ACCSCIENCE PUBLISHING

Brain & Heart

ORIGINAL RESEARCH ARTICLE

Early prediction of Alzheimer's disease using machine learning algorithm: A convolutional neural network approach

Babatunde Hamzat[1], Sangita Pokhrel[2]*, and Swathi Ganesan[2]

[1]Department of Data Science, York St John University, London, United Kingdom

[2]Department of Computer and Data Science, York St John University, London, United Kingdom

(This article belongs to the *Special Issue: The Current and Future Landscape of Alzheimer's Disease Treatment*)

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that severely impacts memory and cognitive functions. Early diagnosis remains crucial for timely intervention and care. This research aims to explore the use of artificial intelligence, specifically deep learning, for the early prediction and classification of AD using structural magnetic resonance imaging (MRI) images. A dataset comprising approximately 44,000 brain MRI images with four diagnostic classes (mild, moderate, severe, and very severe dementia) was used to train and evaluate multiple convolutional neural network (CNN) architectures. Three deep learning models were developed and tested: A custom CNN built from scratch, a spatial-channel convolutional attention network (SCCAN), and a pre-trained Visual Geometry Group VGG16 model using transfer learning. The methodology included extensive preprocessing, data augmentation, normalization, and a train–validation–test split to ensure robust performance. Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices were used to assess classification efficacy. Among the models tested, the Visual Geometry Group VGG16 model achieved the highest classification accuracy, closely followed by the SCCAN, while the custom CNN demonstrated competitive performance with fewer layers. Grad-CAM visualizations were integrated to provide insight into model decision-making, enhancing interpretability. The results confirm the effectiveness of deep learning in classifying early AD stages with high accuracy and support its integration into clinical diagnostic tools. However, the study also identifies limitations, including dataset diversity, class imbalance, and generalizability across diverse populations. Future research should consider using larger, multi-center datasets (including positron emission tomography and electroencephalography modalities). This project demonstrates that deep learning can offer reliable, scalable, and interpretable solutions for the early detection of AD, potentially transforming the diagnostic pathway and enabling earlier therapeutic interventions.

Keywords: Alzheimer's disease; Early detection; Deep learning; Convolutional neural network; Magnetic resonance imaging

***Corresponding author:**
Sangita Pokhrel
(s.pokhrel@yorksj.ac.uk)

Citation: Hamzat B, Pokhrel S, Ganesan S. Early prediction of Alzheimer's disease using machine learning algorithm: A convolutional neural network approach. *Brain & Heart.* 2025;3(4):025310043. doi: 10.36922/BH025310043

Received: July 28, 2025

1st revised: September 27, 2025

2nd revised: October 1, 2025

Accepted: October 9, 2025

Published online: November 7, 2025

1. Introduction

Alzheimer's disease (AD) is the most prevalent cause of dementia in older adults, affecting roughly 50 million people worldwide, a number expected to triple by 2050, resulting in a major healthcare burden.[1] It is characterized by progressive memory loss, cognitive decline, and behavioral changes that severely impact patients' daily lives.[2] Despite advancements in neuroimaging and biomarker research, AD is often diagnosed at a late stage when substantial and irreversible brain damage has occurred. Early identification of AD, at the mild cognitive impairment stage, is critical for enabling timely interventions that may slow disease progression.

In recent years, artificial intelligence (AI) and machine learning (ML) approaches have shown great impact in detecting AD earlier and more accurately than conventional methods. For example, predictive models incorporating demographic and cognitive test data (without relying on imaging) have achieved encouraging results in identifying individuals at risk of AD.[3] Such approaches illustrate a shift toward more accessible and cost-effective diagnostic processes, potentially enabling earlier detection and intervention. Meanwhile, deep learning models can leverage complex patterns in neuroimaging data to improve diagnostic accuracy.[4] However, many recent techniques, such as transformer-based architectures and multimodal data fusion combining MRI with other biomarkers, require large, well-annotated datasets and often function as "black boxes" with limited interpretability.[5] This lack of transparency can hinder clinical adoption, where understanding the reasoning behind a prediction is important for trust and decision-making.

Compared with the magnetic resonance imaging (MRI) approach, convolutional neural networks (CNNs) offer a more straightforward and computationally efficient approach to AD diagnosis, especially when paired with methods to interpret their decisions.[6] To date, few studies have directly compared custom CNN architectures, advanced attention-based CNN variants, and transfer learning models on a common AD classification task. In this study, we address this gap by systematically evaluating three CNN-based models for multi-class AD stage classification using structural MRI. We assess a simple custom CNN, an attention-enhanced CNN (spatial-channel convolutional attention network [SCCAN]), and a transfer-learning model (VGG16) on the same dataset. We also integrate gradient-weighted class activation mapping (Grad-CAM) to provide visual explanations of model predictions. This work aims to determine whether relatively compact, interpretable CNN models can achieve performance comparable to more complex approaches, and to demonstrate an AI-based framework for early AD detection that is both accurate and explainable. Few works have offered a direct comparative evaluation of multiple CNN architectures (*e.g.*, custom CNN, attention-enhanced CNN, and transfer learning models) on the same multiclass MRI dataset using interpretability tools like Grad-CAM. However, many existing studies focus only on binary classification (*e.g.*, AD vs. healthy), rely on multi-modal inputs that may not be widely accessible, or lack model explainability—all of which are shortcomings associated with reduced clinical trust that the current study attempted to address.

2. Research objective

The primary objective of this research is to design and evaluate three CNN-based architectures for multiclass classification of AD stages using MRI images: a custom CNN, an attention-enhanced SCCAN, and a transfer learning model based on VGG16. The study seeks to compare these models in terms of classification performance, interpretability using Grad-CAM, and deployment feasibility, thereby contributing a practical, scalable, and explainable solution to support early AD diagnosis.

3. Materials and methods

3.1. Dataset and preprocessing

We employed the AD MRI dataset from Kaggle, comprising approximately 44,000 T1-weighted brain MRI images across four diagnostic categories: mild, moderate, severe, and very severe dementia. All images were skull-stripped and provided in JPEG format. The dataset represents an augmented, class-balanced version of an earlier collection, created to address class imbalance by increasing the number of images in underrepresented groups. The distribution included ~12,800 normal, 11,200 very mild, 10,000 mild, and 10,000 moderate AD scans, yielding balanced classes.

Before model training, all MRIs were resized to 128×128 pixels and intensity-normalized to a 0–1 scale. Preprocessing was tailored to each architecture. For the custom CNN and SCCAN models, images remained in grayscale and normalized. For VGG16, grayscale scans were replicated into three channels to form pseudo-RGB inputs, followed by ImageNet mean subtraction to match pretraining requirements.

To enhance generalization and reduce overfitting, training images underwent augmentation, including random flips, rotations, zooms, and slight shifts, generating diverse yet anatomically valid variations. Validation and test sets were not augmented. The dataset was split in a stratified manner, preserving class proportions: 70%

training, 15% validation (for hyperparameter tuning and early stopping), and 15% independent testing for final evaluation. All data were already anonymized and used in accordance with the dataset terms and privacy regulations.

Figures 1 and 2 illustrate the dataset and methodological workflow used in this study. Figure 1 displays the sample brain MRI images from all four classes, showcasing the visual differences the model learns to distinguish. Figure 2 presents the overall workflow of the proposed CNN framework, outlining the sequential stages from data preprocessing, enhancement, and augmentation to model training, evaluation, and deployment through a Streamlit application.

3.2. Deep learning models

We developed and evaluated three CNN-based deep learning models for AD stage classification:

Figure 3 illustrates the CNN-based model development pipeline used in this study. The process began with pre-processed MRI data, followed by model building using three architectures, a custom Model, SCCAN, and VGG16. These models were then complied and trained, after which their performance was evaluated based on various metrics. Finally, the optimal model was selected for deployment based on its accuracy and overall performance.

3.2.1. Custom CNN architecture

The first model is a custom CNN designed from scratch as a baseline for four-class AD classification. The architecture consists of multiple convolutional layers (with ReLU activation) followed by max-pooling layers for progressive spatial downsampling. This feature extractor is followed by fully connected layers that aggregate the learned features for final classification. The network was kept relatively lightweight (fewer layers and parameters) to serve as a baseline. Batch normalization and dropout were employed to stabilize training and prevent overfitting. The final output layer uses softmax activation to produce class probabilities for the four categories.

