International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

Improved Alzheimer's Disease Detection Using an Adaptive Spatial Attention-Based Twin Neural Network

[1]Shaik Mahaboob Basha, [2]S. Swarnalatha, [3]B. Shoban Babu

[1]Ph.D. Scholar, Department of ECE, Sri Venkateswara University College of Engineering, Tirupati, Andhra Pradesh, India.
[2]Professor, Department of ECE, Sri Venkateswara University College of Engineering, Tirupati, Andhra Pradesh, India.
[3]Professor, Department of ECE, Sri Venkateswara College of Engineering, Tirupati, Andhra Pradesh, India.
[1]shaik001122@gmail.com, ORCID ID: 0009-0006-3390-3267
[2]swarnasvu09@gmail.com, ORCID ID: 0000-0001-6002-1389
[3]bshobanbabu@gmail.com, ORCID ID: 0000-0002-3110-7113

**Corresponding Author: Shaik Mahaboob Basha*

Abstract: Alzheimer's disease (AD) is one of the most prevalent forms of dementia, typically initiating with mild memory loss and progressively leading to severe cognitive decline. Early diagnosis is critical for effective intervention and improving patient outcomes. However, existing diagnostic approaches face challenges such as long prediction times, limited feature extraction capability, and the requirement for extensive training data, particularly in Machine Learning (ML)-based systems. To address these issues, this study proposes a novel computer-aided diagnostic framework called Adaptive Spatial Attention-based Twin Neural Network (ASTNet) for early AD detection using MRI data. Unlike traditional methods, ASTNet integrates several components: a biased fuzzy clustering-based segmentation technique that enhances tissue segmentation by emphasizing brain-relevant regions while suppressing skull and irrelevant tissues; Possibilistic C-Means (PCM) to improve segmentation robustness; a Convolutional Bidirectional LSTM (CBi-LSTM) network for context-aware feature extraction that captures temporal dependencies and reduces high-dimensional noise; and an adaptive spatial attention-based Twin Neural Network (TNN) for disease classification that improves spatial feature discrimination across cognitively normal (CN), mild cognitive impairment (MCI), and AD categories. The proposed approach was evaluated on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and achieved a classification accuracy of 0.9909, outperforming the compared Deep Learning (DL) models. Furthermore, the model demonstrates computational efficiency under the evaluated experimental setting and shows potential for computer-aided clinical decision support.

Keywords: Alzheimer's disease diagnosis, Skull stripping, Biased fuzzy clustering, Bidirectional Long Short-Term Memory, Twin Neural Network.

1 INTRODUCTION

Alzheimer's disease is a neurological disorder that affects brain cells and leads to memory loss. Therefore, AD patients are unable to complete daily tasks [1]. AD causes nerve cell damage, leading to brain shrinkage, enlargement of the cerebral ventricles, and reduction in hippocampal volume. Mild cognitive impairment (MCI) is considered a prodromal stage of AD, while cognitively normal (CN) subjects are commonly included as a reference group. Distinguishing between these categories remains challenging for neurologists [2]. Although AD can occur before 65 years of age, the late-onset form is the most common clinical presentation [3]. AD primarily affects specific regions of the brain that have complex anatomical structures.

Diagnosing AD from brain images is a challenging task for neurologists. For this reason, classification techniques are needed for diagnosing AD using brain images [4]. Various kinds of dementia are diagnosed based on the disease symptoms, and AD is the most common type to affect older adults. It causes episodic memory loss and complexities in learning new information. People suffer from behavioral abnormalities, increased memory loss, and cognitive impairments as the condition worsens. AD progressively spreads neuronal damage across different brain regions. Hence, early diagnosis is required to reduce the rate of deterioration. Also, early diagnosis is required for patients to get the required treatment and improve their quality of life [5]. The challenge of diagnosing AD in its early stages arises from several factors, such as the slow and subtle appearance of symptoms, the absence of distinct biomarkers, and the similarity of symptoms to other cognitive disorders.

Furthermore, the high costs and limited accessibility of advanced diagnostic methods, along with ethical and psychological considerations surrounding the diagnosis, add complexity to early detection efforts. Overcoming these obstacles demands improvements in diagnostic technology, greater awareness, a deeper understanding of biomarkers, and better access to healthcare services. Addressing these issues will necessitate progress in diagnostic technology, a deeper understanding of biomarkers, and greater accessibility to reliable diagnostic tools in both healthcare facilities and broader community environments [6].

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

MRI is one of the primary imaging modalities used for AD diagnosis the damaged cells and nerves in the brain. Imaging modalities such as functional MRI (fMRI), structural MRI (sMRI), and positron emission tomography (PET) have been widely investigated for AD diagnosis [7], [8]. Since no definitive cure for AD exists, accurate early diagnosis remains essential. Traditional AD diagnosis involves performing numerous tests and collecting large amounts of diverse, heterogeneous data. Therefore, analysing large volumes of heterogeneous data leads to delays and lower accuracy [9].

ML is a subset of artificial intelligence (AI), a very useful technology for the gradual development of healthcare. ML techniques assist neurologists by supporting and validating clinical diagnoses. The AD syndromes are identified through the use of a Computer-Assisted Diagnosis (CAD) system developed through ML techniques such as Decision Tree Classifier with Hyperparameter Tuning (DTC-HPT) [10]. Neuroimaging statistics are used to create automatic MRI classification and segmentation tools. Traditionally, feature extraction from MRI images has relied on manually designed techniques. Using these MRI features, ML algorithms such as Logistic Regression, Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Artificial Neural Networks (ANN) [11] are used to classify brain diseases [12].

Due to the complex anatomical structure of the brain, automatic AD detection using conventional ML approaches remains challenging, resulting in lower classification accuracy [13]. DL techniques generally achieve higher classification performance than conventional ML methods. Consequently, imaging-based AD diagnosis using DL techniques has received significant attention [14]. To classify the AD-MRI images, the Convolutional Neural Network (CNN) [15], InceptionNet, Siamese neural networks [16], Constraint Fusion Network [17], and Visual Geometry Group (VGG) 16 are used to diagnose the diseases [18]. Performance has also been improved using hyperparameter tuning [19], ensemble-based techniques [20], and ROI-based DL approaches. These DL models perform feature extraction and classification to improve AD detection performance [21].

1.1. Novelty of the Proposed Approach

Clinical assessment of AD faces challenges in accurately distinguishing between its various stages due to the complexity and variability of the disease. Current diagnostic processes are not only expensive but also time-consuming. The skull region often influences classification results, as the feature extraction process inadvertently captures non-brain tissues, leading to misclassifications and reducing classification performance. Additionally, model performance often depends on the characteristics of the MRI dataset, requiring adaptation when new datasets are introduced, which may reduce generalization performance. The goal is to enhance the classification process and improve model generalization across different imaging datasets, ultimately ensuring more accurate and efficient AD prediction while preserving the integrity of the original data. These challenges are addressed using CBi-LSTM-based feature extraction and the proposed ASTNet model. CBi-LSTM reduces feature dimensionality by automatically learning discriminative representations. ASTNet adopts a spatial attention module that considers the relevant spatial information. The model's capacity to generalize is improved by taking into account important features, which reduces the impact of noisy features. Moreover, the adaptive learning procedure updates the model parameters during training.

The novelty of the proposed approach is based on the integration of ASTNet, biased fuzzy clustering-based segmentation, context-aware feature extraction with CBi-LSTM, and end-to-end diagnosis with computational efficiency under the evaluated experimental setting. Unlike conventional fuzzy clustering methods, the proposed biased fuzzy clustering technique prioritizes brain-relevant regions by reducing the influence of non-informative tissues such as the skull. This bias-based approach enhances image clarity and segmentation precision, which is critical in MRI-based AD diagnosis. The integration of the PCM algorithm offers enhanced robustness in tissue segmentation by effectively handling noise and outliers in medical imaging data. It ensures better feature consistency for downstream classification tasks. Traditional CNN models may not effectively capture temporal or sequential dependencies in extracted features. The proposed use of a CBi-LSTM network enables the model to retain information from both past and future image contexts, improving the discriminative power of the extracted features. While attention mechanisms have been used in prior work, the ASTNet introduces a twin neural network structure with spatial attention specifically adapted for AD-related MRI data.

This architecture emphasizes the most relevant spatial features and improves classification performance across cognitively normal, MCI, and AD classes. The proposed framework is designed to be computationally efficient under the evaluated experimental setting. It makes ASTNet a practical framework for computer-aided diagnosis and future clinical evaluation. These design choices contribute to improved accuracy, interpretability, and computational efficiency compared with the evaluated methods. Unlike conventional pipeline-based approaches, the proposed framework is designed such that each module addresses a specific limitation of the subsequent processing stage. The biased fuzzy clustering and PCM modules improve tissue representation by suppressing non-brain regions before feature learning. This enables the CBi-LSTM network to learn more discriminative spatial-contextual features with reduced interference from irrelevant structures. The extracted features are subsequently refined by the adaptive spatial attention mechanism, allowing the twin neural network to emphasize disease-relevant anatomical regions during classification. Therefore, the contribution of the proposed framework is not merely the combination of existing techniques but the complementary interaction among pre-processing, feature extraction, attention-guided representation learning, and classification within a unified end-to-end diagnostic framework.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

The main contributions of the proposed AD diagnosis framework are as follows.

- To develop a novel computer-aided diagnosis system, ASTNet, for early AD diagnosis, leveraging an adaptive spatial attention-based twin neural network that improves diagnostic accuracy by specifically addressing limitations in conventional attention-based networks, such as insufficient spatial discrimination and ineffective cross-region interaction.
- To propose a biased fuzzy clustering-based segmentation method that outperforms traditional fuzzy clustering by incorporating bias towards relevant brain tissue regions, effectively filtering out skull and non-brain areas to enhance segmentation robustness, adaptability, and interpretability.
- To implement a CBi-LSTM network for efficient feature extraction that handles high input dimensionality while preserving temporal dependencies, capturing both past and future contextual cues critical for distinguishing between cognitively normal, MCI, and AD states.
- To design a refined attention-based twin neural network classifier for multi-class AD classification, offering improved inter-class separability by dynamically weighting features and supporting fine-grained MRI classification of different categories.
- To reduce computational complexity and diagnosis latency through optimized pre-processing and attention-guided feature refinement, supporting efficient implementation of AD diagnostic systems for future clinical applications.

The remainder of the paper is organized as follows. Section 2 reviews recent research on AD detection. Section 3 presents the proposed methodology. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the paper.

2 Related Works

Kruthika et al. [22] developed a multi-stage classifier comprising SVM, KNN, and NB classifiers. The framework was designed for AD classification. In this model, MRI images were processed using FreeSurfer for brain image normalization and feature extraction prior to classification. Particle swarm optimization (PSO) was used to obtain the important characteristic of AD as the best feature among many feature vectors. The AD Neuroimaging Initiative (ADNI) database was considered to provide better performance. This method achieved an accuracy of 96.31±1.22%, sensitivity of 91.27±1.44%, specificity of 89.90±1.14%, and precision of 96.05±1.21%. However, the feature selection strategy did not adequately identify the most discriminative features.

Lei et al. [23] suggested a framework for estimating the clinical score with longitudinal point data. Based on a deep polynomial network and entropy-regularized joint learning, feature encoding and feature selection were accomplished. Longitudinal score prediction was achieved by scenario 1, and scenario 2 was used to predict the next time point score with the help of time point data. This scenario was utilized to enhance the predicted score accuracy. ADNI was used in this proposed framework to reveal the interaction between clinical score and MRI data. However, the feature encoding and selection strategy remained a limitation of the proposed approach.

Naz et al. [24] proposed a transfer learning framework based on frozen feature extraction for binary and ternary classification using ImageNet-pretrained convolutional networks. This method was used to classify the performance demonstrated, such as CN, MCI, CN versus AD, and AD versus MCI. This method was compared with the VGG architecture and Convolution network (ConveNet) (GoogLeNet, DenseNet, AlexNet, Inceptionv3, InceptionResNet, ResNet). This method achieved better accuracy as 98.89% (MCI/CN), 99.27% (MCI/AD), and 97.06% (AD/CN). However, the feature extraction capability of the convolutional network remained limited for certain classification tasks.

Park et al. [25] predicted AD with Deoxyribonucleic acid (DNA) methylation data and large-scale gene expression using the DL technique. This method used a multi-omics dataset, which included various integrated omics data, minimal sample size data, and high-dimensional data. This model was used to reduce feature dimensionality with various genes and methylated positions in the multi-omics dataset. They invoked the deep neural network (DNN) based detection-assisted diagnosis approach for enhancing the performance of other ML algorithms. This integration of DNA methylation data and gene expression improves the prediction performance. This method provides an accuracy of 93.90%. However, the study focused only on two integrated molecular layers of the multi-omics dataset.

Bi et al. [26] presented unsupervised CNN-based feature estimation and unsupervised AD classification. The CNN technique with TOP data achieved accuracies of 97.01% and 92.60%. However, the method involved relatively high computational complexity, and its performance varied across datasets. The MRI images are used to classify AD-related cortical atrophy and its stages. With MRI-based features, cognitive scores are predicted using extensive information about morphologically based features. Zhao et al. [27] presented disease progression prediction based on a generative adversarial network (GAN). From the brain image, the clinical stages are estimated using DenseNet optimized for focal loss. With mi-GAN, performance improved by 0.64% over the GAN network.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

Nguyen et al. [28] suggested a Recurrent neural network (RNN)-based prediction of AD to solve the problem related to missing data and longitudinal data. The missing data was considered a pre-processing problem, and the data was processed using forward filling and linear padding. To provide generalization ability, Bron et al. [29] developed a deep CNN and SVM for AD prediction. Modulated grey matter maps were created from MRI images using extensive pre-processing techniques. Most studies on AD prediction have been based on the classification of current disease status rather than disease evolution. The previous information and current data are connected through a Long Short-Term Memory (LSTM) network. Hong et al. [30] integrated past and current information using an LSTM network. The time step data was obtained using the data pre-processing framework. DL approaches generally require large volumes of training data and careful optimization of neural network architectures.

Khan et al. [31] used transfer learning to solve the problem related to weight initialization. The VGG16 architecture was used for weight initialization and layer-by-layer tuning. The size of the training data was minimized using image entropy to select informative image slices. Finally, the discriminative image regions were focused on using Class Activation Maps (CAM) for decision-making. Auriemma et al. [32] presented the gene ontology visualization based on a similarity graph and a dynamic distance-graph model. In order to visualize the similarity data, a new human-interaction system (HIS) was developed, considering gene ontology functions such as biological processes, cellular components, and molecular functions. A comparative analysis of traditional AD detection approaches is presented in Table 1.

