

# Efficient Alzheimer's disease classification using transfer learning with efficientNetV2S

Mojgan Askarizade<sup>1,\*</sup>

<sup>1</sup>Department of Computer Engineering,  
Faculty of Engineering, Ardakan  
University, Ardakan, Yazd, Iran

\*Author for correspondence:  
Email: maskari@ardakan.ac.ir

Received date: August 07, 2025  
Accepted date: August 28, 2025

Copyright: © 2025 Askarizade M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that affects millions of individuals worldwide, leading to memory loss, cognitive decline, and functional impairments. Early and accurate detection of AD is critical for effective management and treatment planning. This paper presents an efficient approach for Alzheimer's disease classification using a deep learning model based on the EfficientNetV2S architecture, leveraging transfer learning to enhance performance. EfficientNetV2S, an evolution of the EfficientNet model, is designed to balance speed and accuracy by combining fused-MBConv and MBConv layers, making it highly suitable for tasks requiring both high performance and computational efficiency. In this study, we fine-tune a model initialized with ImageNet-pretrained weights on a domain-specific Alzheimer's dataset. Furthermore, we rigorously validate the model's performance using k-fold cross-validation, confirming its reliability and generalizability across diverse data subsets. The proposed model achieved an accuracy of 98.1%, a precision of 98.9%, recall of 98.3%, and F1-score of 98.6%. These results demonstrate significant improvements in performance, outperforming other state-of-the-art models. Transfer learning allows the model to adapt pretrained features to the Alzheimer's domain, speeding up training and improving generalization. Our findings highlight the potential of EfficientNetV2S for high-performance applications in medical image classification, where both computational efficiency and accuracy are crucial.

**Keywords:** Alzheimer's disease, Machine learning, Deep learning, Image preprocessing, Magnetic resonance imaging (MRI)

## Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disorder, presents a growing challenge to healthcare systems worldwide due to its complex nature and devastating impact on patients and their families. Early detection of Alzheimer's is crucial as it enables timely intervention and management, potentially slowing disease progression and improving quality of life. One of the most effective tools for AD detection is Magnetic Resonance Imaging (MRI), which allows non-invasive visualization of brain structure and abnormalities.

However, interpreting MRI scans for Alzheimer's detection is highly complex and requires expertise in distinguishing subtle changes in brain morphology across different stages of the disease. In recent years, computer-aided diagnostic (CAD) techniques, powered by artificial intelligence (AI) and machine learning (ML), have gained significant attention for their ability to automate and enhance this process. These methods have shown promise in analyzing MRI images to detect Alzheimer's-related changes, but challenges remain due to issues such as small datasets, imbalanced class distributions, and the difficulty of capturing complex brain features.

The field of Alzheimer's detection has seen a variety of approaches utilizing both machine learning (ML) and deep learning (DL) techniques. In traditional ML methods, features are manually extracted from MRI images, and algorithms such as support vector machines (SVMs), random forests, and logistic regression are applied for classification. While these methods have

demonstrated some success, they often depend heavily on feature engineering and expert knowledge, which can be limiting. Deep learning (DL) techniques, particularly convolutional neural networks (CNNs), have revolutionized the field by automating feature extraction and achieving state-of-the-art performance in image classification tasks. These DL models, however, require large amounts of labeled data and substantial computational resources for training. To address this, transfer learning has emerged as a powerful technique, where a model pretrained on a large dataset (such as ImageNet) is fine-tuned on the target dataset. This approach significantly reduces the need for large labeled datasets and computational resources, making it particularly useful in medical imaging applications where labeled data may be scarce.

In this study, we propose a novel approach for Alzheimer's disease classification using MRI scans, by combining transfer learning with the EfficientNetV2S model. EfficientNetV2S is an advanced convolutional neural network known for its efficiency and scalability, achieving high performance while maintaining relatively low computational costs. The model is fine-tuned on a preprocessed and balanced dataset comprising MRI scans categorized into four classes: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The framework of our method involves several key stages: dataset preprocessing, model selection and fine-tuning, and performance evaluation.

First, the MRI dataset undergoes a series of preprocessing steps, including normalization of pixel intensity values to a standard range, spatial resizing of MRI images to ensure consistency, and bilateral filtering for noise reduction while preserving edge details. To enhance model performance and increase the diversity of training data, data augmentation techniques such as random rotations, flips, and scaling are applied during training. These augmentations help the model generalize better to unseen data by simulating variations commonly observed in real-world scenarios. The EfficientNetV2S model, pre-trained on the ImageNet dataset, is selected for the Alzheimer's classification task and fine-tuned on the augmented MRI dataset to learn specific features necessary for distinguishing between different Alzheimer's stages. The model is trained using a classification loss function, such as categorical cross-entropy, and its performance is evaluated using metrics such as accuracy, precision, recall, and F1 score. After training, the model's results are compared with other state-of-the-art methods for Alzheimer's detection, including six other pre-trained models.

The contributions of this study are as follows:

#### **High-performance Alzheimer's classification**

We demonstrate the effectiveness of EfficientNetV2S in classifying Alzheimer's disease from MRI scans, emphasizing its scalability, accuracy, and efficiency in medical image analysis.

#### **Comprehensive model comparison**

Through extensive experiments, we compare the performance of seven pretrained models—EfficientNetB0, MobileNet, ResNet, InceptionResNet, VGG16, InceptionV3, and EfficientNetV2S—to identify the most suitable architecture for Alzheimer's disease detection.

#### **Impact of learning rate adjustments**

We analyze the influence of learning rate modifications on

model performance, underscoring its critical role in optimizing training processes and achieving superior results.

Our findings demonstrate that the proposed approach significantly outperforms other models in terms of accuracy and F1 score, especially in distinguishing between the four classes of Alzheimer's disease severity: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. This study underscores the importance of integrating advanced deep learning models with careful experimental analysis, contributing to the growing potential of AI in the automated diagnosis of Alzheimer's disease.

The structure of this paper is organized as follows: **Section 2** reviews related works, focusing on machine learning, deep learning, and transfer learning methods for Alzheimer's disease (AD) detection. **Section 3** describes the dataset and preprocessing steps and outlines the proposed framework, highlighting the use of EfficientNetV2S. **Section 4** presents the experimental results, showcasing the performance comparison of seven pretrained models, including EfficientNetV2S, and analyzing the impact of learning rates on model performance. **Section 6** concludes the paper by summarizing the contributions and providing directions for future research.

