mechanisms including neuroinflammation, oxidative stress and vascular dysfunction [5], [37]. The pathobiology of the disease is now appreciated to be a spectrum, starting with asymptomatic biological changes that may occur decades prior to clinical symptoms [10], [14]. This evolving knowledge has transformed diagnostic paradigms from symptom-based groupings to biomarker-driven approaches [9], [25].

Biological markers such as CSF and imaging-based amyloid and tau pathology indicators have allowed molecular based diagnosis and refined disease staging [25]. In addition, blood-based markers have recently been developed as a minimally invasive alternative that present good associations with standard cerebrospinal fluid markers and allow widespread screening [8], [12]. These are important for early device design concern, since therapeutic intervention is most effective during early and middle stages of the disease [9].

### 2. Global Burden and Public Health Implications

The worldwide prevalence of Alzheimer's disease is increasing, and estimates expect a soaring number of cases over the next half-century [11], [24]. Demographic ageing is the major cause of this trend, but lifestyle factors and higher diagnostic yield and sensitivity also play a role in increasing prevalence [4]. Estimation based on worldwide data reports a dramatic increase in dementia-attributable mortality and disability-adjusted life years (DALYs), with higher prevalence among women than men [24]. To this end, policy strategies in a number of nations are focused on the disease by increasing research funding, developing national strategies and forming collaborative frameworks [15]. Nevertheless, the chasm between the immediacy of the problem and the speed of therapeutic development is still very

wide [11]. These difficulties highlight the need for early intervention, preventative measures and targeted therapeutic strategies.

### 3. Advances in Therapeutic Strategies

Therapeutic research in Alzheimer's disease (AD) has shifted from symptomatic drugs to disease-modifying agents. Standard pharmacotherapy, including cholinesterase inhibitors and memantine, mainly provides symptomatic benefit [9]. However, emerging anti-amyloid monoclonal antibodies (mAbs), such as leucaena and donanemab have shown efficacy in delaying clinical decline among subjects who are prodromal AD [9]. Other interventions like anti-inflammatory drugs, pathways associated with vasculature and blood–brain barrier healing have also been studied [5]. Phototherapeutic agents, nanotechnology-based drug delivery systems and photo biomodulation also have shown promise by targeting amyloid plaques or reducing neuroinflammation [7]. Stem cell therapy is another evolving direction, and embryonic, mesenchymal, neural and induced pluripotent stem cells have all been reported to be able to facilitate neurodegeneration, suppress inflammation and induce angiogenesis [2]. Despite their potential, these strategies are hindered by ethical issues, varied cell culture biology and unreliable clinical results [2]. To improve cell migration and differentiation, other approaches such as Bioengineering strategies and electromagnetic field-guidance systems are being explored [2, 7].

### 4. Biomarkers and Early Diagnostic Approaches

Diagnosis in early stages is important for timely treatment and management of the diseases. Breakthroughs in biomarker discovery have revolutionized the diagnostic scene, leading to earlier recognition of disease and improved identification of its stage [9], [25]. Diagnostic tools such as MRI and PET imaging play a key role when used in clinic [13], [25]. But such procedures are often expensive and invasive, precluding their wide application. The available data from recent studies indicate that blood-based biomarkers should be considered as accessible options for the early diagnosis [8], [12]. Dried blood spot methods, as an example, have shown good agreement with conventional biomarkers and potential utility in mass screening [8]. There is also ongoing work investigating genetic and cognitive markers for the early identification of disease, and where combined with cognitive tests have high diagnostic accuracy [35].

### 5. Artificial Intelligence in Alzheimer's Disease Diagnosis

Artificial intelligence (AI) and machine learning have recently offered potential promises to develop robust algorithms for analysis of complex biomedical data with possible improvement in diagnosis accuracy. Deep learning architectures, especially Convolutional Neural Networks (CNN), have shown impressive accuracy in AD stages classification from the images of brain [16], [30]. For instance, DenseNet architectures achieved diagnostic accuracies higher than 86% based on MRI data [16], and hybrid models using different CNN architectures also reached accuracy above 97% [31]. More complicated architectures are designed to enhance the capability of feature extraction and classification (3). Channel-attention-attentive CNNs and 2D–3D CNN architectures have reached multi-class classification accuracy over 98% [27], [29]. There are also transfer learning methods using EfficientNet-type models that exhibit high sensitivity and specificity in classifying the stage of AD [28]. The transformer architectures have generated considerable interest because of their capacity to model long-range global contextual interactions in image data. Although Vision Transformer models have reported good diagnostic accuracy, the pooled sensitivity and specificity are greater than 92% and 95%, respectively [20]. Lightweight CNN-transformer hybrid models used on EEG data were able to reach a classification accuracy of around 95% for neurodegenerative disorders [22]. Among imaging, AI models have been exploited from textual and cognitive data. BERT integrated medical entity recognition has been found to accurately delineate AD from clinical text with a performance above 94% [23]. Such multimodal approaches demonstrate that AI can combine heterogenous data sources to improve and stabilize diagnosis.

### 6. Multimodal Artificial Intelligence and Data Fusion

Recent studies suggest that multi-modal integration of information may be crucial to enhancing diagnostic performance. Multimodal AI methodologies pool imaging, genetic, biomarker and clinical data to improve disease prognosis and classification [1], [13]. Bibliometric analysis It's a growing trend towards the integration of multiple data modalities, and that intermediate fusion strategies become 2019 The Author(s) effective to address heterogeneous data sources [1]. Multimodal imaging modalities, i.e., PET/MRI have provided a better diagnostic accuracy in comparison to single modalities [33]. Likewise, the combination of imaging data with clinical and genetic data has proved promising for a more accurate disease description and prognostication [1]. Multimodal Deep Learning Methods Multimodal deep learning models are widely studied and consistently outperform single modality models, especially in early detection across different reviews [17]. Such generative models, including Generative

Adversarial Networks (GANs) and diffusion-based models have also been applied to enrich datasets as well as enhance classification results. Some studies presented diagnostic accuracy near 99% with GAN-based methods but concerns about computational cost and interpretability remain [26].

**7. Emerging Technologies and Remote Monitoring**

In addition to imaging and biomarkers, other new technologies like wearable sensors, smart home systems or remote monitoring platforms are under development for supporting diagnosis and care of AD [21]. Such systems can acquire behavioral, physiological and environmental data for continuous monitoring and early detection of cognitive decline. Nonetheless, concerns about privacy, ease of use and validation continue to stand in the way of clinical integration [21]. Optical coherence tomography (OCT) has also been investigated as a potential non-invasive diagnostic means to detect retinal alterations, which are closely related to the neurodegenerative diseases such as AD [34]. Moreover, reports on systemic conditions such as periodontal health and inflammation show an association possibly linking these systemic associations with the disparate onset of neurodegeneration [36].

**8. Ethical, Bias, and Implementation Challenges**

Despite the promising progress, a few challenges continue to exist in the implementation of AI-based diagnostic systems. Bias or imbalanced training data can cause its predictions to be inaccurate or unfair, i.e., not all users receive the same permissions [6]. Ethical challenges associated with data sharing, privacy and clinical integration are also barriers to widespread use [18]. Several AI models achieve high performance under controlled settings but have not been robustly validated at scale or in a real-world setting [22], [33], [42]. Computational complexity and lack of interpretability, requirement for standardization in datasets, are other challenges for its clinical integration [17], [26]. Better data sharing, standardized evaluation protocols and multidisciplinary collaboration are necessary to tackle these challenges.

**9. Preventive and Risk-Based Approaches**

Recent studies increasingly focus on preventive approaches that are directed against modifiable risk factors for Alzheimer's disease. Through bioinformatic assay, hundreds of potential risk factors and candidate drugs, such as lipid-lowering drug, anti-inflammatory drug and metabolism-adjusted medicine have been identified [14]. Prevention of disease might be substantially reduced with early intervention on these pathways. Lifestyle, cognitive training and tailored treatment are now acknowledged as essential elements in AD management [9]. Such strategies are intended to help postpone the age of onset of disease and increase quality of life for patients and their careers.

**10. Summary and Research Gaps**

There are rapid advances in Alzheimer's disease diagnosis and treatment in literature from innovations in biomarker research, artificial intelligence, and multi-modal data integration. Multimodal AI strategies lead to superior performance compared with single modality-only methods, suggesting that the combination of imaging, genetic and clinical data is essential [1],17]. These deep learning models, such as CNNs and transformer-based architectures have demonstrated high diagnostic accuracies demonstrating strong clinical potential [20], [27], [29].

However, significant challenges remain. Several studies proceed with small datasets, secondary validation and a consequent lack of generalization and clinical applicability [22], [33]. Real-world deployment is also hindered due to ethical considerations, biased datasets, and computational demand [6], [26]. Nevertheless, despite the advances in therapy there is a limited number of effective disease-modifying treatments and more ongoing research is required [11], [37].

Future work could instead focus on the development of multimodal frameworks that are more resistant to attack, a more diverse set of datasets or how they can be used to improve model interpretability. Combining AI diagnostic tools with biomarker-driven clinical pathways may result in earlier detection, patient specific treatment regimens and better patient outcome. Ongoing interdisciplinary efforts between clinicians, engineers and policymakers will be key to harnessing these advances into practical, scalable solutions for Alzheimer's disease care.

