I.J. Intelligent Systems and Applications, **2024, 1, 52-69**
Published Online on February 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2024.01.05

Enhancing Early Alzheimer's Disease Detection: Leveraging Pre-trained Networks and Transfer Learning

Naveen. N.*
Department of Computer Science & Engineering, M. S. Ramaiah University of Applied Sciences, Visvesvaraya Technological University, Belagavi, 590018, India
E-mail: naveensetty@sjbit.edu.in
ORCID iD: https://orcid.org/0000-0001-9146-3251
*Corresponding Author

Nagaraj. G. Cholli
Department of Information Science & Engineering, RV College of Engineering, Visvesvaraya Technological University, Belagavi, 590018, India
E-mail: nagaraj.cholli@rvce.edu.in
ORCID iD: https://orcid.org/ 0000-0001-7409-8272

Received: 19 September 2023; Revised: 26 October 2023; Accepted: 07 December 2023; Published: 08 February 2024

Abstract: Alzheimer's Disease (AD) is a progressive neurodegenerative disorder affecting millions worldwide. Early and accurate AD detection is crucial for timely intervention and improving patient outcomes. Lately, there have been notable advancements in using deep learning approaches to classify neuroimaging data associated with Alzheimer's disease. These methods have shown substantial progress in achieving accurate classification results. Nevertheless, the concept of end-to-end learning, which has the potential to harness the benefits of deep learning fully, has yet to garner extensive focus in the realm of neuroimaging. This is attributed mainly to the persistent challenge in neuroimaging, namely the limited data availability. This study employs neuroimages and Transfer Learning (TL) to identify early signs of AD and different phases of cognitive impairment. By employing transfer learning, the study uses Magnetic Resonance Imaging (MRI) images from the Alzheimer's Disease Neuroimaging (ADNI) database to classify images into various categories, such as Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD). The classification task involves training and testing three pre-trained networks: VGG-19, ResNet-50, and Inception V3. The study evaluates the performance of these networks using the confusion matrix and its associated metrics. Among the three models, ResNet-50 achieves the highest recall rate of 99.25%, making it more efficient in detecting the early stages of AD development. The study further examines the performance of the pre-trained networks on a class-by-class basis using the parameters derived from the confusion matrix. This comprehensive analysis provides insights into how each model performs for different classes within the AD classification framework. Overall, the research underscores the potential of deep learning and transfer learning in advancing early AD detection and emphasizes the significance of utilizing pre-trained models for this purpose.

Index Terms: VGG, Inception, Residual Network (ResNet), Convolutional Neural Network (CNN), Transfer Learning (TL), Magnetic Resonance Imaging (MRI), Deep Neural Networks (DNN).

1. Introduction

Alzheimer's disease (AD) is a progressive neurological illness that gradually impairs cognitive functions, leading to the deterioration of thought processes and consciousness in affected individuals. AD directly affects the advancement of mental capabilities and the overall functioning of neurocognition [1]. AD is the most prevalent form of dementia, accounting for approximately 60%–80% of all dementia cases. Nearly ten million fresh dementia patients are reported annually, and as of 2020, the global population of individuals living with dementia exceeded 55 million. Projections indicate that this number will nearly double every 20 years, reaching 78 million in 2030 and further escalating to 139 million by 2050 [2]. Nearly 60-90% of neurodegenerative ailments are categorized as AD dementia subtype [3], which cannot be cured.

Dementia typically follows a chronic or progressive course. This progressive process often begins with the Cognitively Normal (CN) stage, leading to Mild Cognitive Impairment (MCI) and ultimately culminating in AD. Identifying symptoms of Alzheimer's disease through clinical measures is feasible, but it is a labor-intensive procedure demanding specialized expertise. Early diagnosis often proves challenging for experts unless symptoms are prominently evident due to the time-consuming nature of the process. The early detection of Alzheimer's can play a pivotal role in mitigating the risk of neurological disorders [4, 5]. A timely diagnosis empowers patients with the awareness of the significance of adopting preventive measures to reduce the chances of the disease progressing from MCI to AD. MCI is an initial phase of AD and can lead to a subtle yet detectable and quantifiable reduction in cognitive functions. It is possible to delay the development of AD if MCI is detected early and treated appropriately [6]. Considering the various available treatment options, current research focuses on identifying MCI individuals who have not yet progressed to dementia so that the disease progression can be reduced or even prevented. Also, AD is more likely to develop in individuals with MCI than in others [7].