Table 1. Analysis of Traditional Alzheimer's Disease Detection Approaches

Study	Techniques	Advantages	Disadvantages	Research Gap
Kruthika et al. [22]	Multi-stage classifier, SVM, KNN, NB	Well suited for the classification process.	Feature significance was not adequately addressed.	Limited deep feature learning
Lei et al. [23]	Support vector regression	Provides improved longitudinal score prediction using regression.	Feature encoding is improper.	No end-to-end classification
Naz et al. [24]	VGG	Classifies different stages of AD.	Performance is limited by the shallow network architecture.	Limited feature representation
Park et al. [25]	DNN	Accurate prediction performance.	Requires a large amount of training data.	No MRI feature integration
Bi et al. [26]	CNN	Used for unsupervised classification problems.	High computational complexity limits practical implementation.	No attention mechanism
Zhao et al. [27]	3D patch-based Mi-GAN	Accurate multi-stage prediction using multi-point information.	Limited to disease progression prediction.	No attention-guided learning
Nguyen et al. [28]	RNN, linear filling, forward filling	Prediction is possible with cross-sectional data.	Not suitable for real-time disease prediction.	Limited spatial modeling
Bron et al. [29]	DNN, SVM	Better generalization can be achieved.	Diagnosis relies primarily on clinical assessments rather than pathological confirmation.	No end-to-end framework
Hong et al. [30]	LSTM	Predicts disease progression using the cortical thickness average (CTA).	The current disease stage is not classified.	Limited spatial feature learning
Khan et al. [31]	VGG16, CAM	Accuracy is increased with layer-wise feature learning.	Probabilistic image measures were not provided.	No spatial-temporal modeling

While numerous studies have explored AD detection using ML and DL techniques, each comes with notable limitations that highlight the need for a more robust and adaptive approach. Kruthika et al. [22] employed traditional classifiers with particle swarm optimization for feature selection but lacked adaptive feature learning, reducing classification robustness. Lei et al. [23] focused on longitudinal clinical score prediction using MRI data rather than direct end-to-end AD classification. limiting accuracy. Naz et al. [24] used frozen features from pre-trained convolutional networks, which restricted domain-specific learning from MRI scans. Park et al. [25] focused on multi-omics data, rather than considering the crucial spatial imaging information necessary for anatomical assessment.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

Bi et al. [26] proposed a fully unsupervised CNN framework for automatic AD diagnosis using MRI images, while Zhao et al. [27] proposed a disease progression prediction framework using a 3D multi-information GAN and DenseNet; however, the study focused on future MRI generation and progression prediction rather than direct feature refinement for MRI classification. Nguyen et al. [28] proposed a minimal recurrent neural network for longitudinal AD prediction with integrated missing-data handling, whereas Bron et al. [29] investigated the external generalizability of CNN- and SVM-based MRI classification models. Hong et al. [30] modeled temporal sequences with LSTM but did not explicitly preserve the spatial structure of MRI data. Similarly, Khan et al. [31] applied transfer learning with VGG16 and CAMs without deeper integration of spatial-temporal features. Finally, Auriemma et al. [32] focused on gene ontology visualization without leveraging imaging data for AD classification.

These limitations collectively underscore the need for a solution like ASTNet, which integrates adaptive spatial attention, Convolutional Bidirectional-LSTM for temporal and spatial feature learning, PCM-based segmentation for noise reduction, and a twin neural network for accurate multi-class classification. The proposed framework is designed to improve classification performance and interpretability by integrating these complementary components. Although deep residual networks have demonstrated strong feature learning capability for medical image classification, conventional single-stream architectures primarily learn global representations from the input image. In AD classification, subtle anatomical variations among closely related stages such as EMCI, LMCI, and MCI require more discriminative feature representations. The proposed framework addresses this limitation by integrating adaptive spatial attention with a shared-weight twin architecture, allowing disease-relevant structural regions to receive greater emphasis while suppressing irrelevant background information. This complementary design improves feature discrimination without substantially increasing computational complexity under the evaluated experimental setting.

3 Proposed Methodology

DL techniques have recently been developed for accurate AD diagnosis. Although AD is currently incurable, early diagnosis enables timely clinical intervention and appropriate disease management. However, ML techniques generally require extensive training data for effective recognition. On the other hand, researchers have increasingly adopted DL models for early AD detection. However, if MRI images are directly input to a DL model, non-brain regions such as the skull may influence feature learning, leading to incorrect classifications. Therefore, non-brain tissues and image noise should be removed before feature extraction. Bias field correction is performed to reduce intensity inhomogeneity, followed by skull removal and tissue segmentation. Therefore, the proposed computer-aided diagnosis system called ASTNet is introduced for accurate AD detection. The proposed framework classifies MRI images into cognitively normal (CN), MCI-related categories, and AD classes. The overall architecture of the proposed AD approach is given in Fig. 1.

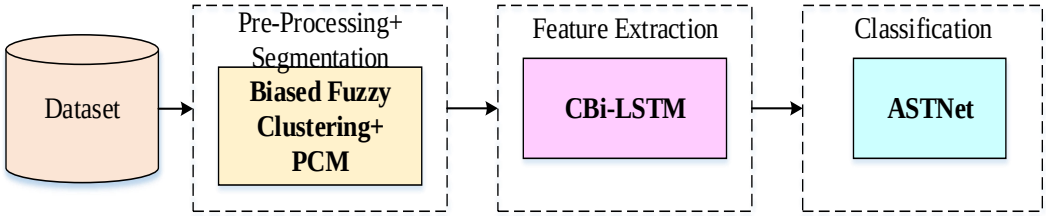


Fig. 1. Architecture of the proposed Alzheimer's Disease diagnosis

3.1. Pre-processing with Biased Fuzzy Clustering

The proposed Fuzzy C-Means (FCM) approach models the data structure where the objects within the cluster have higher similarity. To resolve the issue related to generalization, fuzzy clustering equations are modified with bias correction [33], and the effect of fuzzy initialization is adjusted. For fuzzy partitioning, the bias correction term is utilized as the total information, which resolves the effect of poor initialization. Additionally, biased fuzzy clustering integrates the flexibility of fuzzy clustering, in which data points may belong to multiple clusters and have different degrees of membership. This approach permits the data points related to more than one cluster, and the sum of membership values across the entire cluster is equal to one. Moreover, this approach aims to minimize the weighted sum of the squared distances between the cluster center and data points. Biased fuzzy clustering is affected by contextual information, prior knowledge, or domain-specific constraints. The bias has several forms, including data characteristics, class prior, and external constraints. It can be introduced with the objective of enhancing meaningfulness, stability, and accuracy. In dimensional Euclidean space, data points are defined to be positive integers that are greater than one. The FCM algorithm is defined with the following equation

$$K_n(\eta, b) = \sum_{l=1}^{p} \sum_{j=1}^{d} \eta_{jl}^{n} \, ||Y_l - b_j||^2 \quad (1)$$

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

where Y_l denotes the l^{th} data point, b_j denotes the center of the j^{th} cluster, d is the number of clusters, p is the number of data points, and $n > 1$ is the fuzzification coefficient. The membership function η_{jl} associated with the data point y_l and cluster j is represented with

$$\eta = [\eta_{jl}]_{d\times p} \in N_{FCM} = \left\{ \eta = [\eta]_{d\times p} | \sum_{j=1}^{d} \eta_{jl}, \eta \geq 0, 0 < \sum_{l=1}^{p} \eta_{jl} < p \right\} \quad (2)$$

where N_{FCM} denotes the feasible fuzzy partition space. The partition matrix $\eta = [\eta_{jl}]_{d\times p}$ is obtained by developing the FCM algorithm. The objective function $k_n(\eta, b)$ is minimized by considering the cluster center $b = \{b_1 \dots b_d\}$. For minimizing $k_n(\eta, b)$, the necessary conditions are developed with a Lagrange multiplier using the following equation,

$$b_j = \frac{\sum_{l=1}^{p} \eta_{jl}^n Y_l}{\sum_{l=1}^{p} \eta_{jl}^n} \quad (3)$$

$$\eta_{jl} = \frac{||Y_l - b_j||^{\frac{-2}{n-1}}}{\sum_{k=1}^{d} ||Y_l - b_k||^{\frac{-2}{n-1}}} \quad (4)$$

The bias-based fuzzy clustering algorithm is defined using the following equation:

$$k_n(\eta, b, q) = \sum_{l=1}^{p} \sum_{j=1}^{d} \eta_{jl}^n ||Y_l - b_j|| - x \sum_{l=1}^{p} \sum_{j=1}^{d} \eta_{jl}^n ln(q_j) \quad (5)$$

Here, $\sum_{j=1}^{d} q_j = 1$, and $\sum_{j=1}^{d} \eta_{jl} = 1$. Hence, the clustering is processed with the necessary conditions by the given update.

$$b_j = \frac{\sum_{l=1}^{p} \eta_{jl}^n Y_l}{\sum_{l=1}^{p} \eta_{jl}^n} \quad (5a)$$

$$\eta_{jl} = \frac{(||y_l - b_j||^2 - x\, ln(q_j))^{\frac{-1}{n-1}}}{\sum_{k=1}^{d} (||y_l - b_k||^2 - x\, ln(q_k))^{\frac{-1}{n-1}}} \quad (5b)$$

$$q_j = \frac{\sum_{l=1}^{p} \eta_{jl}^n}{\sum_{k=1}^{d} \sum_{l=1}^{p} \eta_{kl}^n} \quad (6)$$

The parameter update equation is given as follows.

$$x^{(u)} = (0.99)^u \quad (7)$$

Algorithm 1. Biased Fuzzy Clustering for Skull Removal

Input: Brain MRI image data

Step 1: The values are initialized as $\boldsymbol{\varepsilon > 0}$, $\boldsymbol{2 \leq d \leq p}$, $\boldsymbol{u = 0}$, $\boldsymbol{x^{(0)} = 1}$, and $\boldsymbol{q^{(0)}} = \left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\right)$.
Step 2: By using equation (7), the parameter for $\boldsymbol{x^u}$ is updated.
Step 3: The membership function $\boldsymbol{\eta^{u+1}}$ and $\boldsymbol{q^{(u)}}$, $\boldsymbol{b^{(u)}}$ is calculated with equation (5b).
Step 4: Estimate the weighted probability $\boldsymbol{q_j^{u+1}}$ with equation (6).
Step 5: The cluster center $\boldsymbol{b^{u+1}}$ is updated with $\boldsymbol{\eta^{(u+1)}}$ using equation (5a).
Step 6: For a convenient matrix norm $||.||$, $\boldsymbol{b^{(u)}}$ it is compared with $\boldsymbol{b^{(u+1)}}$.
If $||\boldsymbol{b^{(u+1)} - b^{(u)}}|| < \boldsymbol{\varepsilon}$, then the process is terminated.
Else, $\boldsymbol{u = u + 1}$ and continue with step 2.

Output: Clustered non-brain tissues

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

For the given cluster center b_j, the probability mass $q_j$is computed using the following equation. In the proposed BFCA, the conventional squared Euclidean distance $||y_l - b_j||^2$ is modified by incorporating the adaptive bias term $-x\,ln(q_j)$, resulting in the biased distance measure used during fuzzy partitioning. The BFCA updates the weight factor x according to $x^{(u)} = (0.99)^u$, enabling gradual bias adaptation over successive iterations. This adaptive update reduces the influence of poor initialization during the clustering process. Next, unwanted tissues are segmented from the MRI images using the PCM algorithm. Using this algorithm, cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM) are segmented from the brain image. The brain tissues are separated using PCM-based fuzzy partitioning, which satisfies the following constraints:

$$\begin{cases} \forall j \in \{1..C\}\, \forall i \in \{1..N\}\, \eta_{ji} \in [0,1] \\ \forall j \in \{1..C\}\, 0 < \sum_{i=1}^{N} \eta_{ji} < N \\ \forall i \in \{1..N\}\, \max_j \eta_{ji} > 0 \end{cases} \tag{8}$$

The minimization process is expressed as

$$J_{PCM}(\eta, b) = \sum_{j=1}^{C}\sum_{i=1}^{N} \eta_{jl}^{n}\left(Y_i, b_j\right) + \sum_{l=1}^{N}\sum_{j=1}^{C}\left(1-\eta_{jl}\right)^n \tag{9}$$

With this objective function, the fuzzy maps for MRI images are created as {WM_{MRI}, GM_{MRI}, CSF_{MRI}}. Finally, the segmented image is obtained using these three fuzzy maps. The PCM segmentation process is given in Algorithm 2.

Algorithm 2. Segmentation with PCM

Input: Pre-processed image

Step 1: The PCM parameters, including the number of clusters, are initialized to segment brain tissue and the skull region.
Step 2: In the feature space, initialize the cluster centroids to indicate the center of each cluster.
Step 3: Define the fuzziness parameter to estimate the cluster boundaries.
Step 4: Define the possibility parameter to control the degree of uncertainty within each cluster.
Step 5: Initialize fuzzy membership matrix and possibilistic membership matrix, in which each value represents the degree of membership. The sum of membership values across the entire cluster for each pixel is equal to one.
Step 6: Update the membership matrices and cluster centroids iteratively. The update procedure is given in steps 6.1 to 6.3.
Step 6.1: For each pixel, the distance between the pixel and the cluster center is estimated using Euclidean distance.
Step 6.2: Update the fuzzy membership based on distance and fuzziness parameters.
Step 6.3: Update the cluster centroid using fuzzy and membership functions.
Step 7: Repeat the steps from 6.1 to 6.3 until the convergence criteria are reached. Convergence can be reached when the membership matrices and cluster centroids are changed below a predefined threshold.
Step 8: Segmentation results are obtained based on the clustering regions.

Output: Segmented clusters

3.2. Convolutional Bi-LSTM-based Feature Extraction

Feature extraction is essential for improving the classification accuracy. In the proposed approach, image features are extracted using CBi-LSTM. Bi-LSTM layers receive the features as input from convolutional layers and provide richer temporal feature representations. From the MRI data, discriminative features are extracted using the CBi-LSTM network. Fig. 2 illustrates the CBi-LSTM architecture. The CNN is used for spatial feature extraction using kernels that detect local patterns from images. The convolution layer of CBi-LSTM captures structural variations and local spatial patterns of brain regions, including hippocampal and gray matter atrophy. These spatial details are fed to a Bi-LSTM, in which temporal dependencies are modelled across sequential MRI slices. In the training procedure, these patterns are learned to extract high-level semantic details. The convolution layers are responsible for detecting local features, and non-linear feature representations are learned through the Rectified Linear Unit (ReLU) activation function. In addition to that, the pooling layer downsamples and reduces the dimensionality of spatial features. For each subject, the MRI volume is considered a set of slices.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

LSTM captures long-range dependencies from sequential data. It enables the model to capture previous and future context by dealing with the temporal sequence in both directions. The forward LSTM processes the sequence from first to last, and the backward LSTM processes in the reverse direction. This bidirectional processing permits the network to capture the contextual dependencies in both directions across slices. From both directions, the hidden states are integrated to form a richer representation and permit the model to capture contextual information before and after each time step. By combining bidirectional temporal modeling with convolutional feature extraction, the sequential and spatial patterns are effectively captured. It enables robust classification of cognitive states from MRI data. In the proposed framework, the Bi-LSTM does not model the temporal progression of disease over multiple clinical visits. Instead, it models the sequential relationship among ordered MRI slices belonging to the same brain volume. Although each slice is ultimately classified independently, the sequential learning mechanism enables the network to capture contextual information from neighboring slices, thereby preserving structural continuity and improving the discriminative capability of the extracted features.