## **Related Works**

Alzheimer's disease (AD) detection using MRI images has become a crucial area of research, given the increasing importance of early diagnosis in managing the disease. Over the years, various machine learning (ML) and deep learning (DL) techniques have been employed to identify and classify brain abnormalities indicative of Alzheimer's. This section reviews key advancements in these approaches, starting with traditional machine learning methods, followed by deep learning techniques, and finally, the utilization of transfer learning with pretrained models.

### **Machine learning techniques for Alzheimer detection**

Classical machine learning techniques have been widely used for Alzheimer's Disease (AD) detection from MRI images. Methods such as Support Vector Machines (SVM), Decision Trees (DT), and Random Forest (RF) have been applied to classify AD and Normal Control (NC) images. For example, Vaithinathan K employed Random Forest, KNN, and linear SVM, achieving 89.58% sensitivity and 85.82% specificity using the ADNI dataset [1]. Similarly, Kavitha *et al.* applied decision trees, random forests, SVM, and XGBoost, achieving an accuracy of 83% for early AD detection [2]. These traditional machine learning methods often rely on manually engineered features and are effective when the dataset is well-structured and the features are properly selected, such as through correlation or information gain techniques. Additionally, advanced models like XGBoost, when combined with feature extraction techniques such as Discrete Wavelet Transform (DWT), have demonstrated even higher performance, with one study achieving an accuracy of 97.88% [3].

Recent work, such as that by Rao *et al.*, highlights the application of 3D MRI technology in AD detection, where 2D slices of white and grey matter are taken from coronal, sagittal, and axial orientations, followed by feature extraction using Multi-Layer Perceptron (MLP) and SVM classifiers [4]. This

approach, evaluated through Precision, Recall, Accuracy, and F1-Score, emphasizes the critical role of machine learning in providing high-accuracy, early-stage AD diagnosis, which has the potential to alleviate the extensive healthcare burdens of this condition. Despite the promising outcomes, traditional machine learning methods have limitations. They heavily depend on the expertise of the researcher to select relevant features and are often unable to capture the complex relationships within the data, which is particularly important in medical imaging tasks like Alzheimer's detection.

### **Deep learning techniques for Alzheimer detection**

On the other hand, neural networks, especially Convolutional Neural Networks (CNN), have gained significant attention for their ability to automatically learn complex features from MRI images. Several studies have leveraged CNNs for AD detection, including Weimingling *et al.*, who achieved an accuracy of 81.4% and an AUC of 87.8% by using an Extreme Learning Machine (ELM) for classification after extracting features from patch images [5]. Other studies, like those by Abolbahr *et al.* [6] and Amir Ebrahimi *et al.* [7], used deep neural networks (DNN) and combinations of 2D/3D CNNs with Recurrent Neural Networks (RNN) to classify AD with accuracies above 90%. A further CNN-based framework achieved 99.6% accuracy for binary AD classification and 97.5% for multi-class classification on the ADNI dataset, highlighting deep learning's potential to improve AD diagnosis significantly [8].

Further advancements include a VGG-16-based approach, where the images were preprocessed by converting 3D to 2D, resizing, and then passed through VGG-16 for feature extraction, followed by classification using various methods like SVM, Linear Discriminant Analysis, and K-means clustering. This method achieved a remarkable 99.95% accuracy on fMRI datasets and 73.46% on PET datasets, demonstrating the advantages of using CNNs alongside traditional classifiers [9]. Recently, El-Assy *et al.* introduced a dual-CNN architecture, achieving over 99% accuracy across multiple AD categories by combining distinct CNN models to capture both local and global MRI features [10]. These approaches have shown strong classification accuracy and robustness compared to classical methods, enhancing the effectiveness of early AD detection. However, deep learning techniques require large amounts of annotated data and substantial computational resources for training, which can be a limiting factor in clinical settings where annotated MRI datasets are often limited.

### **Transfer learning for Alzheimer detection**

To overcome the limitations of large annotated datasets and high computational demands, transfer learning using pretrained models has emerged as a powerful solution. Transfer learning allows models to leverage knowledge learned from large-scale datasets in one domain and apply it to a different, often smaller, dataset. This approach has been particularly beneficial for Alzheimer's detection using MRI images, where pretrained models on general image datasets such as ImageNet have been fine-tuned for specific medical imaging tasks. The use of deep learning in medical imaging has gained momentum, with MRI playing a crucial role in Alzheimer's disease (AD) diagnosis—a progressive neurodegenerative disorder affecting memory and

cognitive function. MRI-based studies classify early AD stages, including cognitively normal (CN), mild cognitive impairment (MCI), and AD, using CNN-based transfer learning models. In one study, 2,182 MRI images from the ADNI database were processed using various CNN architectures, achieving top results with EfficientNet models, particularly EfficientNetB0, which reached a 92.98% accuracy rate [11]. EfficientNetB3 further excelled in precision, sensitivity, and specificity. Another study using EfficientNet-b0, combined with both end-to-end and transfer learning, achieved up to 95.29% accuracy for classifying stable mild cognitive impairment (sMCI) versus AD and 87.38% accuracy for multiclass AD stages classification [12]. Different studies have employed additional pre-trained networks like AlexNet, ResNet-18, and GoogleNet on ADNI datasets, with classification accuracies between 94% and 97.5% [13]. The AlexNet model demonstrated particularly high sensitivity (100%) and specificity (98.21%), making it promising for computer-aided diagnostics (CAD) in AD detection. Beyond MRI, combining data from multiple imaging sources, such as fused CT-MRI and EEG signals, has been explored with the HEMRDTL model, using VGG-19 and robust principal component analysis (RPCA) to enhance accuracy [14]. This hybrid model showed notable effectiveness, underscoring the benefit of integrating structural and functional brain data for AD detection. Some studies have also investigated the use of lightweight neural networks like MobileNet for mobile and resource-constrained environments, which yielded favorable results with over 96% accuracy in multi-class AD classification [15,10]. This model's minimal computational demand offers practical benefits for mobile diagnostics, potentially expanding accessibility in clinical settings. Other studies highlighted Xception and other CNN architectures for multi-class MRI classification, with the Xception model achieving an accuracy of 99.6%, demonstrating the promise of deep learning and transfer learning for scalable, non-invasive AD screening [16]. Additionally, a transfer learning-based approach using MRI scans from the ADNI database was proposed to classify AD stages, including normal control (NC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and AD. This method involves extracting gray matter and fine-tuning a pre-trained VGG model with a stepwise freezing strategy, demonstrating superior classification performance [17]. A similar deep learning framework utilizing convolutional neural networks (CNN) for AD classification was developed, incorporating pre-processing, data augmentation, cross-validation, and feature extraction [18]. Two methods were explored: a simple CNN model and a fine-tuned VGG16 model with transfer learning on various datasets. The framework achieved significant performance gains with minimal labeled data and prior domain knowledge, with the CNN model achieving 99.95% accuracy and the fine-tuned VGG16 model achieving 97.44% accuracy, showing promising results with low computational complexity and minimal overfitting.