Advancements in neuroimaging techniques, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), have been instrumental in identifying AD-related structural and molecular biomarkers [8]. Indeed, MRI is extensively utilized in analyzing, detecting, and classifying Alzheimer's disease. It provides detailed structural information about the brain, allowing researchers and clinicians to identify abnormalities, track disease progression, and aid in the accurate diagnosis of AD. Brain imaging technology has advanced rapidly, leading to abundant high-dimensional multimodal data. This has led to an increase in the popularity of computer-aided machine-learning methodologies for integrative analysis. Medical imaging and Computer-aided techniques are the most reliable ways to identify AD in its early stages. Although clinical screening methods are sometimes successful, they are frequently unreliable and expensive. Despite the potential advantages of imaging methods over traditional clinical diagnosis, analysis of the enormous data volume produced by imaging approaches calls for sophisticated computing algorithms. The High Dimensionality and Low Sample Size (HDLSS) dilemma, sometimes known as the "curse of dimensionality," is the major issue with handling imaging data. This problem is also known as the "big G and small n" problem, where "G" and "n" represent the feature and sample counts, respectively. The HDLSS challenge calls for pattern recognition techniques, and Artificial Intelligence (AI) systems fueled by deep learning algorithms provide the answer.

Diverse statistical and machine learning techniques, such as Support Vector Machine (SVM) [9], are employed for fully automated identification and diagnosis of Alzheimer's disease. Deep learning methods like Convolutional Neural Networks (CNN) and sparse auto-encoders have recently outperformed traditional SVM learning methods [10]. Deep Learning techniques do come with certain limitations. Educating deep learning methods from inception demands significant computation and a substantial dataset of annotated medical images [11]. Researchers need help collecting such many medical images. The Transfer Learning (TL) approach is particularly advantageous for medical imaging modalities due to its ability to address challenges stemming from the need for extensive datasets and computational complexities associated with intricate methods. Transfer learning remains valuable even in cross-domain applications, where a model initially trained on general objects can be effectively employed for medical images, encompassing X-rays, CT scans, MRIs, and similar medical imagery. Through transfer learning, an existing pre-trained model is repurposed to work with a fresh and smaller set of images that belong to a different category or domain [12]. Emerging as a deep learning strategy, transfer learning involves repurposing an architecture initially designed for one specific task and adapting it for utilization in a different task.

This study uses CNN-based transfer learning to classify brain image scans into five distinct categories: Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer’s disease (AD). The central motivation for utilizing transfer learning was to harness the feature representations derived from natural brain images and apply them to Alzheimer's disease images. This approach aimed to explore a novel method for categorizing AD, potentially offering valuable assistance to clinicians in making accurate diagnoses and informed decisions. The patient can decrease the likelihood of MCI advancing to AD by taking precautionary measures. Despite working with a limited dataset size, the central focus is delivering better outcomes. The advantage of transfer learning is that pre-trained models can be reused for various purposes, thereby reducing training time drastically and producing better outcomes without leading to over-fitting. With the proposed model, an architecture based on deep learning involves re-training the final convolutional layers and integrating them with fully connected layers. The transfer learning approach and re-training of the final layers led to enhancements in multi-class categorizations like AD, LMCI, EMCI, MCI, and NC.

The remainder of the paper is structured as follows: Section 2 delves into the related work, discussing various techniques for Alzheimer's disease classification. Section 3 outlines the materials and methods utilized in our study. In Section 4, we elaborate on the implementation and present the results of our experiments. Section 5 is dedicated to discussions, while Section 6 presents the research's conclusion and outlines potential future directions.