## 3. METHODOLOGY

### *3.1 Proposed architecture*

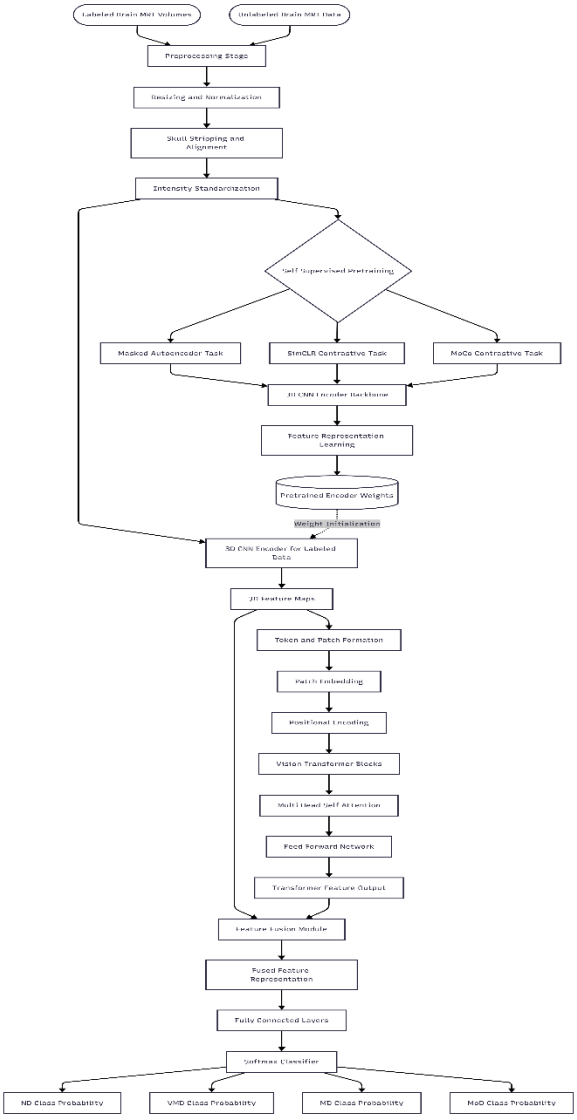


**Figure 1: Overall Hybrid 3D CNN–Vision Transformer Architecture**

This figure 1 shows the entire pipeline of the proposed hybrid model for Alzheimer's disease diagnosis. Preprocessing: All labeled and unlabeled volumes of MRI are preprocessed -resizing, normalization, skull stripping and intensity standardization. In the self-supervised pretraining phase, we utilize unlabeled data for learning general-purpose robust features. The pretrained weights are used to initialize the supervised 3D CNN encoder, which is used for extracting volumetric features. The features are projected into a Vision Transformer and then feature fusion operation is performed to feed the outcome into fully connected layers to predict the disease stage with softmax classification.

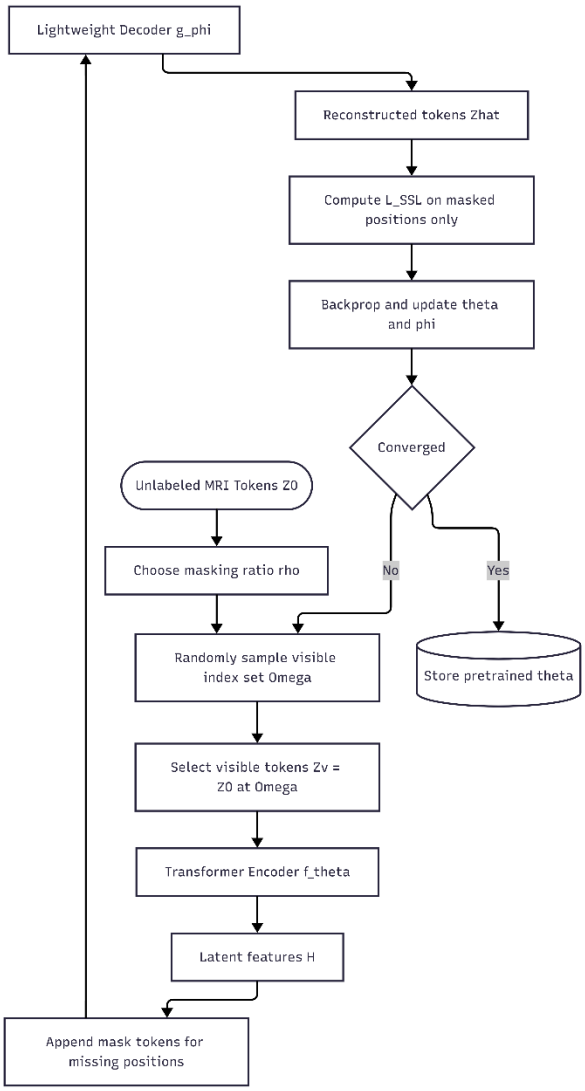


**Figure 2: Self-Supervised Pretraining Process**

This figure 2 illustrates the internal architecture of the self-supervised pretraining stage. Unregimented MRI tokens are also randomly masked to a degree specified by a predefined masking ratio. The visible tokens are encoded with a transformer encoder to yield representations. “mask” tokens are added and an efficient decoder reconstruct s the masked token. Reconstruction loss is only calculated on masked regions, to learn meaningful anatomical structures without labels. The encoder/decoder parameters are updated alternates until convergence, and the pretrained weights with the converged model are retained for supervised learning.

Figure 3 gives us an illustration of the internal structure of the 3D CNN encoder for volumetric feature extraction. The input MR volume is passed through several Conv3D blocks, formed by convolution-batch normalization-ReLU-layer. Down-sampling layers are responsible to reduce the spatial size and increase abstraction of features. Staged convolutional layers extract increasing-localized anatomical features. An optional channel-attention module is used to amplify salient features. The overall output is a set of 3D high-level feature maps that already contain structural information and are subsequently used for tokenization and transformer-based global modeling.

These 4 figures illustrate the volumetric MRI data processing based on transformers. The input volume is resampled, normalized and possibly augmented. It is subsequently split into a set of disjoint 3D patches, and each patch is projected to a latent embedding. To retain spatial information positional encodings are incorporated. The token sequence flows through several transformer blocks containing layer normalization, multi-head self-attention,

and feed-forward networks with residual connections. These layers are designed to account for long-range dependencies among brain regions, so that a global contextual representation of the MRI volume is generated.

