Journal of Computer Science

Research Paper

Using Deep Learning to Support Clinical Decision-Making: The Case of Alzheimer's Disease Diagnosis

Nawal Mohamed Bahy Eldin[1,2], Ghada A. El Khayat[2], Abeer A. Amer[2,3]

[1] Department of Management Information Systems, Egyptian Institute of Alexandria Academy for Management & Accounting, Alexandria, Egypt
[2]Department of Information Systems and Computers, Faculty of Business, Alexandria, Egypt
[3]Department of Computer Science and Information Systems, Faculty of Management Science, Sadat Academy for Management Sciences

Article history
Received: 27 February 2025
Revised: 17 May 2025
Accepted: 27 May 2025

*Corresponding Author:
Nawal Mohamed Bahy Eldin
Department of Information Systems & Computers, Faculty of Business, Alexandria, Egypt
Email: nawal55mohamed55@gmail.com

Abstract: Alzheimer's disease is a chronic, progressive brain disorder that leads to a gradual decline in memory and cognitive functions. In this study, N-VGG16, an advanced deep learning model, is proposed. The model builds upon the VGG16 architecture, incorporating key enhancements to improve its ability to classify neurodegenerative conditions. The model processes structural neuroimaging data using a refined pipeline that applies adaptive histogram equalization for image enhancement and employs data augmentation techniques to address class imbalance issues. A major contribution of this work is the use of gradient-based localization, which allows the model's predictions to be linked to specific brain regions affected by the disease. Evaluation using a standardized dataset showed that the model achieved a high classification accuracy of 99.69%, successfully distinguishing between different clinical stages of Alzheimer's disease. Furthermore, visual interpretation confirmed that the model consistently focused on brain areas commonly associated with the disease. These findings highlight the model's potential to support clinical decision-making by offering both accurate diagnoses and interpretable insights.

Keywords: Alzheimer's Disease (AD), Deep Learning (DL), Convolutional Neural Networks (CNNs), Magnetic Resonance Imaging (MRI), Image Preprocessing, Transfer Learning

Introduction

Alzheimer's disease is characterized by progressive memory loss, cognitive decline, and behavioral changes (Alzheimer's Association, 2019). Symptoms develop gradually and worsen over time, significantly impairing daily functioning. Current diagnostic procedures, ranging from physical and cognitive assessments to neuroimaging and biomarker analysis, are often time-consuming, costly, and inaccessible in standard clinical environments. These challenges are magnified when detecting early-stage Alzheimer's, especially mild cognitive impairment (MCI), underscoring the urgent need for efficient and affordable diagnostic tools (Bootun et al., 2025). Early detection plays a crucial role in enabling timely interventions and providing support to both patients and caregivers. In 2019, Alzheimer's disease affected approximately 5.8 million Americans across all age groups (Alzheimer's Association, 2019). Traditional diagnostic tools heavily rely on expert interpretation and often lack consistency in identifying disease stages. In response, computer-aided diagnosis (CAD) systems have emerged as promising technologies, assisting physicians in making more accurate assessments and facilitating the development of reliable and highly accurate prediction models for early Alzheimer's detection (Sarakhsi et al., 2022). These tools are designed to enhance diagnostic decision-making for radiologists, clinicians, and caregivers. By optimizing CAD systems, healthcare institutions can improve operational efficiency, reduce medical costs, and promote better healthcare outcomes for individuals (Salehi et al., 2020). Deep learning plays a vital role in machine learning and draws inspiration from human cognitive processes, enabling intelligent systems to

SCIENCE Publications

analyze data and solve complex problems (Alsadhan, 2023). Convolutional Neural Networks (CNNs) have become the gold standard in medical image interpretation, demonstrating superior performance in recognizing radiomic patterns and outperforming traditional methods. CNNs are particularly effective in classifying medical images and assisting in manual annotation processes (Mehmood et al., 2020). Recent advances have shown the effectiveness of transfer learning in diagnostic image classification tasks. This research employs the VGG16 architecture, pretrained on the ImageNet dataset, as the foundation for Alzheimer's disease detection. This approach leverages the model's proven feature extraction capabilities while reducing computational demands. The deep convolutional structure of VGG16 is particularly well-suited for medical imaging due to its hierarchical feature learning ability and capacity to capture complex pathological patterns. Through systematic fine-tuning, this architecture is adapted to optimize performance for neuroimaging tasks, addressing the specific challenges posed by limited medical datasets.

This study presents a novel, tailored approach, N-VGG16, designed to optimize performance specifically for neuroimaging tasks. The primary contributions of this work are as follows:

1. Development of N-VGG16: A tailored variant of the VGG16 model with 12 fine-tuned convolutional layers optimized for structural neuroimaging analysis.
2. Design of a Robust Data Processing Pipeline: Incorporation of SMOTE-based class balancing (expanding the dataset from 5,154 to 7,770 samples), Contrast-Limited Adaptive Histogram Equalization (CLAHE) for MRI enhancement, and anatomically-aware augmentation techniques.
3. Generation of Clinically Interpretable Outputs: Utilizing Grad-CAM heatmap visualizations to highlight affected neuroanatomical regions, thereby providing transparency for diagnostic decision-making.

This paper provides a comprehensive overview of existing diagnostic methods, outlines the data sources, preprocessing techniques, and the proposed N-VGG16 model architecture. We present the experimental findings and comparative evaluations and summarize the key outcomes; suggesting future directions for enhancing model applicability in clinical settings.

Literature Review

Deep learning evolved as a highly effective method for improving diagnostic accuracy and patient outcomes. In recent years, various studies have investigated its potential applications in medical imaging, particularly in processing MRI scans to diagnose and classify Alzheimer's disease. These tests seek to identify modest structural abnormalities in the brain that may suggest the existence or progression of the disease. This section provides a summary of previous studies.