2. Related Works

Hon et al. [13] proposed a transfer learning approach employing VGG and Inception deep CNN architectures re-trained on ImageNet data by training the final fully connected layers on a tiny MRI dataset. Image entropy was applied to determine which portions of MRI images are the most informative. VGG-16 and inception with transfer learning

provided promising accuracies of 92.30% and 96.25%, respectively. In the work by Sarraf et al. [14], the study focused on classifying brains with AD and normal brains using functional MRI 4D data. The approach involved converting the 4D data into 2D images utilizing neuroimaging tools Nibabel and OpenCV. These 2D images were labeled as AD versus Normal Control (NC). For binary classification, the LeNet model based on CNN was employed. The outcomes were compared with those of a support vector machine model, with the proposed model outperforming it by achieving an impressive accuracy of 96.86%.

Three 2D CNNs were suggested by Jyoti et al. [15] to produce three unique MRI views. Each CNN in their architecture comprised three convolutional layers and four dense blocks, with a majority vote used to determine the outcome. Oh et al. [16] suggested a convolutional auto-encoder-based solution for the AD and CN classification challenge, solving the pMCI data limitation with transfer learning. Jain et al. [17] established an Alzheimer's Disease detection framework by employing CNN and VGG-16 models, integrating ImageNet feature extraction methods. Their model achieved a notable accuracy of 95.73% for predicting Alzheimer's disease. A multi-task deep CNN architecture and 3D densely linked CNNs were used by Liu et al. [18] to propose a coarse-to-fine hierarchical ensemble learning strategy for concurrent hippocampal segmentation and Alzheimer's disease categorization. This technique involves initially slicing an MRI image into several slices, after which pre-trained deep neural networks are used to extract information from the slices. The ensemble learning process was then used to refine each slice's findings using the coarse predictions. To account for a global feature map obtained from the entire brain and regional feature maps extracted using a segmentation model, Zhao et al. [19] created a region ensemble model with three sequential subnetworks. The feature representations were combined in their approach, and an attention-based strategy was employed for classification.

In Ebrahimi et al.'s study [20], a CNN-based deep learning model was employed for predicting Alzheimer's Disease in MRI images. They contrasted the CNN model's outcomes with the experimental accuracy of an RNN model, showcasing that the CNN outperformed the RNN, achieving superior accuracy. Khan et al.'s work [21] introduced a hybrid Machine Learning (ML) based model to discern AD and MCI in the ADNI dataset. They integrated nineteen classifiers into their study. XGBoost, random forest, and SVM were utilized for the first experiment. The second experiment focused solely on the random forest classifier. In the third experiment, they implemented the hybrid model incorporating all classifiers, which yielded the highest accuracy among the three experiments.

Lei et al. [22] employed a multi-layered independent RNN model to discern the correlation between data and brain regions. They utilized the ADNI dataset, conducting feature extraction and image processing to derive clinical scores from the dataset. In another study, Revathi et al. [23] introduced a model for early AD identification consisting of two stages. The first stage involved SVM and random forest models to evaluate the influence of blood pressure and diabetes on cognitive impairment. The second stage employed a multinomial logistic regression model to assess the risk of AD, categorizing labels as "definite Alzheimer's", "uncertain Alzheimer's", and "no Alzheimer's". Mahendran et al. [24] implemented an EDRNN (Ensemble Deep Recurrent Neural Network) model for identifying AD in a DNA methylation dataset. Their results, including accuracy (88.70%), specificity (0.874), sensitivity (0.879), F1-score (0.884), and AUC (0.876), were presented. The proposed EDRNN model's performance surpassed that of RNN, CNN, and DRNN models, emerging as the superior choice for AD identification.

Ghazal et al. [25] developed a transfer learning model for AD detection. Their approach involved a multi-class dataset comprising MRIs with varying levels of AD, such as Moderate Dementia (MOD), Mild Dementia (MD), Very Mild Dementia (VMD), and Non-Dementia (ND). Remarkably, the model achieved an accuracy of 91.70% for early disease detection. In the work by Odusami et al. [26], a randomized concatenated model was introduced for identifying AD using an MRI dataset. They harnessed two models, ResNet-18 and DenseNet121, for feature extraction. The MRI images were categorized into five types: MCI, AD, EMCI, LMCI, and NC. Their reported results included a precision of 0.9894, a recall of 0.9889, and an impressive accuracy of 98.86%.