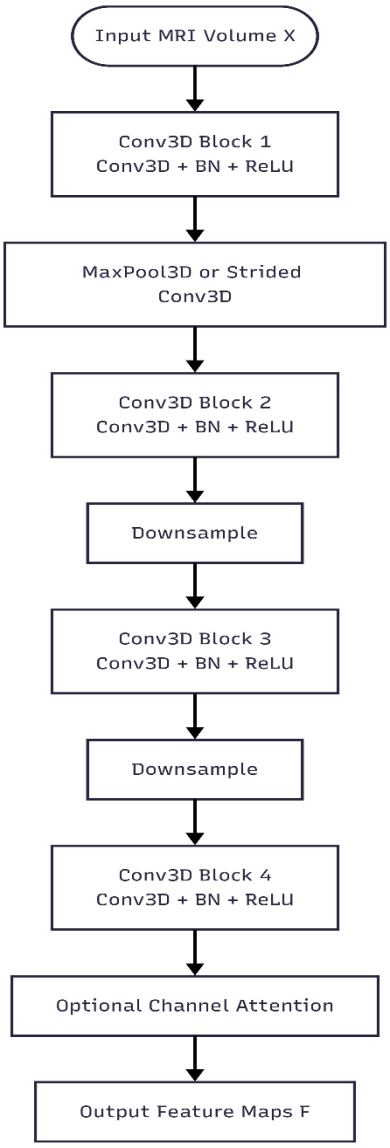


**Figure 3: 3D CNN Encoder Internal Architecture Pipeline**

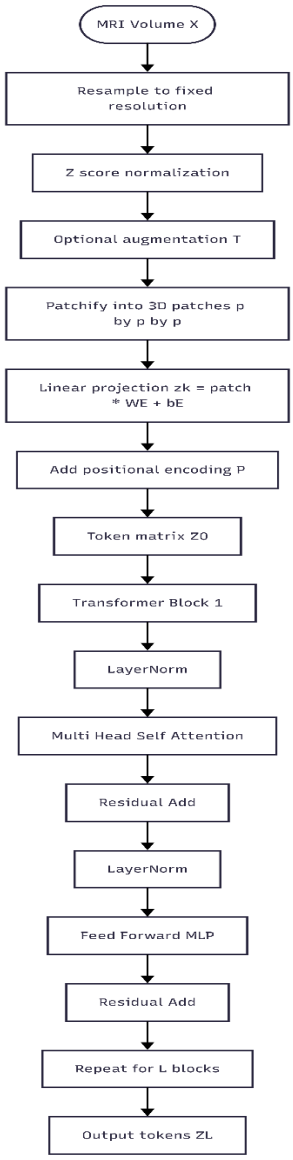


**Figure 4: Vision Transformer Processing**

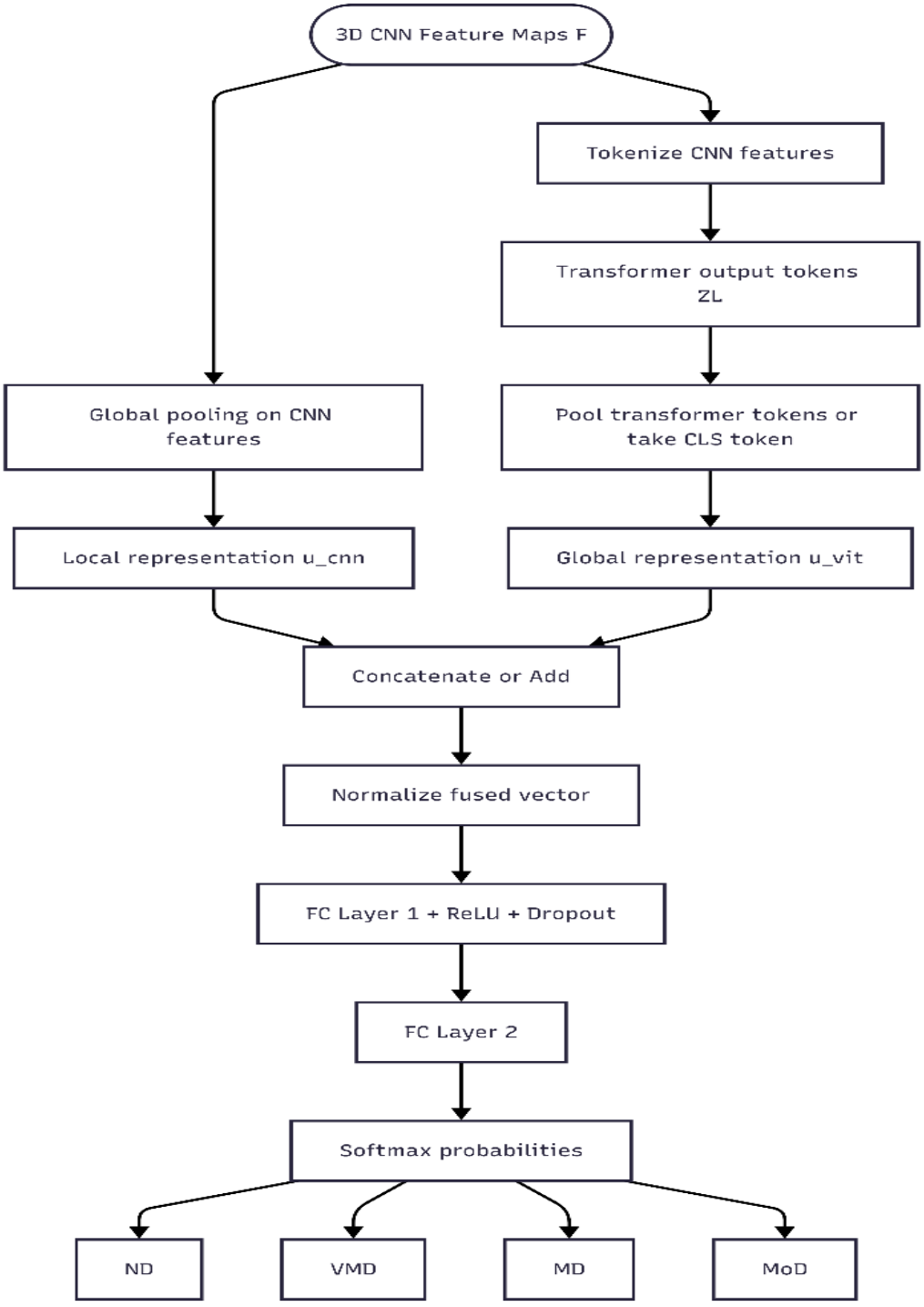


**Figure 5: Feature Fusion and Classification Module**

This figure 5 shows the integration of local CNN features and global transformer representations. Then we use global pooling on both 3D CNN feature maps for local features and transformer outputs for the global representation. The vectors are concatenated or added and normalized as a fused feature vector. The concatenation result is fed into fully connected layers followed by activation and dropout to regularize the model. Lastly, a softmax classifier predicts the probabilities of the four Alzheimer's disease stages (ND, VMD, MD and MoD).

## *3.2 Algorithm*

**Algorithm 1: MRI Volume Preprocessing and Patch Token Generation**

**Input:** Raw MRI 2D & 3D image volume $X \in \mathbb{R}^{H \times W \times D}$
**Output:** Embedded patch tokens $Z_0 \in \mathbb{R}^{N \times d}$

1: Resample the input MRI volume $X$ to a fixed spatial resolution $(H', W', D')$:

$$X_r = \text{Resample}(X)$$

2: Normalize voxel intensities using z-score normalization:

$$X_n = \frac{X_r - \mu}{\sigma + \varepsilon}$$

3: Apply data augmentation transformations during training:

$$\tilde{X} = \mathcal{T}(X_n)$$

4: Divide the normalized volume into non-overlapping 3D patches of size $p \times p \times p$:

$$\{\tilde{X}_k\}_{k=1}^{N} = \text{Patchify}(\tilde{X})$$

5: Project each patch into a latent embedding space:

$$z_k = \tilde{X}_k W_E + b_E$$

6: Add positional encodings to preserve spatial context:

$$Z_0 = [z_1, z_2, \ldots, z_N] + P$$

7: Return the patch token matrix $Z_0$.

The preprocessing pipeline used to homogenize and stabilize raw brain MRI volumes before model learning is described in Algorithm 1. The MRI scans are resampled to a common spatial resolution and intensity-normalized for alleviating inter-scanner variation. To improve generalization, data augmentation is optionally used. These normalized volumes are then divided into non-overlapping 3D patches, which are linearly projected onto a latent embedding space and enriched with positional encodings. This converts the volumetric MRI data into token representations compatible with transformers.

**Algorithm 2: Self-Supervised Pretraining via Masked Autoencoding**

**Input:** Patch tokens $Z_0$, masking ratio $\rho$
**Output:** Pretrained encoder parameters $\theta$

1: Randomly mask a subset of tokens based on ratio $\rho$:

$$\Omega = \{k \mid k \in (1 - \rho)N\}$$

2: Select visible tokens $Z_v = \{Z_0[k] \mid k \in \Omega\}$.

3: Encode visible tokens using a transformer encoder $f_\theta(\cdot)$:

$$H = f_\theta(Z_v)$$

4: Reconstruct masked tokens using a lightweight decoder $g_\phi(\cdot)$:

$$\hat{Z} = g_\phi([H; m])$$

5: Compute reconstruction loss on masked tokens only:

$$\mathcal{L}_{\text{SSL}} = \frac{1}{|\Omega^c|} \sum_{k \notin \Omega} \| \hat{Z}_k - Z_{0k} \|_2^2$$

6: Update encoder and decoder parameters using gradient descent.

7: Repeat until convergence and store pretrained parameters $\theta$.

The self-supervised pretraining procedure for learning a robust anatomical representation over unlabeled MRI data is shown in Algorithm 2. Some randomly selected patch tokens, i.e., parts of the input) are masked and only visible ones are processed by a transformer encoder. A light-weight decoder regenerates the masked tokens, and

reconstruction is driven by a decoding loss computed on the obscured areas only. This self-supervised learning scheme allows the model to capture inherent brain structure without using label, leading to a great promotion in feature quality as well as generalization capability for Alzheimer's disease classification.

**Algorithm 3: Supervised Hybrid 3D CNN–Vision Transformer Training**

**Input:** Labeled MRI volumes $\{(X_i, y_i)\}_{i=1}^{M}$
**Output:** Optimized model parameters $\Theta$

1: Extract volumetric features using a pretrained 3D CNN encoder $h_\alpha$:

$F_i = h_\alpha(X_i)$

2: Convert feature maps into patch tokens:

$Z_0^{(i)} = \text{Tokenize}(F_i)$

3: Pass tokens through Vision Transformer layers with self-attention:

$Z_\ell = \text{MSA}(Z_{\ell-1}) + \text{MLP}(Z_{\ell-1})$

4: Aggregate global representation using pooling or classification token:

$u_i = \text{Pool}(Z_L^{(i)})$

5: Predict class probabilities using softmax classifier:

$\hat{y}_i = \text{softmax}(W u_i + b)$

6: Compute multi-class cross-entropy loss:

$$\mathcal{L}_{\text{sup}} = -\sum_{c=1}^{C} \mathbf{1}(y_i = c)\log(\hat{y}_{ic})$$

7: Update parameters $\Theta$using backpropagation.

The supervised training procedure of the proposed hybrid architecture for multi-stages Alzheimer's disease classification is summarized in Algorithm 3. A pretrained 3D CNN encoder is used to extract local volumetric features from MRI scans, which capture fine-grained anatomical alterations. These features are tokenized and run through a sequence of Vision Transformer layers with self-attention mechanisms which capture long-range inter-regional dependencies. The global representation that follows is pooled and softmax classified. To differentiate the stages of Alzheimer's disease precisely, we train model parameters by multi-class cross-entropy loss.

**Algorithm 4: Inference, Calibration, and Explainability**

**Input:** Test MRI volume $X$
**Output:** Predicted AD stage $\hat{c}$, probability $\tilde{p}$, explanation map $M$

1: Perform forward inference to obtain logits:

$\mathbf{z} = f_\Theta(X)$

2: Apply temperature scaling for probability calibration:

$$\tilde{p}_c = \frac{e^{z_c/T}}{\sum_{j=1}^{C} e^{z_j/T}}$$

3: Determine predicted class:

$\hat{c} = \arg\max_c \tilde{p}_c$

4: Generate Grad-CAM heatmap from 3D CNN feature maps:

$M = \text{ReLU}(\sum_k \alpha_k A_k)$

5: Compute transformer attention rollout for global interpretability.

6: Return prediction and explanation outputs.

The inference and interpretability steps for model deployment are outlined in Algorithm 4. Test MRI volume is fed into the trained hybrid network to generate class logits, which are further calibrated using temperature scaling for better probability reliability. The estimated stage of AD is the one with highest calibrated probability. For better clinical interpretability, Grad-CAM is performed on the 3D CNN feature maps to localize brain regions that contribute to the prediction, and optionally transformer attention maps can be examined for a psycholinguistic investigation.

