Mesopotamian journal of Big Data
Vol. **(2025)**, 2025, **pp**. 313–328
DOI: https://doi.org/10.58496/MJBD/2025/020 , ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

Research Article

Recognition of Alzheimer's Disease Stages Via Inceptionv3 and Resnet50

Iman Aljubouri [1,*], Mostafa Ragheb [2], Mohamad Hamady [2]

1 Biomedical Information College, University of Information Technology and Communications, Baghdad, Iraq
2 Arts, Sciences and Technology University (AUL), Computer and Communication Engineering Department, Cornich Al Mazraa, Beirut, Lebanon

ARTICLE INFO

Article History
Received 07 March 2025
Revised 01 Jul 2025
Accepted 07 Sep 2025
Published 27 Oct 2025

Keywords
Alzheimer's disease
Deep learning model
Image processing
InceptionV3
ResNet50

ABSTRACT

Early and precise detection of Alzheimer's disease (AD) is essential for successful treatment. This research presents a system that autonomously detects and categorizes the phases of Alzheimer's disease via brain scans and sophisticated deep learning techniques, including InceptionV3 and ResNet50. These models started with pretrained weights and were augmented by including bespoke classification layers, which consisted of dropout, batch normalization, and dense layers to increase performance and mitigate overfitting. The preprocessing processes included scaling the picture to 224 by 224 pixels, using average filtering for denoising, and converting the color space to guarantee compatibility with the models. Evaluations of the OASIS dataset illustrate the efficacy of the proposed approaches in accurately differentiating among the various phases of Alzheimer's disease, including four classifications: nondemented, slightly demented, very mildly demented, and moderately demented. ResNet50 outperforms InceptionV3, achieving an accuracy of 93.9%, a micro F1 score of 94%, and a macro F1 score of 96%, demonstrating its efficacy and consistent performance in detecting and classifying all categories. Compared with current models, the suggested technique is more effective.

1. INTRODUCTION

A major challenge for healthcare providers has been Alzheimer's disease (AD) in the last decade. The global prevalence of AD is estimated to be approximately 5.5 million in the 65+ age group. It is the sixth leading cause of death in the country. The deterioration of cognitive capacity and the lack of confirmed treatments that alter the disease are hallmarks of AD, an incurable neurodegenerative brain ailment [1]. Much research and development has led to the development of ways to diagnose the disease in its presymptomatic stages in the hopes of slowing or stopping its course. Consequently, advanced neuroimaging techniques, such as positron emission tomography and magnetic resonance imaging (MRI), have been developed and used to identify structural and genetic indicators associated with AD. Due to the great improvement in neuroimaging equipment, integrating large-scale multimodal neuroimaging data with high dimensions has become more challenging [2]. Consequently, there has been a substantial increase in the demand for computer-assisted systems capable of performing integrative analysis. Recent advancements in artificial intelligence and other intelligent technologies have enabled the utilization of several innovative techniques, including machine learning (ML), to assist physicians in diagnosing Alzheimer's disease. Several well-known pattern analysis methods have been used for early AD detection and are now promising for the prediction of AD progression [3, 4]. These methods include logistic regression (LR), support vector machines (SVMs), linear discriminant analysis, recursive feature elimination via support vector machines, and the linear programming boosting method.

Classification techniques, such as SVMs, often encompass four stages: feature extraction, feature selection, dimensionality reduction, and selection of feature-based approaches. Time, specialized knowledge, and even several optimization rounds may be necessary for these processes [5]. The area of large-scale, high-dimensional medical imaging analysis has attracted much interest in deep learning (DL), a subfield of ML, as a way to overcome these challenges. With raw neuroimaging data, deep learning creates features via "on-the-fly" machine learning. Compared with more traditional ML techniques, deep learning approaches such as convolutional neural networks (CNNs) perform far better [6, 7].

Recent advancements in machine learning and deep learning have enhanced the diagnosis and categorization of Alzheimer's disease. Researchers have used many methodologies for neuroimaging, cognitive, and behavioral data. The approaches include support vector machines (SVMs), artificial neural networks (ANNs), decision trees, convolution neural

*Corresponding author. Email: emanaljubouri88@gmail.com

networks (CNNs), and recurrent models such as long short-term memory (LSTM) and bidirectional long short-term memory (BiLSTM). Despite notable advancements, many gaps remain in the literature that this project intends to address:

- Many existing studies depend on traditional machine learning techniques (such as SVM and decision trees) or rudimentary deep learning models (such as shallow CNNs or basic LSTM variations). Limited research has been conducted on advanced architectures such as ResNet50 and InceptionV3, which are prevalent in computer vision but underexploited in medical imaging for Alzheimer's disease diagnosis.
- Direct, comprehensive comparisons of modern deep learning models on identical Alzheimer's disease datasets are lacking. It is crucial to ascertain which models exhibit optimal performance across various AD categorization tasks.
- Many contemporary deep learning applications concentrate only on binary classification (AD vs non-AD). Limited research has assessed the capacity of modern models, such ResNet50 and InceptionV3, to differentiate between phases such as mild cognitive impairment (MCI), moderate Alzheimer's disease (AD), and severe AD. This ability is crucial for prompt diagnosis and action planning.

ResNet50 and InceptionV3 are used for their proficiency in feature extraction and transfer learning. Many studies connected to AD, however, do not capitalize on these advantages.

This work aims to develop a deep learning-based system capable of detecting and classifying the beginning or severity of Alzheimer's disease. Consequently, this article intends to do the following:

- The various phases of Alzheimer's disease were analysed, and contemporary machine learning and deep learning techniques were applied to neuroimaging and clinical data.
- We introduce sophisticated deep learning architectures using InceptionV3 and ResNet50, which exhibit robust feature extraction and transfer learning capabilities.
- We propose a deep learning framework using these models for the precise classification of Alzheimer's disease stages, employing pretrained weights and fine-tuning techniques to optimize the extraction of critical discriminative characteristics relevant to disease development.
- We evaluate the efficacy of these models on a typical AD dataset and systematically contrast them with conventional approaches.

The following four primary sections structure this manuscript. Section 2 describes AD, illustrates the importance of using DL techniques, and discusses some related work in the literature. In Section 3, the method proposed in this research is described, followed by the methodology outcomes and discussion in Section 4. Section 5 concludes the work in this research paper.

2. BACKGROUND

2.1 Alzheimer Disease

Alzheimer's disease is the most prevalent form of dementia when cognitive deterioration is examined [8]. AD is characterized by a gradual start, initially presenting as slight memory deficits before progressing to a deterioration in cognitive functions, including communication and environmental perception [9]. The brain areas involved in cognitive functions, including memory, language, and reasoning, are impacted. A decline in an individual's ability to perform normal activities may result in significantly negative consequences [10]. AD is categorized as a progressive neurological condition.

This signifies alterations in the brain's structure caused by the accumulation of specific proteins. It causes atrophy and neuronal loss, which are later pathological alterations in the brain. Initial manifestations of AD frequently include difficulties in recalling previous conversations or events. This occurrence may lead to significant cognitive deficits and a marked decline in individuals' ability to perform daily activities [11]. The progression of symptoms can be either mitigated or accelerated by pharmacological medications. A remedy for AD has not yet been discovered. In advanced phases, a substantial decline in brain function may result in detrimental outcomes such as dehydration, malnutrition, or illness [12]. The previously listed issues may result in fatalities.