3. Materials and Methods

The dataset used, network design, and the parameter set for training constitute the three main aspects of any DL model. The effective synergy of these three aspects forms the foundation of a successful DL model, ensuring optimal task completion. The model's design is verified through numerous trials involving training with initial parameters and performance assessment. This process is iterated multiple times, training the model with initial settings and evaluating its performance. The model's design is finalized based on this evaluation. If the performance falls short, adjustments are made to the dataset, network design, or training settings until the desired level of performance is achieved.

In this article, we propose utilizing pre-trained networks with minimal modifications to achieve a wide performance range with minimal complexity. The subsequent sections detail the primary classification phases and describe the adopted pre-trained networks.

3.1. Dataset Description

The dataset includes brain DICOM images obtained through MRI scanning. These scans were gathered as part of the Alzheimer's Disease Neuroimage Initiative (ADNI) project, established to conduct AD therapy clinical trials. The Alzheimer's Therapeutic Research Institute (ATRI) at the University of Southern California oversees the concept and development of ADNI. ADNI 1, five-year research, was the database's initial release. The ADNI GO online version,

which was later updated to ADNI 2 and ADNI 3, was created as a result of the additional 2-year extension. The primary goal of the ADNI database is to create biomarkers that allow for conducting clinical studies on AD in its early phases. These indicators are employed to forecast cognitive deterioration. Various imaging modalities, including MRI, fMRI, PET, and others, are used to create the images in ADNI. According to ADNI's current state, enormous data has been gathered, including neuroimage scans, genetic attributes, and biomarkers found in cerebrospinal fluid. This study employs structural MRI data and transfer learning to classify the AD stages.

3.2. Network Architecture

A. Convolution Neural Networks (CNN)

Three crucial steps comprise conventional machine learning techniques: feature extraction, reduction, and classification. Standard CNN combines all these phases, and the feature extraction process can be automated with CNN. Sample CNN structure is shown in Fig. 1 In its initial stages, the weights of the layers function as feature extractors, evolving through repetitive learning. The CNN structure comprises three layers: 1) The convolution layer does feature extraction, 2) The pooling layer performs dimensionality reduction, and 3) The fully connected layer in a CNN takes the two-dimensional matrices from the previous layers and performs the classification task.

In a Convolutional Neural Network, the convolutional layer applies a learnable filter (kernel) to an input image, which allows the network to extract relevant features from the data. Here is a breakdown of the dimensions involved:

Input Image: H × W × C, where H is the height, W is the width, and C is the number of channels (e.g., 3 for RGB images).

*Filter (Kerne*l)*:* FH × FW × FC, where FH is the filter height, FW is the filter width, and FC is the number of filter channels (should match the number of input channels C).

Output Activation Map: AH × AW, where AH is the activation height and AW is the activation width. The convolution operation and the chosen hyperparameters, such as the stride and padding, determine the size of the output activation map.

The output activation map is obtained by sliding the filter over the input image, computing the element-wise dot product between the filter and the corresponding input image patch, and then applying an activation function. The process is repeated across the entire input image, producing the output activation map. The stride and padding affect the dimensions of the output activation map, and they allow control of the spatial dimensions of the output feature map. The following formulas determine the spatial size of the output activation map:

$$\mathrm{AH} = (\mathrm{H} - \mathrm{FH} + 2 * \mathrm{padding}) / \mathrm{stride} + 1 \tag{1}$$

$$AW = (W - FW + 2 * padding) / stride + 1 \tag{2}$$

The stride determines the step size of the filter as it slides over the input image, and the padding determines the number of zero-padded pixels added to the borders of the input image to ensure that the filter can be applied to the edges.

Activation functions introduce non-linearity into the neural network by transforming the output of each neuron (or unit) non-linearly. Without activation functions, the neural network would be equivalent to a linear model, regardless of its depth, because a linear combination of linear functions remains a linear function.