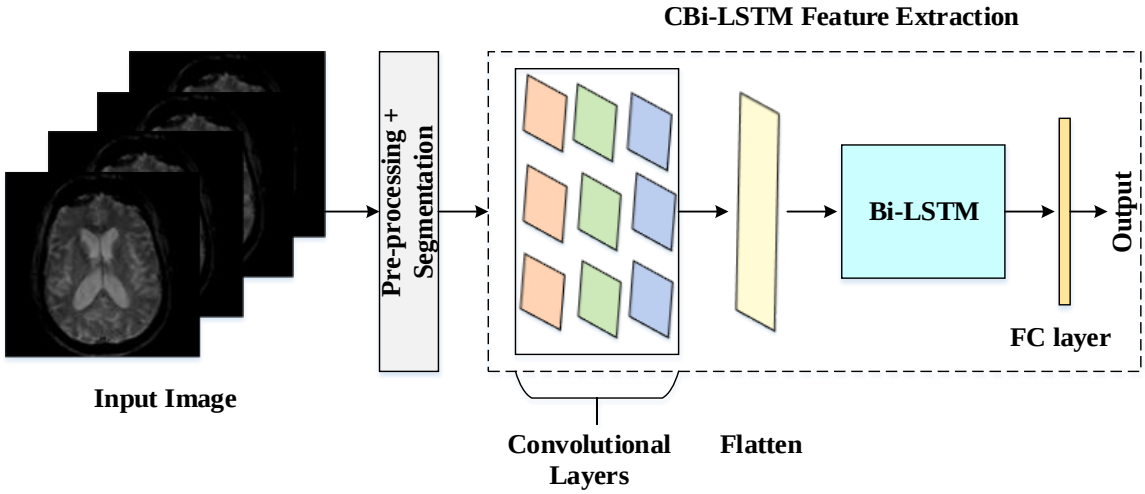


Fig. 2. Architecture of CBi-LSTM

Bi-LSTM is a kind of RNN in which the functions use sigmoid and tanh activation functions. If the input data is large, the learning process is complicated, leading to dependency issues. Hence, gate functions are introduced for an efficient learning process. The internal structure of Bi-LSTM varies from that of the RNN. It consists of four interacting components, including gate mechanisms, and interacts with each other. It has an input gate, a forget gate, a cell state, and an output gate, in which each cell is connected with two states, namely the input and hidden states. Bi-LSTM can add or delete information using the sigmoid gate operation. The LSTM and Bi-LSTM architectures are given in Fig. 3(a) and 3(b). The flow of information from the sigmoid gate is optional, based on sigmoid gate outputs ranging between 0 and 1. The equation used for the LSTM network is described with the following equations.

$$g_u = \sigma(X_g y_u + x_g i_{u-1} + c_g) \tag{10}$$

$$j_u = \sigma(X_j y_u + x_j i_{u-1} + c_d) \tag{11}$$

$$D_u = tanh(X_d y_u + x_d i_{u-1} + c_j) \tag{12}$$

$$D_u = g_u D_{u-1} + j_u * \widetilde{D}_u \tag{13}$$

$$\sigma_u = \sigma(X_0 y_u + x_0 i_{u-1} + c_0) \tag{14}$$

$$i_u = p_u * tanh(D_u) \tag{15}$$

Here, j_u represents the output with time u, i_u denotes the hidden state, p_u denotes the output gate activation, the weight matrices are represented with X_g, cell state is denoted with D_u, each state associated with the cell state is denoted with D_{u-1}, at the time u, the input is y_u, σ represents the sigmoid function, forget gate bias is denoted with c_g, g_u represents the forget gate output, whose values range from 0 to 1, and the output sigmoid is denoted by p_u and $\widetilde{D}_u$ is candidate cell state.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

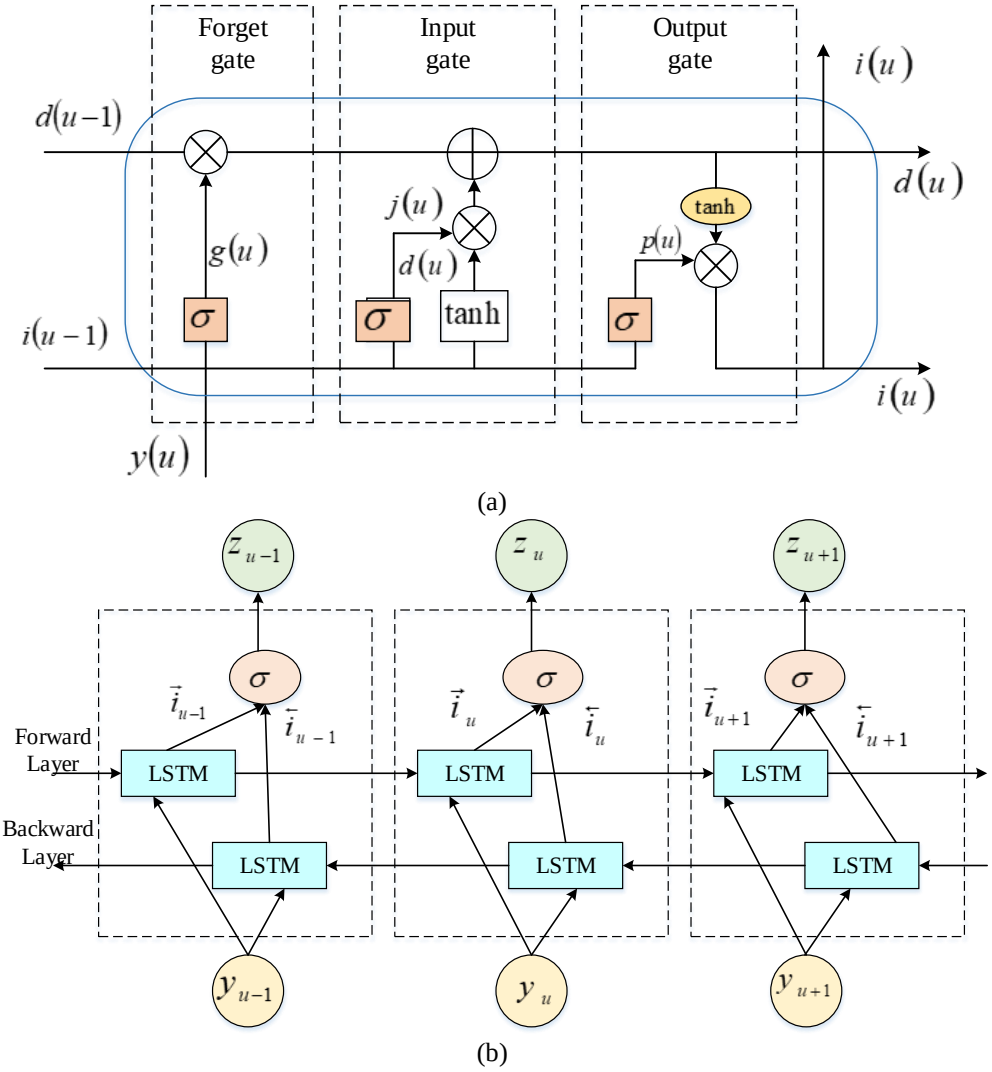


Fig. 3. (a) LSTM and (b) Bi-LSTM architecture

The LSTM performance is enhanced through Bi-LSTM, in which two LSTM networks are used for more effective sequence learning. The Bi-LSTM processes information in two directions: past to future and future to past. It is accomplished by adding hidden states to the forward sequence. The hidden sequence of i_u and then the backward hidden sequence of i_u is estimated. The final product z_u is detected by combining forward and backward sequences with the following equations.

$$\vec{\imath}_u = I(X_{yi} y_u + X_i \vec{\imath}_{u-1} + c_i) \tag{16}$$

$$\overleftarrow{\imath}_u = I(X_{yi} y_u + X_i \overleftarrow{\imath}_{u+1} + c_i) \tag{17}$$

$$z_u = X_f \vec{\imath}_u + X_{iw} \overleftarrow{\imath}_u + c_z \tag{18}$$

where $\vec{\imath}_u$ and $\overleftarrow{\imath}_u$ denote the forward and backward hidden states, respectively; X_f and X_{iw} are trainable weight matrices used to combine the hidden states; c_z denotes the bias vector; and z_u is the final Bi-LSTM output representation.

The deep image features are extracted using CBi-LSTM, and then a flattening layer is applied. The flattening layer reshapes the extracted feature maps into a vector sequence suitable for Bi-LSTM processing. Feature vectors derived from the Bi-LSTM and Convolutional approach characterize the primary goal of the CBi-LSTM technique. In order to identify the various types of Alzemer's Disease, these extracted feature vectors, which include more prominent discriminative information, are input into the classification layer. The CBi-LSTM approach is given in Algorithm 3.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

Algorithm 3. Feature extraction based on CBi-LSTM

Input: Segmented image
Step 1: Resize the image to the required input size.
Step 2: Apply convolution layers, ReLU, and pooling for image feature extraction with non-linearity, as well as dimensionality reduction while preserving important features.
Step 3: Reshape the feature map for sequential processing.
Step 4: Apply reshaped features to Bi-LSTM layers for extracting sequential information and long-range dependencies.
Step 5: Apply fully connected layers and obtain feature vectors for high-level extracted features of the image.
Output: Image features

This framework is designed to process intricate spatial and temporal relationships within data, such as 2D images, making the classification task highly effective. The process begins with a 2D input image, defined by its height, width, and number of channels. The image first passes through one or more convolutional layers, where a set of filters moves across the image, extracting local spatial features. After convolution, the extracted features are reshaped into a 2D format that is compatible with the subsequent LSTM layers. In the Bidirectional LSTM phase, the model processes the input in two directions. The bidirectional approach permits the extraction of discriminative features, which are essential for understanding complex relationships, such as temporal patterns or dependencies between distant parts of the image. The prediction of the Bi-LSTM layer is a set of hidden states that effectively captures both the spatial and temporal relationships in the image.

3.3. AD Classification with ASTNet

For AD classification, an adaptive spatial attention mechanism is integrated with the twin neural network. The extracted image features are then provided to the twin neural network through the attention module. The structure of the residual block and ResNet-based twin neural network is given in Fig. 4 and 5. ASTNet processes spatiotemporal data, where the input consists of spatial features extracted by the CBi-LSTM module. Initially, spatial attention is applied to detect the spatial location within the essential data. For brain images, the relevant spatial regions are highlighted using the adaptive spatial attention mechanism. By using convolution layers, spatial features are extracted and applied to capture temporal dependencies across each time step. Then, the spatial and temporal features are fused to generate a unique feature representation. Finally, these representations are fed to the dense layer for the prediction tasks.

The adoption of a twin neural network is motivated by the need to improve feature consistency and discriminative representation learning in MRI-based AD classification. Brain MRI images belonging to adjacent disease stages, particularly EMCI, LMCI, and MCI, often exhibit subtle anatomical differences that are difficult to distinguish using a single-stream architecture. The shared-weight twin structure enables the network to learn consistent feature embeddings while suppressing variations caused by normal anatomical differences and imaging conditions. This shared representation improves inter-class discrimination and complements the adaptive spatial attention mechanism by emphasizing disease-relevant structural variations before the final classification stage.

In the proposed multiclass classification framework, the twin neural network has two parallel branches that can process the features extracted by the CBi-LSTM module independently. These two branches are identical in structure and share the same weights. The shared-weight design enables the network to learn consistent feature embeddings while reducing variations caused by anatomical differences and imaging conditions, thereby improving the discrimination of closely related AD stages. The objective is for both branches to learn comparable feature representations from the input features. These features are then fed into both branches of the ASTNet simultaneously. The network learns to quantify the similarity between the two branches by feeding the shared feature representations into each.

This twin architecture enables consistent feature learning while improving representation robustness through shared-weight processing. This representation improves the model's ability to classify different disease classes. After independent processing, the output of the twin branches is combined to generate the prediction for the final multi-class classification. This modelling enables the classifier to learn complementary representations of the same input features. Hence, the discrimination capability and robustness are improved across all classes. It is particularly important for tasks like multi-class image classification, classification, or comparison. Each branch independently processes these features, but because the branches share weights, the network is able to effectively learn relationships between the processed feature representations. This method helps the network to better recognize patterns or relationships within the features of the image. The input is mapped to the feature space of the intermediate hidden layer $\phi(.)$. In the final layer, the linear classifier is learned with each network.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

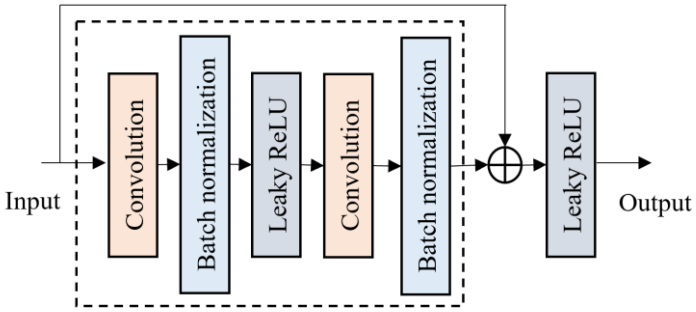


Fig. 4. The structure of the residual block

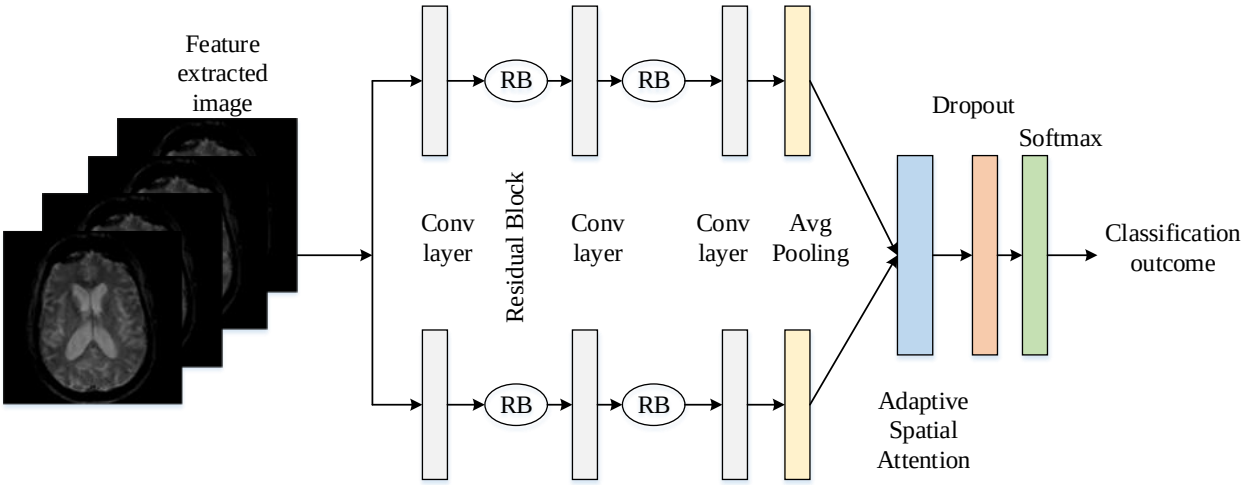


Fig. 5. Architecture of the proposed ResNet-based Twin Neural Network

As CNN depth increases, gradient-related optimization problems may occur during backpropagation. By updating the weights, the network becomes unstable, and the process is termed vanishing gradients and network degradation. If the gradient is low, then the global optimal has not converged. By introducing residual blocks, this kind of degradation is resolved. Batch normalization is included with the convolution operation for output normalization and for enhancing efficiency. Next, the residual block is used for each convolution layer with the ReLU activation function. The input and output of the blocks are represented with y and z.