The principal manifestation correlated with AD is memory impairment. Initial indicators include challenges in recollecting previous conversations or occurrences [11]. Nevertheless, as the disease progresses, cognitive ability decreases, and additional symptoms manifest. Those who are affected by the previously mentioned condition may initially manifest symptoms such as cognitive dysfunction and memory impairment. As symptoms progress, one might observe a family member or friend exhibiting heightened awareness of the issues at hand [13]. Brain abnormalities characterize AD, contributing to the gradual development of cognitive and functional declines [14]. The issues mentioned above can be categorized as shown in Fig. 1 [15].

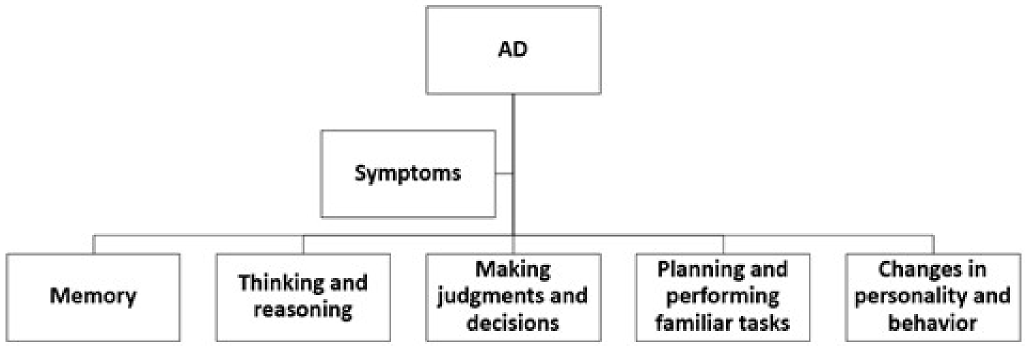

Fig. 1. Major symptoms detected in AD patients.

As illustrated in Fig. 1, the aforementioned problems can be categorized into five separate groups:

- Memory: This is a prevalent phenomenon that individuals may encounter on an intermittent basis.
- Thinking and reasoning: Pupils who have received a diagnosis of attention deficit (AD) frequently encounter challenges with concentration and reasoning, specifically concerning abstract notions such as numbers.
- Making Judgments and Decisions: An individual may wear inappropriate clothing for the current weather conditions or display poor judgment in social situations.
- Planning and Performing Familiar Tasks: The execution of routine tasks becomes more challenging for individuals with AD as a result of the subsequent sequential stages: planning and performing familiar tasks.
- Changing Personality and Behavior: Personality and Behavior Alterations: AD is associated with neurobiological changes that can significantly influence an affected individual's disposition and conduct.

1.2 Deep Learning

ML is an area of artificial intelligence in which computer systems gain knowledge through experience [16]. ML refers to the implementation of methods by computers to autonomously learn from data and perform tasks without explicit programming. DL employs sophisticated algorithms whose architecture was inspired by the human brain. It is an artificial intelligence (AI) methodology that instructs computers to process data in a manner analogous to that of the human brain. Models employing DL are capable of discerning intricate patterns in text, audio, images and other data types to produce accurate analyses and predictions [17]. The primary objective of AI is to develop intelligent systems capable of human-like thought and action. Recently, several AI systems, such as voice-activated television remotes, digital assistants, fraud detection mechanisms, and automatic facial recognition technologies have been developed via DL.

DL is also an essential component of innovative technologies such as autonomous vehicles and virtual reality. DL models are intelligent systems in which data scientists train to perform tasks via an algorithm or a predetermined sequence of operations. Businesses employ DL models across a range of applications to assess data quality and produce forecasts [19]. DL is currently utilized extensively in numerous industries, including the automotive, aerospace, manufacturing, electronics, and medicinal sectors [20].

Neural networks are typically composed of multiple layers. In the initial stratum, referred to as the input layer, input signals are received and transmitted to the subsequent stratum. A number of computations and feature extractions follow the hidden layer. Frequently, numerous concealed layers exist. To provide input signals to the concealed nodes, a weight is applied to each signal, and a bias is added to the result. Each stratum is then assigned an activation function to determine which node should be executed [21].

1.3 Related Work

Researchers and physicians are utilizing ML and DL for the detection of AD. Reference [22] details groundbreaking work that used Eigenbrain on MRI scans of the brain to reveal a detection mechanism. For model training, they combined SVMs with particle swarm optimization. Following their advice helped pinpoint which areas of the brain were most impacted by AD. There is a mention of a nearby study article in [23] that focused on MRI scans. In this study, the scientists used ANN and gradient boost models to differentiate dementia from other distinct traits. The authors achieved comparable outcomes in their study [22]. In [24], the authors proposed a multimodal technique that combined cognitive and linguistic aspects. The scientists trained the model to detect AD and its severity via an ANN. The results from their method outperform those from well-established methods.

To identify AD, the authors of [25] created a new framework using ML and DL techniques. The decision trees, LTSM, CNN, SVM, K nearest neighbor (KNN), and multilayer perceptron (MLP) were among the numerous methods used and evaluated. The detection accuracy has been used to measure the performance of various ML and DL techniques. With a detection accuracy of 91.28%, the experimental findings show that BiLSTM performs better than the ML approaches do. In [26], the authors used DL to analyse imaging, genetic, and clinical test data to identify patients as AD, MCI, or controls. To extract features from genetic and clinical data, they utilized stacked denoising autoencoders; they employed 3D convolutional neural networks for imaging data. Furthermore, an innovative method for interpreting data was developed to identify the best features that the deep models acquired via perturbation analysis and clustering. Deep models, including KNN, SVM, and random forest (RF), outperformed shallow models in the AD neuroimaging initiative dataset. The precision, recall, accuracy, and F1 scores of these models indicate that multimodality data integration outperforms single-modality models.

In [27], scientists devised a model for diagnosing and tracking the progression of AD that is both precise and straightforward to interpret. A two-layer model was constructed by incorporating eleven modalities obtained from the AD neuroimaging initiative dataset through the implementation of the random forest classification method. The initial layer detects Alzheimer's disease patients in the early stages, whereas the subsequent layer distinguishes MCI-to-AD progression. Twenty-two explainers and the SHapley additive explanations feature attribution framework were used to guarantee the model's transparency and explainability.

The authors of [28] introduced a lightweight stacked convolutional neural network with a channel attention network for the categorization of AD based on MR images to overcome existing challenges. The authors progressively connected five CNN modules to form a stacked CNN aimed at generating a hierarchical understanding of features through multilevel extraction, effectively reducing noise and enhancing the efficacy of the weights. This function was later incorporated into a channel attention module to discern pertinent features on the basis of the channel dimension, hence improving the selection of essential features. Consequently, the model exhibited a reduced number of parameters, making it suitable for training on constrained datasets. However, the authors in [29] employed advanced DL methods to increase the diagnostic effectiveness of gait analysis. Datasets acquired from wearable sensors and motion capture devices were employed to construct convolutional neural networks and recurrent neural networks for classifying individuals as healthy or at risk for AD.