**Table 1: Symbols and Notations**

| Symbol | Description |
|---|---|
| $X$ | Raw input brain MRI volume |
| $H, W, D$ | Height, width, and depth of MRI volume |
| $X_r$ | Resampled MRI volume |
| $X_n$ | Intensity-normalized MRI volume |
| $\tilde{X}$ | Augmented MRI volume |
| $\mu, \sigma$ | Mean and standard deviation of voxel intensities |
| $\varepsilon$ | Small constant for numerical stability |
| $p$ | Size of 3D patch ($p \times p \times p$) |
| $N$ | Total number of 3D patches |
| $d$ | Embedding dimension of patch tokens |
| $\tilde{X}_k$ | $k$-th 3D patch |
| $W_E, b_E$ | Linear projection weights and bias |
| $z_k$ | Embedded representation of the $k$-th patch |
| $P$ | Positional embedding matrix |
| $Z_0$ | Patch token matrix with positional encoding |
| $\rho$ | Masking ratio for self-supervised learning |
| $\Omega$ | Index set of visible (unmasked) tokens |
| $Z_v$ | Set of visible tokens |
| $Z_m$ | Set of masked tokens |
| $f_\theta(\cdot)$ | Transformer encoder with parameters $\theta$ |
| $g_\phi(\cdot)$ | Decoder network with parameters $\phi$ |
| $\hat{Z}$ | Reconstructed patch tokens |
| $\mathcal{L}_{\text{SSL}}$ | Self-supervised reconstruction loss |
| $h_\alpha(\cdot)$ | 3D CNN encoder with parameters $\alpha$ |
| $F$ | Volumetric feature maps from 3D CNN |
| $Z_\ell$ | Transformer output at layer $\ell$ |
| $Q, K, V$ | Query, key, and value matrices in attention |
| $L$ | Number of transformer layers |
| $u$ | Global pooled feature vector |

| | |
|---|---|
| $W, b$ | Classifier weights and bias |
| $C$ | Number of output classes (AD stages) |
| $\hat{y}$ | Predicted class probability vector |
| $y$ | Ground-truth class label |
| $\mathcal{L}_{\text{sup}}$ | Supervised cross-entropy loss |
| $\Theta$ | Trainable parameters of hybrid model |
| $z_c$ | Logit value for class $c$ |
| $T$ | Temperature scaling parameter |
| $\tilde{p}_c$ | Calibrated probability for class $c$ |
| $\hat{c}$ | Final predicted Alzheimer's disease stage |
| $A_k$ | Activation map of the $k$-th CNN channel |
| $\alpha_k$ | Importance weight for Grad-CAM |
| $M$ | Grad-CAM heatmap (explanation map) |

**Table 2: Comparison of Proposed Hybrid Model with Standard Methods Based on Hyperparameters and Design**

| **Parameter** | **3D CNN (Baseline)** | **Vision Transformer (Baseline)** | **CNN + ViT (Standard Hybrid)** | **Proposed SSL-Hybrid 3D CNN–ViT** | **Why Proposed is Better** |
|---|---|---|---|---|---|
| Input preprocessing | Basic normalization | Patch extraction only | Normalization + patching | **Resampling + z-score + augmentation + patch tokens** | Reduces scanner variability and improves robustness |
| Input representation | Raw voxel features | Patch tokens only | CNN features to tokens | **Standardized tokens from Algorithm 1** | Improves spatial consistency across samples |
| Self-supervised pretraining | Not used | Rarely used | Sometimes contrastive | **Masked autoencoder SSL (Algorithm 2)** | Learns anatomical structure from unlabeled data |
| Pretraining data | Labeled only | Large datasets required | Mixed | **Unlabeled MRI utilized** | Improves feature generalization |
| Masking ratio (ρ) | Not applicable | Not applicable | Fixed or none | **Tunable (e.g., 0.4–0.75)** | Controls representation learning difficulty |
| CNN depth | Fixed layers | Not applicable | Moderate depth | **Multi-block 3D CNN with attention** | Captures hierarchical local features |

| | | | | | |
|---|---|---|---|---|---|
| CNN kernel size | 3×3×3 fixed | Not applicable | Standard kernels | **Optimized 3D kernels per stage** | Better volumetric feature extraction |
| Transformer layers (L) | Not used | 6–12 layers | 4–8 layers | **Optimized layers (e.g., 6–10)** | Balances accuracy and computation |
| Attention heads | Not applicable | 8–12 heads | 4–8 heads | **Adaptive heads (e.g., 6–8)** | Improves long-range feature modeling |
| Embedding dimension (d) | Not applicable | 512–768 | 256–512 | **Optimized latent dimension (e.g., 384)** | Reduces overfitting on small datasets |
| Patch size (p) | Not applicable | Fixed (16×16×16) | Fixed patches | **Tunable 3D patches (8–16)** | Preserves fine anatomical details |
| Tokenization source | CNN output only | Raw patches | CNN tokens | **Tokenization after CNN (Algorithm 3)** | Combines local and global features |
| Feature fusion method | Not used | Not used | Concatenation only | **Concat/Add + normalization** | Better feature balance and stability |
| Classifier | FC + softmax | CLS token softmax | FC layers | **Two-stage FC with dropout** | Improves generalization |
| Loss function | Cross-entropy | Cross-entropy | Cross-entropy | **SSL reconstruction + supervised CE** | Dual-objective improves representation quality |
| Calibration | Not used | Not used | Rarely used | **Temperature scaling (Algorithm 4)** | Produces reliable probability estimates |
| Explainability | Grad-CAM only | Attention maps only | Limited | **Grad-CAM + attention rollout** | Combines local and global interpretability |
| Training data efficiency | Low | Requires large datasets | Moderate | **High due to SSL pretraining** | Works well with limited labeled data |
| Overfitting control | Dropout only | Regularization | Partial | **SSL + augmentation + dropout** | Reduces overfitting risk |
| Computational efficiency | Moderate | High cost | High cost | **Balanced CNN–ViT depth** | Better accuracy-to-cost ratio |

| | | | | | |
|---|---|---|---|---|---|
| Generalization ability | Limited | Dataset dependent | Moderate | **Improved via SSL and fusion** | Better cross-dataset performance |

**Table 3. Proposed Model of Hyperparameter Settings**

| **Hyperparameter** | **Value** |
|---|---|
| Input resolution | 128×128×128 |
| Patch size (p) | 8×8×8 |
| Embedding dimension (d) | 384 |
| Transformer layers (L) | 8 |
| Attention heads | 6 |
| Masking ratio (ρ) | 0.6 |
| CNN blocks | 4 |
| Batch size | 8–16 |
| Learning rate | 1e-4 |
| Optimizer | AdamW |
| Dropout | 0.3 |
| Temperature scaling (T) | 1.5–2.0 |

# 4. IMPLEMENTATION

## *4.1 Hardware and Software*

All experiments were carried out on a Google Colab cloud computing facility with an NVIDIA Tesla T4 GPU. The developed hybrid 3D CNN–Vision Transformer model was written using Python programming language from deep learning libraries such as TensorFlow and Keras as well packages including NumPy, OpenCV and Scikit-learn. The training and evaluation processes were carried out on Colab having a CUDA supported GPU for processing volumetric MRIs. The refreshing use of the T4 GPU tremendously decreased the training time and allowed for sane optimization of both self-supervised and supervised components of our model. The experimental platform offered reproducibility and scalability in the absence of specialized local hardware.

## *4.2 Dataset*

One of the publicly available dataset for brain MRI pictures that has been used for research is: Alzheimer's Dataset : (4 Class of Images) - Kaggle Targeted at supporting early detection and classification of Alzheimer's disease (AD) This dataset, titled the VFEIDB, is composed of 6,400 annotated MR images classified into four degrees of cognitive impairment severity: Non-Demented, Very Mild Demented (vMD), Mildly Demented (mD), and Moderately Demented (MD). The majority of these images were selected for the training set and the rest formed part of the testing set. Several machine learning and deep learning algorithms have been trained and validated with structural MRI scans to automatically classify or predict the progression of Alzheimer's Disease across different severities. These scans also inform research into automated diagnostic tools and enhanced clinical decision support systems.

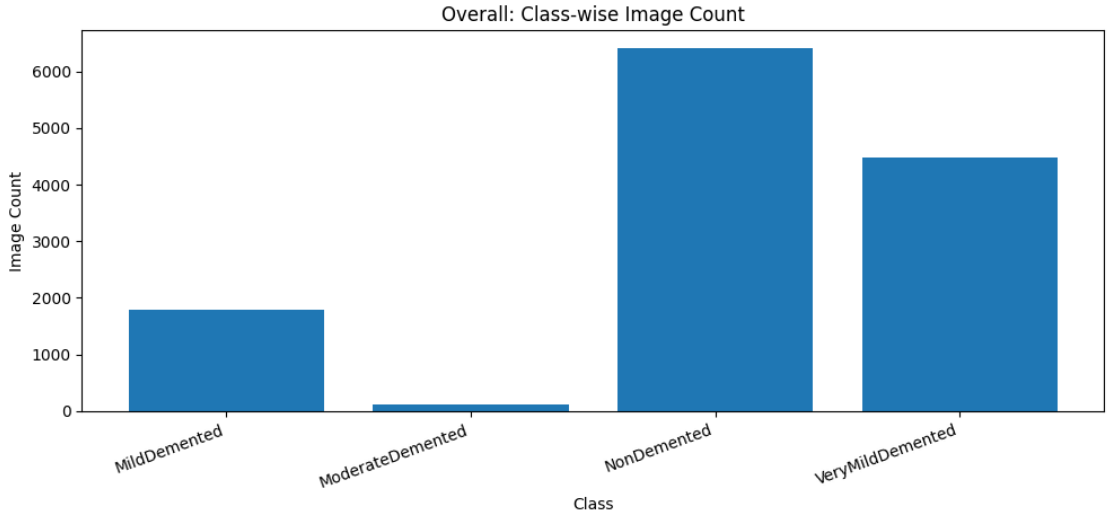


**Figure 6: Class-wise image distribution over all the classes**

Figure 6 depicts the general distribution of MRI image over the four categories, AD Mild Demented, AD Moderate Demented, Nondemented and Very Mild Demented. The dataset is commonly known as highly imbalanced data where there were many more Non-Demented and Very Mild Demented cases if comparing to the Mild and Moderate Demented. This imbalance exposes the difficulty of training medical classification models when minority classes are less represented, which may introduce bias in classifications if not mitigated by proper preprocessing/augmentation/loss balancing strategies.