Deep Learning Models for Alzheimer's Disease Detection

Li et al. (2022) proposed a 3D automated method based on convolutional neural networks to classify AD and MCI using structural MRI scans. Their model accuracy of 94.19% for AD classification versus normal control (NC), and 94.57% for classifying normal control versus normal control (MCI).

Agarwal et al. (2023) study used the EfficientNet-b0 architecture. On the binary classification task of distinguishing stable mild cognitive impairment (sMCI) from Alzheimer's disease, the model achieved a training accuracy of 95.29% and a test accuracy of 93.10%. For the ternary classification (AD vs. normal cognitive impairment (CN) vs. stable MCI), the model achieved a training accuracy of 85.66% and a test accuracy of 87.38%.

Alsadhan (2023) proposed a computer-aided diagnosis system based on neuroimaging and deep learning for the early detection of Alzheimer's disease. The study compared the ResNet50 and VGG16 architectures for disease stage classification. Initial experiments using quadratic classification showed suboptimal performance, with VGG16 achieving 69% accuracy versus 60% for ResNet50. After improving the model by simplifying the tasks to binary classification and comprehensive metric analysis, VGG16 demonstrated superior performance.

Chakravarth & Shivakanth (2025) proposed a system that synergistically combines speech pattern analysis and MRI processing techniques. achieving a classification accuracy of 94.2%. This system relies on a dual deep learning approach, using a combined CNN-RNN model to capture temporal patterns in speech data and Vision Transformer networks to perform comprehensive spatial analysis of neuroimaging features.

Sekar et al. (2025) developed an improved Xception-based early-stage AD using MRI analysis. Their modified architecture demonstrated diagnostic performance with 96% classification accuracy, as well as 92% precision and 93% recall rates. The researchers used a comprehensive dataset sourced from Kaggle, covering the full spectrum of Alzheimer's disease, from non-senile to moderate dementia.

Deenadayalan & Shantharajah (2025) The study proposes a combination of EfficientNetB0 with dual attention mechanisms for the analysis of AD progression from MRI scans. This hybrid architecture demonstrated excellent diagnostic performance, achieving a training accuracy of 99.93% while maintaining strong generalization with a test accuracy of 93.59%.

Sountharajah et al. (2025) developed a weighted graph convolutional neural network (W-GCNN) the stage AD by analyzing brain connectivity using diffusion magnetic resonance imaging (DMI). Their framework successfully distinguished between three clinical categories model classification accuracy of 91%.

An approach utilizing a Convolutional Neural Network (CNN) to classify AD using MRI slices across coronal, sagittal, and axial planes was demonstrated by Ramineni et al. (2025), achieving 91% accuracy.

A multimodal model combining CNN and LSTM for Alzheimer's disease detection was developed by Haq et al. (2025). This method achieved 92.3% accuracy by transforming 3D MRI data into 2D features.

Research by El-Aziz et al. (2025) focused on a deep learning model that enhanced detection from MRI scans by integrating features from VGG16, MobileNet, and InceptionResNetV2, resulting in 97.93% accuracy.

Pre-trained models such as ResNet50, DenseNet121, and VGG19 were utilized by Islam & Uddin (2025) to classify the disease into four different stages, achieving high and consistent accuracies (up to 97.70%).

A 3D CNN model was proposed and evaluated by Rahman et al. (2025) using the ADNI dataset, achieving an accuracy of 92.89%.

For multi-stage AD detection, Sharma et al. (2025) implemented a CNN architecture using structural MRI data, demonstrating a strong test accuracy of 95.16%, which outperformed Inception-v3.

A combined ResNet50, Transformer, and LSTM model was presented by Wu et al. (2025) which processes MRI images from sagittal, coronal, and axial views. Using the ADNI dataset, this model achieved 96.92% accuracy.

The method proposed by Alorf (2025) involved combining the spatial feature extraction capabilities of CNNs and the contextual understanding of transformers using both MRI and clinical data, with the model achieving 96% accuracy.

An improved U-Net model was proposed by Kale & Chavan (2025). Their approach integrated feature extraction using ISIH, MBP, and Multi Texton, and achieved a final classification accuracy of 96.3% using an En-LeCILSTM model.

The deep learning model DHAN-GAN, proposed by Chen et al. (2025), combines heterogeneous dynamic attention networks and competitive generative networks to improve Alzheimer's disease diagnosis. The model effectively integrates three datasets—structural MRI, SNPs, and gene expression data—achieving 92.31% classification accuracy.

Clinical Challenges in AI-Based Alzheimer's Detection

While AI diagnostic tools show significant promise, several barriers hinder clinical adoption. The generalizability of diagnostic models is significantly constrained by two factors:

(1) inadequate representation across patient populations in existing datasets, and

(2) variability in medical imaging acquisition protocols.

Furthermore, the black box nature of many AI systems creates interpretability issues, undermining clinician confidence in these technologies (Schouten et al., 2025). Addressing these limitations requires coordinated efforts among AI developers, healthcare professionals, and regulators to establish standardized validation frameworks and to develop clinically transparent models.

Explainable AI (XAI) has become a critical solution, bridging the gap between complex algorithms and clinical utility. Studies by (Khan et al., 2022) and (Jahan & Khan, 2024) demonstrate that interpretable models not only maintain high diagnostic accuracy but also provide decision-making transparency, a crucial factor for clinician acceptance. Complementary to this, longitudinal datasets like those proposed by (Gkoumas et al., 2024) enable comprehensive disease progression tracking, enhancing predictive capabilities for early intervention.