Moreover, a study used MRI data from Kaggle, including non-dementia, very mild dementia, mild dementia, and moderate dementia categories. Three pre-trained networks—VGG-19, ResNet-50, and InceptionV3—were assessed, achieving classification accuracies of 92.86%, 85.99%, and 91.04%, respectively, demonstrating the efficacy of transfer learning for AD classification [19]. While existing methods have made notable

strides in Alzheimer's disease classification, challenges such as class imbalance, limited data availability, and the need for computational efficiency persist. Traditional ML approaches often depend heavily on manual feature engineering, while CNN-based methods, despite their high performance, require extensive labeled datasets and computational resources. Transfer learning has addressed some of these issues, yet most studies lack robust mechanisms to optimize model scalability. Our approach leverages an advanced deep learning architecture, integrating EfficientNetV2S for Alzheimer's severity classification, providing a more efficient and scalable solution compared to prior works.

## Proposed Methodology

This section presents a transfer learning approach using EfficientNetV2S for Alzheimer's stage classification. As illustrated in **Figure 1**, training images are first augmented and normalized to balance the dataset. The preprocessed images are then used to fine-tune the pretrained EfficientNetV2S model, followed by additional CNN layers for task-specific adaptation. Finally, the trained model is evaluated on test images to assess classification performance across Alzheimer's stages. This approach effectively leverages pretrained features and data augmentation to enhance model accuracy and robustness.

### Dataset description and augmentation

The dataset<sup>1</sup> used in this study is an augmented version of the original Kaggle Alzheimer's dataset<sup>2</sup>, which includes images from four diagnostic categories: "No Impairment," "Very Mild Impairment," "Mild Impairment," and "Moderate Impairment." The original dataset exhibited a significant class imbalance, with class distributions of 3,200 samples for "No Impairment," 2,240 for "Very Mild Impairment," 896 for "Mild Impairment," and only 64 for "Moderate Impairment." This disparity often led to classifiers being biased toward the majority class, which is particularly problematic in early Alzheimer's detection where false negatives can be critical. To address this issue, a data augmentation approach employing Wasserstein Generative Adversarial Networks with Gradient Penalty (WGANs-GP) was

implemented. WGANs-GP effectively mitigates mode collapse, a common problem in traditional DC-GANs, by generating synthetic MRI images for the minority classes, thereby enhancing diversity and rectifying the class imbalance. After augmentation, each class was brought to a balanced distribution, with all classes now containing 2,560 samples. This resulting dataset, which comprises a mix of real and synthetic images, significantly improved model performance, particularly in recognizing minority classes.

<sup>1</sup><https://www.kaggle.com/datasets/lukechugh/best-alzheimer-mri-dataset-99-accuracy>

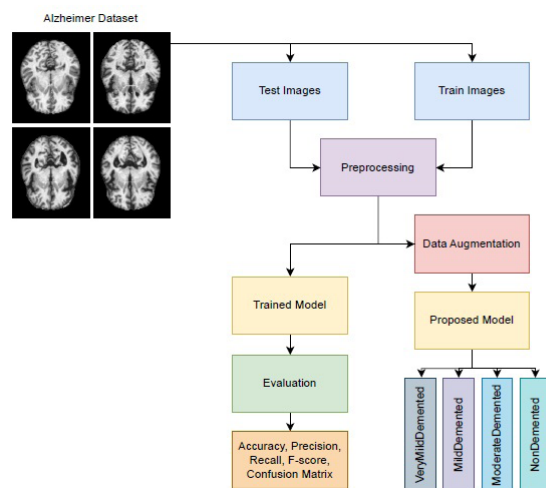
<sup>2</sup><https://www.kaggle.com/datasets/marcopinamonti/alzheimer-mri-4-classes-dataset>

### Dataset pre-processing

The pre-processing pipeline in this study aimed to standardize and enhance the quality of MRI images to optimize model performance for Alzheimer's detection. Each image was processed in grayscale format to reduce computational complexity, with the following sequential steps applied to each image in the dataset:

**Noise reduction:** Non-Local Means (NLM) filtering was employed for noise reduction in MRI images. NLM is a powerful denoising technique that works by averaging similar patches in the image, regardless of their spatial proximity, to preserve fine details and structures while reducing noise. Unlike traditional filters, such as the Bilateral, Median, and Gaussian filters, which rely on local neighborhood information, NLM takes into account the global image content, making it particularly effective in preserving image textures and structures.

The key advantage of NLM over other filters lies in its ability to handle noise without introducing blurring or losing important details, which is crucial in medical imaging where fine structures, such as anatomical boundaries, need to be preserved. In this study, the NLM filter was applied with optimized parameters based on empirical evaluations.



**Figure 1.** Proposed workflow diagram for Alzheimer's stage classification.



**Bilateral filter:** The Bilateral filter smooths images while preserving edges by considering both spatial proximity and intensity similarity. However, it may struggle with complex textures or high levels of noise, often leading to blurring at edges.

**Median filter:** The Median filter is widely used for noise removal by replacing each pixel's value with the median of the intensities within a defined neighborhood. It is particularly effective at preserving edges and removing outlier noise, but it may not perform well for large or dense noise patterns, potentially leading to the loss of fine details in the image.

**Gaussian filter:** The Gaussian filter is a linear smoothing technique that applies a Gaussian function to average pixel intensities over a region. It is effective in reducing random variations and smoothing the image but tends to blur edges and reduce the visibility of fine structures, which can impact the clarity of critical features in applications such as medical imaging. Theoretically, NLM outperforms these filters for medical images due to its ability to preserve intricate structural details while efficiently removing noise. This makes it particularly suited for applications like MRI, where maintaining the clarity of anatomical features is critical for accurate diagnosis and analysis. In the results section, we demonstrate the superiority of the NLM filter through visual comparisons and numerical evaluations against the other filtering techniques.