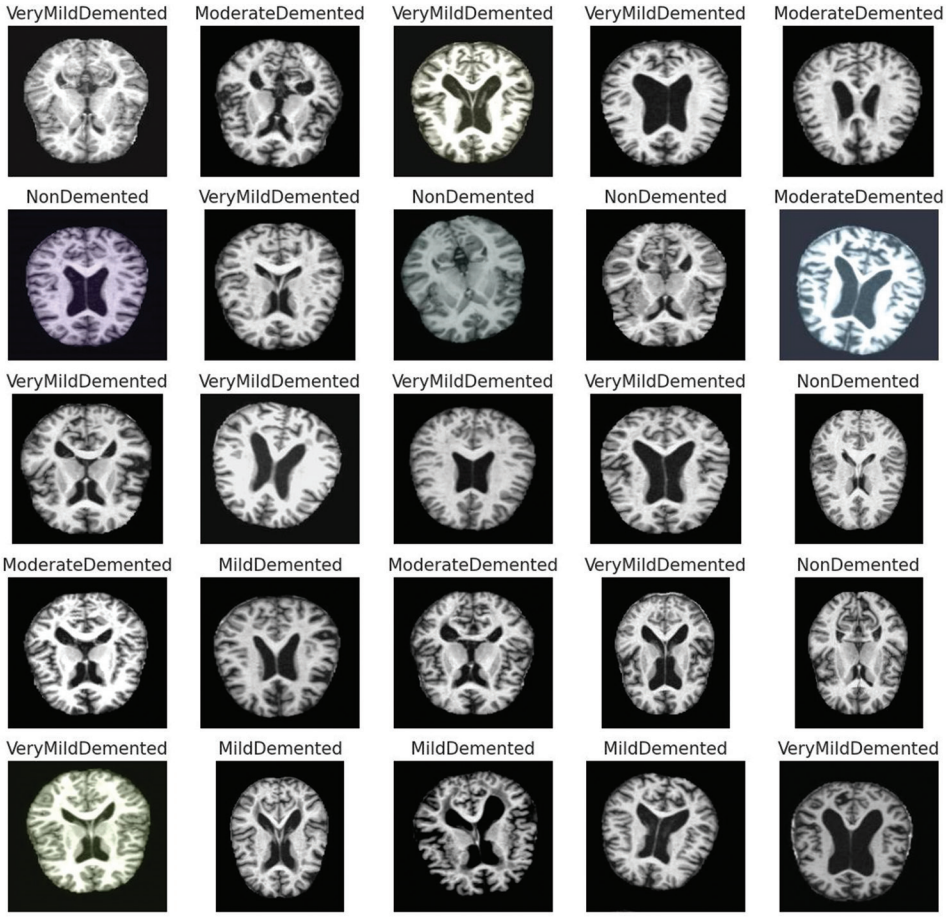

Figure 1. Representative brain MRI images of all classes
Abbreviation: MRI: Magnetic resonance imaging.

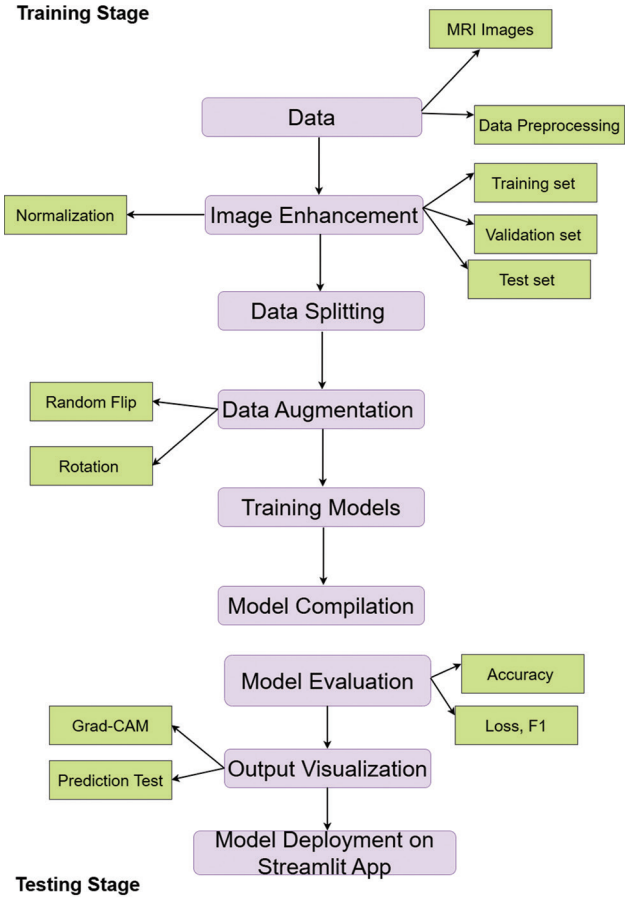

Figure 2. Workflow of the CNN model framework
Abbreviations: CNN: Convolutional neural network; Grad-CAM: Gradient-weighted class activation mapping; MRI: Magnetic resonance imaging.

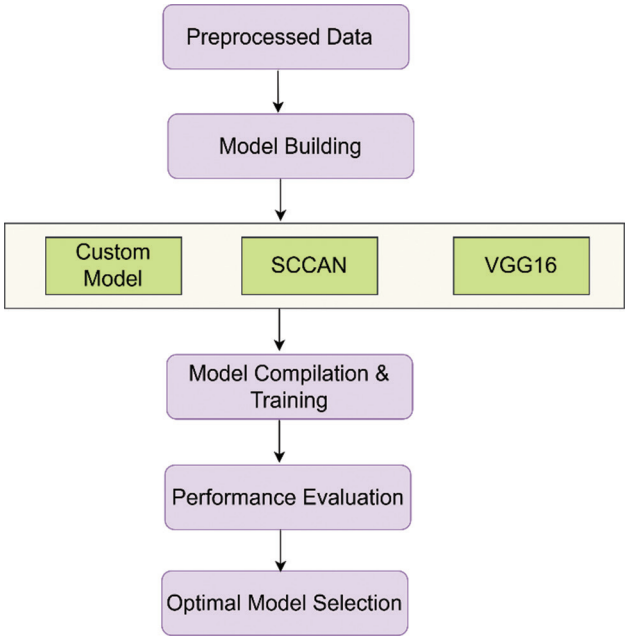

Figure 3. CNN-Based model development pipeline
Abbreviations: CNN: Convolutional neural network; SCCAN: Spatial-channel convolutional attention network.

Figure 4 presents the architecture of the custom CNN for stage classification of AD. The network consists of four convolutional blocks, each with convolution, batch normalization, ReLU activation, and max pooling, enabling hierarchical feature extraction while reducing dimensionality. Filter counts increase progressively (32, 64, 128, 256). A global average pooling layer condenses features into a 256-dimensional vector, followed by a dense layer with dropout for regularization. The final softmax layer outputs probabilities for four AD stages. The model has 457,156 parameters, with 456,196 trainable.

3.2.2. Attention-based CNN (SCCAN)

The second model, termed SCCAN, builds on a CNN architecture by incorporating an attention mechanism to improve feature learning. In this model, convolutional feature maps pass through a channel-attention module inspired by squeeze-and-excitation networks, which adaptively recalibrates feature map importance.[7] This allows the network to emphasize the most informative features (*e.g.*, regions with characteristic AD pathology) while suppressing less relevant information. The SCCAN architecture retains a similar overall structure to the custom CNN but with attention blocks inserted after certain convolutional layers. By explicitly modeling feature importance, SCCAN aims to boost classification performance for subtle early-stage patterns without a substantial increase in model complexity. The final output layer produces class probabilities using softmax.

Figure 5 presents the architecture of the SCCAN model, a lightweight CNN enhanced with channel attention mechanisms for Alzheimer's stage classification. The design adopts a multi-branch structure in which each convolutional block integrates a channel attention module. These modules apply both global average pooling and global max pooling across channels, followed by reshaping and dense layers. The resulting outputs are merged via addition and activation functions to generate attention weights. These weights are then multiplied element-wise with the original feature maps, selectively emphasizing the most informative channels and guiding the model's focus toward discriminative brain regions.