$$G_j(y) = X_j y + c_j \tag{19}$$

$$z = G_2(G_1(y)) + X_t y \tag{20}$$

Here $G_j(y)$ denotes the output of the j^{th} residual block, X_t is the projection matrix used when input and output dimensions differ. The learned weights and biases of j^{th} layer convolution operation is denoted with $X_j y$ and c_j. Each block has a convolution layer, requiring the channel number for the input y is similar to $G_2(G_1(y))$. It is passed through the additional layer and fed to the activation function. If the dimensions are varied, the convolution layer with weight X_t is included for varying the dimensions of y. The test point $y$is predicted with the following equation. In the hidden layer space, the point $\phi(.)$ is mapped, and the label z as below.

$$z = \begin{cases} +1, & \dfrac{x_{(+1)}^T \phi(y) + c_{(+1)}}{||x_{(+1)}||} \leq \dfrac{x_{(-1)}^T \phi(y) + c_{(-1)}}{||x_{(-1)}||} \\ -1, & otherwise \end{cases} \tag{21}$$

Here $\phi(y)$ is feature embedding, $x_{(+1)}$ and $x_{(-1)}$ are classifier weights.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

3.3.1. Adaptive Spatial Attention Module (ASAM)

With the ASAM, the local pixel dependencies are evaluated. Feature representations are learned adaptively, and the nearest spatial features are selected adaptively. The significance of adjacent pixels is focused on the spatial weight matrix creation. The resulting feature map improves the system's accuracy of the DL network. ASAM is utilized to focus on informative spatial features during learning that include more essential details. This enables the model to dynamically adjust the input data at different stages. Furthermore, the computational efficiency of the ASAM module minimizes the need to process less informative regions. The accuracy is enhanced by focusing on relevant parts of the input data. It can be invoked by producing an attention map that represents the input regions. For each region, ASAM generates attention scores using learning techniques. These scores represent the importance of different spatial regions. Based on learned spatial attention weights, the attention map is produced by calculating a weighted sum. After generating attention maps, the input feature map is rescaled and element-wise multiplication. Finally, the processed feature map is applied to the preceding layers for further processing or classification. The ASAM of ASTNet is shown in Fig. 6.

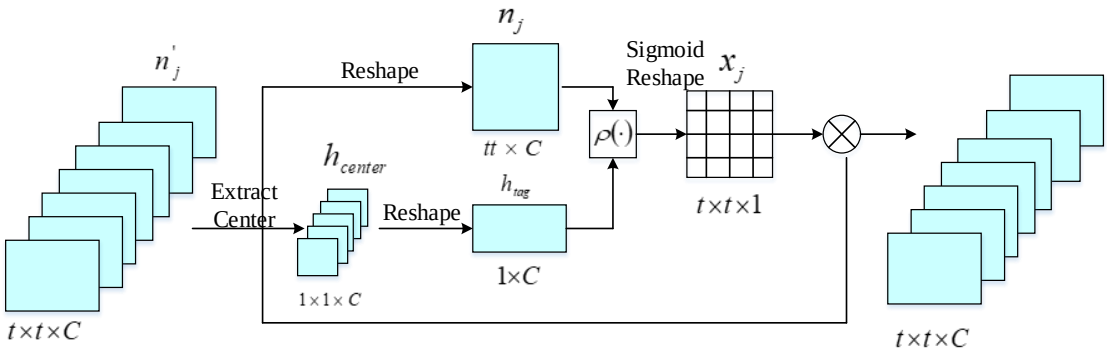


Fig. 6. Architecture of the Adaptive Spatial Attention Module (ASAM) used in ASTNet

The similarity between the center and the nearest pixels is indicated by the computation of reshaping the input n'_j as $h_j \in S^{tt \times C}(tt = t \times t)$. From the center of n'_j, the pixel center $h_{center} \in S^{1 \times 1 \times C}$ and it is reshaped into $h_{tag} \in S^{1 \times C}$. Then, the pixels h_{tag} and h_j are given to the scoring function $v(.)$ for estimating the similarity score. The scoring function is described with the following equation.

$$v(i_j) = \varphi\left(\sum_{j=1}^{tt} i_j X\right) \tag{22}$$

$$i_j = exp\left(-\frac{1}{C}\left\|h_j - h_{tag}\right\|_2^2\right) \tag{23}$$

Here $\varphi(.)$ represents ReLU and C is feature dimension. The correlation between $h_{tag.v(.)}$ and h_j is calculated with i_j which is invoked with the fully connected layer and the weight matrix $X \in S^{tt \times tt}$. By integrating $X$with i_j the similarity scores are obtained with the activation function ReLU $\varphi(.)$. The spatial weight matrix $x_j \in S^{t \times t \times 1}$ is calculated by including the sigmoid function. Finally, element-wise multiplication of n'_j with x_j is performed by modifying the spatial information.

$$n_j^n = x_j \otimes n'_j \tag{24}$$

Here, n_j^n represents the output of the spatial attention layer and $\otimes$ represents element-wise multiplication. The weights of each layer are initialized and updated with the Adam optimizer. The loss function of the model is computed by selecting the categorical cross-entropy (CCE) loss function, which is modeled as,

$$CCE_{loss} = -\sum_{i}^{N} t_i \, log(f(sc)_i) \tag{25}$$

Here $f(sc)_i$ represents softmax probability. The ground truth is denoted with t_i, the score of CNN for each $class_i$ is represented withsc_i, and N represents the number of classes. For multi-class classification, the labels are said to be one-hot, and the positive class N_p is only present. There can be one non-zero element, which is $t_i = t_p$ for the target vector t. By discarding the elements present in summation that are zero because of the target labels. Further, the loss function is simplified with the following equation.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

$$CCE_{loss} = -log\left(\frac{e^{sc_p}}{\sum_j^N e^{sc_j}}\right) \quad (26)$$

CCE loss is otherwise called softmax loss, and it includes the combination of softmax activation with cross-entropy loss. This loss is used to perform multi-class classification and train the neural network to obtain the class probability distribution for every image. The weight and bias values are updated across each iteration until convergence is reached. For estimating the weight in the final layer, the weights of the previous layers are extended with the backpropagation operation.

4 Experimental Results

This section describes the dataset, performance metrics, and experimental results. The proposed diagnosis system is implemented in Python (version 3.10.13) using the Spyder IDE. TensorFlow is used for model construction, training, and evaluation in the Python environment. The experiments were conducted on a system with 16 GB RAM, an Intel® Core™ i7-4790 CPU @ 3.60 GHz, Intel® HD Graphics 4600 (113 MB), 932 GB storage, and a 64-bit operating system. Also, the proposed model was trained using CPU-only execution. No dedicated GPU acceleration was utilized. No cloud-based or external computational resources were used. The implementation code and trained model parameters used in this study are available from the corresponding author upon reasonable request. These resources will be provided to support reproducibility and validation of the reported findings.

The final classification is performed at the image/slice level. During feature extraction, neighboring slices from the same MRI volume are processed sequentially to learn contextual representations, whereas the final prediction is generated for each slice. Multiple MRI slices were obtained from each subject; however, subject-level partitioning was used during data splitting. To ensure unbiased and robust evaluation, the data was partitioned at the subject-level to avoid data leakage between training and testing sets. Specifically, subjects were randomly divided into training (69%), testing (15.5%), and validation (15.5%) subsets, ensuring that all MRI sessions from a given subject appeared only in one subset to maintain independence. The final model evaluation is based on an unseen test set in which the reported results correspond to a single independent test run. In addition, 5-fold cross-validation is performed within the training set to avoid overfitting. The model performance was evaluated on a held-out testing set, which was not used in any stage of model training or validation. This evaluation protocol helps ensure that the reported performance reflects generalization to unseen subjects rather than repeated-measure bias or data leakage.

The average training time per epoch is 58.3 seconds, and the total training time is 1.62 hours for the entire training process. The total training time is computed by multiplying the number of epochs by the average training time per epoch (number of epochs × training time). To ensure fair comparison among different DL architectures, all models were reimplemented under an identical experimental scenario using the same training and validation sets. Subject-level independence is considered for all comparison models to avoid data leakage. Additionally, the same input dimension, pre-processing pipeline, training/testing split, and evaluation metrics were used with the same training protocol and fixed hyperparameters. Also, the same CPU-only hardware environment is utilized for the proposed and existing model training and evaluation. These standardized conditions allow fair comparison by minimizing experimental bias. The hyperparameters used for the proposed approach are given in Table 2. These values are selected through empirical tuning based on validation-set performance.

Table 2. Hyperparameters of the Proposed Approach

Parameter	Values
Learning Rate	0.0001
Optimizer	Adam
Batch Size	32
Epochs	100
Dropout	0.02
Activation	ReLU
Loss function	Categorical Cross Entropy (CCE)
Number of dense units	64
Pooling type	Average pooling
Convolution kernel size	3x3
Attention kernel size	1x1

The workflow from subject acquisition to final prediction is given below.

- Subject acquisition: A total of 58 subjects were included for MRI scan acquisition across five different classes, namely CN, MCI, LMCI, EMCI, and AD.

- Pre-processing: MRI images were pre-processed using intensity normalization (min-max normalization) to

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

minimize inter-scan variability.

- Subject-based grouping: According to the subject IDs, all images were grouped.
- Subject-level split: The subjects are divided into a training set, a testing set, and a validation set without overlap.
- Cross-validation: To reduce model overfitting, five-fold cross-validation was performed on the training set.
- Training: The model training was performed using the training folds.
- Training and validation: Model performance was monitored using the validation set while training.
- Testing: Final evaluation was performed on an independent test set. The class labels are predicted as an output for unseen MRI images/slices.

4.1. Details of the Dataset

The proposed approach is evaluated using the ADNI dataset (https://adni.loni.usc.edu/data-samples/access-data/). The ADNI project was launched in 2003 by Dr. Michael W. Weiner. The dataset consists of 58 subjects and multiple MRI slices per subject. All the experiments are performed from the ADNI-2 phase, and the participants are selected based on the study inclusion criteria described by the ADNI consortium. For model development and evaluation, Structural brain MRI scans obtained during the ADNI-2 phase were utilized. The initial dataset has 2,937 nonaccelerated and 2,925 accelerated 3T sagittal T1-weighted scans. From this scan, subjects are identified in the same session, and subjects are removed with scanner changes. The ages of these individuals range from 63 to 82.

All participants underwent clinical, cognitive and neuropsychological evaluations during scan acquisition. Each subject received an accelerated T1-weighted scan immediately after a nonaccelerated scan without leaving the scanner. From 55 ADNI cites, high resolution structural brain MRI images are captured through 3T MRI scanners. Accelerated scanning time is 5:12 to 5:34 minutes, and nonaccelerated scan time is 9:06 to 9:26 minutes. Also, the image quality has been evaluated by the ADNI MRI quality control center of Mayo Clinic. Some of the scans are rejected due to motion, clinical abnormalities, and technical problems. For image correction, the standard processing pipeline 'grinder' is utilized. It includes 3D gradient unwarping for Siemens and GE scans. Furthermore, the 'N3' bias field correct was performed for all scans. There are a total of 1,106 MRI slices (images) belonging to five classes: Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI), Early Mild Cognitive Impairment (EMCI), and Cognitively Normal (CN). Table 3 presents the subject-wise and image-wise distribution of the proposed ADNI dataset across the training, testing, and validation subsets.

Table 3. Proposed ADNI Dataset Statistics and Distribution

Class	Total images	Total Subjects	Training	Testing	Validation
AD	124	12	8	2	2
MCI	180	13	9	2	2
LMCI	55	8	6	1	1
EMCI	275	10	7	1	2
CN	472	15	10	3	2
Total	1106	58	40	9	9

The number of subjects in each class with training, testing and validation split is given in Table 3. Subject-level split was performed for the dataset to ensure each subject was allocated exclusively to either the training, testing or validation set. This distribution confirms that no subject appears in more than one data partition (training, testing, or validation). To ensure experimental reproducibility, all MRI images underwent the same processing pipeline throughout the study. Initially, MRI images were pre-processed using biased fuzzy clustering and PCM-based tissue segmentation. The processed images were then resized and normalized before feature extraction using the proposed CBi-LSTM module. The extracted feature representations were subsequently classified using the ASTNet framework with identical hyperparameter settings for all experiments. Subject-level partitioning was maintained throughout training, validation, testing, cross-validation, ablation analysis, and statistical evaluation to prevent information leakage between different experimental stages.

4.2. Performance Metrics

In this section, the metrics used to evaluate the performance of the proposed approach are discussed. The first metric is accuracy which measures the proportion of correctly detected samples out of all samples.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{27}$$

where TP indicates the true positive, and TN indicates the true negative, FP represents the false positive, and FN represents the false negative. The next parameter is precision which estimates the correctly predicted positive samples across all positive predictions.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

$$Precision = \frac{TP}{TP + FP} \tag{28}$$

The next parameter is recall which estimates the correctly predicted positive samples across all actual positive samples.

$$Recall = \frac{TP}{TP + FN} \tag{29}$$

The next metric is F-measure which estimates the harmonic mean of precision and recall

$$F\_measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{30}$$

The next parameter is specificity which measures the correctly predicted negative samples among all actual negative samples.

$$Specificity = \frac{TN}{TN + FP} \tag{31}$$

The next metric is False Positive Rate (FPR) measures the proportion of negative samples that are incorrectly classified as positive.

$$FPR = \frac{FP}{FP + TN} \tag{32}$$

The last metric used is Dice Similarity Coefficient (DSC) which gauge the similarity between predicted segmentation and ground truth segmentation.

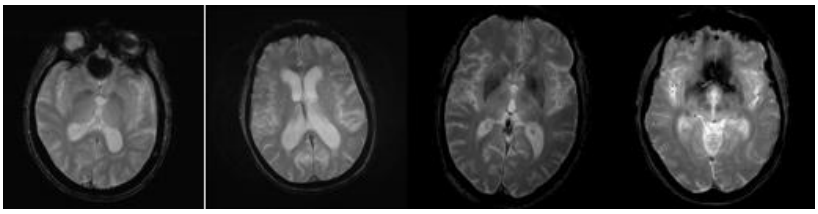
$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{33}$$

where A represents the predicted segmentation, and B represents the ground truth segmentation.