**Resizing:** Each image was resized to a target dimension of 128x128 pixels to achieve a consistent input size across the dataset, ensuring compatibility with the model architecture.

**Normalization:** Pixel values were normalized to a range of [0,1], facilitating faster convergence during training by stabilizing the input distribution and aiding in generalization.

This pre-processing approach effectively standardized image quality, structure, and scale across the dataset, providing the model with optimized inputs for training.

#### EfficientNet model series

The EfficientNet model series, introduced by Tan and Le (2019), marked a significant advancement in the design of convolutional neural networks (CNNs) by optimizing both efficiency and accuracy across various deep learning tasks. Traditional CNN architectures relied on manual tuning of network depth, width, and resolution, but EfficientNet introduced a systematic approach to scaling these dimensions through compound scaling. This innovation enabled EfficientNet to outperform prior CNN models on benchmarks such as ImageNet while using fewer parameters and FLOPs (floating-point operations per second).

**EfficientNetV2S:** EfficientNetV2 is a more recent evolution of the EfficientNet architecture, designed to address efficiency limitations. The EfficientNetV2S model achieves both speed and accuracy by blending fused-MBConv layers in the early stages and MBConv layers in deeper stages. Progressive learning, a strategy where input image resolution is gradually increased, further improves robustness to varying scales, leading to better generalization.

**Transfer learning with efficientNetV2S:** In this study, transfer learning was employed using EfficientNetV2S as the

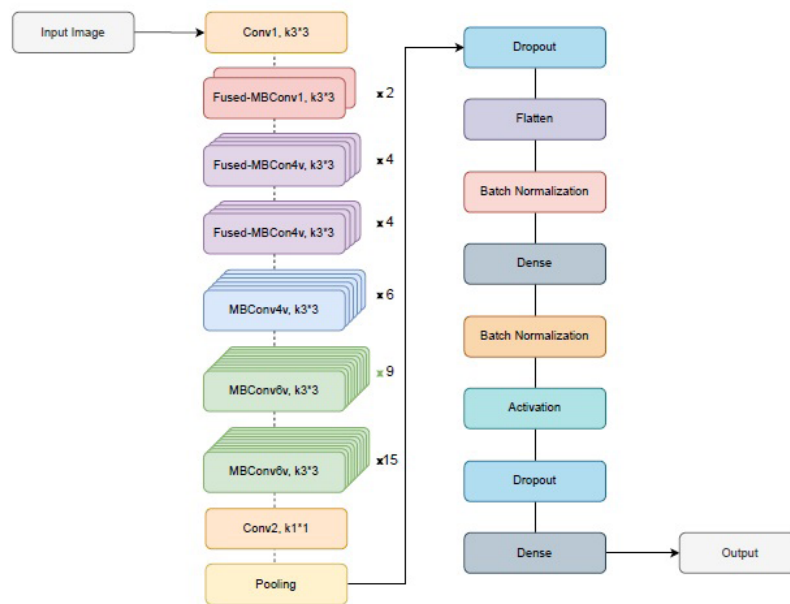
backbone model to classify Alzheimer's disease into four groups. The EfficientNetV2S model was pre-trained on the ImageNet dataset, providing a robust feature extraction framework learned from millions of images. To adapt the model for the Alzheimer's classification task, its top classification layers were excluded. By removing these task-specific layers, the model retained its ability to extract general-purpose features while allowing the addition of custom layers tailored to the new task. The modified EfficientNetV2S model was initialized with an input shape of (130, 130, 3) to match the resolution of the dataset. The output from the feature extraction layers was passed through a Dropout layer (0.5) to prevent overfitting, followed by a Flatten layer that converted the feature maps into a one-dimensional vector. To improve training stability, Batch Normalization was applied to the flattened features. Subsequently, a Dense layer with 128 units, initialized with the He uniform initializer, was added to optimize weight initialization. This layer was batch-normalized and activated using ReLU for efficient learning. An additional Dropout layer (0.5) was incorporated to further reduce overfitting by minimizing neuron dependencies. Finally, the model included a Dense layer with 4 output neurons, corresponding to the four Alzheimer's disease categories, with a softmax activation function to output probabilities for each class.

The proposed model is illustrated in **Figure 2**, highlighting the integration of EfficientNetV2S's feature extraction capabilities with task-specific layers for Alzheimer's group classification. By excluding the original classification layers and introducing customized components, the model was effectively fine-tuned to achieve high accuracy on this specialized task. The input layers, output shape, and parameters of the proposed ensemble model are presented in **Table 1**.

## Results and Discussions

In this section, we present the results of the experiments conducted to evaluate the performance of various pretrained models for Alzheimer's disease detection, with a particular focus on the EfficientNetV2S model under different configurations. All experiments were conducted on an NVIDIA Tesla P100 GPU, utilizing the Keras library in Python for model development and training. The choice of the P100 GPU was made to accelerate the training process and efficiently handle the computational demands of deep learning models.

However, we acknowledge that such high-performance hardware may not always be available in practical scenarios. Therefore, to assess the broader applicability of the models, it is important to consider their computational requirements on less powerful hardware. Lightweight architectures such as MobileNet and EfficientNetB0 are optimized for efficiency and can achieve reasonable performance on consumer-grade GPUs or even CPUs, making them suitable for deployment in resource-constrained environments. On the other hand, deeper models like VGG16 and ResNet, while offering high accuracy, may require substantial computational resources, leading to increased inference time and memory consumption on lower-end devices. The dataset was divided into distinct training and test sets, where the test data was separated based on different datasets to ensure an unbiased evaluation. From the remaining training data, an 80-20 split was applied, allocating 80% for model training and 20% for



**Figure 2.** Proposed method architecture with EfficientNetV2S backbone.

**Table 1.** Total layers, output shape, and parameters of the proposed model.

Layers	Output Shape	Parameters
Input Layers	(300, 300, 3)	0
EfficientNetV2S	(10, 10, 1280)	20,331,360
Dropout	(10, 10, 1280)	0
Flatten	128,000	0
BatchNormalization	128,000	512,000
Dense	128	16,384,128
BatchNormalization	128	512
Activation	128	0
Dropout	128	0
Dense	4	516

validation purposes. Additionally, we employed cross-validation to further assess the robustness and generalization capability of the models.