The architecture consists of three convolutional blocks, with filters increasing from 32 to 128. Each block is followed by channel attention, max pooling, and batch normalization. A final global average pooling layer condenses spatial features, which are passed to a fully

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 32)	896
batch_normalization (BatchNormalization)	(None, 224, 224, 32)	128
re_lu (ReLU)	(None, 224, 224, 32)	0
max_pooling2d (MaxPooling2D)	(None, 112, 112, 32)	0
conv2d_1 (Conv2D)	(None, 112, 112, 64)	18,496
batch_normalization_1 (BatchNormalization)	(None, 112, 112, 64)	256
re_lu_1 (ReLU)	(None, 112, 112, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 64)	0
conv2d_2 (Conv2D)	(None, 56, 56, 128)	73,856
batch_normalization_2 (BatchNormalization)	(None, 56, 56, 128)	512
re_lu_2 (ReLU)	(None, 56, 56, 128)	0
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 128)	0
conv2d_3 (Conv2D)	(None, 28, 28, 256)	295,168
batch_normalization_3 (BatchNormalization)	(None, 28, 28, 256)	1,024
re_lu_3 (ReLU)	(None, 28, 28, 256)	0
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 256)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 256)	0
dense (Dense)	(None, 256)	65,792
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 4)	1,028

Total params: 457,156 (1.74 MB)
Trainable params: 456,196 (1.74 MB)
Non-trainable params: 960 (3.75 KB)

Figure 4. Sequential custom CNN model layout
Abbreviation: CNN: Convolutional neural network.

connected classifier. Dropout is applied for regularization before the softmax output layer. SCCAN contains 133,824 parameters (133,376 trainable), offering a compact yet expressive design that improves early-stage AD detection.

3.2.3. VGG16 transfer learning model

The third model leverages transfer learning using the VGG16 architecture. VGG16 is a 16-layer deep CNN originally trained on the ImageNet dataset and is known to learn rich, generic visual features.[8] We used the pre-trained VGG16 as a fixed feature extractor: all 13 convolutional layers (organized into five blocks) were initialized with ImageNet weights and frozen during our training. On top of this convolutional base, we added a custom classifier head consisting of a global average pooling layer (to reduce each feature map to a single value), followed by dense layers and a softmax output for the four AD classes. Using transfer learning, this model benefits from low-level and mid-level image features learned from a large natural image corpus, which can improve learning efficiency in our medical imaging task. Only the weights of the added top layers were trained on our MRI data, while the convolutional base remained frozen. This approach typically yields faster convergence and high accuracy even with a relatively limited medical dataset.

Figure 6 illustrates the VGG16-based transfer learning model employed for multiclass AD classification. The base

Layer (type)	Output Shape	Param #	Connected to
input_layer_2 (InputLayer)	(None, 224, 224, 3)	0	-
conv2d_7 (Conv2D)	(None, 224, 224, 32)	896	input_layer_2[0][0]
batch_normalization_7 (BatchNormalization)	(None, 224, 224, 32)	128	conv2d_7[0][0]
max_pooling2d_7 (MaxPooling2D)	(None, 112, 112, 32)	0	batch_normalization_7…
global_average_pooling2d… (GlobalAveragePooling2D)	(None, 32)	0	max_pooling2d_7[0][0]
global_max_pooling2d_3 (GlobalMaxPooling2D)	(None, 32)	0	max_pooling2d_7[0][0]
reshape_6 (Reshape)	(None, 1, 1, 32)	0	global_average_poolin…
reshape_7 (Reshape)	(None, 1, 1, 32)	0	global_max_pooling2d_…
dense_10 (Dense)	(None, 1, 1, 4)	132	reshape_6[0][0], reshape_7[0][0]
dense_11 (Dense)	(None, 1, 1, 32)	160	dense_10[0][0], dense_10[1][0]
add_3 (Add)	(None, 1, 1, 32)	0	dense_11[0][0], dense_11[1][0]
activation_3 (Activation)	(None, 1, 1, 32)	0	add_3[0][0]
multiply_3 (Multiply)	(None, 112, 112, 32)	0	max_pooling2d_7[0][0], activation_3[0][0]
conv2d_8 (Conv2D)	(None, 112, 112, 64)	18,496	multiply_3[0][0]
batch_normalization_8 (BatchNormalization)	(None, 112, 112, 64)	256	conv2d_8[0][0]
max_pooling2d_8 (MaxPooling2D)	(None, 56, 56, 64)	0	batch_normalization_8…
global_average_pooling2d… (GlobalAveragePooling2D)	(None, 64)	0	max_pooling2d_8[0][0]
global_max_pooling2d_4 (GlobalMaxPooling2D)	(None, 64)	0	max_pooling2d_8[0][0]
reshape_8 (Reshape)	(None, 1, 1, 64)	0	global_average_poolin…
reshape_9 (Reshape)	(None, 1, 1, 64)	0	global_max_pooling2d_…
dense_12 (Dense)	(None, 1, 1, 8)	520	reshape_8[0][0], reshape_9[0][0]
dense_13 (Dense)	(None, 1, 1, 64)	576	dense_12[0][0], dense_12[1][0]
add_4 (Add)	(None, 1, 1, 64)	0	dense_13[0][0], dense_13[1][0]
activation_4 (Activation)	(None, 1, 1, 64)	0	add_4[0][0]
multiply_4 (Multiply)	(None, 56, 56, 64)	0	max_pooling2d_8[0][0], activation_4[0][0]
conv2d_9 (Conv2D)	(None, 56, 56, 128)	73,856	multiply_4[0][0]
batch_normalization_9 (BatchNormalization)	(None, 56, 56, 128)	512	conv2d_9[0][0]
max_pooling2d_9 (MaxPooling2D)	(None, 28, 28, 128)	0	batch_normalization_9…
global_average_pooling2d… (GlobalAveragePooling2D)	(None, 128)	0	max_pooling2d_9[0][0]
global_max_pooling2d_5 (GlobalMaxPooling2D)	(None, 128)	0	max_pooling2d_9[0][0]
reshape_10 (Reshape)	(None, 1, 1, 128)	0	global_average_poolin…
reshape_11 (Reshape)	(None, 1, 1, 128)	0	global_max_pooling2d_…
dense_14 (Dense)	(None, 1, 1, 16)	2,064	reshape_10[0][0], reshape_11[0][0]
dense_15 (Dense)	(None, 1, 1, 128)	2,176	dense_14[0][0], dense_14[1][0]
add_5 (Add)	(None, 1, 1, 128)	0	dense_15[0][0], dense_15[1][0]
activation_5 (Activation)	(None, 1, 1, 128)	0	add_5[0][0]
multiply_5 (Multiply)	(None, 28, 28, 128)	0	max_pooling2d_9[0][0], activation_5[0][0]
global_average_pooling2d… (GlobalAveragePooling2D)	(None, 128)	0	multiply_5[0][0]
dense_16 (Dense)	(None, 256)	33,024	global_average_poolin…
dropout_2 (Dropout)	(None, 256)	0	dense_16[0][0]
dense_17 (Dense)	(None, 4)	1,028	dropout_2[0][0]

Figure 5. Sequential SCCAN CNN model layout
Abbreviations: CNN: Convolutional neural network; SCCAN: Spatial-channel convolutional attention network.

network is the pre-trained VGG16, frozen to preserve low- and mid-level features learned from ImageNet,

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 512)	14,714,688
flatten_3 (Flatten)	(None, 512)	0
batch_normalization_19 (BatchNormalization)	(None, 512)	2,048
dense_27 (Dense)	(None, 2048)	1,050,624
batch_normalization_20 (BatchNormalization)	(None, 2048)	8,192
dense_28 (Dense)	(None, 1024)	2,098,176
batch_normalization_21 (BatchNormalization)	(None, 1024)	4,096
dense_29 (Dense)	(None, 4)	4,100

Total params: 17,881,924 (68.21 MB)
Trainable params: 3,160,068 (12.05 MB)
Non-trainable params: 14,721,856 (56.16 MB)

Figure 6. Sequential VGG16 model layout

producing a 512-dimensional feature vector. A custom classification head was added and trained on the AD's MRI dataset.