Materials and Methods

The methodological framework, illustrated in Fig.1, follows a systematic path for classifying Alzheimer's disease from neuroimaging data. The workflow begins with MRI data collection and preparation and progresses through three critical preprocessing steps: dimensionality standardization through image resizing, contrast enhancement via CLAHE optimization, and class distribution balancing using SMOTE. The processed data are then fed into a custom N-VGG16 architecture, a modified version of the VGG16 network specifically for three-class neurological classification. To ensure clinical interpretability, the framework includes Grad-CAM visualization, which generates anatomical heatmaps that identify the brain regions most influential in the model's diagnostic decisions.

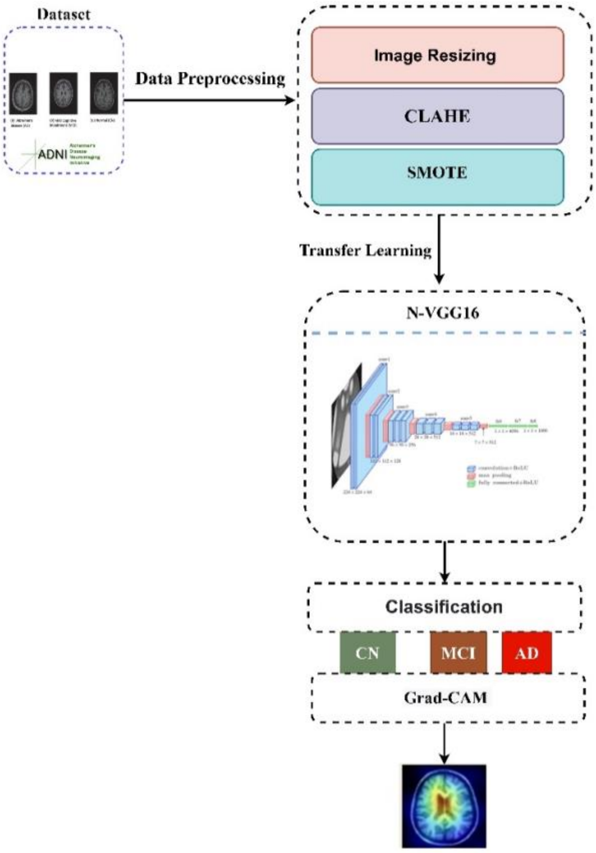

Fig. 1. Working procedure of the entire proposed model

Dataset

This study employed structural MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (Naz et al., 2022). The original dataset consisted of 199 baseline 3D T1-weighted whole-brain scans in NIfTI format, acquired from a North American cohort of older adults (55-90 years) with balanced gender distribution. For computational efficiency in convolutional neural network applications, a preprocessed 2D derivative of this dataset obtained through Kaggle was utilized. It contained axial plane extractions while maintaining traceability to the source ADNI scans. The image processing protocol focused on the medial temporal lobe structures, particularly the hippocampus, due to their established relevance in Alzheimer's pathology. From the volumetric scans, 2D axial slices at consistent anatomical landmarks were systematically extracted, generating 5,154 quality-controlled images stratified across three diagnostic categories (detailed in Table 1). The complete 3D source data remain accessible via the ADNI repository.

Preprocessing

The preprocessing phase involved three main steps to optimize the MRI data for deep learning analysis. First, all images were resized to a uniform resolution of 224×224 pixels. This uniformity ensures consistent input dimensions and enhances the convergence of the stable model. Second, to improve image quality, the Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm was applied to process the MRI images. This technique improves local contrast in images by segmenting them into small 8 x 8-pixel regions, then adjusting the intensity distribution in each region individually, with a maximum contrast limit of 2.0 to prevent noise amplification. This method helps highlight fine details of brain tissue, especially in the hippocampus and cerebral cortex, two important biomarkers for diagnosing Alzheimer's disease. Finally, the synthetic minority oversampling technique (SMOTE) addressed the class imbalance by creating synthetic samples of underrepresented classes, expanding the dataset from 5,154 to 7,770 images. The augmented dataset of 7,770 images was stratified at the patient level to ensure clinical validity and prevent data leakage. Training set (5,439 images, 70%), validation set (1,165 images, 15%), and test set (1,166 images, 15%) were used for final unbiased evaluation.

Table 1. Diagnostic Class Distribution

Class	Description	No. of Samples
AD	Alzheimer's disease	1124
MCI	Mild cognitive impairment	2590
CN	Cognitively Normal	1440

Model Architecture (N-VGG16)

This study is based on a transfer learning technique using the VGG16 architecture, a well-known convolutional neural network pre-trained on the ImageNet corpus, which contains over one million images across 1,000 categories. This architecture, developed by Simonyan and Zisserman of the Optical Engineering Group at Oxford University, has proven highly effective in image classification tasks due to its structural depth and the use of uniform-sized filters. In this work, the VGG16 convolutional baseline, consisting of 13 convolutional layers distributed over five blocks, is retained. Each block is followed by a max-pooling layer that progressively reduces spatial dimensions while preserving the underlying features. To ensure model stability, weight updates for these layers were disabled during training (freezing) to leverage previously acquired knowledge while reducing computational requirements. To adapt to the task of classifying brain MRI images into three stages of Alzheimer's disease (cognitively normal CN, mild cognitive impairment MCI, and Alzheimer's disease AD), the original classification layers were replaced with a custom classification header, resulting in the modified version called N-VGG16.

Global Average Pooling (GAP)

To optimize feature extraction and reduce computational complexity, a GlobalAveragePooling2D (GAP) layer was adopted in place of the conventional flattening operation following the convolutional base. This approach condenses spatial dimensions into a compact, translation-invariant feature vector by averaging each feature map, effectively preserving salient global patterns associated with brain structural changes while disregarding redundant localized details. By eliminating the need for flattening and subsequent dense connections, GAP inherently reduces the parameter count, thereby mitigating overfitting risks and enhancing the model's generalization capability.