The neural network can learn complex patterns and relationships in the data by applying activation functions. The non-linear transformation allows the network to model and approximate non-linear relationships in the input data, enabling it to solve more sophisticated and intricate tasks.

Commonly used activation functions include:

$$\textit{ReLU (Rectified Linear Unit): } f(x) = max(0, x) \tag{3}$$

$$\textit{Sigmoid: } f(x) = (1) / ((1 + exp(-x))) \tag{4}$$

$$\textit{Tanh (Hyperbolic Tangent): } f(x) = (exp(x) - exp(-x)) / (exp(x) + exp(-x)) \tag{5}$$

$$\textit{Leaky ReLU: } f(x) = \max(ax, x) \textit{ where 'a' is a slight positive slope for negative inputs.} \tag{6}$$

These activation functions introduce non-linearity, allowing the neural network to learn and generalize complex relationships in the data, making it more powerful and versatile in handling various tasks.

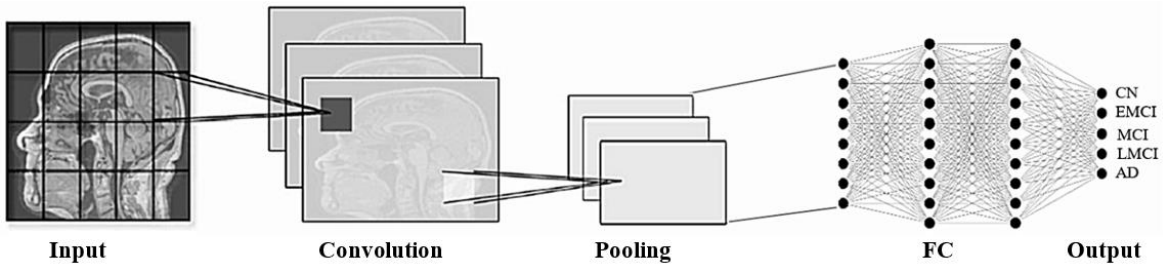


Fig.1. CNN structure

For the proposed multi-classifier, the SoftMax function is used, which returns the probability for a data point belonging to each class, calculated from (7).

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}, \text{ for } i=1,2\ldots\ldots,k \text{ and } x= [x_1,x_2,\ldots\ldots x_k] \quad (7)$$

Where x is the input vector, e^{x}_{i} is the standard exponential function for the input vector, k is the number of classes in the multi-class classifier, and e^{x}_{j} is the standard exponential function for the output vector.

Overfitting occurs when a model learns to perform exceptionally well on the training data but fails to generalize to new, unseen data. Dropout is a regularization technique that addresses overfitting by randomly deactivating (dropping out) a fraction of neurons during each training iteration. This prevents the network from relying too heavily on specific neurons and their interactions, forcing it to learn more robust and general features. By randomly dropping out neurons during training, dropout prevents the network from becoming overly reliant on specific neurons. This encourages the network to learn more diverse and robust features that are not tied to the presence of particular neurons.

While specialized deep learning models are often developed for specific applications, the true strength of deep learning lies in its adaptability to different problems. Over time, there has been a shift toward using pre-trained models as a foundation for new studies. In recent years, many studies have focused on utilizing pre-trained models. These approaches involve directly using the learned weights of pre-trained models or incorporating additional layers onto these models using transfer learning techniques. Our study used pre-trained CNN models like VGG-19, ResNet -50, and Inception V3.