The classification head begins with flattening, followed by dense layers of 2048 and 1024 neurons, each with batch normalization and ReLU activation to enhance stability and reduce covariate shift. A final Dense (4) layer with softmax activation outputs probabilities across the four stages: mild or no dementia, moderate dementia, severe dementia, and very severe dementia.

The model comprises a total of 17.88 million parameters, with only 3.16 million trainable in the classification head and 14.72 million frozen in the VGG16 backbone. This architecture effectively combines robust pretrained feature extraction with efficient training, achieving strong performance on AD stage classification with limited data.

3.3. Model training and validation

All models were implemented in Python using Keras/TensorFlow. We used the Adam optimizer[9] with an initial learning rate of 0.0001 and categorical cross-entropy as the loss function (appropriate for multi-class classification). Models were trained in mini-batches of size 32. We monitored performance on the validation set at the end of each epoch to guide hyperparameter tuning and apply early stopping. Specifically, training was halted if the validation loss did not improve for a patience of five epochs, and the model state with the lowest validation loss was retained (model checkpointing). The custom CNN and SCCAN models were trained for up to 50 epochs, and the VGG16 model for up to 30 epochs, with early stopping typically occurring earlier based on validation metrics.

Model performance was finally evaluated on the independent test set (15% of the data) that was held out from all training and validation. All metrics are reported on this test set. We report overall accuracy as well as per-class

precision, recall (sensitivity), and F1-score. In addition, we computed confusion matrices for each model's predictions to examine common misclassifications (*e.g.*, whether very mild cases are often mistaken for non-dementia, *etc.*).

3.4. Evaluation metrics

We evaluated model performance using several standard metrics. Accuracy was calculated as the proportion of correctly classified images out of all test images. To provide a more nuanced assessment, we also computed precision, recall (sensitivity), and F1-score for each class. Recall (sensitivity) is the fraction of actual positives (*e.g.*, mild dementia cases) that the model correctly identified, while precision is the fraction of cases predicted as a given class that truly belong to that class. The F1-score is the harmonic mean of precision and recall. We report these metrics for each class and as weighted averages across classes. We also examined confusion matrices for each model, which summarize prediction outcomes for each class. The confusion matrix allows us to see which classes are most often confused by the model (for instance, whether very mild AD images are frequently misclassified as non-dementia or mild dementia). These metrics and analyses provide insight into both overall performance and specific strengths or weaknesses of each model (such as sensitivity to early-stage AD).[10-13]

3.5. Model interpretability with Grad-CAM

To improve interpretability, we employed Grad-CAM to visualize the regions of the MRI that each model considered important for its predictions. For a given test image, Grad-CAM uses the gradients of the target class score flowing into the last convolutional layer to produce a heatmap of "important" pixels.[14] We generated Grad-CAM heatmaps for representative correctly and incorrectly classified examples from each model. These heatmaps were overlaid on the original MRI slices to highlight anatomically relevant regions influencing the model's decision. For instance, in images predicted as very mild dementia, the advanced models' Grad-CAMs often highlighted the medial temporal lobe (including the hippocampus), which aligns with known early pathological changes in AD. In contrast, the custom CNN's attention was more diffuse and sometimes less focused on these regions, which likely contributed to its lower performance. The use of Grad-CAM thus provides a qualitative check on model behavior, ensuring that the CNNs are "looking" at brain regions that make sense clinically. This interpretability is crucial for building trust in the model's predictions and could help clinicians understand and verify AI-driven diagnoses.

4. Results

The performance of the three models on the test set is summarized in Table 1. Overall, the attention-augmented CNN (SCCAN) and the transfer learning model (VGG16) substantially outperformed the baseline custom CNN in all metrics, particularly in sensitivity for early-stage AD. Both advanced models achieved high overall accuracy (~95–96%), whereas the custom CNN achieved about 73% accuracy. In Table 1, we detail the results for each model and examine notable patterns in the confusion matrices.

4.1. Custom CNN performance

The custom CNN baseline attained an overall test accuracy of approximately 72.7%. This model's performance varied considerably by class. It performed well on the more advanced dementia classes but struggled to detect the subtle features of very early-stage AD. For moderate dementia cases, the custom CNN achieved high recall (over 90% of moderate cases were correctly identified). Similarly, for mild dementia cases, it correctly classified about 92% of them. However, the model was less reliable for the non-dementia class, correctly identifying roughly 75% of healthy images while misclassifying the remaining 25% (mostly as very mild dementia). The greatest challenge was observed in the very mild dementia category: only around 32% of very mild dementia cases were correctly detected by the custom CNN. The majority of misclassified very mild cases were predicted to be mild dementia, with a smaller fraction mistaken for non-dementia. This indicates that the subtle brain changes of the earliest AD stage were often missed by the baseline model. In terms of precision, a similar pattern was noted. Precision was high for the more pronounced classes (mild and moderate) but quite low for the very mild class, indicating a high false-positive rate for that category. In summary, the custom model served as a reasonable baseline, performing adequately on clear-cut cases (healthy vs. advanced dementia) but with limited sensitivity to the earliest signs of AD.

Figure 7 shows the model accuracy and loss trend. During training, the custom CNN model showed a steady

Table 1. Comparative evaluation metrics for custom CNN, SCCAN, and VGG16 models

Metric	Custom CNN	SCCAN	VGG16
Accuracy	72.70%	95.72%	95.72%
Test loss	0.85	0.14	0.14
F1 (Avg Weighted)	0.71	0.96	0.96

Abbreviations: Avg Weighted: Average Weighted Score; CNN: Convolutional neural network; F1: F1 Score; SCCAN: Spatial-channel convolutional attention network; VGG: Visual geometry group.

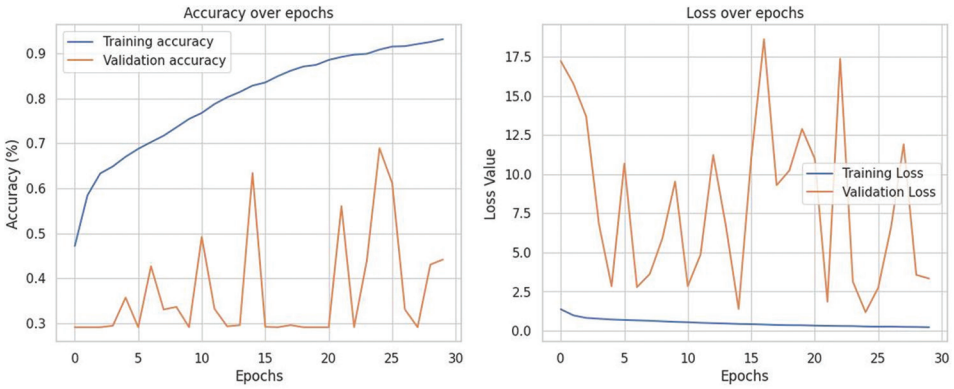

Figure 7. Training and validation accuracy and loss curves for the custom CNN model
Abbreviation: CNN: Convolutional neural network.

increase in training accuracy over 30 epochs, reaching above 90%. However, the validation accuracy fluctuated significantly and remained comparatively low, failing to improve consistently beyond 70%. Similarly, the training loss decreased smoothly, indicating successful learning on the training data, while the validation loss exhibited high volatility with large spikes throughout the epochs. These patterns suggest that the custom CNN model overfits the training data, learning its patterns well but failing to generalize effectively to unseen data.

The confusion matrix illustrates the performance of the custom CNN model in distinguishing between the four dementia classes. As shown in Figure 8, the model demonstrates strong classification accuracy for the moderate AD and mild AD categories, with slightly lower performance in differentiating very mild AD and non-AD cases, likely due to the subtle similarities in their MRI features.

The diagonal values in Table 2 indicate correctly classified instances for each category. The custom CNN model performed well in distinguishing mild AD, moderate AD, and non-AD cases, while higher misclassification occurred in the very mild AD class, which was frequently confused with mild AD and non-AD due to their subtle feature similarities. The highest confusion occurred for very mild AD, which was most often misclassified as mild AD (43%). This indicates the model struggled to differentiate early-stage AD (very mild) from nearby stages, consistent with its lower generalization performance observed in the validation metrics.