Dense Layers and Regularization

The model employs two fully connected (Dense) layers (256 and 128 units) following the GAP layer, both utilizing ReLU activation for non-linear feature learning. To enhance generalization and prevent overfitting, L2 regularization is applied to each layer, constraining weight magnitudes while maintaining model expressivity. This architecture achieves an optimal balance between learning capacity and robustness, particularly important for medical imaging datasets with limited samples.

Batch Normalization and Dropout

Each Dense layer is followed by a Batch Normalization layer to standardize and stabilize the learning process, enhancing training efficiency and convergence. To further prevent overfitting, a Dropout layer with a rate of 0.25 was inserted after each batch normalization layer, which randomly deactivates neurons during training.

Output Layer

A final Dense output layer with three neurons, corresponding to the three Alzheimer's stages (CN, MCI, and AD), is added. It employs the softmax activation function to generate normalized probability scores, allowing the model to predict the most likely stage of Alzheimer's disease for each input MRI scan.

Grad-CAM Integration

To enhance clinical interpretability, the model incorporates Gradient-weighted Class Activation Mapping (Grad-CAM), a technique that generates visual explanations for predictions. This method extracts feature activations from the last convolutional layer (block5_conv3), computes gradient-weighted class-specific importance, and produces heatmaps that highlight the most influential brain regions in the model's decision-making process.

Fine-Tuning Strategy

To allow the model to adapt to the specific characteristics of Alzheimer's disease pathology visible in brain, the last 12 layers of the VGG16 base were fine-tuned. This enabled the model to adjust previously learned features to the new domain while preserving the generalizable knowledge gained from pretraining on ImageNet.

Results and Discussion

Statistical analysis of the confusion matrix, Fig.2 demonstrates that the proposed N-VGG16 model achieved exceptional performance in classifying Alzheimer's disease stages, confirming its high efficacy as a diagnostic aid. The N-VGG16 model attained perfect classification (100% accuracy) for Alzheimer's cases (AD) with 388 correct identifications and zero errors. For Mild Cognitive Impairment (MCI), it reached 99.7% accuracy (387 correct classifications with one misclassification as AD), while normal cognitive cases (CN) showed 99.7% accuracy (388 correct classifications with one misclassification as MCI). These outstanding results highlight N-VGG16's precision in differential diagnosis and its strong potential for clinical decision support, particularly in distinguishing between disease stages and early detection of cognitive decline.

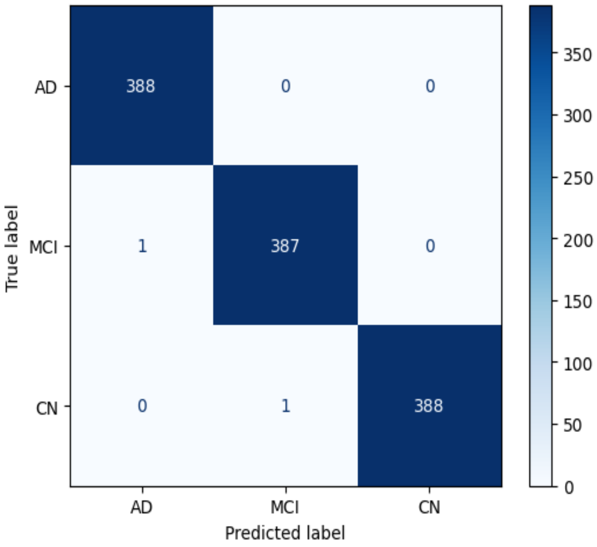

Fig. 2. Confusion matrix for classification results

The model showed improvement in its classification performance throughout the training process, as illustrated in Fig 3. This figure presents the progression of the model's learning based on two key indicators: accuracy and loss. The consistent decline in loss values for both training and validation datasets indicates that the model is effectively reducing prediction errors over time. In parallel, the gradual rise in accuracy reflects the model's increasing effectiveness in detecting Alzheimer's disease cases. Moreover, the close correspondence between training and validation curves suggests strong generalization ability and minimal risk of overfitting. These patterns collectively demonstrate the efficiency of the adopted model architecture and training approach in achieving reliable classification results.

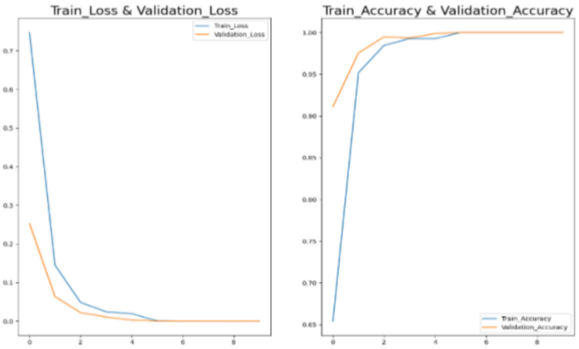

Fig. 3. Training Curves

A key technique for assessing how well predictive models work is the confusion matrix, which shows how well they can differentiate between various data patterns. Four primary indicators form the basis of this matrix: accurate positive and negative classifications, classification mistakes, and so on. Several performance metrics can be computed from these indicators, such as: overall accuracy, which shows the proportion of correct classifications; sensitivity, which gauges how well the model detects positive cases; predictive accuracy, which establishes the dependability of positive classifications; and the F1 metric, which strikes a balance between sensitivity and predictive accuracy to offer a thorough evaluation of model performance (Ruuska et al., 2018).