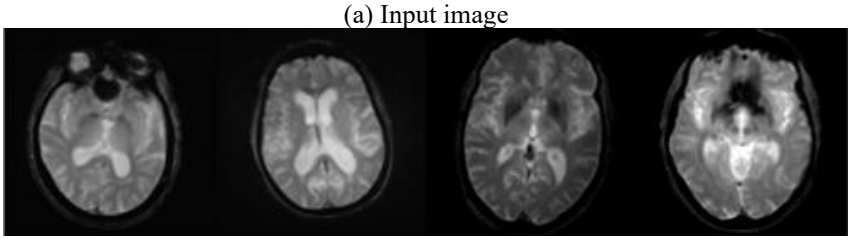
4.3. Performance Analysis

4.3.1. Analysis of Pre-processing Approach

Sample images from the ADNI dataset and the corresponding pre-processing outputs are shown in Fig. 7 and Fig. 8. Initially, the MRI images are processed using biased fuzzy clustering for skull removal. The pre-processed images improve classification performance by removing unwanted non-brain tissues before feature extraction.



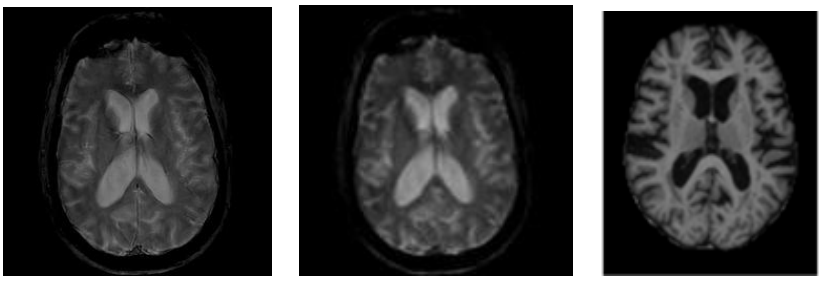
(a) Input image



(b) Pre-processed image

Fig. 7. Sample ADNI MRI Images for AD Classification

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026



(a) (b) (c)

Fig. 8. Comparison of Input MRI Images and Pre-processed Outputs using Fuzzy Clustering and Biased Fuzzy Clustering (a) Input image, (b) Pre-processed image with fuzzy clustering, (c) Pre-processed image with biased fuzzy clustering

The analysis of pre-processing performance is given in Fig. 9. The proposed pre-processing stage employs biased fuzzy clustering and PCM-based segmentation for skull removal and tissue separation. Some of the commonly used pre-processing approaches, including Gaussian Filter (GF) based noise removal, z-score normalization, min-max normalization, Fuzzy k-means clustering, Atlas-based segmentation, and Contrast Limited Adaptive Histogram Equalization (CLAHE), are compared. The proposed approach achieves better performance than the compared pre-processing methods. This improvement is primarily attributed to the biased fuzzy clustering and PCM-based segmentation techniques. These techniques effectively reduce noise and improve tissue segmentation in MRI images. Conventional Fuzzy C-Means clustering may misclassify boundary pixels. Additionally, more accurate boundary estimation can be achieved using the proposed approach.

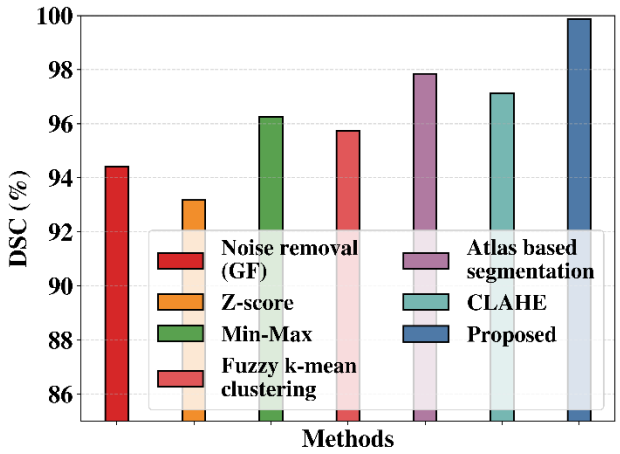


Fig. 9. Analysis of pre-processing performance

4.3.2. Analysis of Classification Performance

Fig. 10 shows the confusion matrix of the ADNI dataset. The number of classes used is five, namely AD, LMCI, CN, EMCI, and MCI. No misclassifications were observed for the EMCI, LMCI, and MCI classes. A small number of misclassifications were observed for the AD and CN classes. Scanner variation, subject variation, and noise are the common issues related to the ADNI MRI dataset. In the proposed approach, the multistage framework resolves this through pre-processing and feature learning. For all MRI images, intensity normalization and alignment are performed to an anatomical space to handle Scanner variation. It ensures the representation of images from different acquisition scenarios in a unique coordinate system.

These variations related to scanner protocols and acquisition parameters are minimized using the intensity normalization technique. Biased fuzzy clustering-based segmentation and PCM resolve subject-level variability by suppressing irrelevant regions while considering the most informative regions. Therefore, it minimizes the anatomical noise caused by non-brain tissue regions. Additionally, the integration of CBi-LSTM learns the feature representation with less noise sensitivity. The adaptive spatial attention module suppresses the influence of non-informative and noisy regions. Therefore, the integration of the proposed framework enhances the MRI image quality and minimizes the variability towards accurate classification. The number of correctly classified and misclassified samples in the final testing set is given in Table 4. The overall accuracy was computed as the ratio of correctly classified samples (219) to the total number of testing samples (221), resulting in an accuracy of 0.9909.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

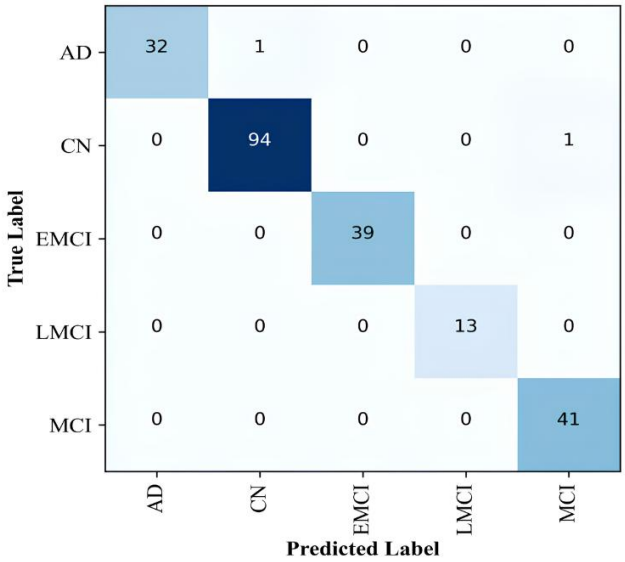
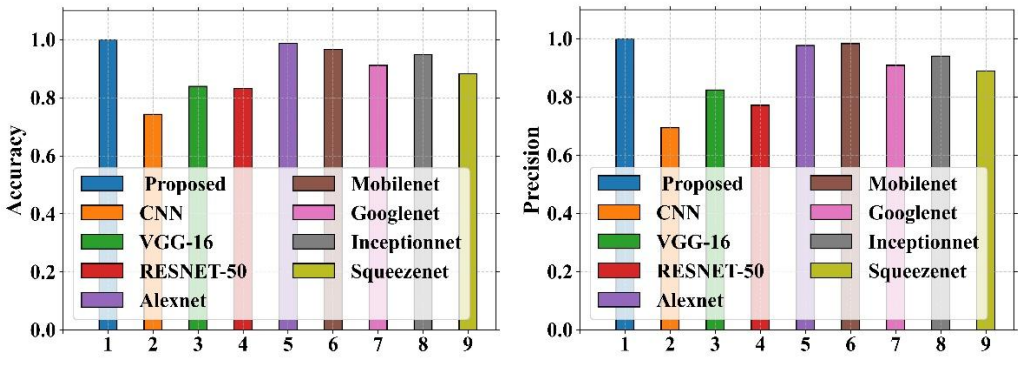


Fig. 10. Confusion Matrix of the ADNI Dataset with AD Prediction

Table 4. Number of Correctly Classified and Misclassified Samples in the Final Testing Set

Classes	Total samples	Number of correctly classified samples	Number of misclassified samples
AD	33	32	1
CN	95	94	1
EMCI	39	39	0
LMCI	13	13	0
MCI	41	41	0
Total	**221**	**219**	**2**

The proposed classification performance based on accuracy and precision metrics is given in Fig. 11. The proposed technique achieves higher results than other existing approaches. The lowest performance was observed with the CNN-based model. All accuracy values are in the range of 0.8 to 1.0. Due to its efficient learning paradigm, the performance is higher with a DL-based approach. ASTNet achieves higher precision than AlexNet and MobileNet. The performance of some existing classification algorithms, including VGG16 and ResNet-50, is below 0.8. The higher accuracy and precision performance are due to the integration of advanced feature extraction approaches. That captures long-range dependencies and local spatial patterns to improve the quality, and it leads to better classification performance. Moreover, the use of ASTNet focuses on learning image similarity to improve decision-making performance. The optimal training process leads to better generalization and resolves overfitting issues by dealing with complex dependencies.



(a) (b)

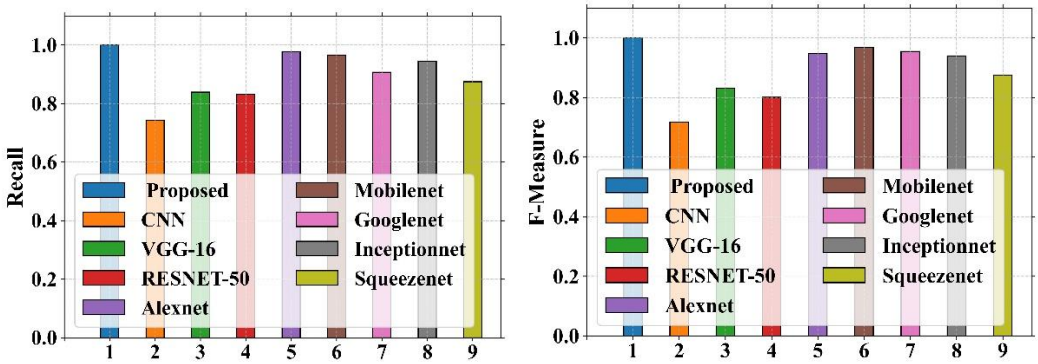
Fig. 11. Classification Performance of the Proposed Approach (a) Accuracy (b) Precision

The performance in predicting AD based on recall and F-measure is shown in Fig. 12. Performance varies between F-measure and recall. The achieved performance is higher for the proposed approach, whereas the lowest performance is achieved with the CNN-based approach. The CNN method simplifies the process and minimizes the prediction performance with less processing power. When considering recall performance, all proposed and existing approaches are between 0.6 and 1.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

Similar to the accuracy and precision performance, the results are lower for the CNN-based approach and other approaches such as GoogleNet, SqueezeNet, AlexNet, and InceptionNet. The performance of the F-measure is analyzed using existing techniques such as CNN and SqueezeNet. Techniques such as VGG16, ResNet50, and CNN have lower performance than other DL-based techniques. A higher result was achieved with the proposed approach of AlexNet, MobileNet, GoogleNet, InceptionNet, and SqueezeNet. The higher recall and F-measure performance of the proposed approach is due to the model classifier, which depends on the context and dataset. The proposed model provides balanced performance across all classes, resulting in high recall and F-measure values. It provides better performance across recall and F-measure. Moreover, recall performance minimizes false negatives, which leads to better F-measure performance.

The proposed specificity performance is given in Fig. 13. When analyzing existing DL-based techniques, a lower result is obtained with CNN, VGG16, and ResNet-based approaches. Other DL-based performances are higher than 0.8. The higher specificity of the proposed value indicates the ability of the model to correctly detect the negative classes with fewer false positives. While reducing the number of false positives, incorrect classification is avoided, which leads to better specificity values. It can be attained by PCM-based segmentation that reduces the effect of noise and enhances structural clarity in MRI images. In addition, the utilization of CBi-LSTM assures better extraction of sequential and spatial patterns. On the other hand, the adaptive spatial attention layer ensures the model focuses on the most relevant brain regions. These components are integrated into the proposed architecture to make more informed and discriminative decisions, leading to improved specificity.



(a) (b)

Fig. 12. AD Prediction Performance with the Proposed approach (a) Recall (b) F-measure

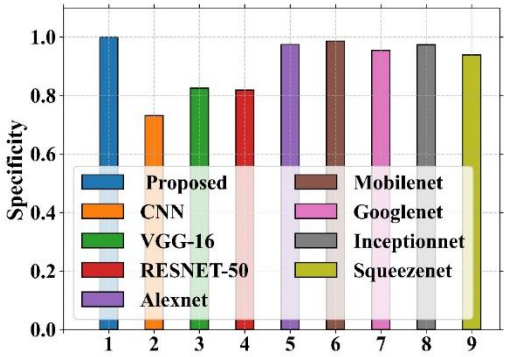


Fig. 13. Specificity Performance Comparison

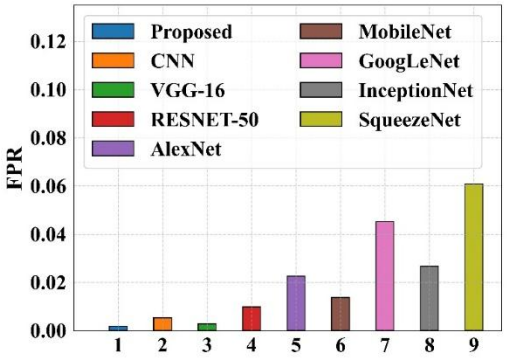


Fig. 14. FPR Performance comparison with existing approaches

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

The FPR performance is compared with traditional DL techniques, as shown in Fig. 14. This FPR is analyzed using existing techniques, including CNN, VGG-16, and ResNet. This performance variation is due to an efficient feature extraction module and better classification of the proposed approach. The spatial attention module emphasizes disease-relevant brain regions while suppressing irrelevant information. The false positive rate is lower for the proposed approach. It minimizes the misclassification and increases the accuracy of classification. Better performance in terms of FPR was achieved due to the proposed approach, which integrates several advanced approaches. Moreover, adaptive attention techniques consider the related brain regions to enhance the detection capabilities for AD patients with a lower FPR. In addition to that, the ASTNet and fuzzy clustering approach improve the model's ability to better classify patterns by handling uncertainty towards lower FPR. The performance is enhanced with feature extraction capabilities and the handling of complex feature representations.

To evaluate the effectiveness of the proposed ASTNet framework, its classification performance is first compared with widely used baseline DL architectures, including CNN, VGG-16, ResNet-50, AlexNet, MobileNet, GoogLeNet, InceptionNet, and SqueezeNet. All models were evaluated under the same experimental setting using the proposed ADNI dataset. The comparative results are presented in Table 5. In addition to the comparison with baseline DL architectures, the proposed ASTNet framework is further compared with recently published state-of-the-art Alzheimer's disease diagnosis methods reported in the literature. These approaches were selected because they use the ADNI dataset under comparable experimental settings. This comparison highlights the competitiveness of the proposed method with respect to recent research and demonstrates its improved classification performance. The comparative results are presented in Table 6.