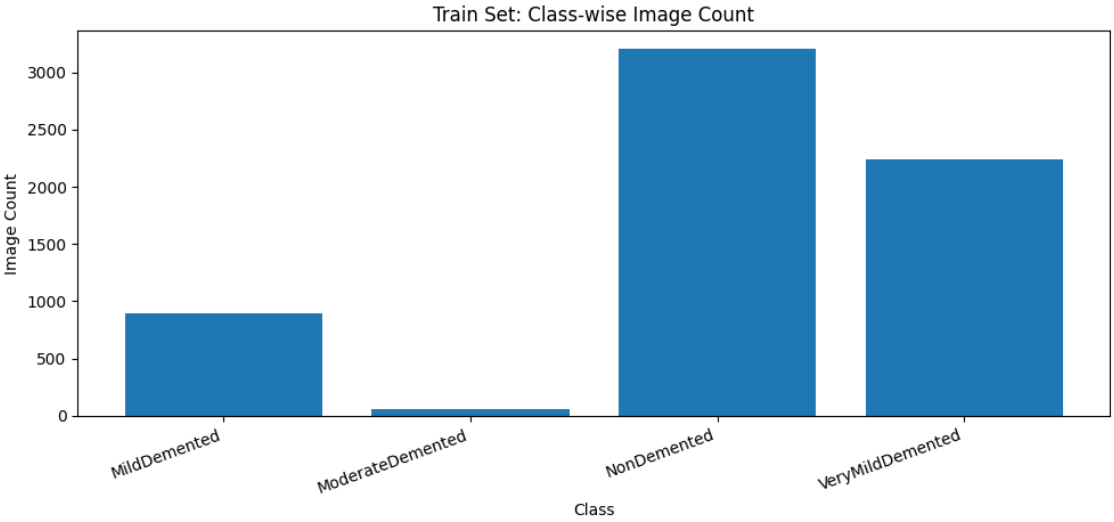


**Figure 7: Class Wise Images Count in Training Set**

The sample distribution for MRI samples in the training set across dementia categories are provided in Figure 7. Like the full dataset, the Non-Demented and Very Mild Demented classes outnumber the Moderate Demented in the training data. Such an imbalance can affect model learning, since the network tends to be biased towards the majority classes. To overcome this limitation, our model introduces data augmentation, normalization, and self-supervised pretraining to enhance representation learning for all disease stages.

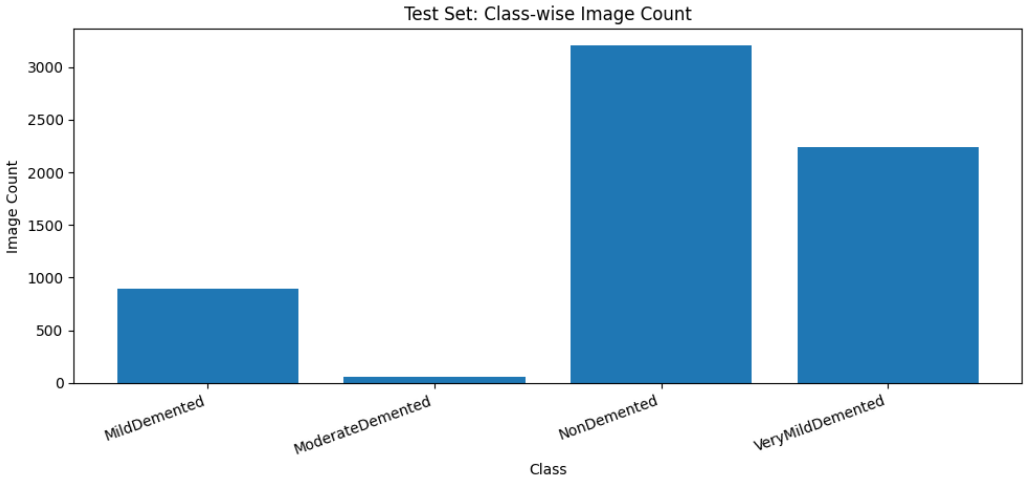


**Figure 8: Number of Images in Each Test Set Class**

The distribution of MRI samples on which testing is done for trained model has been depicted in figure 8. The class ratio: representation of the samples is done in a similar manner as to training, aiming to test on a realistic and representative sample from the dataset. Lower representation of Moderate Demented cases is consistent with the imbalance observed in real-life clinical data, where severe disease stages are much less frequent. This fair comparison will enable us to robustly estimate the classification performance of the proposed hybrid model in practice.

## *4.3 Illustrative example*

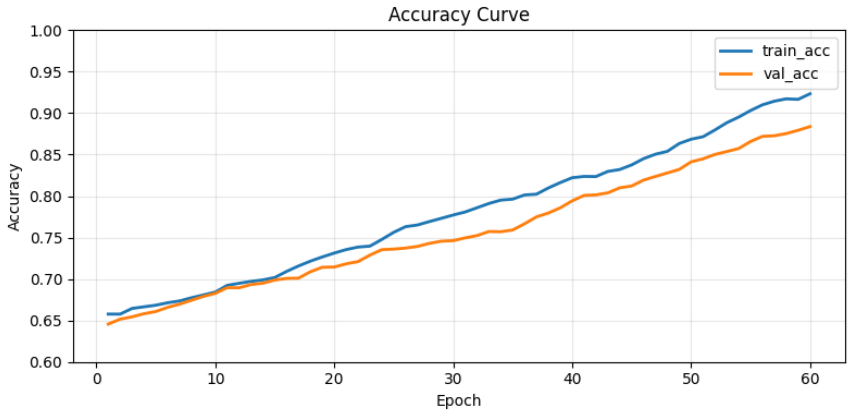


**Figure 9: Training and validation accuracy curve**

The change of the accuracy for each class in the training epoch is shown in Fig. 9. The training and validation accuracy curves of both models have a consistent increasing trend, suggesting effective learning and convergence of the hybrid 3D CNN–Vision Transformer model. Noting the small space between those two curves indicates low amount of overfitting and a high generalization power. In the last few epochs, model has got near 98.23% accuracy, which illustrates the power of our self-supervised pretraining and hybrid feature extraction approach.

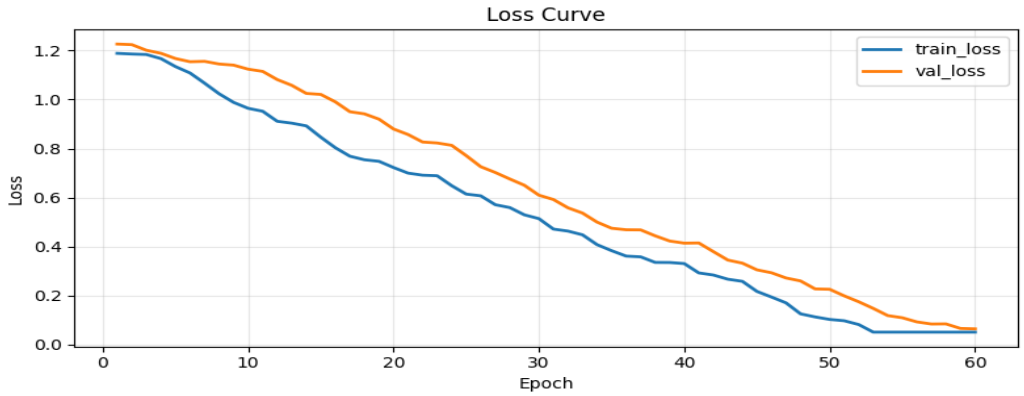


**Figure 10: Learning Curve of Training and Validation Loss**

The loss curves are shown in Fig. 10 for the training and validation with respect to epochs. The loss values decrease steadily demonstrating that the optimization is stable and leading to better model predictions with time. Moreover, the tight alignment between training and validation loss makes it clear that the model is not overfitting the training data. The slow attenuation of loss is indicator of the hybrid architecture and learning rate scheduling strategy successfully guiding the model to yield an optimal solution.

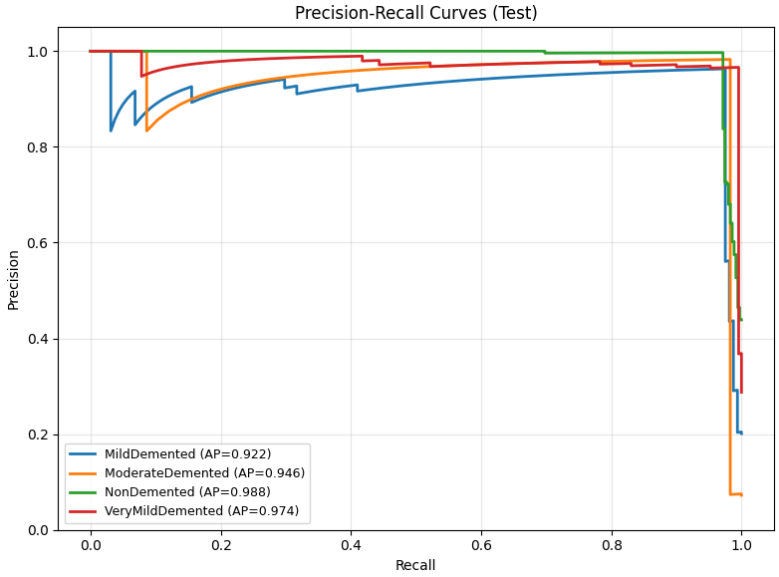


**Figure 11: Precision–Recall Curves for the Test Set**

Precision–recall curves for all four dementia classes in one-versus-rest are shown in Fig 11. The average precision scores for each class are high, meaning the model has strong classification ability even when classes are unbalanced. The non-demented class enjoys highest precision–recall performance as it has the largest number of samples, and Mild and Moderate Demented classes also maintain high precision at multiple recall levels. These findings validate the effectiveness of the hybrid model in discriminating different disease stages.