Accuracy (ACC), as shown in Equation 1, describes how accurate a classification model is in making predictions. It is calculated by comparing.

$$Accuracy\ (ACC) = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \qquad (1)$$

Recall as given in Equation 2 shows how effective a model is at detecting the correct positive outcomes.

$$\mathrm{Recall} \ = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (2)$$

Predictive accuracy (PPV), according to Equation 3 Predictive accuracy indicates how accurately a model classifies cases as positive.

$$\mathrm{Precision\ (PPV)} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad (3)$$

The F1 score, as shown in Equation 4, is used to evaluate the performance of a model by combining precision and recall into a single balanced metric.

$$\mathrm{F1-score} \ = \frac{2*\mathrm{TP}}{2*\mathrm{TP+FP+FN}} \qquad (4)$$

Table 2 shows the performance of the N-VGG16 model in classifying Alzheimer's disease stages. It achieved perfect accuracy (1.00) across all evaluation metrics (precision, recall, F1 score) for Alzheimer's disease (AD) cases, while achieving near-perfect results (0.996) for MCI and CN cases. The model's overall accuracy was 0.998 on the validation set of 1,165 samples, with weighted and overall means of 0.997, confirming the model's high efficiency and ability to accurately distinguish between different disease stages.

Table 2. Performance Metrics and Classification Report

Classification	Precision	Recall	F1-Score	Support
AD	1.00	1.00	1.00	388
MCI	0.996	0.996	0.996	388
CN	0.996	0.996	0.996	389
Averages				
Macro Avg.	0.997	0.997	0.997	1165
Weighted Avg.	0.997	0.997	0.997	1165
Overall Accuracy	-	-	**0.998**	**1165**

Interpretability of deep learning models is particularly critical in medical imaging applications, where clinical trust and decision transparency are paramount. To address this, Gradient-weighted Class Activation Mapping (Grad-CAM) has emerged as a powerful visualization tool, first introduced by Selvaraju et al. (2017) in their seminal work (Selvaraju et al., 2017). As demonstrated in subsequent neuroimaging studies (Loveleen et al., 2023) (Fareed et al., 2023). Grad-CAM acts as a visual explanatory framework that highlights the discriminative image regions influencing a model's predictions. In the context of Alzheimer's disease classification, the current implementation applies Grad-CAM to the final convolutional layers (block5_conv3) of the N-VGG16 model. The Grad-CAM-generated heatmaps provide clinically interpretable visualizations that precisely localize disease-specific neuroanatomical patterns (Loveleen et al., 2023). The method computes gradient-weighted activations, producing spatial attention maps that validate the model's focus areas against clinical knowledge. As shown in Fig.4, these visualizations for AD, MCI, and cognitively normal (CN) subjects not only enhance model interpretability but also provide clinicians with verifiable evidence supporting diagnostic predictions (Fareed et al., 2023). The technique's ability to localize decision-relevant features makes it particularly valuable for validating deep learning systems in medical imaging domains.

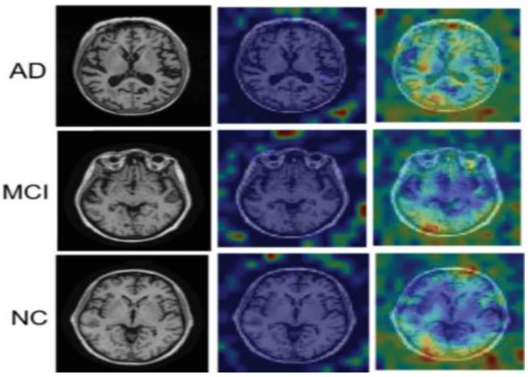

Fig. 4. Grad-CAM algorithm applied to AD, MCI, and CN images to visualize the results.

Deep learning techniques are rapidly advancing in the classification of Alzheimer's disease (AD) stages using MRI, with diverse data preprocessing and sample amplification approaches, as shown in Table 3. Most studies rely on publicly available datasets such as ADNI and Kaggle, using traditional amplification techniques such as horizontal and vertical flipping, rotation, and geometric transformations.

While preprocessing methods vary from rescaling, skull removal, and density normalization, the proposed model introduces an advanced contrast-level enhancement (CLAHE) technique to enhance subtle details of clinical relevance. Models used in the literature range from convolutional neural networks (CNNs) to hybrids, with significant variations in accuracy (60%–97.93%), as shown in the table. Many of them face key challenges such as data imbalance, small samples, and clinical interpretation difficulties. In contrast, the proposed improved model (N-VGG16) offers a comprehensive solution that achieves an exceptional classification accuracy of 99.69%. This is achieved through a series of improvements. First, the model's architecture was improved by replacing traditional layers with advanced factorial aggregation (GAP) layers and adding dense neural layers with advanced regularization mechanisms. Second, integrated strategies were incorporated to interpret diagnostic decisions using thermal imaging techniques (Grad-CAM), enabling clinicians to understand the decision-making logic. Third, Data challenges were overcome by applying advanced pre-processing techniques.