The authors of [29] examined the viability of multitask models for volumetric analysis of magnetic resonance imaging scans in the diagnosis of knee osteoarthritis while considering computational efficiency. To exploit the correlation Between segmentation and classification tasks, two 3D multitask models, OA-MTL (osteoarthritis-multitask learning) and RES-MTL (residual multitask learning), were created to concurrently segment knee structures and classify the incidence of knee osteoarthritis.

Table I provides a summary of the comparison of existing works discussed in this section. Consequently, current deep learning-based anomaly detection systems have focused on improving classification efficacy. Nonetheless, few experimental studies have focused on creating a more compact model that exhibits both decreased complexity and improved accuracy in identification. This work seeks to create an efficient and precise deep learning-based system for identifying and categorizing the beginning and development stages of Alzheimer's disease. A sophisticated deep learning architecture using InceptionV3 and ResNet50, capitalizing on their robust feature extraction skills and transfer learning potential, is shown. The OASIS dataset is used to evaluate and compare these techniques.

TABLE. I. COMPARATIVE SUMMARY OF PREVIOUS STUDIES ON AD DETECTION USING ML/DL

Study	Data Modality	Techniques Used	Main Contribution
[22]	MRI	SVM mixed with Particle Swarm Optimization; Eigenbrain	Detect brain areas affected by AD.
[23]	MRI	ANN with Gradient Boosting	Detect AD.

[24]	Cognitive with Language ability	ANN	Multimodal approach based on cognitive and linguistic features.
[25]	Various type of Data	DT, BiLSTM, CNN, SVM, KNN, and MLP	Comparative study between ML and DL methods.
[26]	Imaging, Genetic and Clinical data	3D-CNN, Stacked Denoising Auto-Encoders, SVM, RF and k-NN	Effective feature selection technique.
[27]	ADNI (11 modalities)	Random Forest with SHapley Additive explanations	Two-layer interpretable model for detection and MCI-to-AD progression tracking
[28]	MRI	Stacked CNN embedded with Channel Attention Network	Lightweight model suitable for small datasets with an improved feature selection technique.
[30]	Gait (wearables, motion capture)	CNN and RNN	Used gait analysis for classifying AD risks.

2 PROPOSED APPROACH

This article presents DL-based approaches for detecting and classifying AD levels. In addition to the normal case, three levels of AD were considered: mild, moderate, and very mild dementia. Advanced approaches to DL are utilized in this paper to distinguish the various phases of AD via the following methodology process (see Fig. 2).

2.1 Data collection

A sequence of photographs is collected for training the DL model. The photographs that are gathered ought to depict both the healthy state and the three potential phases of AD [31]:

- Non-Demented (Non-D) Individuals: This group comprises normal patients and shows no signs of AD.
- Mild Demented (Mild-D): This group includes patients for whom medical professionals and family members can both observe that they are experiencing severe memory and cognitive problems.
- Moderate Demented (Moderate-D): Patients in this group are characterized by increased levels of confusion and forgetfulness. They require progressively more assistance with activities of daily living and personal hygiene.
- Very Mild Demented (VeryMild-D): In this group, the cognitive abilities of the patients decline as the condition progresses. The condition hinders the patient's capacity to engage in physical activities and ambulate. The

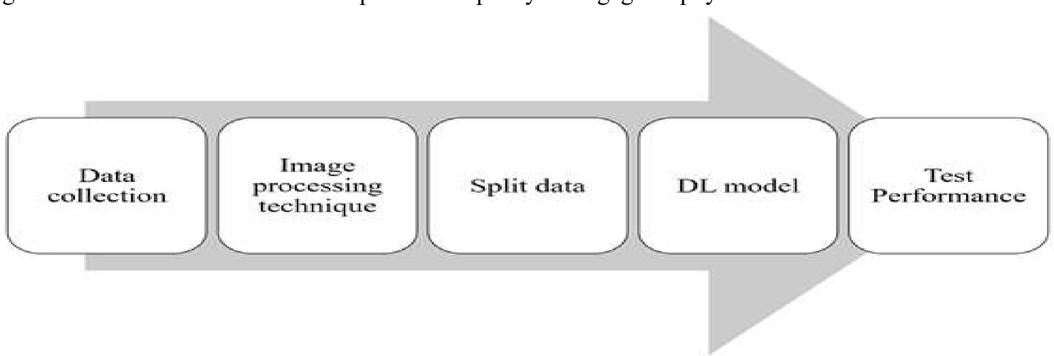

Fig. 2. Proposed approach process from collecting significant data to testing the performance of DL models.

Patients exhibit a loss of communicative ability. Conscious discourse and communication have become impossible. They may seldom express words or sentences. Ongoing support with personal hygiene is needed.

This research uses the OASIS MRI dataset, which includes 40,384 pictures sourced via Kaggle. The images were categorized into four categories according to the progression of Alzheimer's disease. No dementia symptoms (non-D) may suggest mild dementia symptoms (very mild-D), whereas mild dementia symptoms are suggested (mild-D), and moderate dementia symptoms are indicated (moderate-D). The dataset serves as an essential resource for the research and identification of early indicators of Alzheimer's disease [37]. This research analysed 6,430 MR images with the following distributions: non-D: 43.8%, very mild-D: 30.7%, mild-D: 24.7%, and moderate-D: 0.8%.

2.2 Image Preprocessing Technique

By employing image processing techniques, the acquired images can be refined [32]. All the MR images must undergo a preprocessing phase to meet the primary specifications of ResNet50 and InceptionV3, ensuring consistency and compatibility with the deep learning models:

- Resizing: All the images are resized to 224 × 224 pixels to match the input dimensions of InceptionV3 and ResNet50:
- Denoising: A 3 × 3 average filter is applied to smooth high-frequency noise:

$$I_{i,j} = \frac{1}{9}\sum_{m=-1}^{1}\sum_{m=-1}^{1} I_{i+m,j+n}$$

- Color Correction: BGR channels are reordered to RGB for compatibility with ImageNet-based models.
- Normalization: Pixel intensities, called I, are normalized to the [0, 1] range:

$$I_{nom} = \frac{I}{255}$$

- These steps enhance DL model performance and reduce overfitting by creating suitable input features.

2.3 Split Data

A two-step split technique was used to partition the dataset into smaller subgroups, ensuring a comprehensive and equitable evaluation of the suggested deep learning models. The whole dataset was first divided into training and temporary test subsets with the *train_test_split*() function, employing an 80:20 ratio and stratification according to class labels to preserve a uniform class distribution throughout the subsets. The temporary test set was then divided into two equal parts, the validation and final test sets, each comprising 10% of the total data, via stratified sampling once again. To ensure that the results could be repeated, a fixed random seed (*random_state* = 42) was used.

2.4 Deep Learning Models

In this step, InceptionV3 and ResNet50 are implemented. Several changes were made to the standard pretrained models to make the InceptionV3 and ResNet50 architectures better suited to AD images. The descriptions of the learning algorithms are as follows [33]:

2.4.1 InceptionV3

For the multiclass classification challenge of Alzheimer's disease level detection and classification, the InceptionV3 convolutional neural network served as the foundation of our deep learning model. InceptionV3 is a recognized architecture noted for its computational efficiency and proficient multiscale feature extraction using factorized and parallel convolutions.