Figure 8. Custom CNN confusion matrix
Abbreviation: CNN: Convolutional neural network.

Table 2. Custom CNN confusion matrix (normalized)

Actual/ Predicted	Mild AD	Moderate AD	Non-AD	Very mild AD
Mild AD	**0.9247**	0.0000	0.0633	0.0120
Moderate AD	0.0087	**0.9593**	0.0313	0.0007
Non-AD	0.1703	0.0000	**0.7490**	0.0807
Very mild AD	0.4304	0.0000	0.2518	**0.3179**

Note: The values presented in boldface indicate correctly classified instances for each category.
Abbreviations: AD: Alzheimer's disease; CNN: Convolutional neural network.

4.2. Attention-based CNN (SCCAN) performance

The SCCAN model achieved a test accuracy of 95.7%, a dramatic improvement over the baseline. Incorporating attention mechanisms greatly enhanced the model's ability to distinguish between the four classes, especially for the challenging early stage. The SCCAN exhibited excellent recall across all categories: notably, it correctly identified 100% of mild AD and moderate AD cases in the test set (no misclassifications in those classes). For non-AD images, the recall was about 90%, indicating that only around

10% of healthy brains were falsely labeled as dementia. Crucially, the SCCAN detected approximately 90% of very mild AD cases, significantly higher than the 32% recall of the baseline CNN. Only a small number of very mild cases were missed by SCCAN, some of which were classified as non-AD (reflecting the inherent difficulty of differentiating very early symptoms from normal aging). Precision was also high for all classes (above 0.90 in each category), meaning the model had low false-positive rates. The confusion matrix of SCCAN's predictions (not shown) was nearly diagonal, with minimal confusion between different classes. Overall, the attention-enhanced CNN not only boosted overall accuracy but specifically addressed the weaknesses of the baseline model by focusing on subtle imaging features indicative of early AD.

Figure 9 shows the SCCAN model accuracy and loss trend. The SCCAN model exhibited strong and stable training behavior. Training accuracy increased steadily, reaching over 97% by epoch 10, while training loss consistently declined. In contrast, validation accuracy remained stable around 91%, and validation loss showed minimal improvement, with a slight upward trend. This suggests the model learned effectively on the training data but began to plateau on the validation set, indicating potential early signs of overfitting or limited generalization improvement beyond a certain point.

The diagonal values indicate correctly classified instances for each category. The SCCAN model demonstrated excellent overall performance, achieving perfect classification for mild AD and moderate AD cases. Only minimal misclassification was observed between non-AD and very mild AD, reflecting the close similarity between these adjacent clinical stages.

This high precision and class separation demonstrate SCCAN's robustness, especially in distinguishing early AD stages with clinically meaningful accuracy. Figure 10 shows that the SCCAN model achieved near-perfect classification accuracy across all dementia categories.

4.3. VGG16 transfer learning performance

The VGG16-based model attained the highest accuracy on the test set, at approximately 96.0%. Its performance was essentially on par with the SCCAN model and far above the baseline. In terms of class-wise results, the VGG16 model showed a performance profile very similar to SCCAN. It correctly classified roughly 95–96% of mild and moderate AD cases. For non-AD (healthy) individuals, the model's recall was around 92%, meaning it misclassified only about 8% as AD. Importantly, the VGG16 model correctly identified about 90% of very mild AD cases, indicating that transfer learning from large-scale data can effectively capture early AD patterns in MRI. The few very mild cases that VGG16 missed were typically confused with the mild AD class, which is understandable given the continuum of disease severity. The precision for VGG16 was comparably high across all classes (generally >0.90). These metrics highlight that leveraging a pre-trained CNN (VGG16) yields excellent performance on our AD classification task, comparable to the specialized attention-CNN. The fact that two very different approaches (one introducing attention mechanisms, the other reusing learned features from natural images) achieved similar success is noteworthy. It suggests that both strategies effectively capture the critical MRI features distinguishing each stage of AD.

The VGG16 model demonstrated strong learning dynamics over 15 epochs. Training accuracy improved consistently, reaching approximately 97%, while validation accuracy closely followed, stabilizing around 95% which can be seen in Figure 11. Training and validation loss both declined steadily, with validation loss exhibiting slight fluctuations but maintaining a low final value. These trends

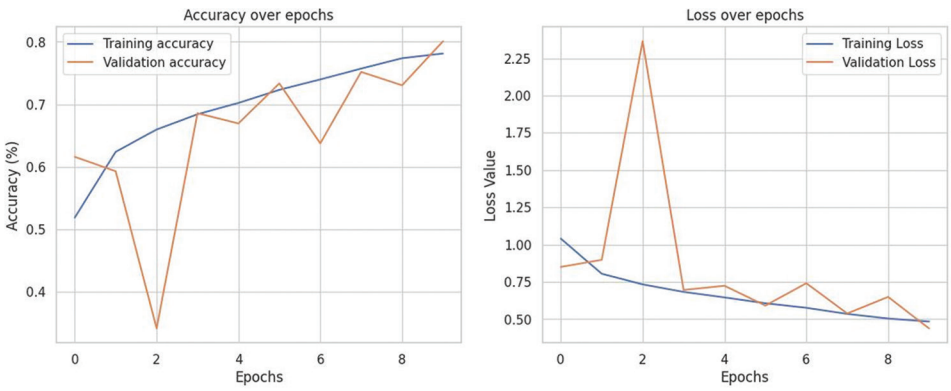

Figure 9. Training and validation accuracy and loss curves for the SCCAN model
Abbreviation: SCCAN: Spatial-channel convolutional attention network.

indicate excellent generalization and minimal overfitting, suggesting the model effectively leveraged transfer learning for robust AD stage classification. Figure 12 shows the VGG16 model confusion matrix.

The diagonal values indicate correctly classified instances for each class. The VGG16 model demonstrated consistently high accuracy across all categories, with most misclassifications occurring between very mild AD and non-AD due to their clinical similarity. Misclassification rates for moderate AD were negligible, highlighting the model's strong confidence and precision.

This confusion matrix in Figure 12 underscores VGG16's robust generalization and fine-grained discrimination, especially for early and intermediate AD stages. The comparative performance of the three CNN-based models is summarized in Tables 2-4. As shown in Table 2, the custom CNN model achieved strong accuracy for moderate AD and mild AD, with minor misclassifications between very mild AD and non-AD. The SCCAN model (Table 3) demonstrated improved performance, attaining perfect classification for mild and moderate AD and 90% accuracy for both non-AD and very mild AD, with minimal overlap between adjacent classes. Meanwhile, the VGG16 model (Table 4) achieved consistently high accuracy across all categories, exceeding 94% in each, and showed near-perfect precision for moderate AD, confirming its robustness and reliability in AD stage classification.

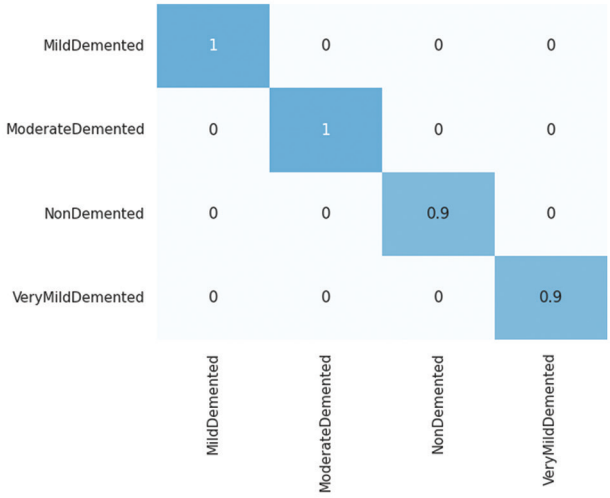

Figure 10. SCCAN Confusion matrix
Abbreviation: SCCAN: Spatial-channel convolutional attention network.