Table 3. Compared Models

Author	Methodology	Accuracy	Limitations	Dataset
(Alsadhan, 2023)	This study developed a computer system based on convolutional neural networks (CNNs) models improved by transferring knowledge from previous VGGNet and ResNet models. The images underwent preprocessing, including format conversion, dimensionality standardization (244×244 pixels), and intensity normalization.	VGG16: accuracy of 69% ResNet50: accuracy of 60%	The study faced several significant limitations that affected the model's performance and practical applicability. The most significant of these challenges was with moderate dementia cases, representing only 1.02% of the total sample (52 images out of 5,121). This imbalance led to a significant decrease in the model's accuracy when classifying multiple categories compared to binary classification. The model also suffers from a lack of clear mechanisms for interpreting decisions (XAI).	Kaggle
(Agarwal et al., 2023)	The research methodology used a combination of end-to-end learning and transfer learning methodologies to classify Alzheimer's disease using MRI data. Specifically, the study applied the EfficientNet-B0 architecture trained on a dataset of 458 T1-weighted brain MRI.	training and 87.38% accuracy and 88.00%	The study encountered several significant limitations that impacted the model's performance. The most notable of these was the limited sample size (229 AD cases, 229 MCI cases, and 245 CN cases). This shortcoming led to a significant decrease in the model's accuracy. The model also lacks clear mechanisms for interpreting decisions (XAI).	ADNI
(K et al., 2025)	This research presents an improved deep learning framework based on Xception. The research used a neuroimaging dataset spanning four cognitive stages. Before training the model, all MRI scans were preprocessed.	96%	The study faces some limitations that affect its results and practical applications. First, the dataset suffers from a significant imbalance in distribution across different disease categories, and the model has not been tested on preclinical cases. A lack of interpretable XAI techniques to explain the model's decisions.	Kaggle

Author	Methodology	Accuracy	Limitations	Dataset
(Deenadayalan & Shantharajah, 2025)	This study used the EfficientNetB0 model enhanced with dual attention mechanisms to extract features from three different levels. SMOTE data balance techniques were also incorporated to address the problem of class misalignment. The images underwent advanced processing.	93.59%	The model demonstrated difficulty in differentiating between intermediate disease stages (MiD and MoD) due to overlapping imaging features. despite employing attention mechanisms, the model's limited interpretability hinders its clinical adoption, as decision-making processes remain inadequately transparent for medical practitioners.	Kaggle
(Sountharajah et al., 2025)	This study adopted a methodology for analyzing structural brain networks using the W-GCNN (Weighted Graph Convolutional Network) model to distinguish between CN, MCI, and AD. The research sample included 358 participants. MRI data were processed, and the proposed model relied on four graph convolutional layers for feature extraction.	91%	Although the model demonstrated accurate classification performance, it lacks interpretable explainable AI (XAI) techniques to clarify its decision-making process.	ADNI
(Ramineni et al., 2025)	This study developed a CNN-based framework for Alzheimer's disease classification using structural MRI data. T1-weighted images were processed through a multi-planar analysis pipeline, where each volumetric scan was segmented into coronal, sagittal, and axial plane slices. These orthogonal views were then systematically processed through the CNN architecture to capture complementary neuroanatomical features.	91%	Although the model demonstrated accurate classification performance, it lacks interpretable explainable AI (XAI) techniques to clarify its decision-making process.	ADNI
(Haq et al., 2025)	This study developed a hybrid model combining convolutional neural networks (CNNs) and long-short-term memory (LSTM) networks. The research relied on 505 scans of Alzheimer's disease (AD) patients (135), mild cognitive impairment (MCI) patients (215), and normal individuals (155). The 3D images were converted into 2D slices.	92.3%	Although the model demonstrated accurate classification performance, it lacks interpretable explainable AI (XAI) techniques to clarify its decision-making process.	ADNI-1
(El-Aziz et al., 2025)	This research presents a deep learning framework for diagnosing Alzheimer's disease from MRI. It employs VGG16, MobileNet, and InceptionResNetV2 were used with an improved weighted fusion mechanism.	97.93%	Although the model has high accuracy, it suffers from data imbalance (only 64 samples for advanced stages compared to 3,200 samples for normal stages). Second, the model lacks clear mechanisms to explain the basis for case classification.	Kaggle
(Islam & Uddin, 2025)	This study developed a methodology for detecting the four stages of AD using transfer learning techniques. The research relied on three pre-trained neural network models (ResNet50, DenseNet121, and VGG19) optimized for the medical classification task.	97.70%	This study faces several methodological limitations: First, the model lacks interpretable AI (XAI) mechanisms, which limits clinicians' ability to understand the rationale behind the diagnostic decisions made by the model. Second, there is significant variation in the sample distribution across categories, with only 64 cases of moderate dementia compared to a total of 3,200 cases.	Kaggle
(Rahman et al., 2025)	In this study, a 3D MRI-based model was designed to detect Alzheimer's disease. The research relied on a dataset comprising 2,182 scans of 221 patients, divided into three main categories: CN, MCI, and AD. A series of advanced data processing processes were applied, including image refinement of non-brain tissues, standardization of image anatomy parameters, image quality enhancement using Gaussian filters, and careful selection of the most diagnostically significant segments.	84.05%	This study faces two major challenges: First, the imbalance between classes, with Alzheimer's disease (453) patients sampled compared to mild cognitive impairment (981), which may affect the model's accuracy in classifying underrepresented cases. Second, the lack of analysis to interpret the model's decisions.	ADNI

Author	Methodology	Accuracy	Limitations	Dataset
Proposed model	A hybrid deep learning framework that combines automated classification with clinically interpretable visualization for Alzheimer's disease diagnosis is developed. Two key architectural innovations to the VGG16 baseline (creating N-VGG16) were introduced: (1) replacement of traditional pooling with generalized average pooling (GAP) to preserve spatial relationships in neuroimaging data, and (2) integration of Grad-CAM to generate intuitive heatmaps highlighting disease-relevant brain regions. For optimal performance, comprehensive preprocessing pipeline including CLAHE-based contrast enhancement and addressed class imbalance through SMOTE oversampling was implemented. The model was trained with early stopping, achieving both high diagnostic accuracy and clinically meaningful explanations through pathology-aligned visual biomarkers.	99.68%	This study overcomes the major limitations of previous research by developing an integrated hybrid model that combines high diagnostic accuracy with clinical interpretability. This approach relies on several innovative strategies: (1) addressing the data imbalance issue through artificial augmentation techniques, (2) enhancing interpretability by incorporating Grad-CAM mechanisms to generate heat maps that highlight key diagnostic regions, and (3) achieving a high accuracy of 99.6% thanks to structural improvements in the enhanced N-VGG16 model. The model also demonstrated practical clinical applicability with reasonable computational requirements. These comprehensive improvements provide an integrated solution that exceeds the limitations of traditional models in terms of accuracy, interpretability, and clinical applicability.	Kaggle