- **Feature Extraction**. An InceptionV3 model pretrained with weights derived from ImageNet was used. The upper classification layers were omitted (include_top =False), and the input picture dimensions were modified to 224 × 224 × 3. All the convolutional layers in the foundational model were immobilized throughout training to maintain the general-purpose features acquired from extensive datasets.

On the basis of the input image, denoted by $X \in \mathrm{R}^{224\times224\times3}$, the InceptionV3 model can be employed as a deep feature extractor as follows:

$$F = f_{\text{Inception}}(X) \quad (3)$$

where $F \in \mathrm{R}^{n\times n\times d}$ represents the extracted feature map.

- **Custom Classification Head**. A custom classification head was applied to the extracted features via the following layers:

Global average pooling that minimizes each feature map to a scalar:

$$g_i = \frac{1}{n^2} \sum_{i=1}^{n^2} F_{i,j}, \; for \; i = 1,....,d \tag{4}$$

where:

- F is the 3D feature tensor,
- n represents the spatial dimension,
- d is the number of feature maps,
- g represents the resulting 1D vector after applying global average pooling across spatial dimensions.

- **Fully connected layers [34]:** The features resulting from the previous step were fed into a series of dense layers on the

basis of ReLU activations, batch normalization, and dropout regularization:

$$h_l = \phi(BN(Dropout(W_l \cdot h_{l-1} + b_l))) \tag{5}$$

where:

- h_l: Output (activation) of the l-th layer.
- h_{l-1}: Output (activation) of the previous layer, layer $l-1$.
- W_l: Weight matrix for the l-th layer (learnable parameters).
- b_l: Bias vector for the l-th layer (learnable parameters).
- $W_l \cdot h_{l-1} + b_l$: Linear transformation, also known as affine transformation.

- **Dropout:** To avoid overfitting, this regularization method "drops" (sets to 0) a subset of units at random during training.

Batch normalization (BN): BN normalizes the inputs to have zero mean and unit variance, assisting in stabilizing and accelerating training.

ϕ: Nonlinear activation function (ReLU).

Output layer represented by a dense layer with Softmax activation. It is used to provide the final class probabilities for the adopted 4 levels:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}, \quad \mathrm{i = 1,......,C} \tag{6}$$

where:

- $\hat{y}_i$ *is* the predicted probability that the input belongs to class *i*.
- z_i is the logit provided by the DL model for class *i*.
- C is the total number of classes (4 levels of AD).
- $\sum_{j=1}^{C} e^{z_j}$ is the normalization term.

Model training. The Adam optimizer is used with the categorical cross-entropy loss function:

$$L_{CCE} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{7}$$

- L_{CCE} represents the categorical cross-entropy loss value.

– C is the total number of classes in the classification task (the adopted four AD levels).
– y_i is the true label for class i in one-hot encoding. It is set to 1 if class i is the correct class and 0 otherwise.
– $\hat{y}_i$ represents the predicted probability that the input belongs to class i. It is usually calculated on the basis of the Softmax function.

The training was conducted over 60 epochs with two callbacks. The first component is ModelCheckpoint. It is used to preserve the optimal model according to the validation performance. The second mechanism is early stopping, which is used to terminate training if the validation performance does not improve for 10 consecutive epochs.

Evaluation Metrics. The accuracy of the model is computed as follows:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} 1_{\{\arg\max(\hat{y}_i) = \arg\max(y_i)\}} \quad (8)$$

Where: N is the available number of samples in the dataset.

- $\hat{y}_i$ represents the predicted output vector for the i-th sample.
- where y_i is the true label value of the i-th sample.
- $\arg\max(\hat{y}_i)$ represents the index of the maximum value in the predicted vector $y\hat{}_i$, representing the predicted class.
- $\arg\max(y_i)$ represents the index of the maximum value in the ground truth vector y_i, representing the true class.
- $1_{\{\arg\max(\hat{y}_i) = \arg\max(y_i)\}}$ is an indicator function that equals 1 if the predicted class matches the true class for sample i and 0 otherwise.

2.4.2 ResNet50 Model

The ResNet50 model serves as a feature extractor to provide a robust multiclass classification model for the phases of Alzheimer's disease. ResNet50 is a deep convolutional neural network recognized for its residual connections. It mitigates the vanishing gradient issue and facilitates the training of deeper networks.

- **Feature Extraction**: The top classification layers were removed from the pretrained ResNet50 (include_top=False).

The supplied picture dimensions were adjusted to 224 × 224 × 3. The training preserves all the convolutional layers in the basic model to keep the obtained hierarchical visual characteristics from enormous natural image datasets.

For an input image, called $X \in \mathrm{R}^{224\times224\times3}$, the ResNet50 model was used as a deep feature extractor:

$$F = f_{\mathrm{ResNet}}(X) \quad (9)$$

where F represents the extracted feature tensor provided by the last convolutional block.

- **Custom Classification Head**: A custom classification head was employed on the extracted feature maps. it contains:

The flattening layer converts the 3D feature map F into a 1D vector:

$$f = \mathrm{Flatten}(F) \quad (10)$$

where $\mathbf{f} \in \mathrm{R}^d$ represents the flattened feature vector.

A completely linked output layer, referred to as a dense layer, incorporates Softmax activation to associate features with class probabilities, as described in equation 6.

- **Model training**: The Adam optimizer was used to compile and train the model on the basis of the categorical cross-entropy loss function given in equation 7. Two training callbacks were utilized: early stopping (patience = 8) to terminate training when the validation loss ceased to improve and restoring_best_weights=True to retain the model with the lowest validation loss. The training generator allowed a maximum of 60 epochs for model training, whereas a separate validation generator confirmed the model's proper functionality.

- **Evaluation Metrics**: The model's performance was evaluated in terms of accuracy, as defined by equation 8.
- The proposed architectural tuning improves feature extraction by decreasing overfitting and improving generalization, especially for differentiating between mild and moderate AD stages. EarlyStopping and ModelCheckpoint callbacks were used to handle overfitting by tracking validation loss and accuracy, and training was stopped when no improvement was observed for 10 epochs.

2.5 Test Performance

The efficacy of the proposed method is assessed by calculating the F1 score, accuracy, confusion matrix (CM), and recall precision. These measures are discussed later in this section.

An evaluation of the effectiveness of a classification algorithm is obtained through the use of CM. The categorization problem investigated in this study is not binary. There are four different output classes that might be used. As a result, the CM should conform to Table II [35].

TABLE. II. A SAMPLE OF CM REPRESENTING FOUR DIFFERENT CLASSES: A, B, C, AND D

		Predicted			
	Class	**A**	**B**	**C**	**D**
Actual	**A**	TP_A	P_{AB}	P_{AC}	P_{AD}
	B	P_{BA}	TP_B	P_{BC}	P_{BD}
	C	P_{CA}	P_{CB}	TP_C	P_{CD}
	D	P_{DA}	P_{DB}	P_{DC}	TP_D

The variable TP_I represents the true positive for Class *I*, whereas the variable P_{IJ} represents the instance when the real class is *I* but the projected class is *J*. The CM that represents a single class is derived according to Table III.