The results presented here highlight the efficiency and effectiveness of the tested models, comparing them based on key metrics such as accuracy, F1-score, precision, and recall across different classes. Each subsection discusses the outcomes of individual models, beginning with a comparison of EfficientNetV2S against other well-known models, followed by a detailed examination of its performance under various learning rates, and concluding with an in- depth analysis of its training dynamics and classification performance.

#### Evaluation metrics

To evaluate the performance of the models, we use several widely accepted metrics, each of which provides insight into different aspects

of the classification results. These metrics are crucial for Alzheimer's disease detection, as they account for class imbalance and the clinical significance of accurate diagnosis:

**Accuracy:** The overall proportion of correctly classified samples out of the total number of samples. While accuracy gives a general idea of model performance, it may not be sufficient in the presence of class imbalance, as a model that predicts the majority class correctly can still have high accuracy while failing to identify the minority class effectively. It is calculated as:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}} \quad (1)$$

**Precision:** The proportion of positive predictions that are actually correct, indicating the accuracy of positive predictions. High precision is critical in medical tasks to avoid false positives, which could lead to unnecessary treatments or interventions. It is calculated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2)$$

**Recall:** The proportion of actual positives that are correctly identified by the model, showing how well the model identifies positive instances. In the context of Alzheimer's detection, high recall ensures that the model does not miss any positive cases, which is crucial for timely diagnosis and intervention. It is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3)$$

**F1-score:** The harmonic mean of precision and recall, providing a balance between the two metrics. It is particularly useful in situations where there is an imbalance between the positive and negative classes. The F1-score is a critical measure in medical diagnosis, as it accounts for both false positives and false negatives, ensuring that both precision and recall are optimized. It is calculated as:

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

These metrics are used to evaluate and compare the performance of the models in terms of their ability to classify Alzheimer's disease correctly, ensuring a comprehensive assessment of both individual class performance and overall model efficacy. The choice of these metrics is especially relevant in Alzheimer's detection tasks, where both false positives and false negatives can have significant clinical consequences. Therefore, precision, recall, and F1-score provide a more nuanced understanding of model performance beyond simple accuracy, making them particularly valuable for evaluating deep learning models in healthcare applications.

### Pretrained models

To evaluate the performance of the proposed Alzheimer's detection model, the pretrained EfficientNetV2S model is compared with several other well-known and widely used deep learning architectures. These models include VGG16, ResNet, MobileNet, InceptionResNet, InceptionV3, and EfficientNetB0. The selection of these models was based on their demonstrated effectiveness in medical imaging tasks, widespread adoption in the deep learning community, and their diverse architectural characteristics that provide different approaches to feature extraction and classification. VGG16 is known for its deep yet simple structure, consisting of uniform convolutional layers that have proven effective in image classification tasks. ResNet incorporates residual connections to tackle the vanishing gradient problem, enabling the training of very deep networks without performance degradation. MobileNet is a lightweight model designed for efficient processing on mobile and embedded devices, utilizing depthwise separable convolutions to reduce computational cost. InceptionResNet and InceptionV3 combine inception modules with residual connections to optimize accuracy and efficiency. EfficientNetB0, a base-line variant of the EfficientNet family, balances depth, width, and resolution using compound scaling to achieve high accuracy with minimal computational overhead. EfficientNetV2S, the focus of this study, builds upon the EfficientNet framework by introducing improved training strategies and computational optimizations. Its design emphasizes both accuracy and efficiency, making it a strong candidate for Alzheimer's detection. By comparing EfficientNetV2S with these established models, we aim to showcase

its advantages in terms of performance and computational efficiency for medical image analysis. Below, we provide a brief overview of each model's architecture and capabilities.

**VGG16:** VGG16 is a deep convolutional neural network characterized by its simple architecture, consisting of 16 layers with 3x3 convolutional filters and max-pooling layers [20]. Its simplicity and uniformity in design enable it to effectively capture hierarchical features from input images, making it suitable for complex tasks like Alzheimer's detection. Although VGG16 has a large number of parameters, it remains a baseline for image classification tasks due to its strong performance on benchmark datasets.

**ResNet:** ResNet, or Residual Networks, introduced the concept of residual connections to address the vanishing gradient problem, enabling the training of very deep networks [21]. The model's ability to learn residual mappings rather than direct mappings allows it to capture more complex features and improves performance in image classification tasks. ResNet is particularly effective in scenarios that require deep networks for fine-grained classification, such as Alzheimer's detection.

**MobileNet:** MobileNet is optimized for mobile and embedded devices by utilizing depthwise separable convolutions, which reduce computational cost while maintaining high accuracy [22]. The model is lightweight, making it ideal for resource-constrained environments, such as real-time medical imaging systems. Its efficiency and compactness make it a suitable alternative for Alzheimer's detection, particularly in scenarios with limited computational resources.

**InceptionResNet:** InceptionResNet combines the strengths of both the Inception and ResNet architectures by integrating residual connections into the Inception model [23]. This hybrid approach enhances feature extraction across multiple scales while mitigating issues such as vanishing gradients. Inception ResNet's ability to capture multi-level features makes it particularly effective for complex tasks, including Alzheimer's detection.

**InceptionV3:** InceptionV3 is an advanced version of the Inception model that introduces factorized convolutions and other optimizations to reduce computational complexity while maintaining high accuracy [24]. It is well-suited for processing large datasets and complex image classification tasks, making it an ideal choice for medical image analysis, such as Alzheimer's detection.

**EfficientNetB0:** EfficientNetB0 employs a compound scaling method that simultaneously scales the depth, width, and resolution of the model to achieve superior performance with fewer parameters than traditional architectures [25]. Its high efficiency in terms of both accuracy and computational cost makes it particularly suitable for large-scale image classification tasks, including Alzheimer's detection.

In the following sections, we will compare the performance of the EfficientNetV2S model with these pretrained models to evaluate their respective effectiveness in Alzheimer's detection. Each model is evaluated based on accuracy, computational efficiency, and their ability to generalize across various medical imaging datasets.