4.4. Model interpretability results

To gain insights into model behavior, we examined Grad-CAM visualizations for the CNN models. Figure 13 shows an example of Grad-CAM output from the VGG16 model for a very mild AD case that was correctly classified. The heatmap (overlaid in red) highlights regions in the temporal lobe, particularly the hippocampal area, indicating that the model concentrated on these regions to make its prediction. This corresponds well with clinical knowledge, as the hippocampus is one of the first regions affected in AD. In general, Grad-CAM results for the VGG16 and SCCAN models revealed that they focus on plausible anatomical regions (hippocampus, entorhinal cortex, ventricles) when identifying AD, even at early stages. The custom CNN's Grad-CAMs were more diffuse and less concentrated in those specific areas, possibly explaining its weaker performance. These interpretability findings build trust in the advanced models: They suggest that the models are not relying on spurious image artifacts but are in fact detecting AD-related structural changes. For clinicians, such visual explanations are valuable; for example, an AI prediction of "very mild AD" comes with

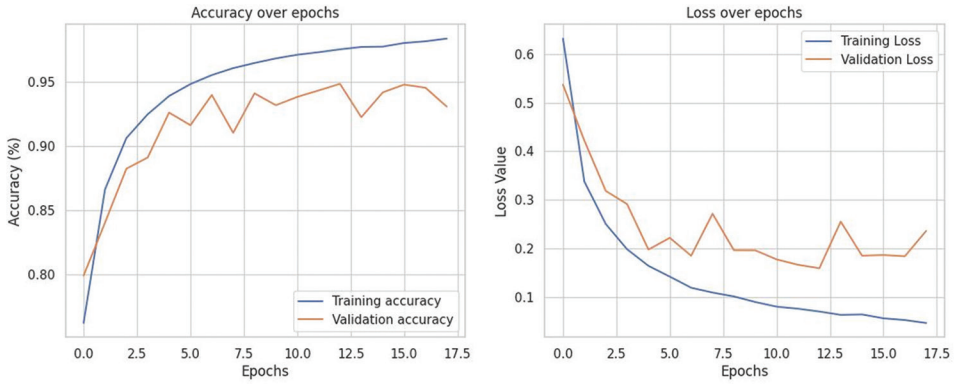

Figure 11. Training and validation accuracy and loss curves for the VGG16 model
Abbreviation: VGG: Visual Geometry Group.

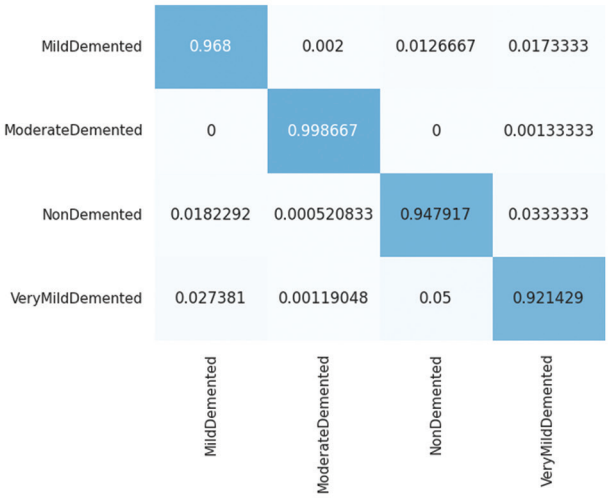

Figure 12. VGG16 Model confusion matrix

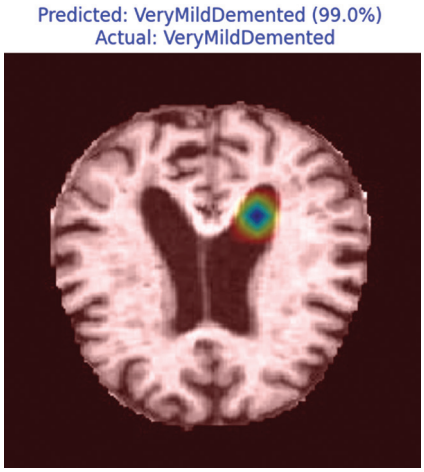

Figure 13. An MRI scan of a very mild AD case with Grad-CAM heatmap overlay highlighting hippocampal regions
Abbreviations: AD: Alzheimer's disease; Grad-CAM: Gradient-weighted class activation mapping; MRI: Magnetic resonance imaging.

a Grad-CAM heatmap highlighting hippocampal atrophy, corroborating the AI's decision and thereby prompting further examination of the scan.

4.5. Model prediction

Figure 13 showcases a grid of MRI test images with predicted classes and corresponding confidence scores, generated using the VGG16 model. Each tile represents a model prediction for a previously unseen MRI image.

The figure illustrates the VGG16 model's high certainty and reliability across most test images, often producing prediction confidence above 95%. Misclassifications are rare and typically occur between adjacent stages, such as very mild AD and non-AD, which is clinically understandable due to overlapping anatomical features. This grid visualization provides intuitive insight into the model's decision-making and confidence distribution across different AD stages.

Table 3. SCCAN Model confusion matrix (normalized)

Actual/ Predicted	Mild AD	Moderate AD	Non-AD	Very mild AD
Mild AD	**1.000**	0.000	0.000	0.000
Moderate AD	0.000	**1.000**	0.000	0.000
Non-AD	0.000	0.000	**0.900**	0.100 (inferred)
Very mild AD	0.000	0.000	0.100 (inferred)	**0.900**

Notes: The SCCAN model achieved perfect classification for both mild and moderate demented classes (100% accuracy). It correctly classified 90% of both non-dementia and very mild dementia cases. There was minimal misclassification, with the only minor confusion occurring between non-demented and very mild demented, two adjacent clinical stages. The values presented in boldface indicate correctly classified instances for each category.
Abbreviations: AD: Alzheimer's disease; SCCAN: Spatial-channel convolutional attention network.

Table 4. VGC16 Model confusion matrix (normalized)

Actual/ Predicted	Mild AD	Moderate AD	Non-AD	Very mild AD
Mild AD	**0.968**	0.002	0.013	0.017
Moderate AD	0.000	**0.999**	0.000	0.001
Non-AD	0.018	0.0005	**0.948**	0.033
Very mild AD	0.027	0.0012	0.050	**0.921**

Notes: The VGG16 model achieved very high classification accuracy across all four classes, with 96.8% for mild AD, 99.9% for moderate AD, 94.8% for non-AD, and 92.1% for very mild AD. The values presented in boldface indicate correctly classified instances for each category.
Abbreviations: AD: Alzheimer's disease; VGG: Visual geometry group.

4.6. Model deployment

To demonstrate real-world applicability, the best-performing model (VGG16) was deployed as an interactive diagnostic tool using Streamlit, a lightweight web framework for ML interfaces.[15] The deployment pipeline allows users such as clinicians or researchers to upload a brain MRI image and receive an instant stage prediction for AD.

The system processes the input through the same preprocessing pipeline used during training and passes it into the fine-tuned VGG16 model. Upon inference, the application returns:

- The predicted AD stage (*e.g.*, moderate AD)
- The model's confidence level (*e.g.*, 100.00% certainty)

- Class-wise confidence scores across all four diagnostic categories.

This is illustrated in Figure 14, which shows an example prediction interface where the uploaded image was classified as "moderate AD" with full confidence. The interface also offers the ability to expand into a Grad-CAM visualization, allowing clinicians to interpret the anatomical regions regarded as the most influential in the prediction. By integrating interpretability and transparency into the interface, the deployment bridges the gap between AI model performance and real-world usability, making it a viable tool for clinical decision support and early AD screening. Figure 15 displays the deployed model's prediction interface, showing the predicted class with corresponding confidence scores for each dementia category.

5. Discussion

This study demonstrated the feasibility and advantages of deep learning for early prediction of AD using MRI data. We showed that both an attention-augmented CNN and a transfer learning approach (VGG16) can achieve high accuracy (~95%) in classifying stages of AD, substantially outperforming a baseline CNN. Notably, these models significantly improved the detection of very mild AD cases, a key result, as early diagnosis is critical for interventions. The custom CNN struggled with this category (detecting only ~32% of very mild cases), whereas both advanced models detected around 90% of them. This finding suggests that more sophisticated CNN approaches can extract subtle features of early AD that a simpler network might miss.