TABLE. III. CM SAMPLE FOR ONE CLASS

		Predicted	
		A	**Not A**
Actual	**A**	TP_A	FN_A
	Not A	FP_A	TN_A

The *FNs* of Class *I* in Table II are computed as follows:

$$FN = P_{AB} + P_{AC} + P_{AD} \tag{11}$$

FP denotes a false positive for class *I* and is computed as follows:

$$FP_A = P_{BA} + P_{CA} + P_{DA} \tag{12}$$

TN is the true negative of class *I* and is computed as follows:

$$TN = TP_B + P_{BC} + P_{BD} + P_{CB} + TP_C + P_{CD} + P_{DB} + P_{DC} + TP_D \tag{13}$$

The performance metrics are calculated via Table II and represented in Table IV [36, 37].

TABLE. IV. . PERFORMANCE METRICS FORMULAS

Metric	Formula
Precision	Precision = TP/(TP + FP)
Recall	Recall = TP/(TP + FN)
F1-score	F1-score = 2 × precision × recall/(precision + recall)
Micro F1-score	MicroF1-score = Σ(Number of class_i * F1-score(class_i))/Total Number of Classes
Macro F1-score	MacroF1-score = Σ(F1-score)/Number of Classes
Accuracy	Accuracy = (TP + TN)/(TP + FP + FN + TN)

Each metric description is listed below:

- Precision denotes the proportion of accurate positive predictions that are realized.
- Recall measures the number of positive observations that are predicted to be positive.
- The F1 score is determined by combining the model's recall and accuracy to complete the calculation. With respect

to recall and precision, it is the harmonic mean of the two.

- The micro F1 score is calculated by adding together the total number of false positives, true positives, and false negatives. In situations where there is an imbalance between the classes in the dataset, the micro F1 score can be useful.
- The macro F1 score is a calculation that determines the average of the F1 scores for each class, treating all classes in the same manner. The F1 score, which is averaged across all classes, is the most accurate measurement for multiclass classification.
- Accuracy is a term that refers to the degree to which the classifier is able to predict the future properly. Additionally, it is a representation of the proportion of accurate forecasts in relation to the total number of predictions.

3 RESULTS AND DISCUSSION

This section presents and discusses the outcomes of applying the InceptionV3 and ResNet50 models to the dataset outlined in Section 3.1. A comparison with the current system is also shown. Fig. 3 shows a comparison of MRI images before and after the use of preprocessing techniques at various stages of Alzheimer's disease. The photos in the right column exhibit enhanced clarity and uniformity due to scaling, denoising, color correction, and normalization, making them more appropriate for deep learning model input.

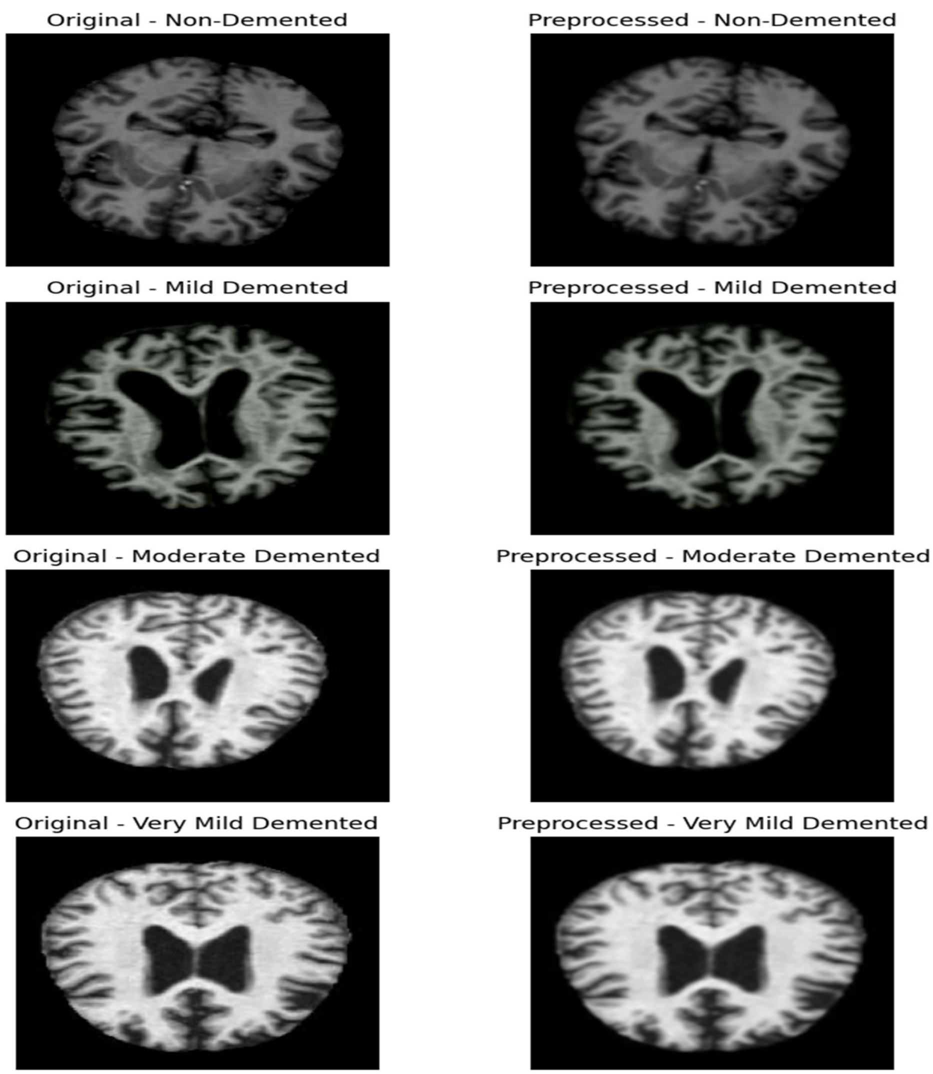

Fig. 3. Comparison between original and preprocessed MR images for different AD stages. The left column shows the original images, and the right column illustrates the corresponding preprocessed versio

3.1 Model performance results

Fig. 4 and Fig. 5 show the confusion matrices for InceptionV3 and ResNet50 for the multiclass classification task of AD stage prediction.

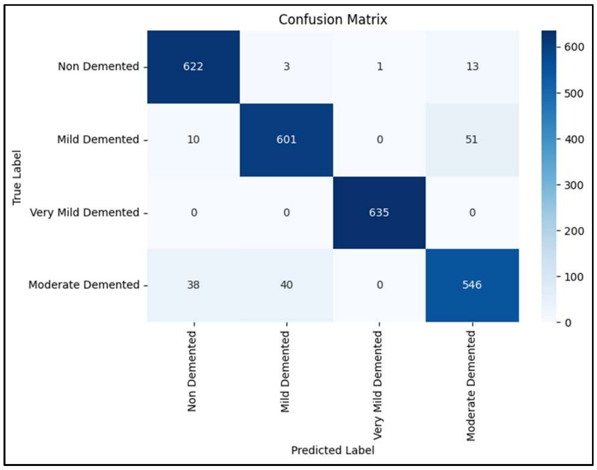

Fig. 4. CM for ResNet50.