Hyperparameter tuning in deep learning often involves a combination of trial and error, domain knowledge, and insights gained from previous studies. There is no one-size-fits-all approach, and finding the correct hyperparameter values is often a blend of experience and experimentation. The hyperparameters utilized in this study were established through an iterative process involving multiple alternatives. The goal was to identify optimal choices that enhance performance without altering the core structure of the pre-trained models. Dropout layers were introduced between these layers to mitigate overfitting, with a dropout rate of 0.5 chosen to ensure stable outcomes. The optimizer selected was Stochastic Gradient Descent Method (SGDM) optimization. For the Loss parameter, categorical cross-entropy was adopted, and accuracy was designated as the chosen metric. The models underwent training processes lasting for a standard 100 epochs. This specific number of epochs was established through repeated experimentation to halt the training and fine-tune hyperparameters before encountering an overfitting scenario. For all models except Inception V3, images were prepared in a 224 × 224 × 3 format for processing. However, the Inception V3 architecture was designed for a 299 × 299 × 3 image structure, requiring the training data to be adjusted accordingly. Throughout training and validation, the models' ultimate weights were saved. These weights were then assessed using a separate set of testing data that had not been exposed to the models before. This rigorous testing procedure was employed to validate the models' test results, ensuring their reliability.

To apply the DL models to MRI data, two platforms were employed: Kaggle, a collaborative platform for data science problem-solving, and Google Colaboratory (Colab), a platform that offers the ability to write and execute Python code, particularly for ML and data analysis tasks, utilizing resources like GPUs. The training of DL models took place using the Keras Applications DL libraries in the Python programming language, utilizing TensorFlow. After model operations, both graphics and weights were logged, and the models' accuracy and loss parameters were graphically presented. To assess model performance, Confusion Matrices were generated.

B. VGG-19

The VGG-19 architecture is a deep convolutional neural network introduced in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" by Karen Simonyan and Andrew Zisserman in 2014. It is an extension of the earlier VGG-16 architecture, both of which were created by researchers at the University of Oxford

Key characteristics of the VGG19 architecture:

Depth: VGG-19 is characterized by its depth. It has 19 layers, including 16 convolutional layers and three fully connected layers. This depth allows it to capture intricate patterns and features in images.

Convolutional Layers: VGG-19 employs a series of convolutional layers, each followed by a max-pooling layer for down sampling. Using small 3x3 filters in the convolutional layers, combined with the max-pooling, leads to a consistent receptive field size and a more accessible architecture to design and visualize.

Filter Size: All convolutional layers in VGG-19 use 3x3 filters, which helps to maintain a compact architecture. This results in fewer parameters than larger filters, making the model computationally efficient.

Fully Connected Layers: VGG-19 has three fully connected layers followed by a softmax layer for classification after the convolutional and pooling layers. Fig. 2 shows the architecture of VGG-19.

VGG-19 is often used as a pre-trained model for transfer learning due to its performance on image recognition tasks. It can be fine-tuned on specific datasets or used as a feature extractor by removing the last layers and adding new ones tailored to a specific task.

C. ResNet-50

ResNet-50 is a convolutional neural network architecture that belongs to the ResNet family (short for Residual Networks). It was introduced by Microsoft Research in 2015 as part of their effort to address the vanishing gradient problem in very deep neural networks. ResNet-50 is characterized by its depth and the use of residual blocks, which allow for training deeper networks without encountering the degradation problem. Fig. 3 shows the architecture of Resnet-50.

Key features of ResNet-50:

Identity Shortcut Connections (Skip Connections): ResNet-50 introduces the concept of identity shortcut connections or skip connections. Instead of the conventional approach where layers stack sequentially, the input to a layer is added to the output of a later layer in a residual block. This identity mapping helps mitigate the vanishing gradient problem and allows for the training of very deep networks.

Bottleneck Architecture: Each residual block in ResNet-50 employs a bottleneck architecture. It consists of three layers: a 1x1 convolutional layer, a 3x3 convolutional layer, and another 1x1 convolutional layer. This architecture reduces the computational complexity by using 1x1 convolutions to reduce and restore the number of channels.

Skip Connections: Skip connections provide a path for the gradient to flow more directly during backpropagation. If a residual block learns an identity function, the gradients can flow unimpeded through the shortcut connections, making training more accessible.

Pooling Layers: ResNet-50 incorporates average pooling before the final fully connected layer. This helps to reduce spatial dimensions and the number of parameters in the model, which is beneficial for generalization.

Global Average Pooling: Like Inception V3, ResNet-50 uses global average pooling to replace fully connected layers for the final classification. This technique reduces overfitting and the number of parameters in the model.