Table 5. Performance Comparison of the Proposed ASTNet with Baseline DL Architectures on the ADNI Dataset

Methods	Accuracy	Recall	Specificity	Precision	F-measure
CNN [34]	0.7433	0.7433	0.7312	0.6941	0.7178
VGG-16 [34]	0.839	0.839	0.8251	0.8249	0.8319
ResNet-50 [34]	0.8317	0.8317	0.8189	0.7715	0.8005
AlexNet [16]	0.9874	0.9768	0.9745	0.9774	0.9483
MobileNet [35]	0.9671	0.9663	0.9863	0.9836	0.9681
GoogLeNet [18]	0.9114	0.9073	0.9548	0.9089	0.9548
InceptionNet [18]	0.9478	0.9452	0.9733	0.9413	0.9398
SqueezeNet [36]	0.8819	0.8749	0.9391	0.8887	0.8758
Proposed	0.9909	0.994	0.998	0.995	0.995

Table 6. Comparative Analysis of the Proposed ASTNet and Recent State-of-the-art Alzheimer's Disease Detection Methods.

Method	Accuracy	Recall	Specificity	Precision	F-measure
ResNet 50 [34]	0.9527	0.9531	0.9763	0.9539	0.9532
VGG-16 [34]	0.9277	0.9315	0.9649	0.9277	0.9275
DCNN [34]	0.9612	0.9499	0.9773	0.955	0.9523
Shallow CNN [37]	0.9968	0.9468	0.9664	0.9432	0.9475
Multi-task DL [38]	0.9600	0.9700	0.9800	0.9321	0.9275
WPBEM [39]	0.9857	0.9881	0.9824	0.9354	0.9254
CNNs-LSTM-with-Aug [40]	0.9969	0.991	0.992	0.989	0.989
Proposed	**0.9909**	**0.994**	**0.998**	**0.995**	**0.995**

The utilization of biased fuzzy clustering removes the non-informative regions, such as background tissues and skull. It captures uncertainty while preserving structural information from data. Clustering improves the feature representation while resolving overlapping issues to enhance the accuracy. Then, precise segmentation of brain tissues is achieved through PCM, which is considered an important factor in AD. PCM captures complex patterns and improves feature discrimination ability. Additionally, hierarchical and local feature extraction (structural and spatial patterns) is achieved with a convolution module. The residual module provides stable training and gradient flow, and it improves the model performance.

The integration of ASAM improves the robustness and generalization of the classification process. CBi-LSTM network extracts temporal and spatial features, ensuring the detection of variation in brain structure. CBi-LSTM captures bidirectional and long-range dependencies from data in which the contextual relationship is learned. Therefore, it improves the accuracy with enhanced feature representation. Moreover, this process is strengthened using an adaptive spatial attention layer that considers the most relevant regions. Based on this, the model differentiates between similar classes. These design stages are combined to ensure comprehensive learning and generalization towards higher performance. In the proposed multiclass classification, the number of correct and incorrect classes is given in the confusion matrix. For each class (AD, CN, EMCI, LMCI, and MCI), the performance is computed in terms of accuracy, precision, recall, and F1-score metrics. The interpretation of class behavior for overall performance is reported using class-wise metric evaluation in Table 7.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

These results indicate better discrimination of the proposed framework among clinically similar categories. Subtle anatomical variation in the brain region is learned using ASTNet to demonstrate overlapping and progressive categories. Additionally, the exact brain regions are captured using PCM and segmentation with fuzzy clustering. These techniques enhance the feature representation while minimizing irrelevant and noisy tissue regions. From segmented data, clinically similar categories are differentiated through deep hierarchical features of the CBi-LSTM module. Moreover, adaptive spatial attention improves the feature representation of informative disease regions, providing accurate interclass categorization.

The overall improvement in classification performance is attributed to the complementary operation of the proposed framework rather than to a single network component. The pre-processing stage suppresses non-brain structures and enhances tissue representation, reducing irrelevant information before feature extraction. The CBi-LSTM module preserves contextual information during feature learning, while the adaptive spatial attention mechanism emphasizes disease-relevant anatomical regions. Finally, the shared-weight twin neural network improves the discriminative representation of closely related disease stages. The combination of these components contributes to improved classification performance across different AD categories under the evaluated experimental setting.

Table 7. Class-wise Performance Analysis

Classes	Accuracy	Precision	Recall	F1-score
AD	1.000	0.9697	0.9846	0. 9697
CN	0.9895	0.9895	0.9895	0.9895
EMCI	1.000	1.000	1.000	1.000
LMCI	1.000	1.000	1.000	1.000
MCI	0.992	1.000	0.999	0.9998

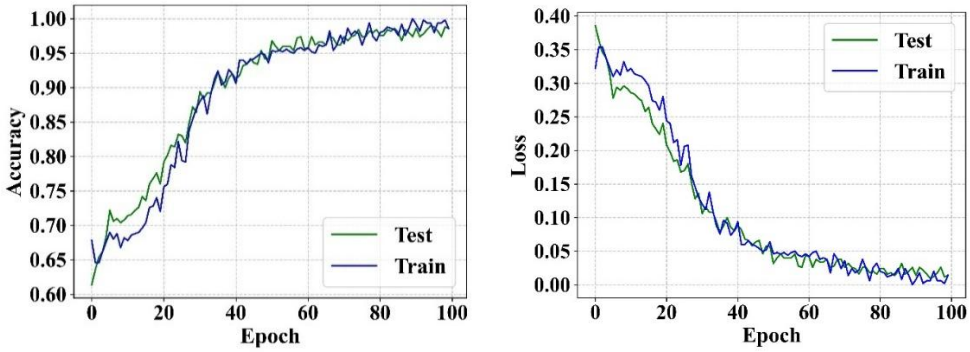


Fig. 15. Performance Comparison with Varying Number of Epochs (a) Accuracy (b) Loss

The performance with different numbers of epochs is shown in Fig. 15 with the training and testing accuracy. With higher epochs, the accuracy increases, and the training loss decreases. The accuracy performance is up to 50 epochs below 0.95. From 50-100, the performance is improved, and the optimal level is reached at epoch 100. Training accuracy is lower than testing accuracy. The lower loss is achieved as the number of epochs increases. The training loss is less than the testing loss of the proposed AD detection. The proposed loss is up to 50 epochs higher. With a value above 50, the loss is reduced, and the optimal performance level can be achieved at 100. In the early stages of epochs, a higher loss is experienced since the solution of the model is not optimal. Therefore, the predictions are inaccurate, which leads to a higher error rate.

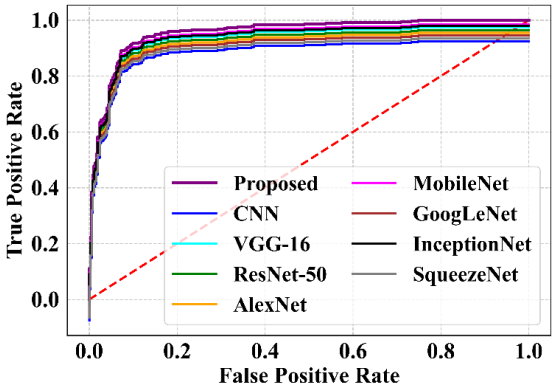


Fig. 16. ROC Curve Comparison

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

During the training process, the weight of the model is adjusted, and the loss is reduced. It indicates the learning and error minimization between actual and predicted values. While reaching an optimal solution, the loss is reduced at a slower rate. When overfitting occurs, the loss starts to increase for the testing set, even if the loss decreases. The proposed ROC result is analyzed using the existing approaches shown in Fig. 16. The ROC curve is generated using the one-vs-rest (OvR) procedure. In this evaluation, each class is considered a positive class, and the remaining classes are combined as a negative class. For each class, ROC is computed, and the overall performance is estimated using macro-average AUC. Therefore, equal weight is assigned to each class. The ROC results are compared with the traditional TL approaches, including VGG16, ResNet-50, AlexNet, and the CNN-based approach. Compared to existing approaches for AD prediction, the proposed approach achieves better performance. The ROC curve demonstrates the trade-off between the true positive rate and the false positive rate.

The proposed algorithm has an efficient learning process and preserves information with a spatial attention layer. The ROC curve for each class given in Fig. 17 shows the discrimination ability of the model across different classes. These curves are closer to the left corner, indicating higher TPR and lower FPR. The ROC values MCI 0.9972, EMCI 1.00, for LMCI 1.00, for CN 0.9908, and for AD 0.9848, representing strong classification performance across the model. These results indicate consistent classification performance across different classes. The difference between each class is represented by the deviation between ROC values in each class. Therefore, this ROC curve validates the robustness of the proposed model for accurate performance.

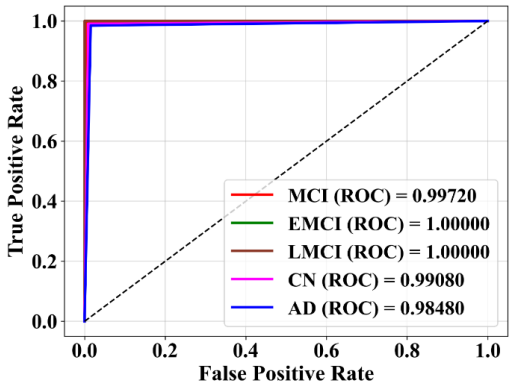


Fig. 17. ROC Curve for Each Class

The relatively consistent ROC performance across classes suggests that the model does not exhibit significant bias toward any particular class. However, slight variations in AUC values indicate that some classes are more easily distinguishable than others, which may be attributed to inter-class similarity and dataset distribution. Overall, the ROC analysis validates the robustness of the proposed approach and complements the accuracy, precision, and recall results, providing a more comprehensive evaluation of the model's performance. The comparative analysis with existing approaches is given in Table 8. The existing AD approaches are limited and have high computational time requirements. Moreover, the decision-making of these approaches is affected by the lack of feature extraction procedures. It can be resolved with the CBi-LSTM feature extraction technique, which captures past and future contextual information. It is significant for predicting complex relationships and dependencies.

The integration of flexibility, robustness, and memory capabilities leads to better disease identification for several applications. The accuracy performance and computational time performance of the proposed approach are better than those of existing AD approaches. This performance justifies the novelty of the proposed approach in all aspects. The proposed approach is significant in terms of robustness, accuracy, and interpretability. Variations in brain structure and function complicate AD prediction. The proposed approach enables the extraction of local and global features for accurate diagnosis. An adaptive twin neural network processes complex features to enhance the robustness and overall accuracy of AD diagnosis. This hybrid process permits the capture of disease-related features for better detection. By introducing biased fuzzy clustering, the model is adaptive to medical data uncertainty, which increases the accuracy and robustness of the process. Moreover, the spatial attention process permits the model to adapt to variations in brain structures, disease progressions, and scans. The integration of spatial attention, fuzzy clustering, and twin neural networks improves the accuracy and sensitivity to outperform conventional approaches. The quantitative comparison with recent state-of-the-art methods is given in Table 9.

In these methods, the ADNI dataset (AD, CN, EMCI, LMCI, and MCI) is used with similar experimental settings. There are several factors contributed for the enhancement of the proposed approach. Initially, the quality of the image is improved with fuzzy clustering-based segmentation. In this process, background noises are removed to get more informative features. PCM improves the robustness and generates a stable and reliable region. Then, contextual dependencies and spatial structures are effectively captured using the CBi-LSTM architecture. Moreover, the clinically accurate categorization of different classes is achieved using ASTNet. The integration of these modules improves the feature quality, discrimination ability, and robustness for better classification.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

Table 8. The Comparative Analysis with Existing Approaches

Approaches	Accuracy	Advantages	Disadvantages
CNN [34]	0.7433	Automatically learns hierarchical image features for MRI classification.	Lowest accuracy (0.7433) and limited capability to capture complex discriminative patterns.
VGG-16 [34]	0.839	Strong feature extraction with a simple and deep architecture.	Moderate accuracy (0.8390), a large number of parameters, and high computational cost.
ResNet-50 [34]	0.8317	Residual connections alleviate the vanishing gradient problem and enable deeper learning.	Moderate accuracy (0.8317) with relatively higher computational requirements.
AlexNet [16]	0.9874	High classification accuracy (0.9874) with a comparatively efficient architecture.	Memory-intensive and lower performance than the proposed model.
MobileNet [36]	0.9671	Lightweight architecture requiring limited computational resources.	Lower accuracy (0.9671) than deeper architectures for complex MRI patterns.
GoogLeNet [18]	0.9114	Efficient multi-scale feature extraction through Inception modules.	Moderate accuracy (0.9114) with increased architectural complexity.
InceptionNet [18]	0.9478	Learns rich multi-scale representations for MRI images.	Lower accuracy (0.9478) than the proposed model and a relatively complex architecture.
SqueezeNet [36]	0.8819	Very small model size with reduced memory requirements.	Lower accuracy (0.8819) and limited feature representation capability.
Proposed	0.9909	Highest classification accuracy (0.9909), effective spatial attention, improved generalization, and relatively lower computational complexity.	Requires sufficient labeled MRI data and further validation on larger multi-center datasets.

Table 9. The comparison with Recent State-of-the-art Methods

References	Techniques	Dataset	Validation	Accuracy (%)
Şener et al. [35]	EfficientNet121	ADNI (AD, CN, EMCI, LMCI, and MCI)	Hold-out validation	98.42
Alayba et al. [41]	Active contour algorithm (ACA), XGBoost, ANN, MobileNet-DenseNet121-Handcrafted, DenseNet121-GoogLeNet-Handcrafted, and MobileNet-GoogLeNet-Handcrafted	ADNI (AD, CN, EMCI, LMCI, and MCI)	5-fold cross-validation	98.8
Sener et al. [42]	vision transformers (ViT), EfficientNetB2 + FPN	ADNI (AD, CN, EMCI, LMCI, and MCI)	10-fold cross-validation	98.89
Proposed	CBi-LSTM	ADNI (AD, CN, EMCI, LMCI, and MCI)	5-fold cross-validation	99.09

4.3.3. Ablation Study

The proposed approach consists of multiple components. To evaluate the contribution of each component, the ablation study is conducted for the proposed approach. The effect of individual components on model accuracy is given in Table 10. By removing each component from the proposed framework, the accuracy is evaluated. When the pre-processing module is removed, the performance is reduced to 0.8302. This reduction is attributed to the lower quality of the raw input images. After removing biased Fuzzy Clustering and PCM, the performance is reduced to 0.8736 and 0.8536. Removing biased fuzzy clustering reduces the model's ability to capture structural relationships and uncertainty in the data. Therefore, the model's ability is reduced with less feature representation to handle overlapping, which leads to lower classification accuracy.

PCM is responsible for capturing the complex patterns. Therefore, the classification accuracy is reduced without PCM due to a lack of feature discrimination ability. In feature extraction, the absence of the convolution module reduces the performance to 0.8932. Removing the CBi-LSTM module reduces the accuracy to 0.7536. The convolution module is responsible for hierarchical and local feature extraction. Therefore, the spatial patterns and structural patterns are not efficiently captured, which leads to lower accuracy. CBi-LSTM captures bidirectional and long-range dependencies from data in which the contextual relationship is learned. Consequently, the reduction in contextual feature representation decreases classification accuracy.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

In ASTNet, the residual block and ASAM play a major role in enhancing the performance. When these modules are removed, the accuracy values are reduced to 0.8036 and 0.8526. The residual module provides stable training and gradient flow, and the absence degrades the model performance. Without ASAM, the robustness and generalization of classification are affected. This quantitative evidence shows that each module is essential for enhancing the performance of the proposed approach.