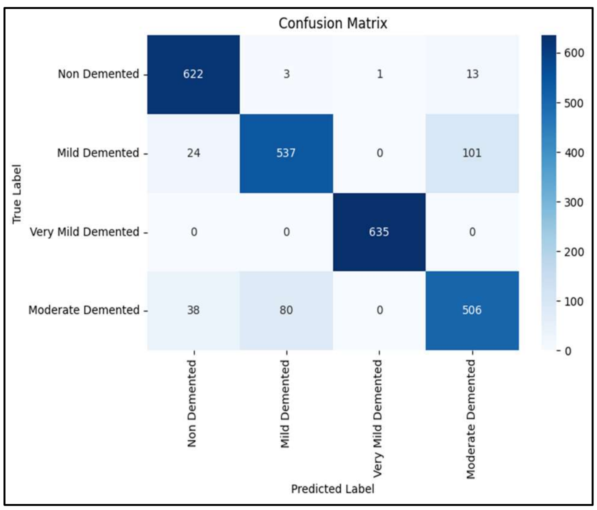

Fig. 5. CM for InceptionV3.

The ResNet50 model excels at classifying Alzheimer's disease categories. It identifies almost all instances of the VeryMild-D classification. The detection rates for the Non-D and Mild-D classes are notably high, with 622 out of 639 and 601 out of 662 accurately categorized, respectively. The Moderate-D class is considered the most challenging class owing to its parallels with other classes and an unequal distribution of patients; however, it attained 546 accurate predictions, despite some ambiguity with the Mild-D and Non-D classes.

The InceptionV3 model excels at identifying the VeryMild-D class (635/635). Nevertheless, the model has poor precision in identifying the mild-D and moderate-D categories. Only 537 out of the 662 Mild-D samples were accurately identified. Nonetheless, 101 samples were erroneously categorized as moderate-D. Furthermore, moderate-D has subpar performance, achieving only 506 accurate predictions and misclassifying 80 samples as mild-D. This indicates a greater degree of misunderstanding among classes than ResNet50 does.

Consequently, both models categorize the very mild-D class with near-perfect precision, indicating their ability to distinguish between early-stage and late-stage symptoms. Nonetheless, ResNet50 outperforms InceptionV3 in distinguishing between the mild-D and moderate-D classes. ResNet50 outperforms the other models in feature extraction during the intermediate phase and in class differentiation. Both models perform well; however, ResNet50 exhibits superior accuracy and excels in distinguishing challenging stages, such as the mild-D and moderate-D classes. These findings indicate that the ResNet-based model is superior for clinical Alzheimer's disease classification tasks where precise diagnosis is critical.

Table V shows a comparison between the performance of InceptionV3 and ResNet50 with and without image preprocessing steps on the basis of precision, recall, and F1 score values for each model.

TABLE. V. PERFORMANCE METRICS FOR EACH AD STAGE USING INCEPTION V3 AND ResNet50 WITH AND WITHOUT IMAGE PROCESSING.

Class	InceptionV3			ResNet50		
	Precision	Recall	F1-score	Precision	Recall	F1-score
With Preprocessing						
Non-Demented	0.91	0.97	0.94	0.97	0.99	0.98
Mild Demented	0.87	0.81	0.84	0.94	0.89	0.91
Moderate Demented	0.82	0.82	0.82	0.96	0.96	0.96
Very Mild Demented	1.00	1.00	1.00	1.00	1.00	1.00
Without Preprocessing						
Non-Demented	0.88	0.93	0.91	0.94	0.96	0.94
Mild Demented	0.83	0.77	0.80	0.91	0.85	0.87
Moderate Demented	0.78	0.78	0.78	0.92	0.91	0.92
Very Mild Demented	0.97	0.97	0.97	0.97	0.97	0.97
Improvement (%)	**3.41**	**4.36**	**3.52**	**3.35**	**3.60**	**3.83**

Table V shows that the application of image preprocessing techniques enhanced all the performance metrics for both InceptionV3 and ResNet50. The enhancement percentages range from 2% to 5%, confirming the importance of image preprocessing techniques such as scaling, denoising, and normalizing. These image processing algorithms enable the production of a more uniform picture with less noise. Consequently, they facilitate the more efficient extraction of features by deep learning models. Despite the minimal margins of enhancement, they underscore the importance of preprocessing in achieving reliable and generalizable classification performance throughout all stages of Alzheimer's disease.

Moreover, Table V indicates that the ResNet50 model has superior performance across almost all the metrics and classes. Both models exhibit flawless performance for the VeryMild-D class, with accuracy, recall, and F1-scores all at 1.00. Consequently, this class is readily identifiable, perhaps due to both models' ability to recognize the distinct early-stage indicators. In the Non-D stage, ResNet50 achieves an accuracy of 0.97 and a recall of 0.99, surpassing InceptionV3's scores of 0.91 for precision and 0.97 for recall. The data indicate that ResNet50 excels at differentiating cognitively normal subjects while maintaining a low rate of false positives. For the Mild-D class, ResNet50 achieves an accuracy of 0.94 and a recall of 0.89, whereas InceptionV3 attains a precision of 0.87 and a recall of 0.81. The F1 score of 0.91 for ResNet50 surpasses the 0.84 achieved by InceptionV3. Consequently, ResNet50 excels in achieving an equilibrium between sensitivity and accuracy. The moderate-D class exhibited the most significant improvement, with ResNet50 attaining a score of 0.96 across all three criteria, demonstrating dominance over InceptionV3, which scored 0.82. Consequently, ResNet50 demonstrates superior accuracy in categorizing advanced-stage dementia patients, exhibiting a reduced incidence of false positives and false negatives. This enhancement aligns with the confusion matrix, demonstrating that ResNet50 committed fewer errors in classifying the Mild-D and Moderate-D categories.

A comparison of the two models reveals exceptional performance characteristics on the basis of their design. The architecture of ResNet50 enables it to handle gradients better and train deeper networks, which leads to its constant high accuracy in identifying all stages of AD. This allows ResNet50 to recognize subtle and complex traits essential for distinguishing between different illness stages. Conversely, InceptionV3 demonstrates proficiency in detecting very mild demented and nondemented stages, although it shows decreased accuracy in the moderate and mild demented phases.

The complicated portions of InceptionV3, which are meant to increase efficiency and capture information at multiple sizes, may not identify different essential aspects as ResNet50 does. ResNet50 generally has a greater quantity of trainable parameters, allowing it to capture a wider range of characteristics. The macro- and micro-F1 scores were used for an extensive evaluation, which demonstrated that ResNet50 displayed enhanced performance and increased robustness.

3.2 Comparison

Owing to the multiclass and imbalanced dataset, macro- and micro-F1 scores are preferable for detecting any biases. Table VI shows a comparison of InceptionV3 and ResNet50 with other extant models that use the same dataset. In addition, this table also includes a comparison with the DenseNet model applied with the same preprocessing step of the proposed approach. The comparison emphasizes accuracy, macro F1 score, and micro F1 score measurements, as well as assessments of diverse methodologies identified in the literature.