Conclusion

This study presents a model for Alzheimer's disease (AD) staging using magnetic resonance imaging (MRI). The proposed improved N-VGG16 model, enhanced with transfer learning capabilities, demonstrates exceptional diagnostic reliability with a test accuracy of 99.69%, supported by robust training (99.86%) and validation (99.98%) results. Grad-CAM imaging ensures the clinical interpretability of the model, which accurately identifies known biomarkers of AD, including hippocampal atrophy, with remarkable anatomical alignment. By combining high diagnostic accuracy with transparent decision-making processes, this framework represents a significant advance in neuroscience. Future applications include integrating multimodal data (such as PET and cerebrospinal fluid biomarkers) and optimizing the model for use on portable medical devices.

Acknowledgement

The authors acknowledge Kaggle repository for providing accessible MRI datasets for Alzheimer’s disease, which were essential for conducting and validating this research. Sincere gratitude is due to the Department of Information Systems and Computers, Faculty of Business, Alexandria University for their continuous support, valuable resources, and guidance throughout the course of this study.

Funding Information

The authors declare no financial support or funding to disclose.

Author Contributions

Nawal Mohamed Bahy Eldin: Conceptualization, methodology design, algorithm implementation, and manuscript writing.
Ghada A. El Khayat: Contributed to the refinement of the research problem, supervised and guided the analysis and the results presentation and contributed to writing and revising the manuscript.
Abeer A. Amer: Collected the data, contributed analysis, and participated in manuscript writing.

Ethics

This study used publicly available, anonymized datasets and did not involve any human or animal participants directly. Hence, ethical approval was not required. All procedures followed are in accordance with institutional guidelines and data usage policies.

References

Agarwal, D., Berbís, M. Á., Luna, A., Lipari, V., Ballester, J. B., & Torre-Díez, I. d. (2023). Automated Medical Diagnosis of Alzheimer´s Disease Using an Efficient Net Convolutional Neural Network. *Journal of Medical Systems, 47*, 1-57. https://doi.org/10.1007/s10916-023-01941-4

Alorf, A. (2025). Transformer and Convolutional Neural Network: A Hybrid Model for Multimodal Data in Multiclass Classification of Alzheimer's Disease. *Mathematics, 13*(10), 1-34. https://doi.org/10.3390/math13101548

Alsadhan, N. (2023). Image-Based Alzheimer's Disease Detection Using Pretrained Convolutional Neural Network Models. *Journal of Computer Science, 19*(7), 877-887. https://doi.org/10.3844/jcssp.2023.877.887

Alzheimer's Association. (2019). 2019 Alzheimer's disease facts and figures. *15*(3), 321-387. https://doi.org/10.1016/j.jalz.2019.01.010

Bootun , D., Auzine, M. M., Ayesha, N., Idris, S., Saba, T., & Khan, M. H.-M. (2025). ADAMAEX—Alzheimer's disease classification via attention-enhanced autoencoders and XAI. *Egyptian Informatics Journal, 30*, 1-16. https://doi.org/10.1016/j.eij.2025.100688

Chakravarth, B. A., & Shivakanth, G. (2025). Integrating Multimodal AI Techniques and MRI Preprocessing for Enhanced Diagnosis of Alzheimer's Disease: Clinical Applications and Research Horizons. *IEEE Access, 13*, 63519-63531. https://doi.org/10.1109/ACCESS.2025.3557533

Chen, X., Wang, S., & Kong, W. (2025). Multimodal data fusion for Alzheimer's disease based on dynamic heterogeneous graph convolutional neural network and generative adversarial network. *Array, 26*, 1-17. https://doi.org/10.1016/j.array.2025.100415

Deenadayalan, T., & Shantharajah, S. P. (2025). Prognostic Survival Analysis for AD Diagnosis and Progression Using MRI Data: An AI-Based Approach. *IEEE Access, 13*, 89059-89078. https://doi.org/10.1109/ACCESS.2025.3564611

El-Aziz, A. A., Mostafa, A. M., Ezz, M., Mostafa, E., Alsayat, A., & Abd El-Ghany, S. (2025). Ensemble deep learning for Alzheimer's disease diagnosis using MRI: Integrating features from VGG16, MobileNet, and InceptionResNetV2 models. *PloS one, 20*(4), 1-29. https://doi.org/10.1371/journal.pone.0318620

Fareed, M. S., Zikria, S., Ahmed, G., Mui-Zzud-Din, Mahmood, S., & Aslam, M. (2023). ADD-Net: An Effective Deep Learning Model for Early Detection of Alzheimer Disease in MRI Scans. *IEEE Access, 10*, 96930-96951. https://doi.org/10.1109/ACCESS.2022.3204395

Gkoumas, D., Wang, B., Tsakalidis, A., Wolters, M., Purver, M.,Zubiaga , A., & Liakata, M. (2024). A longitudinal multi-modal dataset for dementia monitoring and diagnosis. *Language Resources and Evaluation, 58*(3), 883-902. https://doi.org/10.1007/s10579-023-09718-4