## Comparison of Denoising Methods

MRI images are often affected by various types of noise introduced during the scanning process, including salt and pepper noise, speckle noise, and random noise, as illustrated in **Figure 3**. Effective noise removal is crucial to ensure accurate diagnosis while preserving essential anatomical structures. In this study, the denoising process addresses these three types of noise by applying four different filters to the corrupted images, aiming to remove noise without degrading the original image content, as shown in **Figure 3**. To evaluate the effectiveness of the proposed denoising methodology, the performance of each filter is assessed based on visual quality and quantitative metrics. The Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE) are computed to measure the effectiveness of noise reduction while maintaining image fidelity. The Mean Squared Error (MSE) is a widely used metric to measure the average squared difference between the original image  $I_{\text{original}}(x, y)$  and the denoised image  $I_{\text{denoised}}(x, y)$ . It is calculated as follows:

$$MSE = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [I_{\text{original}}(x, y) - I_{\text{denoised}}(x, y)]^2$$

where  $M$  and  $N$  represent the dimensions of the image. A lower MSE value indicates better denoising performance, as it means the denoised image is closer to the original.

The Peak Signal-to-Noise Ratio (PSNR) is another important metric used to assess image quality. It is expressed in decibels (dB) and is derived from the MSE as follows:

$$PSNR = 10 \cdot \log_{10} \frac{(MAX)^2}{MSE}$$

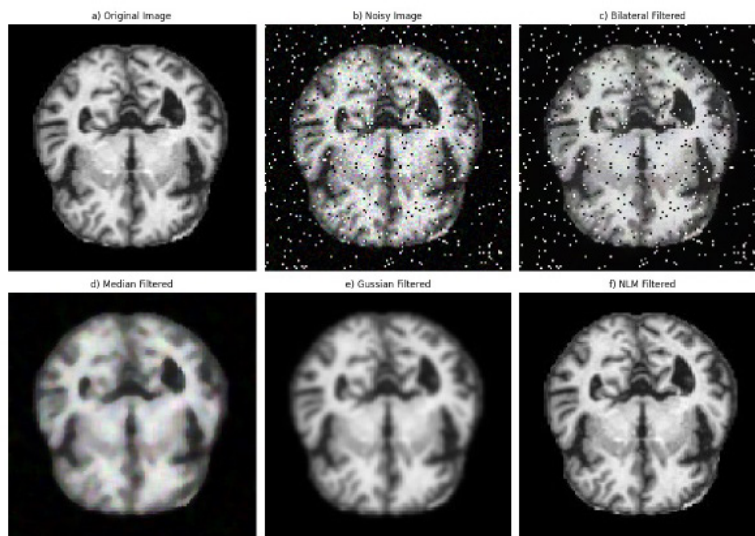
where  $MAX$  is the maximum possible pixel value in the image (e.g., 255 for an 8-bit grayscale image). A higher PSNR

value indicates better image quality, as it signifies a lower level of distortion introduced by the denoising process.

**Table 2** presents a comparative analysis of image quality using these metrics for the applied filters: 2D Median Filter, 2D Bilateral Filter, 2D Gaussian Filter, and 2D Non-Local Means (NLM) Filter. The results demonstrate that the 2D NLM filter outperforms the other methods, achieving the highest PSNR of 55.29 dB and the lowest MSE of 0.19. This indicates its superior capability in reducing noise while preserving fine structural details. The Gaussian filter also shows reasonable performance, whereas the bilateral and median filters, while effective to some extent, introduce more residual noise and less detail preservation. These findings support the selection of the NLM filter as the optimal choice for MRI image denoising.

## Results of pre-trained models

The comparison results for all models, sorted by accuracy, are presented in **Table 3**. This organized presentation allows for a clearer understanding of each model's strengths and limitations. The evaluation results highlight the clear outperformance of EfficientNetV2S, which achieved the highest accuracy of 0.981 and a macro average F1-score of 0.986. This model demonstrates a strong balance across all classes, with exceptional recall for Class 2 (0.987) and Class 3 (0.975), coupled with high precision values, making it the most reliable model for this classification task. In comparison, the other models performed well but did not reach the same level of accuracy or macro average F1-score. EfficientNetB0 achieved the second-highest accuracy of 0.960 and a macro F1-score of 0.967, followed by MobileNet and ResNet, with accuracies of 0.948 and 0.947, respectively. InceptionResNet showed an accuracy of 0.936 and performed strongly in Class 2 but faced challenges with Class 0. VGG16 and InceptionV3 had the lowest accuracies at 0.905 and 0.901, respectively, with both models showing limitations in certain class-specific recall scores.



**Figure 3.** Visual comparison of denoising methods: (a) Original MRI image, (b) Noisy image, (c) Bilateral Filter, (d) Median Filter, (e) Gaussian Filter, (f) NLM Filter.



**Table 2.** Image quality measurements using PSNR and MSE for various filters.

Filter	PSNR (dB)	MSE
2D Bilateral Filter	16.85	1342.67
2D Median Filter	24.53	229.10
2D Gaussian Filter	28.26	97.06
2D NLM Filter	55.29	0.19

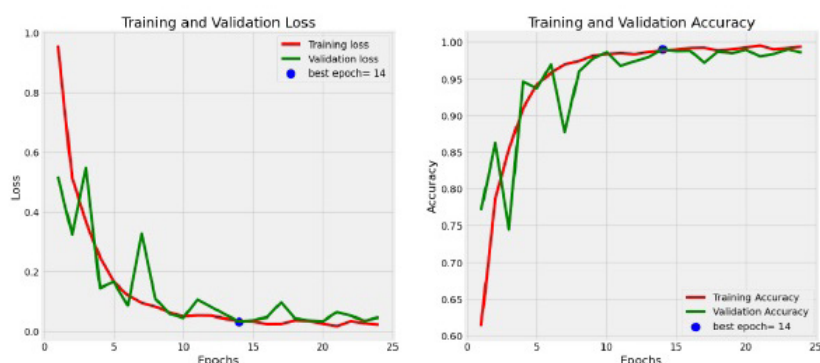
**Table 3.** Performance of different pre-trained models.