Table 10. Ablation study of the proposed ASTNet model

Model Variants	Accuracy
Without biased Fuzzy Clustering	0.8736
Without PCM	0.8536
Without pre-processing	0.8302
Without convolution	0.8932
Without CBi-LSTM	0.7536
Without ASAM	0.8526
Without Residual block	0.8036
Proposed	0.9909

4.3.4. Statistical Validation and Model Evaluation

The statistical significance of the proposed approach is evaluated based on the p-value. The p-values are evaluated based on two approaches, namely the paired t-test and the Wilcoxon signed-rank test. These are computed based on the performance score obtained from 5-fold cross validation. To compute p-value results, the proposed approach is compared to other approaches like ResNet50 [27], VGG-16 [27], DCNN [27], Shallow CNN [31], MTDL [32], WPBEM [33], and CNNs-LSTM-with-Aug [34]. This evaluation is based on the accuracy performance of each model on the ADNI dataset. Cross-validation results were used to estimate performance variability. The p-values of the proposed approach against other approaches are given in Table 11. P-value less than 0.05 represents a statistically significant improvement of ASTNet over the other compared model. Smaller p-values represent stronger evidence against the null hypothesis. The reported confidence intervals indicate the variability of the evaluated methods. These results indicate that the proposed model achieves competitive and statistically significant performance.

Furthermore, the p-value indicates the statistical significance of the proposed model. The statistical analysis demonstrates that the observed improvement is unlikely to be due to random variation under the evaluated experimental setting. The consistently low p-values obtained from both the paired t-test and the Wilcoxon signed-rank test indicate agreement between parametric and non-parametric statistical analyses. From the same cross validation results, 95% confidence intervals were computed using the standard error of the mean. Furthermore, the narrow confidence intervals suggest stable model performance across repeated evaluations. Together, these statistical measures support the reliability and consistency of the proposed framework while complementing the experimental performance metrics. All these statistical analyses were performed using the SciPy library.

Table 11. P-value of the Proposed Approach against Other Existing Approaches

Approaches	P-value (Paired t-test)	P-value (Wilcoxon signed-rank test)	95% Confidence Interval
ResNet50 [34]	0.012	0.013	[94.90–95.60]
VGG-16 [34]	0.008	0.006	[92.30–93.20]
DCNN [34]	0.015	0.012	[95.80–96.40]
Shallow CNN [37]	0.001	0.002	[99.55–99.78]
MTDL [38]	0.025	0.014	[95.70–96.30]
WPBEM [39]	0.010	0.001	[98.40–98.75]
CNNs-LSTM-with-Aug [40]	0.005	0.006	[98.60–98.78]
Proposed	-	-	[99.02-99.14]

The k-fold cross-validation performance of the proposed approach is given in Fig. 18. To ensure model robustness and prevent overfitting, k-fold cross-validation is performed within the training set. Therefore, the training data is split into k equal size folds. One-fold is used for validation, and the remaining folds are used for training in each iteration. This process is repeated 5 times, and each fold serves as a validation set exactly once. To select the optimal configuration, the average performance across all folds is considered. This approach permits the utilization of limited data and enhances the reliability and generalizability of the training model. The classification results of k-fold cross-validation with mean and standard deviation are given in Table 12. The low standard deviation across the five folds indicates stable and consistent model performance.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

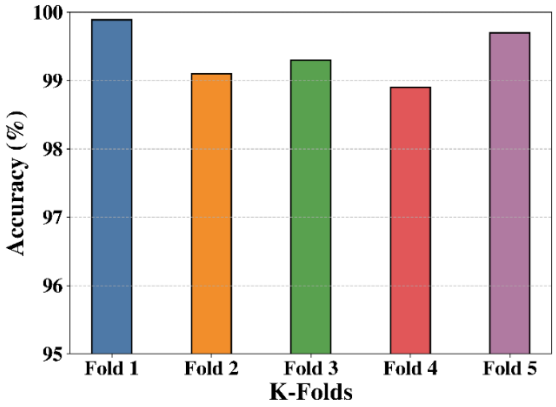


Fig. 18. K-fold cross-validation performance

Table 12. Classification Results of k-fold Cross-validation with Mean and Standard Deviation

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Fold 1	99.92	99.60	99.50	99.55
Fold 2	99.86	99.45	99.35	99.40
Fold 3	99.88	99.50	99.42	99.46
Fold 4	99.90	99.55	99.38	99.47
Fold 5	99.89	99.40	99.35	99.48
Mean	**99.89**	**99.50**	**99.40**	**99.47**
Std. Dev.	**0.02**	**0.07**	**0.06**	**0.05**

The proposed approach was evaluated on a held-out test set consisting of 15.5% of ADNI subjects that are completely independent of the cross-validation and training data. It ensures unbiased performance for the evaluation. In this independent test set, the accuracy of 0.9909 is achieved, and it represents a strong generalization ability. The performance achieved using the held-out test set is given in Table 13. When analyzing this performance, the EMCI and LMCI performance is lower than that of the CN and AD classes. This is likely due to the subtle anatomical differences between these disease stages. These results show that high performance across all disease types while considering class imbalance. The reported accuracy shows the generalization ability with unseen subjects. Accuracy values of cross validation, and independent test-set differ due to their role in evaluation. Cross validation is performed only on the training set during model development. An independent held-out test is performed on the training set for the final model evaluation. Differences between these values are noted because they are computed on different subsets of data.

Table 13. Performance on the Held-out Test Set

Classes	Precision	Recall	F1-score
CN	0.94	0.96	0.95
EMCI	0.90	0.87	0.87
LMCI	0.91	0.89	0.91
AD	0.95	0.97	0.96

The robustness of the proposed approach is evaluated using the Leave-One-Session-Out (LOSO) strategy under longitudinal deviation. In this evaluation, each session corresponds to a different acquisition time point for the same subject. Each subject contains multiple imaging sessions, and the number of sessions differs across each subject. In each LOSO iteration, MRI samples from one session are utilized for testing, and the remaining sessions are used for training. This process is continued for all available sessions to ensure different acquisition time points. The LOSO results are given in Table 14. It indicates that the proposed approach achieved higher performance across different imaging sessions. Although the number of samples varies across sessions, stable performance is achieved. It indicates the model's robustness to temporal variations in MRI scans. It ensures the generalization ability of the model under longitudinal clinical imaging scenarios.

Table 14. LOSO Performance of the Proposed Approach on the ADNI Dataset

Left-Out Session	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Session 1 (Baseline)	0.9896	0.9882	0.9885	0.9883
Session 2 (Subsequent)	0.9912	0.9897	0.9900	0.9898
Session 3 (Subsequent)	0.9905	0.9890	0.9893	0.9891
Session 4 (Subsequent)	0.9908	0.9894	0.9896	0.9895
Mean	0.**9905**	0.**9891**	0.**9893**	0.**9892**

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

4.3.5. Complexity Analysis

The time complexity comparison with existing approaches is shown in Fig. 19. The computational complexity of the proposed technique is compared with DL approaches, including CNN, VGG16, and ResNet. The time requirement for the proposed approach is 50 s, and for the other approaches, these values are above 60 s. The complexity analysis in terms of parameters, FLOPs, and training time is presented in Table 15. This table indicates that the complexity of the proposed approach is lower than that of the existing approach. The higher complexity of the compared models is primarily due to their larger number of parameters and deeper network architectures. The proposed ASTNet model achieves low time complexity due to several optimized components in its architecture. Initially, the pre-processing stage removes non-brain tissues and reduces irrelevant information, thereby reducing the computational burden of subsequent processing. PCM algorithm enhances segmentation efficiency by focusing only on relevant brain tissues, eliminating redundant computations.

Additionally, the CBi-LSTM network processes the features in parallel across spatial and temporal dimensions, improving processing speed. The adaptive spatial attention layer minimizes computation by concentrating only on important regions of the image, reducing the computational burden for the classification stage. Finally, the ASTNet effectively reuses the learned feature representations, further reducing the computational burden. Together, these factors contribute to a faster and more efficient AD process with lower time complexity.

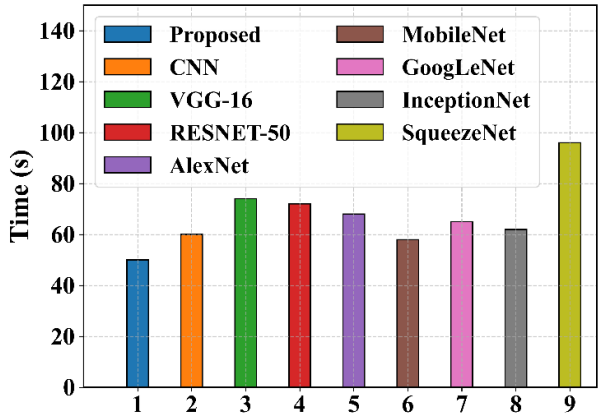


Fig. 19. Comparison of Time Complexity with Existing Approaches

Table 15. Computational Complexity Comparison of the Proposed Model and Existing Approaches in Terms of Parameters, FLOPs, and Training Time.

Model	Parameters (M)	FLOPs (G)	Training Time (s/epoch)
ResNet-50 [34]	25.6	4.1	120
VGG-16 [34]	138.0	15.5	210
DCNN [34]	10.2	2.8	95
Shallow CNN [37]	3.5	0.9	60
MTDL [38]	12.8	3.2	110
WPBEM [39]	8.4	2.1	85
CNN-LSTM-with-Aug [40]	14.6	3.9	130
Proposed	7.32	2.0	58.3

4.3.6. Gradient-weighted Class Activation Mapping (Grad-CAM) Visualization

To improve the interpretability of the proposed classification framework, Grad-CAM is employed. The model was trained on the proposed ADNI MRI dataset, in which the visualization results improve model interpretability. To achieve this, heatmaps are generated for the important brain regions that contribute to the model's predictions. The Grad-CAM Visualization of the proposed approach for the ADNI dataset is given in Fig. 20. This table contains original images, attention maps, and highlighted regions. In highlighted regions, the anatomically significant regions for AD are visualized. These regions are shown for all five classes, including AD, CN, EMCI, LMCI, and MCI. Each heatmap highlights different regions according to their contribution to the final prediction. Therefore, the important regions contributing to disease classification are visualized for a better understanding of the model.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

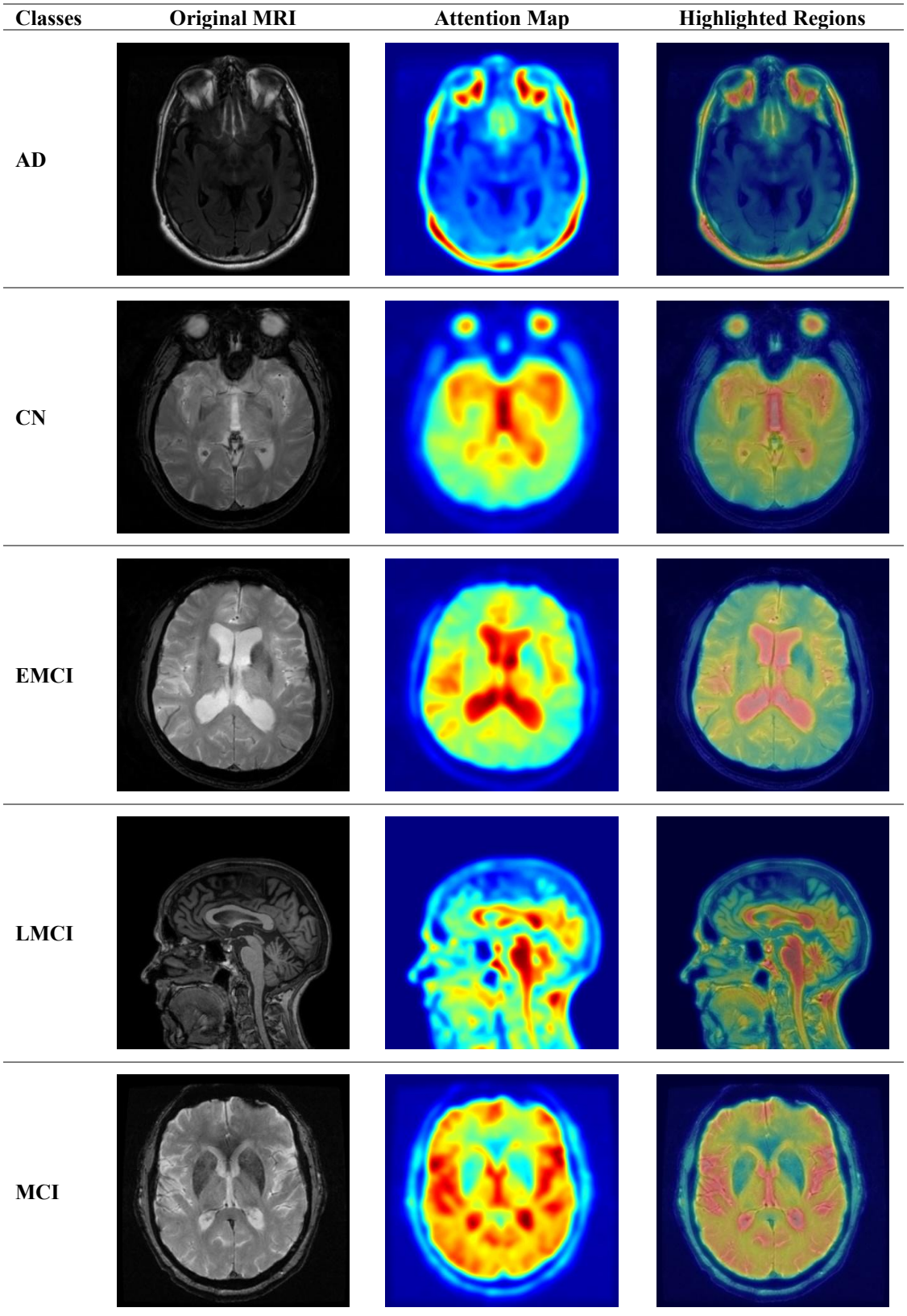


Fig. 20. Grad-CAM Visualization of the Proposed Approach (ADNI Dataset)

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

4.4. Practical Applicability in Clinical Settings

Integration with Current Clinical Workflow: The proposed model serves as a clinical support system for decision-making. For ongoing monitoring, the system could support longitudinal monitoring of AD progression. As patients return for follow-up visits and new imaging data is collected, clinicians can input the updated images, allowing the model to update its predictions based on the latest information.

Automated diagnosis: The model can automatically analyze and process medical images to reduce the burden on clinicians, support faster image analysis and assist clinicians during diagnosis.

Decision Support System: The model is considered a tool for decision-making regarding the early symptoms of AD, supporting timely clinical decision-making.

Real-Time Processing: The computational efficiency suggests the potential for near-real-time implementation on suitable hardware.

User-Friendly Interface: The proposed framework can be integrated into a user-friendly software interface for MRI-based diagnosis.