TABLE. VI. ACCURACY, MICRO F1-SCORE, AND MACRO F1-SCORE COMAPRISON VALUES COMPARED TO MODELS IN THE LITERATURE

	Model	Accuracy (%)	Macro F1-score	Micro F1-score
Proposed Approach	InceptionV3	89.9	0.90	0.90
	ResNet50	93.9	0.94	0.96
	DenseNet	87.1	0.8	0.79
Study [28]	ResNet50	93.1	-	-
	InceptionResNetV2	89.1	-	-
Study [39]	AlexNet	82.2	-	-

These findings unequivocally demonstrate that the proposed ResNet50-based model outperforms other models in the classification of Alzheimer's disease stages. It achieves an optimal accuracy of 93.9%, a macro F1 score of 0.94, and a micro F1 score of 96%. This finding indicates that the model excels not only in predicting Alzheimer's disease overall but also in distinguishing levels (macro F1) and labels (micro F1). The suggested InceptionV3-based model has good performance, achieving an accuracy of 89.9% along with macro- and micro-F1 scores of 90%. These findings surpass those of current methodologies, albeit somewhat inferior to ResNet50. As illustrated in this table, DenseNet reached an accuracy of 87.1%. However, its macro- and micro-F1 scores are lower than those of the proposed ResNet50 and InceptionV3 models, indicating relatively weaker performance in classifying AD stages.

The research in [28] indicated an accuracy of 93.1% for ResNet50 and 89.1% for InceptionResNetV2. The proposed ResNet50 model has an accuracy that is 0.8% superior to that of the model presented in Study [28]. This enhancement results from the preprocessing, model optimization, and dataset organization conducted in this work. The research in

[38] reported that the lowest accuracy of 82.2% was achieved via AlexNet. This result demonstrates that shallow designs inadequately capture the intricate feature hierarchies essential for multiclass AAD stage classification. The proposed ResNet50 architecture demonstrates superior and more consistent performance across all primary criteria, validating its appropriateness for the accurate and reliable classification of Alzheimer's disease progression stages in clinical imaging datasets [40].

4 CONCLUSIONS

The predominant type of dementia is Alzheimer's disease (AD). Diagnosis may be achieved via several methods. The development of a dependable approach for the early detection of Alzheimer's disease is increasingly crucial due to the increasing incidence of diagnoses. Deep learning approaches have exceptional efficacy in several domains, including pattern recognition and medical diagnosis. This research aims to develop an automated method for the identification and categorization of Alzheimer's disease (AD). Advanced deep learning models integrated with image processing methods were used to provide a precise and resilient solution for Alzheimer's disease identification and classification. Two deep learning models were used and compared: InceptionV3 and ResNet50. This work integrates a robust picture preprocessing pipeline that includes scaling, denoising, color correction, and normalization to guarantee compatibility and improve the input quality of deep learning models. To achieve multiclass classification of Alzheimer's disease stages, InceptionV3 and ResNet50 were developed with distinct classification heads and trained via categorical cross-entropy loss. A strategy that meticulously partitions the data and employs comprehensive assessment metrics ensures that the model is adequately tested and effectively identifies early phases. The OASIS MRI dataset, which includes four phases of Alzheimer's disease—non-D, mild-D, very mild-D, and moderate-D—is used to evaluate the suggested methodology. Performance indicators, including precision, recall, accuracy, and F1 score, were used to evaluate the efficacy of the suggested technique. The findings indicate that the suggested technique effectively identifies Alzheimer's disease. The ResNet50 model outperforms the InceptionV3 model, with an accuracy of 93.9%, a micro F1 score of 0.94, and a macro F1 score of 0.96.

The proposed deep learning models exhibit strong performance in identifying all four phases of Alzheimer's disease. ResNet50 attains the maximum accuracy of 93.9% and exhibits flawless classwise precision, particularly in identifying extremely mild and moderate dementia patients. InceptionV3 yields commendable results, particularly in detecting early-stage Alzheimer's disease with elevated recall. These findings indicate that the models are dependable and beneficial in a clinical context for facilitating early and precise identification of multistage Alzheimer's disease. The suggested ResNet50 and InceptionV3 architectures outperform existing systems in the literature in terms of accuracy and F1-scores. These findings demonstrate that our framework excels in achieving balanced and robust categorization across all phases of Alzheimer's disease.

The proposed DL models, especially ResNet50, achieve strong performance in distinguishing between mild and moderate stages of AD. The distinctions between these stages are often complex and pose diagnostic challenges. These results illustrate the potential of the proposed approach to support clinicians in making earlier and more accurate diagnoses and thus enabling timely intervention and improved patient outcomes.

This investigation is conducted using a single dataset. Consequently, the generalizability of these findings is restricted to alternative therapeutic environments, despite promising outcomes. The models may overlook domain-specific features pertinent to AD imaging owing to their reliance on pretrained architectures. Moreover, despite the use of preprocessing measures, class imbalance may continue to influence model sensitivity, especially in the latter stages of Alzheimer's disease. Consequently, the suggested methodology must be evaluated on further datasets, and the problem of unbalanced datasets should be addressed. Knowledge distillation may be used to develop a deep learning model that is more suitable for environments with constrained resources. Furthermore, explainable AI may be used to elucidate the primary determinant in the beginning of Alzheimer's disease.

Conflicts of interest

The authors declare that they have no conflicts of interest related to this article.

Funding

This research paper was not funded by any organization.

Acknowledgment

We would like to thank all the reviewers for the helpful comments and suggestions they gave us while we were planning and making this research. Their generosity with their time is very much appreciated.

References

[1] D. Strooper and B. Karran, 'The cellular phase of alzheimer's disease', Cell, vol. 164, no. 4, pp. 603–615, 2016.

[2] D. P. Veitch, M. W. Weiner, P. S. Aisen, L. A. Beckett, N. J. Cairns, and R. C. Green, 'Understanding disease progression and improving alzheimer's disease clinical trials: Recent highlights from the alzheimer's disease neuroimaging initiative', Alzheimer's & Dementia, vol. 15, no. 1, pp. 106–152, 2019.

[3] A. Sharma, S. Kaur, N. Memon, A. J. Fathima, S. Ray, and M. W. Bhatt, 'Alzheimer's patients detection using support vector machine (svm) with quantitative analysis', Neuroscience Informatics, vol. 1, no. 3, 2021.

[4] W. Zheng et al., 'Application of generalized split linearized bregman iteration algorithm for alzheimer's disease prediction. Aging (Albany NY)', Aging (Albany NY), vol. 12, no. 7, 2020.

[5] P. N. Kumar and L. P. Maguluri, 'Svm-based classifier for early detection of alzheimer's disease', Theory and Practice, vol. 30, no. 5, pp. 1120–1131, 2024.

[6] M. Mamun, S. B. Shawkat, M. S. Ahammed, M. M. Uddin, M. I. Mahmud, and A. M. Islam, 'Deep learning based model for alzheimer's disease detection using brain mri images', in 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 510–0516.

[7] A. Shehri, 'Alzheimer's disease diagnosis and classification using deep learning techniques', PeerJ Computer Science, vol. 8, no. 1, 2022.

[8] C. Silvestri, V. Almici, I. Libri, I. Mattioli, M. Cosseddu, and R. Turrone, 'Sex differences in the severity and progression of neuropsychiatric symptoms across different dementia types', Neurology: Clinical Practice, vol. 14, no. 3, 2024.

[9] Y. Qin, H. Han, Y. Li, J. Cui, H. Jia, and X. Ge, 'Estimating bidirectional transitions and identifying predictors of mild cognitive impairment', Neurology, vol. 100, no. 3, pp. e297–e307, 2023.