Model	Metric	Class 0	Class 1	Class 2	Class 3	Macro Avg
<b>InceptionV3</b>	F1-Score	0.930	1.000	0.921	0.849	0.925
	Precision	0.896	1.000	0.865	0.974	0.934
	Recall	0.966	1.000	0.984	0.752	0.926
	Accuracy					0.901
<b>VGG16</b>	F1-Score	0.875	0.957	0.929	0.880	0.910
	Precision	0.96	1.000	0.901	0.890	0.938
	Recall	0.804	0.917	0.958	0.871	0.887
	Accuracy					0.905
<b>InceptionResNet</b>	F1-Score	0.918	1.000	0.947	0.927	0.948
	Precision	0.926	1.000	0.995	0.871	0.948
	Recall	0.911	1.000	0.903	0.991	0.951
	Accuracy					0.936
<b>ResNet</b>	F1-Score	0.920	1.000	0.959	0.940	0.955
	Precision	0.936	1.000	0.945	0.954	0.959
	Recall	0.905	1.000	0.973	0.926	0.951
	Accuracy					0.947
<b>MobileNet</b>	F1-Score	0.907	1.000	0.967	0.938	0.953
	Precision	0.842	1.000	0.972	0.965	0.945
	Recall	0.983	1.000	0.963	0.913	0.965
	Accuracy					0.948
<b>EfficientNetB0</b>	F1-Score	0.948	1.000	0.969	0.951	0.967
	Precision	0.921	1.000	0.964	0.970	0.964
	Recall	0.978	1.000	0.973	0.933	0.971
	Accuracy					0.960
<b>EfficientNetV2S</b>	F1-Score	0.986	1.000	0.982	0.977	<b>0.986</b>
	Precision	1.000	1.000	0.976	0.979	<b>0.989</b>
	Recall	0.972	1.000	0.987	0.975	<b>0.983</b>
	Accuracy					<b>0.981</b>

### Results of efficientNetV2S model

This section examines the performance of the EfficientNetV2S model in detail, focusing on its training and validation progression and classification accuracy across different classes. The EfficientNet model's training dynamics are depicted in **Figure 4**. Notably, the validation curves experience an initial spike in the early epochs, suggesting rapid initial learning, after which they gradually converge. This behavior indicates that the model quickly learns generalizable

features in the early stages. As the training progresses, the loss curves (both training and validation) show a steady decline, with the validation loss stabilizing towards the later epochs. This suggests effective learning without significant overfitting. Moreover, the model's convergence speed is relatively fast, as evidenced by the rapid decline in loss early on, followed by a slower, more stable reduction in later epochs. There are no significant signs of underfitting or over-fitting, as the training and validation



**Figure 4.** Training and validation loss and accuracy curves for EfficientNetV2S model.

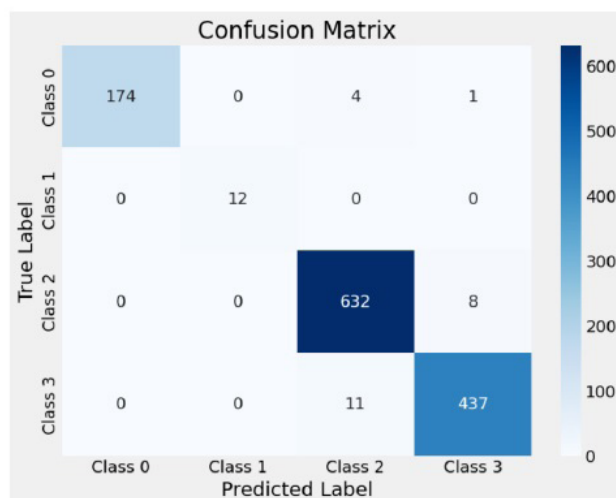
performance align closely over time, indicating that the model is able to generalize well to unseen data. The absence of significant gaps between the training and validation metrics further supports the model's robustness and its ability to avoid overfitting during training. The confusion matrix in **Figure 5** further highlights the model's classification effectiveness across all classes. Based on the confusion matrix, we observe that the model performs very well on most classes, with a few exceptions. Class 0 (healthy) is correctly identified with high accuracy, showing 174 true positives, but there are small misclassifications with 4 instances incorrectly predicted as class 2 and 1 instance as class 3. For Class 1 (mild cognitive impairment), the model struggles more, with 12 true positives and no instances misclassified into other classes, indicating that the model might find it more challenging to distinguish mild cognitive impairment from other stages. Class 2 (Alzheimer's) is identified with a high degree of accuracy, with only 8 instances misclassified as class 3, which is a relatively low number compared to the total of 632 true positives. Similarly, Class 3 (severe Alzheimer's) is also well recognized, with 437 true positives and 11 instances incorrectly classified as class 2. The difficulties observed with class 1 and the

small misclassifications between classes 2 and 3 could stem from the overlap in symptoms or features between the different stages of Alzheimer's, which might make these stages harder for the model to differentiate clearly.

#### Results of efficientNetV2S at different learning rates

**Table 4** presents the performance metrics of the EfficientNetV2S model across three learning rates: 0.01, 0.001, and 0.0001. The metrics include F1- score, precision, recall for each class, macro averages, and overall accuracy.

At a learning rate of 0.01, the model achieved a respectable accuracy of 0.940, with Class 1 attaining perfect precision and recall scores. However, performance in Classes 0, 2, and 3 was slightly lower, resulting in a macro average F1-score of 0.956. Reducing the learning rate to 0.001 yielded improved accuracy at 0.977, alongside enhancements in macro average F1- score and precision, indicating the model's increased stability and consistency across classes. The optimal performance occurred at a learning rate of 0.0001, where the model achieved the highest accuracy of



**Figure 5.** Confusion matrix for EfficientNetV2S model classification results.

**Table 4.** Performance of EfficientNetV2S at different learning rates.

Learning Rate	Metric	Class 0	Class 1	Class 2	Class 3	Macro Avg	Accuracy
0.01	F1-Score	0.951	1.000	0.947	0.924	0.956	0.940
	Precision	0.976	1.000	0.961	0.897	0.959	
	Recall	0.927	1.000	0.933	0.953	0.953	
0.001	F1-Score	0.986	1.000	0.980	0.967	0.983	0.977
	Precision	0.989	1.000	0.975	0.973	0.984	
	Recall	0.983	1.000	0.984	0.962	0.982	
0.0001	F1-Score	0.986	1.000	0.982	0.978	0.986	0.981
	Precision	1.000	1.000	0.977	0.980	0.989	
	Recall	0.972	1.000	0.988	0.975	0.984	

0.981, with consistently high F1- scores across all classes. This setting also resulted in the best macro average precision (0.989) and F1-score (0.986), demonstrating that a lower learning rate allowed the model to converge more effectively and achieve superior performance across all evaluation metrics. This section comprehensively evaluates the effectiveness of various pre- trained deep learning models, focusing particularly on the EfficientNetV2S model for Alzheimer's disease detection. Metrics such as accuracy, precision, recall, and F1-score are used to benchmark model performance, offering insights into both overall efficiency and class-specific outcomes. EfficientNetV2S emerges as the best-performing model, achieving the highest accuracy (0.981) and macro F1-score (0.986). Its strong precision-recall balance across all classes underlines its robustness for this medical classification task.