ResNet-50 is often used as a pre-trained model for transfer learning. The model is trained on massive image datasets like ImageNet, and then the lower layers are retained for feature extraction, while the upper layers are customized for specific tasks. ResNet-50 was a significant milestone in deep learning due to its depth and the innovative use of residual blocks, which led to development of even deeper architectures. It demonstrated that increasing the depth of neural networks does not necessarily result in poorer performance, as was previously believed, but can lead to better results when designed properly.

D. Inception V3

Inception V3 is a convolutional neural network (CNN) architecture that builds upon the earlier Inception models, designed to improve the efficiency and performance of deep learning models for image recognition tasks. Google researchers developed it as part of the Inception architecture family. Fig. 4 presents the architecture of inception V3.

Key features of Inception V3:

Multiple Pathways (Parallel Convolutions): Inception V3 employs a unique structure called "Inception modules." These modules consist of multiple convolutional layers with different filter sizes (1x1, 3x3, and 5x5) and max-pooling layers. These pathways run parallel, capturing features at multiple spatial scales and abstraction levels. The idea is to allow the network to learn fine-grained and high-level features simultaneously.

1x1 Convolutions: Inception V3 heavily uses 1x1 convolutions to reduce the dimensionality of the input feature maps. This helps to reduce the computational complexity and allows for efficient parallel processing.

Bottleneck Layers: To further improve efficiency, Inception V3 often incorporates bottleneck layers, which use 1x1 convolutions to reduce the number of input channels before applying larger filters. This reduces the computational load without sacrificing the model's ability to capture complex features.

Auxiliary Classifiers: Inception V3 includes auxiliary classifiers at intermediate layers during training. These auxiliary classifiers serve as extra branches to the main classification layer. They help to combat the vanishing gradient problem during training, ultimately improving convergence and gradient flow.

Global Average Pooling: Instead of fully connected layers at the end, Inception V3 employs global average pooling. This technique replaces the traditional flattening and fully connected layers with a single average pooling layer that generates predictions directly from the spatial feature maps.

Regularization Techniques: Inception V3 incorporates batch normalization and dropout techniques to enhance generalization and reduce overfitting.

Inception V3 is often used as a pre-trained transfer learning model, where the network's lower layers, which capture general features, are retained, and the upper layers are replaced to suit a specific task. It has achieved state-of-the-art results on various image recognition benchmarks and competitions due to its computational resources and performance efficiency.