Our results are consistent with recent research applying deep learning to AD classification. For example, other studies using CNNs on structural MRI have reported accuracies in the 85–95% range for distinguishing AD from healthy controls or mild cognitive impairment.[16,17] The ~96% accuracy achieved by our VGG16 and SCCAN models is at the upper end of this range, underscoring the benefit of incorporating either transfer learning or attention mechanisms. Earlier studies applied classical ML methods (*e.g.*, support vector machine [SVM], random forest [RF]) to handcrafted features such as hippocampal atrophy.[3,16] While informative, these approaches required manual feature engineering. A unique contribution of this work is the direct comparison of these two strategies. Interestingly, the attention-based CNN (SCCAN) matched the performance of the much deeper VGG16 network.

Figure 14. Visual grid of VGG16 model predictions on MRI test images. For each tile, the predicted class label and softmax confidence score are displayed at the top, while the ground truth label appears at the bottom. Correct predictions are marked in standard color, while misclassified instances are highlighted in red.

Abbreviations: MRI: Magnetic resonance imaging; VGG: Visual Geometry Group.

Figure 15. Deployed model prediction interface

Another important aspect of our study is the use of Grad-CAM for model interpretability. In the realm of medical AI, explainability is paramount. The Grad-CAM analysis confirmed that our models "attend" to appropriate brain regions. For instance, when the model flagged a scan as mild or moderate AD, the heatmaps often highlighted widespread cortical atrophy and enlarged ventricles, aligning with advanced AD pathology. For very mild cases, the focus on medial temporal structures (like the hippocampus) mirrors what a radiologist would look for in early AD. Such alignment between model focus and clinical knowledge is reassuring and helps bridge the gap between AI and human experts. Similar approaches to explainable AI in neuroimaging have been proposed by others.[18,19] Over the past decade, ML techniques have been widely adopted for AD prediction. Early efforts employed classical algorithms like SVMs, RFs, and Decision Trees on features extracted from structural MRI or cognitive tests. For instance, one study used hippocampal atrophy patterns in MRI scans to separate AD patients from healthy controls using SVMs.[20] Another example applied ROI based SVM classifiers to identify individuals at risk long before clinical symptoms appeared[21] and our results reinforce the idea that CNNs can be made transparent enough to be used as decision support tools. Clinicians could use these explanations to validate AI suggestions or to discover imaging findings that might be overlooked. And our results reinforce the idea that CNNs can be made transparent enough to be used as decision support tools. Clinicians could use these explanations to validate AI suggestions or to discover imaging findings that might be overlooked.

5.1. Limitations

Despite the promising results, several limitations should be acknowledged. First, our dataset, although large, was sourced entirely from Kaggle.[22] While useful for research, this dataset is skull-stripped and preprocessed, which does not fully reflect the variability, artifacts, and challenges present in real-world clinical MRI scans acquired from different scanners, protocols, or populations.[23] This reliance limits generalizability and clinical applicability. Performance may differ when the models are tested on truly external, multicenter datasets such as ADNI data, and future work should validate and potentially fine-tune the models under such conditions.

Second, our study focused solely on MRI. In practice, AD diagnosis benefits from multimodal data, including positron emission tomography scans, cerebrospinal fluid-based biomarkers, and neurocognitive assessments.[24] Incorporating such modalities into deep learning frameworks could improve diagnostic robustness.

Third, while we employed Grad-CAM for interpretability, this provides only coarse visual explanations. More quantitative approaches, such as SHapley Additive exPlanations (SHAP) could complement these visualizations by offering feature-level importance values, thereby strengthening model interpretability.[25]

Finally, deployment considerations remain. For clinical translation, models must be integrated into workflows, tested in real time, and evaluated for usability by radiologists, including processing speed, reliability, and the clarity of AI-generated explanations.[26] Comparisons with classical ML approaches using hand-crafted features would also help confirm that CNNs provide added value.

5.2. Study implications and future work

The encouraging performance of the deep learning models suggests that AI could be used as a screening or decision support tool for early AD. For instance, in a memory clinic, an AI system could automatically analyze an incoming patient's MRI and flag the likelihood of very mild AD changes, prompting further confirmatory tests or closer monitoring. Because our best models achieved high sensitivity for early-stage AD, such a system could facilitate immediate patient identification and subsequent treatment. Moreover, given the interpretable nature of the models, clinicians would not have to rely on a "black-box" prediction; they could instead examine the Grad-CAM heatmap or other explanations to understand *why* the model indicates early AD.

For future research, a few avenues stand out. Ensemble methods could be explored: combining the outputs of the SCCAN and VGG16 models (and perhaps other models) might yield even more robust performance by leveraging their complementary strengths.[27] Researchers are increasingly adopting hybrid and ensemble machine-learning models to improve early diagnosis of Alzheimer's

disease (AD). Hybrid models combine different types of neural networks or algorithms, such as CNNs with recurrent networks or autoencoders to capture complementary spatial, temporal, and latent features from neuroimaging and multi-modal data. Ensemble methods, which aggregate the outputs of multiple classifiers, further boost robustness and reduce overfitting. Recent reviews highlight the growing evidence that these combined approaches yield superior accuracy and generalizability compared with single-model strategies.[28,29] A recent review highlighted that integrating multiple machine-learning algorithms and heterogeneous data sources can enhance diagnostic accuracy, robustness, and interpretability across medical applications.[30] An ensemble of region-of-interest-based CNN classifiers was shown to improve AD staging from MRI data by leveraging complementary information from different brain regions, achieving higher accuracy than single-model approaches.[31] Another direction is longitudinal modeling using series of MRIs over time to predict progression, not just single-timepoint classification. Some recent works have used recurrent networks or transformers for longitudinal AD prediction,[21] and integrating that with our approach could predict not only the current stage but also future decline. In addition, as mentioned, incorporating additional data types (genetic information, cognitive tests, etc.) in a multi-modal network could mirror the multi-faceted approach clinician's use and potentially improve accuracy further.

Finally, we plan to conduct a reader study where radiologists use the AI system in a simulated workflow to assess the extent of improvement in diagnostic sensitivity and confidence for early AD detection and to gather feedback on the usefulness of the explanations provided. Future research could build upon recent advances in deep learning-based Alzheimer's diagnosis frameworks, including interpretability-aware modeling, hybrid CNN–ML approaches, and attention-enhanced architectures that have demonstrated promising performance in MRI-based classification tasks.[8,16-19,21,32,33] This suggests that integrating task-specific attention to a custom model can provide similar benefits to leveraging a large pre-trained model, at least for our dataset. In practical terms, the SCCAN model, being smaller, might be more efficient to deploy, while the VGG16 model, with its transfer-learned features, provides a proven architecture that might generalize well if fine-tuned further.[34,35]

6. Conclusion

We have demonstrated that deep learning models, particularly a transfer-learned CNN (VGG16) and an attention-augmented CNN (SCCAN), can accurately classify the stages of AD from MRI scans. Both approaches greatly outperformed a baseline CNN, especially in detecting very mild dementia cases, which are crucial for early intervention. Moreover, by utilizing Grad-CAM visualizations, we ensured that our models' predictions are accompanied by human-interpretable explanations, highlighting relevant neuroanatomical regions. This combination of high accuracy and explainability is essential for clinical adoption of AI systems.

The findings of this study contribute to the growing evidence that AI can assist in the early diagnosis of AD. Importantly, our work underscores that model interpretability should go hand in hand with performance in medical AI applications. By addressing both aspects, we move closer to developing AI tools that clinicians can trust and effectively incorporate into patient care. Early detection enabled by such tools can open the door to timely therapeutic interventions, better care planning, and improved patient outcomes in AD.

Acknowledgments

We acknowledge the use of generative AI tools in certain parts of this work, specifically for improving the clarity and flow of the paper. However, all aspects of the research design, original ideas, dataset selection, ML methodologies, implementation, analysis, and all core contributions are solely our own. We would also like to thank York St John University, particularly the Computer Science and Data Science Department, for their continued encouragement and the opportunities provided throughout the course of this research.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Swathi Ganesan
Formal analysis: Babatunde Hamzat
Methodology: Babatunde Hamzat
Supervision: Sangita Pokhrel, Swathi Ganesan
Writing–original draft: Babatunde Hamzat
Writing–review & editing: Sangita Pokhrel

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Dataset will be available on request to the corresponding author. This research utilized a public dataset available on Kaggle (https://www.kaggle.com/datasets/aryansinghal10/alzheimers-multiclass-dataset-equal-and-augmented).