[10] Z. Breijyeh and R. Karaman, ‘Comprehensive review on alzheimer’s disease: causes and treatment’, Molecules, vol. 25, no. 24, 2020.

[11] M. P. Braz De Oliveira, M. Padovez Rdfc, M. A. De Noronha, C. Nodc, and A. Lpd, ‘Effectiveness of physical exercise at improving functional capacity in older adults living with alzheimer’s disease: a systematic review of randomized controlled trials’, Disability and rehabilitation, vol. 45, no. 3, pp. 391–402, 2023.

[12] P. Borkar, V. A. Wankhede, D. T. Mane, S. Limkar, J. Ramesh, and S. N. Ajani, ‘Deep learning and image processing- based early detection of alzheimer disease in cognitively normal individuals’, Soft Computing, vol. 2023, pp. 1–23.

[13] J. W. Culberson, J. Kopel, U. Sehar, and P. H. Reddy, ‘Aaurgent needs of caregiving in ageing populations with alzheimer’s disease and other chronic conditions: Support our loved ones’, Ageing Research Reviews, vol. 2023.

[14] H. Feng, P. Hu, Y. Chen, H. Sun, J. Cai, and X. He, ‘Decreased mir-451a in cerebrospinal fluid, a marker for both cognitive impairment and depressive symptoms in alzheimer’s disease’, Theranostics, vol. 13, no. 9, 2023.

[15] G. Plascencia-Villa and G. Perry, ‘Roles of oxidative stress in synaptic dysfunction and neuronal cell death in alzheimer’s disease’, Antioxidants, vol. 12, no. 8, 2023.

[16] J. Wang, Encyclopedia of Data Science and Machine Learning. IGI Global, 2023.

[17] Y. Lecun, Y. Bengio, and G. Hinton, ‘Deep learning’, Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[18] M. El Samad, G. Nasserddine, and A. Kheir, Introduction to artificial intelligence. Artificial Intelligence and Knowledge Processing 2023. CRC Press, pp. 1–14.

[19] I. Goodfellow, Y. Bengio, A. Courville, Deep learning. MIT press, 2016.

[20] A. N. Muhammad, A. M. Aseere, H. Chiroma, H. Shah, A. Y. Gital, and I. Hashem, ‘Deep learning application in smart cities: recent development, taxonomy, challenges and research prospects’, Neural Computing and Applications, vol. 33, no. 1, pp. 2973–3009, 2021.

[21] S. Cong and Y. Zhou, ‘A review of convolutional neural network architectures and their optimizations’, Artificial Intelligence Review, vol. 56, no. 3, pp. 1905–1969, 2023.

[22] Y. Zhang, Z. Dong, P. Phillips, S. Wang, J. G. Yang, and J. Yuan, ‘Detection of subjects and brain regions related to alzheimer’s disease using 3d mri scans based on eigenbrain and machine learning’, Frontiers in Computational Neuro- science, vol. 9, no. 1, 2015.

[23] R. M. Rawat, M. Akram, and S. S. Pradeep, ‘Dementia detection using machine learning by stacking models’, in 5th International Conference on Communication and Electronics Systems (ICCES), pp. 849–854.

[24] U. Sarawgi, W. Zulfikar, N. Soliman, and P. Maes, Multimodal inductive transfer learning for detection of alzheimer’s dementia and its severity. .

[25] K. Dashtipour, W. Taylor, S. Ansari, A. Zahid, M. Gogate, and J. Ahmad, Detecting alzheimer’s disease using machine learning methods. Ser. Lecture Notes of the Institute for Computer Sciences. Netherlands, Springer, 2022, pp. 89–100.

[26] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, ‘Multimodal deep learning models for early detection of alzheimer’s disease stage’, Scientific Reports, vol. 11, no. 1, 2021.

[27] S. El-Sappagh, J. M. Alonso, S. R. Islam, A. M. Sultan, and K. S. Kwak, ‘A multilayer multimodal detection and prediction model based on explainable artificial intelligence for alzheimer’s disease’, Scientific Reports, vol. 11, no. 1, 2021.

[28] N. Hassan, A. S. M. Miah, K. Suzuki, Y. Okuyama, and J. Shin, ‘Stacked CNN-based multichannel attention networks for Alzheimer disease detection’, Scientific Reports, vol. 15, no. 1, pp. 5815, 2025.

[29] P. S. Q. Yeoh, S. Goh, K. Hasikin, X. Wu and K. W. Lai, "3D Efficient Multi-Task Neural Network for Knee Osteoarthritis Diag- nosis Using MRI Scans: Data From the Osteoarthritis Initiative," in IEEE Access, vol. 11, pp. 135323-135333, 2023, doi: 10.1109/ACCESS.2023.3338379.

[30] M. Sarkar, ‘Integrating machine learning and deep learning techniques for advanced Alzheimer’s disease detection through gait analysis’, Journal of Business and Management Studies, vol. 7, no. 1, pp. 140–147, 2025.

[31] T. Lfds, V. Fdacd, A. Gratão, L. Kusumota, and R. Rodrigues, ‘Alzheimer’s disease: functional decline and stage of dementia’, Acta Paulista de Enfermagem, vol. 26, no. 1, pp. 219–225, 2013.

[32] A. Isjn and C. Direkog˘lu, ‘Review of mri-based brain tumor image segmentation using deep learning methods’, Procedia Computer Science, vol. 102, no. 1, pp. 317–324, 2016.

[33] A. Nivrito and M. Wahed, Comparative analysis between inception-v3 and other learning systems using facial expressions detection. 2016.

[34] S. A. Hasanah, A. A. Pravitasari, A. S. Abdullah, I. N. Yulita, and M. H. Asnawi, ‘A deep learning review of resnet architecture for lung disease identification in cxr image’, Applied Sciences, vol. 13, no. 24, 2023.

[35] M. Heydarian, T. E. Doyle, and R. Samavi, ‘Mlcm: Multilabel confusion matrix’, IEEE Access, vol. 10, no. 1, pp. 19083–19095, 2022.

[36] S. Kaddoura, A. E. Arid, and M. Moukhtar, ‘Evaluation of supervised machine learning algorithms for multiclass intrusion detection systems’, in Proceedings of the Future Technologies Conference (FTC) 2021, vol. 3, Springer, pp.1–16.

[37] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, ‘Optimal thresholding of classifiers to maximize f1 measure’, in Machine Learning and Knowledge Discovery in Databases: European Conference, Nancy, France, 2014, pp. 225–239.

[38] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, ‘Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults’, Journal of Cognitive Neuroscience, vol. 19, no. 9, pp. 1498-–1507, 2007.

[39] M. H. Al-Adhaileh, ‘Diagnosis and classification of Alzheimer’s disease by using a convolution neural network algorithm’, Soft Computing, vol. 26, no. 16, pp. 7751–7762, 2002.

[40] M. A. Habeeb, Y. L. Khaleel, R. D. Ismail, Z. Al-Qaysi, & F. N. Ameen , ‘Deep Learning Approaches for Gender Classification from Facial Images ’, Mesopotamian Journal of Big Data, vol 2024, pp185-198, 2024.