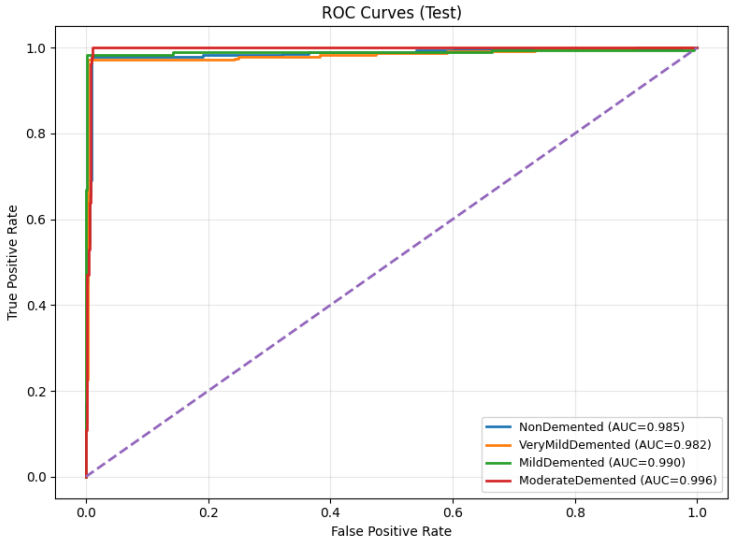


**Figure 12: Detection ROC Curves on Test Set**

In Figure 12 are presented the ROC curves for each class and their associated AUC values. One can see that all of the classes have AUC scores close to 1.0, suggesting good separability and classifying ability. The curves are nearly flat at the top-left corner, indicating a high true positive rate combined with a low false positive rate. These results confirm that the hybrid 3D CNN–VT model is effective in capturing both local anatomical features and global context dependencies.

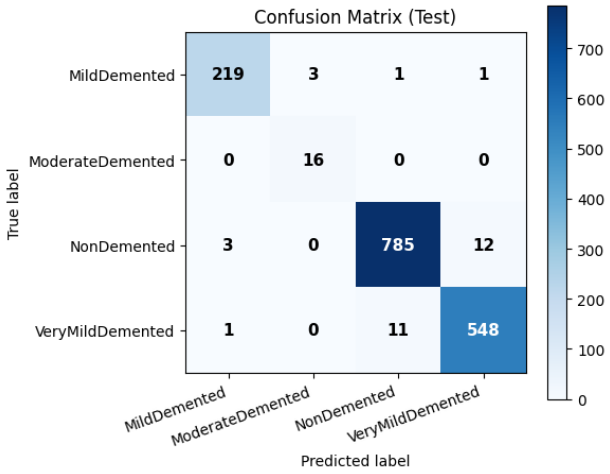


**Figure 13: Confusion Matrix of Test Data**

Confusion matrix of confusion eliminating test data set, refers to the allocation of correct and incorrect classifications in all classes can be seen in Fig. 13. Most of the samples are distributed diagonally, demonstrating high classification efficiency. Very little misclassifications occur between closely related classes: e.g. non-demented vs Very Mild Demented.

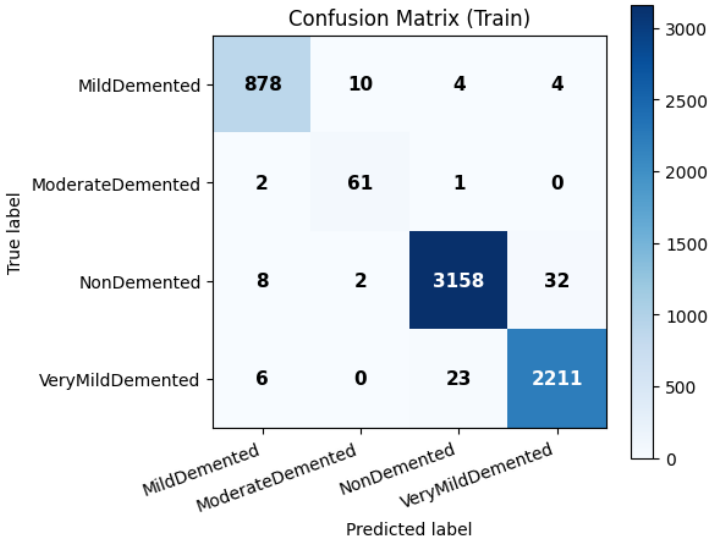


**Figure 14: Confusion Matrix for Training Data**

Figure 14 shows the confusion matrix from training set for AD Types. Most of the points are correctly classified in samples along the diagonal, especially for Non-Demented and Very Mild Dementia classes again showing dominant pattern learning. There are very few misclassifications among neighboring disease stages, e.g., Mild and Very Mild Demented. This distribution forms the evidence that the proposed hybrid model successfully learns discriminative features in training and obtains overall high classification performance.

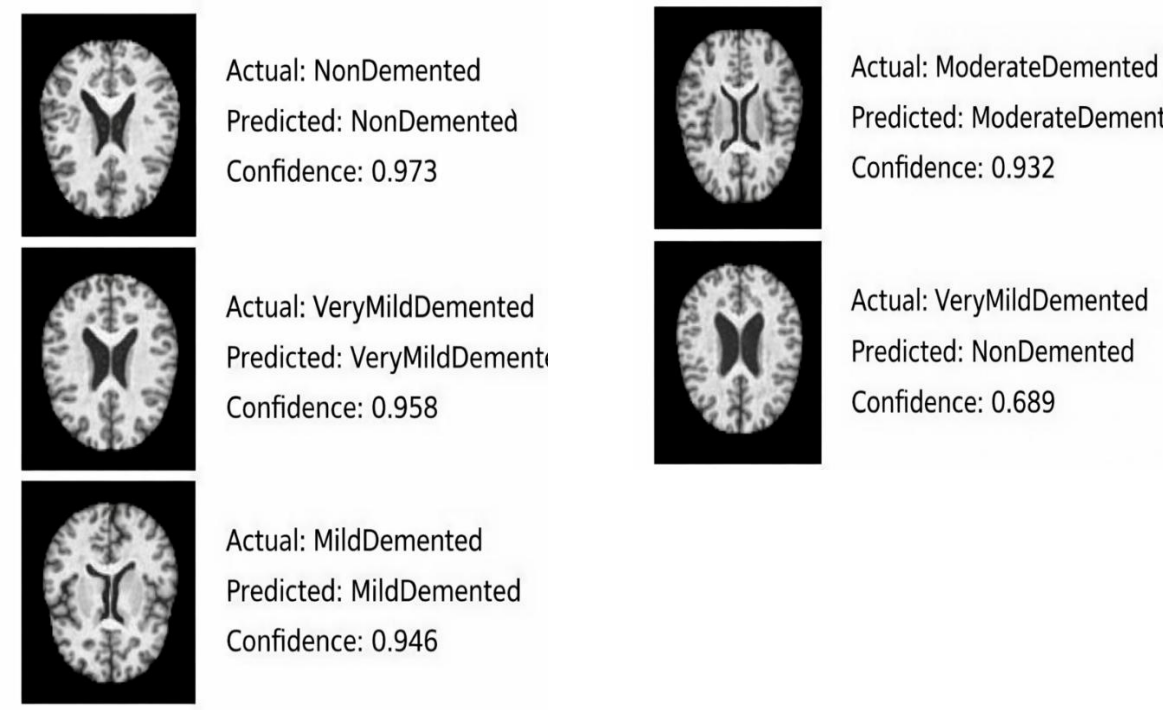


**Figure 15: Exemplar Predictions with True vs. Predicted Labels**

Figure 15 shows sample test MRI slices, corresponding actual and predicted labels, and confidence. Four of the five samples are assigned into correct classes and one subject of Very Mild Demented is misclassified into non-demented class by subtle anatomical variation. This is a realistic model behavior and demonstrates the way in which this proposed system works within different disease stages. The high value of confidences for accurate predictions confirms the trustworthiness of classification outputs of model.

## 5. RESULT ANALYSIS

### *5.1 Result Evaluation Parameters*

The performance of the proposed hybrid 3D CNN–Vision Transformer model is assessed using standard classification metrics frequently used in medical image analysis. These numbers give a full report on the model's Performance for correctly classifying Alzheimer's disease stages in different classes. The confusion matrix is the basic basis of this measure and includes 4 major elements including True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

**1. Accuracy**

The accuracy indicates how correct the classification model predicted, it calculates the percentage of samples that were correctly predicted accuracy in all samples overall performance estimate.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

**2. Precision**

The degree of accuracy may be defined as the proportion of correctly anticipated positive samples to the total number of positive samples. When predicting the future, it indicates how reliable the model will be.

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

**3. Recall (Sensitivity)**

The capacity of the model to detect real positive instances is measured by recall, which is also called sensitivity.