The comparative analysis highlights the relative strengths and weaknesses of other models like ResNet, MobileNet, and EfficientNetB0, which show commendable but comparatively lower performance, while simpler architectures like VGG16 and InceptionV3 lag behind in accuracy and class-specific metrics. The detailed results emphasize the efficiency of EfficientNetV2S in feature extraction and generalization. This study not only demonstrates the superiority of EfficientNetV2S but also showcases its potential applicability in real-world Alzheimer's detection scenarios, setting a high benchmark for future research in medical imaging classification.

## Conclusions

This study presents a novel approach to Alzheimer's stage classification using the EfficientNetV2S model with transfer learning. The model achieved impressive performance, with an overall accuracy of 98.1%, a macro-average F1-score of 98.6%, and precision and recall values of 98.9% and 98.3%, respectively. These results highlight the ability of the EfficientNetV2S architecture to handle the complexities of Alzheimer's stage prediction effectively. Moreover, the model outperformed well-known architectures such as VGG16, ResNet, MobileNet, and EfficientNetB0, with consistent improvements in classification accuracy across all stages of the disease. The results confirm the strength of using advanced deep learning architectures in medical imaging tasks. Future studies could investigate integrating multimodal data, such as clinical history and biomarkers, to complement MRI-based predictions. Additionally, longitudinal analysis of MRI scans to track disease progression and refine staging predictions offers a

promising direction. Exploring cross-domain applications of the EfficientNetV2S model for other neurodegenerative diseases or medical imaging challenges could further validate its versatility and effectiveness.

## References

1. Vaithinathan K. Random forest, knn, and svm for ad classification using adni images. In: Proceedings of the International Conference on Artificial Intelligence. 2019, pp. 1101–6.
2. Kavitha C, Mani V, Srividhya SR, Khalaf OI and Tavera Romero CA. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Front. Public Health. 2022;10:853294.
3. George CM, Menon S. Machine learning for alzheimer detection: A comprehensive approach. J Theor Appl Inf Technol. 2024 Feb 29;102(4):2024.
4. Rao KN, Gandhi BR, Rao MV, Javvadi S, Vellela SS, Basha SK. Prediction and classification of Alzheimer's disease using machine learning techniques in 3D MR images. In: 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS) 2023 Jun 14 pp. 85–90
5. Lin W, Gao Q, Yuan J, Chen Z, Feng C, Chen W, Du M, Tong T. Predicting Alzheimer's Disease Conversion From Mild Cognitive Impairment Using an Extreme Learning Machine-Based Grading Method With Multimodal Data. Front Aging Neurosci. 2020 Apr 1;12:77.
6. Abolbaher et al. CNN and DNN models for ad classification from structural MRI images. Journal of Medical Imaging. 2021;8(3):045003.
7. Ebrahimi A, Luo S; Alzheimer's Disease Neuroimaging Initiative. Convolutional neural networks for Alzheimer's disease detection on MRI images. J Med Imaging (Bellingham). 2021 Mar;8(2):024503.
8. AbdulAzeem Y, Bahgat WM, Badawy M. A CNN based framework for classification of Alzheimer's disease. Neural Comput Appl. 2021 Aug;33(16):10415–28.
9. Janghel RR, Rathore YK. Deep convolution neural network based system for early diagnosis of Alzheimer's disease. Irbm. 2021 Aug 1;42(4):258–67.
10. Mohiuddin dar G, Bhagat A, Ansarullah SI, Othman MT, Hamid Y, Alkahtani HK, et al. A novel framework for classification of different Alzheimer's disease stages using CNN model. Electronics. 2023 Jan 16;12(2):469.
11. Savaş S. Detecting the stages of Alzheimer's disease with

- pre-trained deep learning architectures. *Arab J Sci Eng*. 2022 Feb;47(2):2201–18.
12. Agarwal D, Berbís MÁ, Luna A, Lipari V, Ballester JB, de la Torre-Díez I. Automated Medical Diagnosis of Alzheimer's Disease Using an Efficient Net Convolutional Neural Network. *J Med Syst*. 2023 May 2;47(1):57.
  13. Shanmugam JV, Duraisamy B, Simon BC, Bhaskaran P. Alzheimer's disease classification using pre-trained deep networks. *Biomedical Signal Processing and Control*. 2022 Jan 1;71:103217.
  14. Leela M, Helenprabha K, Sharmila L. Prediction and classification of Alzheimer disease categories using integrated deep transfer learning approach. *Meas. Sens*. 27, 100749.
  15. Rajesh Khanna M. Multi-level classification of Alzheimer disease using DCNN and ensemble deep learning techniques. *Signal, Image and Video Processing*. 2023 Oct;17(7):3603–11.
  16. Li S, Qu H, Dong X, Dang B, Zang H, Gong Y. Leveraging deep learning and xception architecture for high-accuracy mri classification in alzheimer diagnosis. *arXiv preprint arXiv:2403.16212*
  17. Khan R, Akbar S, Mehmood A, Shahid F, Munir K, Ilyas N, et al. A transfer learning approach for multiclass classification of Alzheimer's disease using MRI images. *Front Neurosci*. 2023 Jan 9;16:1050777.
  18. Arafa DA, Moustafa HE, Ali HA, Ali-Eldin AM, Saraya SF. A deep learning framework for early diagnosis of Alzheimer's disease on MRI images. *Multimed Tools Appl*. 2024 Jan;83(2):3767–99.
  19. Ayoub A, Achraf B, Abdelilah J. Comparative Classification of Alzheimer's Disease. In: 2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC) 2024 May 21:1–6
  20. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
  21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* 2016. 770–8.
  22. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Wey, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. 2017 Apr 17.
  23. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the AAAI conference on artificial intelligence*. 2017 Feb 12. (Vol. 31, No. 1).
  24. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* 2015. 1–9.
  25. Tan M, Le QE. Rethinking model scaling for convolutional neural networks. In: *Proceedings of the International conference on machine learning*. Long Beach, CA, USA; 2019 Jun 9. (Vol. 15).