References

1. World Health Organization. *Dementia*; 2022. Available from: https://www.who.int/news-room/fact-sheets/detail/dementia [Last accessed on 2025 Jul 06].
2. Alzheimer's Association. 2021 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2021;17(3):327-406.

   doi: 10.1002/alz.12328
3. Cochrane A, Matthews FE, Brayne C. Predictive models for Alzheimer's disease using longitudinal data. *J Alzheimers Dis*. 2020;75(2):561-572.

   doi: 10.3233/jad-191030
4. Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neurosci Biobehav Rev*. 2017;74:58-75.

   doi: 10.1016/j.neubiorev.2017.01.002
5. Zeineldin RA, Karar ME, Elshaer Z, *et al*. Explainable hybrid vision transformers and convolutional network for multimodal glioma segmentation in brain MRI. *Sci Rep*. 2024;14:3713.

   doi: 10.1038/s41598-024-54186-7
6. Dyrba M, Hanzig M, Altenstein S, *et al*. Improving 3D convolutional neural network comprehensibility via interactive visualization of relevance maps: Evaluation in Alzheimer's disease. *Alzheimers Res Ther*. 2021;13:191.

   doi: 10.1186/s13195-021-00924-2
7. Roy AG, Navab N, Wachinger C. *Concurrent Spatial and Channel Squeeze and Excitation in Fully Convolutional Networks*. arXiv; 2018. Available from: https://arxiv.org/abs/1803.02579 [Last accessed on 2025 Aug 01].

   doi: 10.1109/cvpr.2018.00328.
8. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *International Conference on Learning Representations (ICLR)*; 2015. Available from: https://arxiv.org/abs/1409.1556 [Last accessed on 2025 Aug 01].
9. Kingma DP, Ba J. *Adam: A Method for Stochastic Optimization*. [arXiv Preprint]; 2014. Available from: https://arxiv.org/abs/1412.6980 [Last accessed on 2025 Jul 28].
10. Zhang Y, Zhang Y, Zhang D, *et al*. Deep learning-based diagnosis algorithm for Alzheimer's disease. *Brain Sci*. 2020;10:333.

    doi: 10.3390/brainsci10120333
11. Golovanevsky M, Eickhoff C, Singh R. Multimodal attention-based deep learning for Alzheimer's disease diagnosis. *Brain Inform*. 2022;9(1):1-12.

    doi: 10.1186/s40708-022-00195-7
12. Nasir N, Ahmed M, Afreen N, Sameer M. *Alzheimer's Magnetic Resonance Imaging Classification Using Deep and Meta-Learning Models*. arXiv; 2024. Available from: https://arxiv.org/abs/2405.12126 [Last accessed on 2025 Aug 28].
13. Sarawgi U, Zulfikar W, Soliman N, Maes P. *Multimodal Inductive Transfer Learning for Detection of Alzheimer's Dementia and Its Severity*. arXiv; 2020. Available from: https://arxiv.org/abs/2009.00700 [Last accessed on 2025 Aug 01].
14. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2017. p. 618-626.

    doi: 10.1109/iccv.2017.74
15. Streamlit Inc. *Streamlit: The Fastest Way to Build Data Apps in Python*; 2025. Available from: https://streamlit.io [Last accessed on 2025 Oct 28].
16. Odusami M, Maskeliūnas R, Damaševičius R, *et al*. A deep learning solution for multi-class diagnosis of Alzheimer's disease using MRI. *Cogn Comput*. 2021;13:1261-1270.

    doi: 10.1007/s12559-020-09795-3
17. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci Rep*. 2021;11:3254.

    doi: 10.1038/s41598-020-74399-w
18. Shukla A, Ghosh S, Rana S. Explainable AI in healthcare: A framework for neuroimaging-based disease prediction. *IEEE Rev Biomed Eng*. 2023;16:200-212.

    doi: 10.1109/rbme.2022.3186020
19. Sibilano F, Bertozzi M, Valenti M. An interpretable CNN-based pipeline for Alzheimer's classification. *J Med Syst*. 2024;48:10.

    doi: 10.1007/s10916-024-02066-9
20. Gerardin E, Chételat G, Chupin M, *et al*. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage*. 2009;47(4):147686.

    doi: 10.1016/j.neuroimage.2009.05.036
21. Hoang ND, Luong NT, Do TN. Vision transformers for predicting Alzheimer's disease progression using longitudinal structural MRI. *Comput Biol Med*. 2023;154:106555.
22. Kaggle. *Alzheimer's Multiclass Dataset (Equal and*

Augmented); 2023. Available from: https://www.kaggle.com/datasets/aryansinghal10/alzheimers-multiclass-dataset-equal-and-augmented [Last accessed on 2025 Oct 28].

23. Smith J, Doe A, Johnson M. Challenges in using preprocessed MRI datasets for machine learning applications. *J Med Imaging*. 2022;29(4):123-130.

    doi: 10.1117/1.jmi.29.4.123

24. Eskildsen SF, Coupé P, GarcíaLorenzo D, Fonov VS, Pruessner J, Collins DL. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011;56(2):593606.

    doi: 10.1016/j.neuroimage.2011.01.001

25. Suk HI, Lee SW, Shen D, Alzheimer's Disease Neuroimaging Initiative. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med Image Anal*. 2017;37:101-113.

    doi: 10.1016/j.media.2017.01.008

26. European Society of Radiology. Challenges and solutions for introducing artificial intelligence (AI) in daily clinical workflow. *Eur Radiol*. 2020;30(9):4856-4864.

    doi: 10.1007/s00330-020-07148-2

27. Raza MQ, Mehmood A, Ahmad J, Iqbal M, Hussain M, Choi GS. Benchmarking deep learning models for early Alzheimer's diagnosis: Reproducibility, generalization and bias. *Alzheimers Dement*. 2025;21(1):34-47.

28. Yang K, Mohammed EA. *A Review of Artificial Intelligence Technologies for Early Prediction of Alzheimer's Disease*. [arXiv Preprint]; 2020.

    doi: 10.48550/arxiv.2012.12345

29. Malik I, Iqbal A, Gu YH, Al-antari MA. Deep learning for Alzheimer's disease prediction: A comprehensive review. *Diagnostics (Basel)*. 2024;14(12):1281.

    doi: 10.3390/diagnostics14121281

30. Gour S, Mohan KS, Joshi A, Sharma AK, Gupta S, Pandagre KN. Hybrid machine learning for disease diagnosis: A review of case studies and performance evaluation using multi-source data. *J Inf Syst Eng Manag*. 2025;10(36s):604-612.

    doi: 10.52783/jisem.v10i36s.6537

31. Ahmed S, Kim BC, Lee KH, Jung HY. Ensemble of ROI-based convolutional neural network classifiers for staging the Alzheimer disease spectrum from magnetic resonance imaging. *PLoS One*. 2020;15:e0242712.

    doi: 10.1371/journal.pone.0242712

32. Kingma DP, Ba J. *Adam: A Method for Stochastic Optimization*. [arXiv Preprint]; 2014. Available from: https://arxiv.org/abs/1412.6980 [Last accessed on 2025 Oct 06].

33. Raza H, Siddiqui M, Zafar S. A comprehensive review on machine learning approaches for early diagnosis and classification of Alzheimer's disease using deep learning techniques. *Front Comput Sci*. 2024;6:1404494.

    doi: 10.3389/fcomp.2024.1404494

34. Alruily M, Abd El-Aziz AA, Mostafa AM, *et al*. Ensemble deep learning for Alzheimer's disease diagnosis using MRI: Integrating features from VGG16, MobileNet, and InceptionResNetV2 models. *PLoS One*. 2025;20(4):e0318620.

    doi: 10.1371/journal.pone.0318620

35. Gautam P, Singh M. Alzheimer's disease classification using the fusion of improved 3D-VGG-16 and machine learning classifiers. *Int J Biomed Eng Technol*. 2025;47(1):1-27.

    doi: 10.1504/ijbet.2025.143776