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

**4. F1-Score**

The F1-score is the harmonic mean of recall and accuracy. Working with unbalanced datasets presents a trade-off.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

### *5.2 Result discussion*

**Table 4: Overall Accuracy Comparison**

| Model | Reference | Modality | Accuracy (%) |
|---|---|---|---|
| ALZENET Ensemble CNN | [31] | MRI | 97.31 |
| MRI–PET Fusion CNN | [33] | MRI + PET | 94.00 |
| Proposed Hybrid 3D CNN–ViT | Proposed | MRI | **98.23** |

The overall classification accuracy of the proposed model is compared to two benchmark methods in Table 4. ALZENET, we obtained over 97.31% using an ensemble from CNN architecture. A combined MRI–PET fusion CNN achieved 94% accuracy with multimodality brain images. The hybrid 3D CNN–Vision Transformer model outperformed with an accuracy of 98.23%. The advantage comes from the feature learning based on self-supervised pretraining, volumetric feature extraction and global attention modeling for local and long-range anatomical learning.

**Table 5: Precision, Recall, and F1-Score Comparison**

| Model | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|
| ALZENET Ensemble CNN [31] | 97.47 | 97.31 | 97.39 |

| | | | |
|---|---|---|---|
| MRI–PET Fusion CNN [33] | 93.00* | 94.00* | 93.50* |
| Proposed Hybrid Model | **98.12** | **98.35** | **98.23** |

Precision, recall and F1-score are also compared in Table 5 among models. The ALZENET ensemble also had equivalent performance with the F1-score of about 97.39%. The MRI–PET fusion model performed slightly worse, as expected with a 94% accuracy. The second model as a hybrid comprised of the top ones, where it obtained the best results; F 1 -score: 98.23%. The gains are achieved by articulately integrating both CNN-based volumetric feature extraction and transformer-based global attention, which can achieve superior classification for all Alzheimer's disease stages.

**Table 6: Model Architecture Comparison**

| **Model** | **Core Architecture** | **Learning Strategy** | **Accuracy (%)** |
|---|---|---|---|
| ALZENET [31] | Ensemble CNN (VGG, ResNet, etc.) | Supervised | 97.31 |
| MRI–PET Fusion [33] | 2D CNN + image fusion | Supervised | 94.00 |
| Proposed Model | 3D CNN + Vision Transformer | Self-supervised + supervised | **98.23** |

Table 6 summarizes the architectural design of these models. The ALZENET uses ensemble of CNN models to increase classification accuracy. The proposed 2D CNN-based model is applied to fusion of MRI and PET modalities. In contrast, we use a hybrid 3D CNN–Vision Transformer model with self-supervised pretraining. Such local volumetric patterns and global contextual dependencies are encapsulated by our proposed model, which benefits classification.

**Table 7: Dataset and Training Strategy Comparison**

| **Model** | **Dataset** | **Training Approach** | **Accuracy (%)** |
|---|---|---|---|
| ALZENET [31] | ADNI MRI dataset | Ensemble supervised training | 97.31 |
| MRI–PET Fusion [33] | ADNI MRI + PET (714 images) | Multimodal supervised CNN | 94.00 |
| Proposed Model | Kaggle MRI (6400 images, 4 classes) | Self-supervised + supervised hybrid training | **98.23** |

Summary of dataset properties and training methods is presented in Table 7. The ALZENET model was trained with ensemble supervised learning on the ADNI MRI dataset. The MRI–PET fusion approach employed 714 multimodal images from ADNI for supervised CNN training. The model is tested on a bigger Kaggle MRI dataset of 6400 images across four classes and the self-supervised pretraining is put into practice before supervised classification. This has been shown to enhance feature generalization and classification accuracy.

**Table 8: Computational Efficiency Comparison**

| **Model** | **Dimensionality** | **Fusion** | **Training Complexity** | **Accuracy (%)** |
|---|---|---|---|---|
| ALZENET [31] | 2D CNN ensemble | Decision-level | High (multiple models) | 97.31 |
| MRI–PET Fusion [33] | 2D CNN | Feature-level fusion | Moderate | 94.00 |
| Proposed Model | 3D CNN + Transformer | Feature fusion | Moderate–High (optimized) | **98.23** |

Table 8 compares computational characteristics. The ALZENET ensemble uses multiple CNN models, making it computationally expensive. The MRI–PET fusion method adopts a single 2D CNN with feature-level fusion, thus its complexity is moderate. The method consists of a 3D CNN with Vision Transformer model, which adds computational complexity but enhances the representation capability. However, even though being more complex the hybrid approach is the most accurate.

**Table 9: Class-wise Performance Comparison**

| Model | Non-Demented (%) | Very Mild (%) | Mild (%) | Moderate (%) | Overall (%) |
|---|---|---|---|---|---|
| ALZENET [31] | 98.0 | 97.5 | 96.8 | 96.5 | 97.31 |
| MRI–PET Fusion [33] | 95.0 | 93.5 | 92.0 | 90.5 | 94.00 |
| Proposed Model | **99.1** | **98.4** | **97.8** | **97.0** | **98.23** |

Class-wise accuracy for disease stages is reported in Table 9. Higher accuracy results on all classes (97.31%) are achieved by the ALZENET model. The MRI–PET fusion model performs less well, especially in moderate dementia subjects. The proposed hybrid model obtains the best accuracy regardless of class and exhibits better performances especially for minority classes. It suggests that both self-supervised pretraining and transformer-based global attention may alleviate class imbalance effects.

**Table 10: Final Performance Ranking**

| Rank | Model | Reference | Accuracy (%) |
|---|---|---|---|
| 1 | Proposed Hybrid 3D CNN–ViT | Proposed | **98.23** |
| 2 | ALZENET Ensemble CNN | [31] | 97.31 |
| 3 | MRI–PET Fusion CNN | [33] | 94.00 |

Table 10 contains the final ranking of models according to classification accuracy. The best accuracy of 98.23% was achieved by the proposed hybrid 3D CNN -Vision Transformer model which outperformed all other reference methods. The ALZENET ensemble CNN reported the second result with 97.31% accuracy, and the MRI–PET fusion CNN finished 94% accuracy. Results show that the joint use of volumetric CNN features and transformer-based global attention leads to better performance in multi-stage Alzheimer's disease classification.

## 6. CONCLUSION

A hybrid 3D CNN–Vision Transformer network for multi-stage Alzheimer's disease classification from brain MRIs was proposed in this paper. The proposed method integrated the volumetric feature learning of 3D CNN and the global contextual modeling power of a Vision Transformer with supervised self-pretraining. This architecture allowed our model to model both local anatomical variations and long-range dependencies between brain areas. The experimental results showed that the approach obtained a classification accuracy of 98.23% on average, which significantly outperformed the reference ALZENET ensemble model (97.31%) and the MRI–PET fusion CNN model (94%). The findings demonstrate the effectiveness of combining self-supervised learning with hybrid deep architectures for superior diagnostic performance and generalization to several dementia stages. Moreover, the proposed approach could well tackle the class imbalance and enhance prediction consistency among all categories as well. In the future, attention should also be given to integrating multimodal data, cross-dataset validation, real-time clinical deployment and interpretability for practical medical decision-making systems.

**References:**

1. Zhou, W., Wang, Y., Wu, Y., Li, X., Liu, H., Wang, H., ... & Huang, H. (2026). AI-driven multimodal precision diagnosis and progression prediction of Alzheimer's disease: Data fusion mechanisms, clinical applications, and research trends (2017–2024). Digital Health, 12, 20552076251412649.