4.5. Impact on Early Diagnosis and Treatment

High Diagnostic Accuracy: The proposed approach demonstrates high diagnostic accuracy by detecting subtle variations in brain structure. It has the potential to reduce false-positive and false-negative classifications. From a clinical perspective, reliable differentiation among CN, EMCI, LMCI, MCI, and AD is particularly important because disease progression is gradual and the anatomical differences between adjacent stages are often subtle. Accurate identification of these stages may support clinicians in monitoring disease progression, planning follow-up evaluations, and selecting appropriate therapeutic interventions. Although the proposed framework is intended to function as a computer-aided decision support system rather than a replacement for clinical judgment, its ability to identify disease-related structural changes from MRI images demonstrates its potential value in assisting routine diagnostic workflows.

Personalized Treatment: An accurate early diagnosis may support the development of a more personalized treatment plan. By detecting AD, earlier interventions are performed, resulting in better patient outcomes.

Cost-Effective: The proposed approach minimizes the diagnosis cost, and hence, the burden on healthcare providers is reduced. It leads to efficient resource utilization by focusing on immediate intervention or efficient monitoring.

4.6. Limitations of the Proposed Study

Although the proposed framework demonstrates promising classification performance on the ADNI dataset, several limitations should be acknowledged. First, the experimental evaluation was conducted using a relatively limited number of subjects from a single public dataset. Therefore, further validation using additional multi-center and multi-scanner datasets is required to confirm the generalizability of the proposed framework. Second, the present study focuses only on MRI-based diagnosis and does not incorporate complementary clinical information such as cognitive scores, demographic variables, or genetic biomarkers, which may further improve diagnostic performance. Finally, although the proposed framework demonstrates high computational efficiency under the evaluated experimental setting, prospective clinical validation is necessary before deployment in routine clinical practice.

5 CONCLUSIONS

In this paper, AD classification from brain MRI images is accomplished using the proposed ASTNet framework. Skull removal was effectively achieved using the proposed biased fuzzy clustering algorithm while preserving the brain tissue regions. Long-term dependencies are handled during feature extraction with the CBi-LSTM network. The proposed framework integrates a ResNet-based twin neural network with an adaptive spatial attention mechanism for robust multi-class classification. The improved performance is attributed to the identity mapping of the residual blocks and the adaptive spatial attention mechanism, which emphasizes disease-relevant brain regions. The proposed model was evaluated using the ADNI dataset, and its performance was assessed using standard classification metrics. The proposed framework demonstrates that the integration of complementary pre-processing, feature extraction, attention-guided representation learning, and classification modules can effectively improve MRI-based AD classification. Rather than relying on a single DL architecture, the proposed design exploits the strengths of each component to enhance tissue representation, contextual feature learning, and disease-specific feature discrimination within a unified diagnostic framework.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

These findings indicate the potential of the proposed architecture as a computer-aided decision support tool for MRI-based AD assessment under the evaluated experimental setting. The accuracy, precision, recall, specificity, and F-measure values obtained with the proposed approach are 0.9909, 0.995, 0.994, 0.998, and 0.995. The proposed model demonstrates high classification performance with relatively low computational complexity under the evaluated experimental setting. The improved performance is attributed to effective feature extraction, adaptive spatial attention, and the proposed classification framework. Future work will focus on validating the proposed framework using multiple public datasets and exploring further optimization of each processing stage to improve generalization.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ETHICS STATEMENT

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

STATEMENT OF CONFLICT OF INTERESTS

The authors declare no conflicts of interest related to this study.

REFERENCES

[1] T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, "Multi-class Alzheimer's disease classification using image and clinical features," *Biomedical Signal Processing and Control*, vol. 43, pp. 64–74, Mar. 2018, doi: 10.1016/j.bspc.2018.02.019.

[2] S. Basheera and M. S. S. Ram, "Convolution neural network–based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation," *Alzheimer S & Dementia Translational Research & Clinical Interventions*, vol. 5, no. 1, pp. 974–986, Jan. 2019, doi: 10.1016/j.trci.2019.10.001.

[3] C. A. Valdez-Gaxiola, F. Rosales-Leycegui, A. Gaxiola-Rubio, J. M. Moreno-Ortiz, and L. E. Figuera, "Early- and Late-Onset Alzheimer's Disease: two sides of the same coin?," *Diseases*, vol. 12, no. 6, p. 110, May 2024, doi: 10.3390/diseases12060110.

[4] R. A. Hazarika, A. Abraham, S. N. Sur, A. K. Maji, and D. Kandar, "Different techniques for Alzheimer's disease classification using brain images: a study," *International Journal of Multimedia Information Retrieval*, vol. 10, no. 4, pp. 199–218, May 2021, doi: 10.1007/s13735-021-00210-9.

[5] J. M. F. Montenegro, B. Villarini, A. Angelopoulou, E. Kapetanios, J. Garcia-Rodriguez, and V. Argyriou, "A survey of Alzheimer's Disease Early Diagnosis Methods for Cognitive Assessment," *Sensors*, vol. 20, no. 24, p. 7292, Dec. 2020, doi: 10.3390/s20247292.

[6] N. V. Dokholyan, R. C. Mohs, and R. J. Bateman, "Challenges and progress in research, diagnostics, and therapeutics in Alzheimer's disease and related dementias," *Alzheimer S & Dementia Translational Research & Clinical Interventions*, vol. 8, no. 1, p. e12330, Jan. 2022, doi: 10.1002/trc2.12330.

[7] F. Li, D. Cheng and M. Liu, "Alzheimer's disease classification based on combination of multi-model convolutional networks," *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, Beijing, China, 2017, pp. 1-5, doi: 10.1109/IST.2017.8261566.

[8] V. Sathiyamoorthi, A. K. Ilavarasi, K. Murugeswari, S. T. Ahmed, B. A. Devi, and M. Kalipindi, "A deep convolutional neural network-based computer aided diagnosis system for the prediction of Alzheimer's disease in MRI images," *Measurement*, vol. 171, p. 108838, Dec. 2020, doi: 10.1016/j.measurement.2020.108838.

[9] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network-based Alzheimer's disease classification from magnetic resonance brain images," *Cognitive Systems Research*, vol. 57, pp. 147–159, Jan. 2019, doi: 10.1016/j.cogsys.2018.12.015.

[10] S. Naganandhini and P. Shanmugavadivu, "Effective Diagnosis of Alzheimer's Disease using Modified Decision Tree Classifier," *Procedia Computer Science*, vol. 165, pp. 548–555, Jan. 2019, doi: 10.1016/j.procs.2020.01.049.

[11] G. Battineni, N. Chintalapudi, F. Amenta, and E. Traini, "A comprehensive Machine-Learning model applied to magnetic resonance imaging (MRI) to predict Alzheimer's disease (AD) in older subjects," *Journal of Clinical Medicine*, vol. 9, no. 7, p. 2146, Jul. 2020, doi: 10.3390/jcm9072146.

[12] B. C. Simon, D. Baskar and V. S. Jayanthi, "Alzheimer's Disease Classification Using Deep Convolutional Neural Network," *2019 9th International Conference on Advances in Computing and Communication (ICACC)*, Kochi, India, 2019, pp. 204-208, doi: 10.1109/ICACC48162.2019.8986170.

[13] T. Jo, K. Nho, and A. J. Saykin, "Deep Learning in Alzheimer's Disease: Diagnostic classification and prognostic prediction using Neuroimaging data," *Frontiers in Aging Neuroscience*, vol. 11, p. 220, Aug. 2019, doi: 10.3389/fnagi.2019.00220.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

[14] J. Wen *et al.*, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Medical Image Analysis*, vol. 63, p. 101694, Apr. 2020, doi: 10.1016/j.media.2020.101694.
[15] A. Khalid, E. M. Senan, K. Al-Wagih, M. M. A. Al-Azzam, and Z. M. Alkhraisha, "Automatic analysis of MRI images for early prediction of Alzheimer's disease stages based on hybrid features of CNN and handcrafted features," *Diagnostics*, vol. 13, no. 9, p. 1654, May 2023, doi: 10.3390/diagnostics13091654.
[16] S. Mirzapour, M. Mohazzebi, and M. A. Asadi, "Automated Alzheimer's detection using MRI data: A novel approach with Siamese neural networks," *Biomedical Signal Processing and Control*, vol. 118, p. 109722, Feb. 2026, doi: 10.1016/j.bspc.2026.109722.
[17] G. Sahu, M. Nevendra and M. Karnati, "A Lightweight Dual-Branch Multi-Level Attentive Attributes with Constraint Fusion Network for EEG-Based Alzheimer's Detection," *IEEE Transactions on Automation Science and Engineering*, vol. 23, pp. 2427-2439, 2026, doi: 10.1109/TASE.2026.3654990.
[18] Z. Ullah and M. Jamjoom, "A deep learning for Alzheimer's stages detection using brain images," *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 74, no. 1, pp. 1457–1473, Sep. 2022, doi: 10.32604/cmc.2023.032752.
[19] J. Biswas *et al.*, "Performance-optimized Alzheimer's detection using machine learning with SMOTE and randomized hyperparameter tuning," *Discover Artificial Intelligence*, vol. 6, no. 1, Jan. 2026, doi: 10.1007/s44163-025-00758-z.
[20] 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, Feb. 2024, doi: 10.1016/j.bspc.2024.106079.
[21] A. S. Raj, C. Gunasundari, S. Senthilkumar, and S. Sivamani, "An optimized hybrid deep learning model to detect Alzheimer disease," *Scientific Reports*, vol. 15, no. 1, p. 34081, Sep. 2025, doi: 10.1038/s41598-025-14169-8.
[22] K. R. Kruthika, Rajeswari, and H. D. Maheshappa, "Multistage classifier-based approach for Alzheimer's disease prediction and retrieval," *Informatics in Medicine Unlocked*, vol. 14, pp. 34–42, Dec. 2018, doi: 10.1016/j.imu.2018.12.003.
[23] B. Lei *et al.*, "Deep and joint learning of longitudinal data for Alzheimer's disease prediction," *Pattern Recognition*, vol. 102, p. 107247, Jan. 2020, doi: 10.1016/j.patcog.2020.107247.
[24] S. Naz, A. Ashraf, and A. Zaib, "Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset," *Multimedia Systems*, vol. 28, no. 1, pp. 85–94, May 2021, doi: 10.1007/s00530-021-00797-3.
[25] C. Park, J. Ha, and S. Park, "Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset," *Expert Systems with Applications*, vol. 140, p. 112873, Aug. 2019, doi: 10.1016/j.eswa.2019.112873.
[26] X. Bi, S. Li, B. Xiao, Y. Li, G. Wang, and X. Ma, "Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology," *Neurocomputing*, vol. 392, pp. 296–304, May 2019, doi: 10.1016/j.neucom.2018.11.111.
[27] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang and S. Li, "Prediction of Alzheimer's Disease Progression with Multi-Information Generative Adversarial Network," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 711-719, March 2021, doi: 10.1109/JBHI.2020.3006925.
[28] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, and B. T. T. Yeo, "Predicting Alzheimer's disease progression using deep recurrent neural networks," *NeuroImage*, vol. 222, p. 117203, Aug. 2020, doi: 10.1016/j.neuroimage.2020.117203.
[29] F. the A. D. N. Initiative, "Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease," *EUR Research Repository (Erasmus University Rotterdam)*, Jan. 2021, doi: 10.1016/j.nicl.2021.102712.
[30] X. Hong *et al.*, "Predicting Alzheimer's Disease Using LSTM," *IEEE Access*, vol. 7, pp. 80893-80901, 2019, doi: 10.1109/ACCESS.2019.2919385.
[31] N. M. Khan, N. Abraham and M. Hon, "Transfer Learning with Intelligent Training Data Selection for Prediction of Alzheimer's Disease," *IEEE Access*, vol. 7, pp. 72726-72735, 2019, doi: 10.1109/ACCESS.2019.2920448.
[32] A. Auriemma, "Gene Ontology Terms Visualization with Dynamic Distance-Graph and Similarity Measures (S)," *Proceedings*, vol. 2021, pp. 85–91, Jun. 2021, doi: 10.18293/dmsviva21-013.
[33] M.-S. Yang and Y.-C. Tian, "Bias-correction fuzzy clustering algorithms," *Information Sciences*, vol. 309, pp. 138–162, Mar. 2015, doi: 10.1016/j.ins.2015.03.006.
[34] F. U. R. Faisal and G. -R. Kwon, "Automated Detection of Alzheimer's Disease and Mild Cognitive Impairment Using Whole Brain MRI," *IEEE Access*, vol. 10, pp. 65055-65066, 2022, doi: 10.1109/ACCESS.2022.3180073.
[35] B. Şener, K. Acici, and E. Sümer, "Categorization of Alzheimer's disease stages using deep learning approaches with McNemar's test," *PeerJ Computer Science*, vol. 10, p. e1877, Feb. 2024, doi: 10.7717/peerj-cs.1877.
[36] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv.org*, Feb. 24, 2016, doi: 10.48550/arXiv.1602.07360.
[37] M. El-Geneedy, H. E.-D. Moustafa, F. Khalifa, H. Khater, and E. AbdElhalim, "An MRI-based deep learning approach for accurate detection of Alzheimer's disease," *Alexandria Engineering Journal*, vol. 63, pp. 211–221, Aug. 2022, doi: 10.1016/j.aej.2022.07.062.

International Journal of Emerging Research in Engineering, Science, and Management
Vol. 5 | Issue 3 | pp. 14-43 | Jul-Sep 2026.
www.ijeresm.com | eISSN – 2583-4894
Received: 22 Mar 2026 | Revised: 24 Jun 2026 | Accepted: 03 Jul 2026 | Published: 21 Jul 2026

[38] S. Venkatasubramanian, J. N. Dwivedi, S. Raja, N. Rajeswari, J. Logeshwaran, and A. P. Kumar, "Prediction of Alzheimer's disease using DHO-Based pretrained CNN model," *Mathematical Problems in Engineering*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/1110500.

[39] S. Fathi, A. Ahmadi, A. Dehnad, M. Almasi-Dooghaee, M. Sadegh, and F. the A. D. N. Initiative, "A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images," *Neuroinformatics*, vol. 22, no. 1, pp. 89–105, Dec. 2023, doi: 10.1007/s12021-023-09646-2.

[40] S. E. Sorour, A. a. A. El-Mageed, K. M. Albarrak, A. K. Alnaim, A. A. Wafa, and E. El-Shafeiy, "Classification of Alzheimer's disease using MRI data based on Deep Learning Techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 2, p. 101940, Jan. 2024, doi: 10.1016/j.jksuci.2024.101940.

[41] A. M. Alayba, E. M. Senan, and J. S. Alshudukhi, "Enhancing early detection of Alzheimer's disease through hybrid models based on feature fusion of multi-CNN and handcrafted features," *Scientific Reports*, vol. 14, no. 1, p. 31203, Dec. 2024, doi: 10.1038/s41598-024-82544-y.

[42] B. Şener, K. Açıcı, and E. Sümer, "Improving early detection of Alzheimer's disease through MRI slice selection and deep learning techniques," *Scientific Reports*, vol. 15, no. 1, p. 29260, Aug. 2025, doi: 10.1038/s41598-025-14476-0.