Haq, E. U., Yong, Q., Yuan, Z., Huarong, X., & Haq, R. U. (2025). Multimodal fusion diagnosis of the Alzheimer's disease via lightweight CNN-LSTM model using magnetic resonance imaging (MRI). *Biomedical Signal Processing and Control, 104*. https://doi.org/10.1016/j.bspc.2025.107545

Islam,M., & Uddin, J. (2025). Transfer Learning for Detecting Alzheimer's Disease in Brain Using Magnetic Resonance Images. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 13*(1), 45-56. https://doi.org/10.52549/ijeei.v13i1.5069

Jahan, Z., & Khan, S. B. (2024). Early dementia detection with speech analysis and machine learning techniques. *Discover Sustainability, 5*, 1-18. https://doi.org/10.1007/s43621-024-00217-2

K, S., R, M., S, M., T, D., K, N. A., & Borah, M. D. (2025). Early Detection Of Alzheimer's Disease using Transfer Learning on MRI Data. *3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, (pp. 1-6). Raichur, India. https://doi.org/10.1109/ICICACS65178.2025.10968359

Kale ,S. J., & Chavan, P. U. (2025). Deep ensemble architecture with improved segmentation model for Alzheimer's disease detection. *Journal of Medical Engineering & Technology, 49*(4), 97-121. https://doi.org/10.1080/03091902.2025.2484691

Khan, Y. F., Kaushik, B., Imam Rahmani, M. K., & Ahmed, M. E. (2022). Stacked Deep Dense Neural Network Model to Predict Alzheimer's Dementia Using Audio Transcript Data. *IEEE Access, 10*, 32750 - 32765. https://doi.org/10.1109/ACCESS.2022.3161749

Li, J., Wei, Y., Wang, C., Hu, Q., Liu, Y., & Xu, L. (2022). 3-D CNN-Based Multichannel Contrastive Learning for Alzheimer's Disease Automatic Diagnosis. *IEEE Transactions on Instrumentation and Measurement, 71*, 1-11. https://doi.org/10.1109/TIM.2022.3162265

Loveleen, G., Mohan, B., Shikhar, B. S., Nz, J., Shorfuzzaman, M., & Masud, M. (2023). Explanation-Driven HCI Model to Examine the Mini-Mental State for Alzheimer's Disease. *ACM Transactions on Multimedia Computing, Communications and Applications, 20*(2), 1-16. https://doi.org/10.1145/3527174

Mehmood, A., Maqsood, M., Bashir, M., & Shuyuan, Y. (2020). A deep Siamese convolution neural network for multi-class classification of Alzheimer disease. *Brain sciences, 10*(2), 1-15. https://doi.org/10.3390/brainsci10020084

Naz, S., Ashraf , A., & Zaib , A. (2022). Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. *Multimedia Systems, 28*, 85–94. https://doi.org/10.1007/s00530-021-00797-3

Rahman, A. U., Ali , S., Saqia , B., Halim, Z., Al-Khasawneh, M. A., AlHammadi, D. A., Khan, M. Z., Ullah, I., & Alharbi, M. (2025). Alzheimer's disease prediction using 3D-CNNs: Intelligent processing of neuroimaging data. *SLAS technology, 32*, 1-10. https://doi.org/10.1016/j.slast.2025.100265

Ramineni, V., Khan, F. F., Pyun, J.-Y., & Kwon, G.-R. (2025). Convolutional Inception v4 for Alzheimer's Disease Diagnosis Using Multi-Plane MRI Data. *2025 IEEE International Conference on Consumer Electronics (ICCE).* Las Vegas. https://doi.org/10.1109/ICCE63647.2025.10929903

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., & Mononen, J. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behavioural processes, 148*, 56-62. https://doi.org/10.1016/j.beproc.2018.01.004

Salehi, A. W., Sharma, B. B., Gupta, G., Baglat, P., & Upadhya, A. (2020). A CNN Model: Earlier Diagnosis and Classification of Alzheimer Disease using MRI. *International Conference on Smart Electronics and Communication (ICOSEC)*, (pp. 156-161). Trichy. https://doi.org/10.1109/ICOSEC49089.2020.9215402

Sarakhsi, M., Haghighi, S. S., Fatemi Ghomi, S. M., & Marchiori, E. (2022). Deep learning for Alzheimer's disease diagnosis: A survey. *Artificial intelligence in medicine, 130*. https://doi.org/10.1016/j.artmed.2022.102332

Schouten, D., Nicoletti, G., Dille, B., Chia, C., Vendittelli, P., Schuurmans, M., Litjens, G., & Khalili, N. (2025). Navigating the landscape of multimodal AI in medicine: a scoping review on technical challenges and clinical applications. *Medical Image Analysis, 105*, 1-17. https://doi.org/10.1016/j.media.2025.103621

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra , D. (2017). Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618-626.

Sharma, P., Verma, A., & Khari, M. (2025). Advancing Humanitarian Efforts in Alzheimer's Diagnosis Using AI and MRI Technology. In *AI for Humanitarianism* (pp. 139-154).

Sounthararajah, J., Kumaralingam, L., Jegatheeswaran, T., Srivishagan, S., Thanikasalam, K., & Ratnarajah, N. (2025). Classification of Alzheimer’s Disease Stages using Weighted Brain Connectivity based Graph Convolutional Neural Network. *5th International Conference on Advanced Research in Computing (ICARC)*, (pp. 1-6). Sri Lanka. https://doi.org/10.1109/ICARC64760.2025.10963110

Wu, Q., Wang, Y., Zhang , X., Zhang, H., & Che, K. (2025). A hybrid transformer-based approach for early detection of Alzheimer’s disease using MRI images. *BioImpacts: BI, 15*(1). https://doi.org/10.34172/bi.30849