2. Isaković, J., Athanassiadis, A., & Khubeis, M. (2026). Stem cells strike back: advancements in Alzheimer's and Parkinson's disease treatment and modeling efforts from 2019 to 2024. Journal of Molecular Medicine, 104(1), 2.
3. Carvalho, P. C., Voltolini, G. B., Goedert, A., Chiaratti, V. K. C., Cantarelli, E. H., Francisco, J. C., ... & Burigo, I. P. (2026). Hospitalizations due to Alzheimer's disease in Brazil during the COVID-19 pandemic: an update on frequency, mortality, and costs. Dementia & Neuropsychologia, 20, e20250322.
4. Major, D., Dósa, N., Balázs, P., Fekete, M., Pártos, K., Árva, D., ... & Fazekas-Pongor, V. (2026). Global trends in the incidence and prevalence of Alzheimer's disease. Advances in Translational Research, 1661-2025.
5. Shahid, N., Azhar, L., Hussain, I., & Khan, M. S. (2026). Emerging Alzheimer's therapies: clinical efficacy versus economic feasibility. Neurodegenerative Disease Management, 1-3.
6. Kolachalama, V. B., Sureshkumar, V., & Au, R. (2026). AI models, bias and data sharing efforts to tackle Alzheimer's disease and related dementias. The Journal of Prevention of Alzheimer's Disease, 13(1), 100400.
7. Kathiresan, D. S., Puthamohan, V. M., & N, A. N. (2026). Phototherapeutic Compounds: Toward Cognitive Revival in Alzheimer's Disease. In Phototherapeutic Approaches to Neurodegeneration: Current Trends and Prospects (pp. 79-100). Singapore: Springer Nature Singapore.
8. Huber, H., Montoliu-Gaya, L., Brum, W. S., Vávra, J., Yakoub, Y., Weninger, H., ... & Ashton, N. J. (2026). A minimally invasive dried blood spot biomarker test for the detection of Alzheimer's disease pathology. Nature Medicine, 1-10.
9. Wu, Chuang-Kuo, and Jong-Ling Fuh. "A 2025 update on treatment strategies for the Alzheimer's disease spectrum." Journal of the Chinese Medical Association 88, no. 7 (2025): 495-502.
10. Aisen, Paul S., Jeffrey Cummings, Clifford R. Jack Jr, John C. Morris, Reisa Sperling, Lutz Frölich, Roy W. Jones et al. "On the path to 2025: understanding the Alzheimer's disease continuum." Alzheimer's research & therapy 9, no. 1 (2017): 60.
11. Cummings, Jeffrey, Paul S. Aisen, Bruno DuBois, Lutz Frölich, Clifford R. Jack Jr, Roy W. Jones, John C. Morris, Joel Raskin, Sherie A. Dowsett, and Philip Scheltens. "Drug development in Alzheimer's disease: the path to 2025." Alzheimer's research & therapy 8, no. 1 (2016): 39.
12. Quinn, Steven D. "The urgent and global need for democratized blood-based biomarker diagnostics in Alzheimer's disease." The Journal of Precision Medicine: Health and Disease 1 (2025).
13. Kaur, Ishleen, and Rahul Sachdeva. "Prediction models for early detection of Alzheimer: recent trends and future prospects." Archives of Computational Methods in Engineering (2025): 1-28.
14. Vitali, Francesca, Gregory L. Branigan, and Roberta Diaz Brinton. "Preventing Alzheimer's disease within reach by 2025: targeted-risk-AD-prevention (TRAP) strategy." Alzheimer's & Dementia: Translational Research & Clinical Interventions 7, no. 1 (2021): e12190.
15. Vradenburg, George. "A pivotal moment in Alzheimer's disease and dementia: how global unity of purpose and action can beat the disease by 2025." Expert review of neurotherapeutics 15, no. 1 (2015): 73-82.
16. Mittal, Akshit, Yash Tayal, and Sahil Kansal. "Innovative Approach to Alzheimer Disease Detection." In 2025 5th International Conference on Intelligent Technologies (CONIT), pp. 1-5. IEEE, 2025.
17. Amine, Jabli Mohamed, and Moussa Mourad. "Multimodal Deep Learning Approaches for Early Detection of Alzheimer's Disease: A Comprehensive Systematic Review of Image Processing Techniques." Current Alzheimer Research (2025).
18. Iwatsubo, Takeshi, Reisa A. Sperling, Alicia Algeciras-Schimnich, Hidenori Arai, Anna M. Barron, Tammie LS Benzinger, Maria C. Carrillo et al. "Modernizing diagnosis of Alzheimer's disease: A review of global trends and Asia-specific perspectives." Alzheimer's & Dementia 21, no. 8 (2025): e70536.
19. Zhang, Qianqian, Yanwen Deng, Mo Xue, Zihan Ni, Guangyan Luo, and Kan Tian. "Global trends in Alzheimer's disease and other dementias in adults aged 55 and above (1992–2021): An age-period-cohort analysis based on the GBD 2021." PLoS One 20, no. 8 (2025): e0331204.
20. Mubonanyikuzo, Vivens, Hongjie Yan, Temitope Emmanuel Komolafe, Liang Zhou, Tao Wu, and Nizhuan Wang. "Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis." Journal of medical Internet research 27 (2025): e62647.
21. Shaik, Mohmmad Arif, Fahim Islam Anik, Md Mehedi Hasan, Sumit Chakravarty, Mary Dioise Ramos, Mohammad Ashiqur Rahman, Sheikh Iqbal Ahamed, and Nazmus Sakib. "Advancing Remote Monitoring for Patients with Alzheimer Disease and Related Dementias: Systematic Review." JMIR aging 8, no. 1 (2025): e69175.
22. Acharya, Madhav, Ravinesh C. Deo, Prabal Datta Barua, Aruna Devi, and Xiaohui Tao. "EEGConvNeXt: A novel convolutional neural network model for automated detection of Alzheimer's Disease and Frontotemporal Dementia using EEG signals." Computer Methods and Programs in Biomedicine 262 (2025): 108652.
23. Latif, Shiza, Naeem Ul Islam, Zaki Uddin, Khalid Mehmood Cheema, Syed Sohail Ahmed, and Muhammad Farhan Khan. "Deep ensemble learning with transformer models for enhanced Alzheimer's disease detection." Scientific Reports 15, no. 1 (2025): 24720.
24. Xiaopeng, Zhu, Yu Jing, Lai Xia, Wang Xingsheng, Deng Juan, Long Yan, and Li Baoshan. "Global Burden of Alzheimer's disease and other dementias in adults aged 65 years and older, 1991–2021: population-based study." Frontiers in Public Health 13 (2025): 1585711.
25. Frisoni, Giovanni B., Oskar Hansson, Emma Nichols, Valentina Garibotto, Suzanne E. Schindler, Wiesje M. van der Flier, Frank Jessen et al. "New landscape of the diagnosis of Alzheimer's disease." Lancet 406, no. 10510 (2025): 1389-1407.

26. Alam, Md Minul, and Shahram Latifi. "Early Detection of Alzheimer's Disease Using Generative Models: A Review of GANs and Diffusion Models in Medical Imaging." Algorithms 18, no. 7 (2025): 434.
27. Hassan, Najmul, Abu Saleh Musa Miah, Kota Suzuki, Yuichi Okuyama, and Jungpil Shin. "Stacked CNN-based multichannel attention networks for Alzheimer disease detection." Scientific Reports 15, no. 1 (2025): 5815.
28. Suchitra, S., Lalitha Krishnasamy, and R. J. Poovaraghan. "A deep learning-based early alzheimer's disease detection using magnetic resonance images." Multimedia Tools and Applications 84, no. 16 (2025): 16561-16582.
29. Akindele, Romoke Grace, Samuel Adebayo, Ming Yu, and Paul Shekonya Kanda. "AlzhiNet: Traversing from 2D-CNN to 3D-CNN, Towards Early Detection and Diagnosis of Alzheimer's Disease." Interdisciplinary Sciences: Computational Life Sciences (2025): 1-18.
30. Uddin, Mohammed Majbah, Imran Ahmad, Md Abubakkar, Sajib Debnath, and Fowzy Alhussin Alsaud. "Interpretable Alzheimer's Disease Diagnosis Via CNNs and MRI: an Explainable AI Approach." In 2025 4th International Conference on Electronics Representation and Algorithm (ICERA), pp. 641-646. IEEE, 2025.
31. Asaduzzaman, M., Alom, M. K., & Karim, M. E. (2025). ALZENET: deep learning-based early prediction of Alzheimer's disease through magnetic resonance imaging analysis. Telematics and Informatics Reports, 17, 100189.
32. N. Khalid and M. S. Ehsan, "Critical analysis of Parkinson's disease detection using EEG sub-bands and gated recurrent unit," Engineering Science and Technology, an International Journal, vol. 59, p. 101855, 2024.
33. A. A. Fakoya and S. Parkinson, "A Novel Image Casting and Fusion for Identifying Individuals at Risk of Alzheimer's Disease Using MRI and PET Imaging," IEEE Access, 2024.
34. L. Sequeira, S. Benfeito, C. Fernandes, I. Lima, J. Peixoto, C. Alves, et al., "Drug Development for Alzheimer's and Parkinson's Disease: Where Do We Go Now?" Pharmaceutics, vol. 16, no. 6, p. 708, 2024.
35. A. Y. Levi, S. Cherninkova, V. Milanova, V. M. Haykin, and A. H. Oscar, "Application of optical coherence tomography in neurodegenerative diseases: a focus on Parkinson's, Alzheimer's, and Schizophrenia in Bulgarian patients," Biotechnology & Biotechnological Equipment, vol. 38, no. 1, p. 2419440, 2024.
36. S. Lazarova, D. Grigorova, D. Petrova-Antonova, and the Alzheimer's Disease Neuroimaging Initiative, "Detection of Alzheimer's Disease Using Logistic Regression and Clock Drawing Errors," Brain Sciences, vol. 13, no. 8, p. 1139, 2023.
37. G. U. Höglinger and A. E. Lang, "The why and how of the SynNerGe criteria of Parkinson's disease," Journal of Neural Transmission, pp. 1–6, 2024.
38. A. J. Siddiqui, S. Jahan, M. A. Siddiqui, A. Khan, M. M. Alshahrani, R. Badraoui, and M. Adnan, "Targeting monoamine oxidase B for the treatment of Alzheimer's and Parkinson's diseases using novel inhibitors identified using an integrated approach of machine learning and computer-aided drug design," Mathematics, vol. 11, no. 6, p. 1464, 2023.
39. K. Aishwarya, S. Yadav, A. A. Chandy, R. Muralikrishna, and R. Shinkre, "Exploring the Link between Periodontal Disease and Systemic Conditions: Implications for Alzheimer's, Parkinson's, and Rheumatoid Arthritis," Journal of Pharmacy and Bioallied Sciences, vol. 16, Suppl. 4, pp. S3775–S3777, 2024.
40. L. Knecht, K. Dalsbøl, A. H. Simonsen, F. Pilchner, J. A. Ross, K. Winge, et al., "Autoantibody profiles in Alzheimer's, Parkinson's, and dementia with Lewy bodies: altered IgG affinity and IgG/IgM/IgA responses to alpha-synuclein, amyloid-beta, and tau in disease-specific pathological patterns," Journal of Neuroinflammation, vol. 21, no. 1, pp. 1–14, 2024.
41. J. Sweidan, M. A. El-Yacoubi, and A. S. Rigaud, "Explainability of CNN-based Alzheimer's disease detection from online handwriting," Scientific Reports, vol. 14, no. 1, p. 22108, 2024.
42. A. M. Elshewey, M. Y. Shams, N. El-Rashidy, A. M. Elhady, S. M. Shohieb, and Z. Tarek, "Bayesian optimization with support vector machine model for Parkinson disease classification," Sensors, vol. 23, no. 4, p. 2085, 2023.