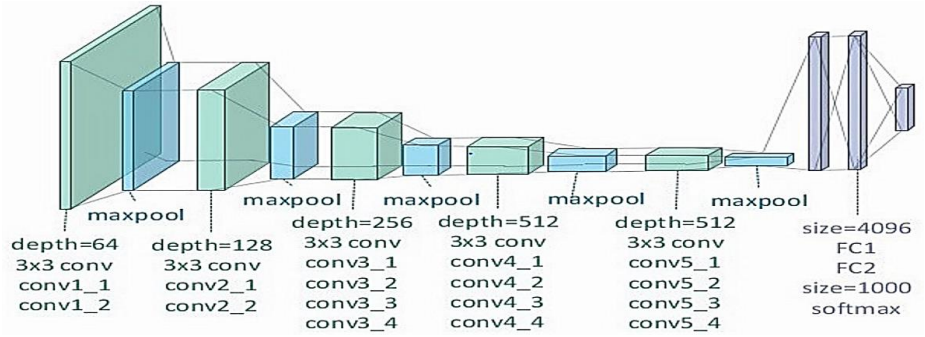


Fig.2. VGG-19 architecture

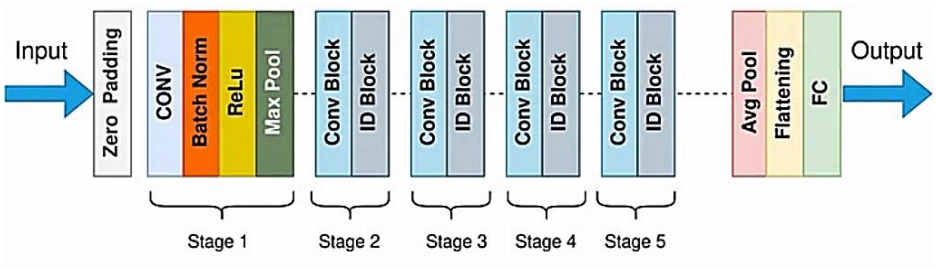


Fig.3. ResNet-50 architecture

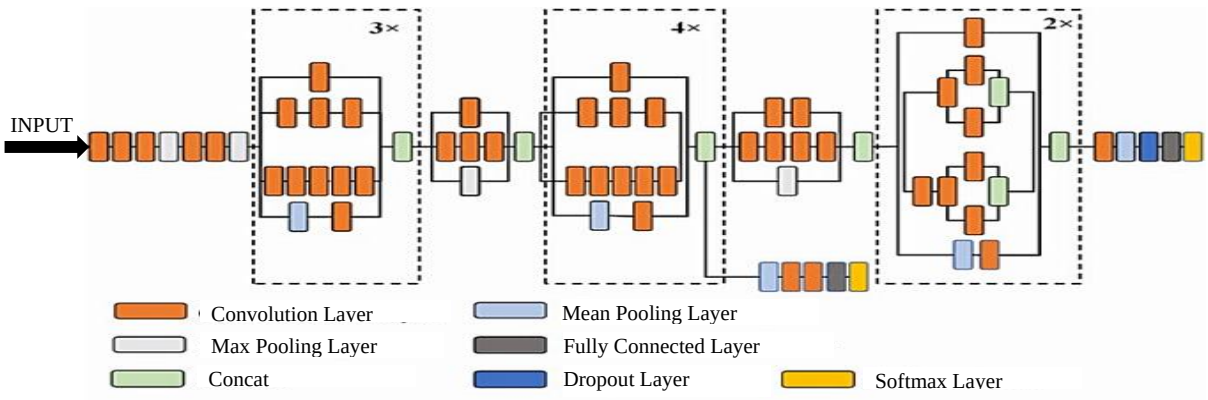


Fig.4. Inception V3 architecture

3.3. Transfer Learning

Transfer learning is a machine learning technique that involves leveraging knowledge gained from one task to improve the performance on another related but different task. In mathematical terms, transfer learning can be described as follows.

Domain: A domain

$$\mathfrak{D} = \{X, (X)\} \tag{8}$$

is defined by two components:

- A feature space X
- A marginal probability distribution (X) where $X = \{x_1, x_2, x_3, \ldots, x\text{n}\} \in X$

If two domains are different, then they either have different feature spaces (Xt $\neq$ Xs) or different marginal distributions ((Xt) $\neq$ (Xs)).

Task: Given a specific domain $\mathfrak{D}$, a task $\mathcal{T}$ = {Y, (.)} consists of two parts:

- A label space Y
- A predictive function (.), which is not observed but can be learned from training data

$$\{(x_i, y_i) \mid i \in \{1, 2, 3, \ldots, N\}, \text{where } x_i \in X \text{ and } y_i \in Y\}.$$

From a probabilistic viewpoint (x_i) can also be written as p($y_i|x_i$), so we can rewrite task $\mathcal{T}$ as

$$\mathcal{T} = \{Y, (Y|X)\} \tag{9}$$

In general, if two tasks are different, then they may have different label spaces (Yt $\neq$ Ys) or different conditional probability distributions (P(Yt|Xt) $\neq$ P(Ys|Xs)).

Given a source domain $\mathfrak{D}$s and corresponding learning task $\mathcal{T}$s, a target domain $\mathfrak{D}$t and learning task $\mathcal{T}$t, transfer learning aims to improve the learning of the conditional probability distribution $P(Y$t|Xt) in $\mathfrak{D}$t with the information gained from $\mathfrak{D}$s and $\mathcal{T}$s, where

$$\mathfrak{D}t \neq \mathfrak{D}s \text{ or } \mathcal{T}t \neq \mathcal{T}s \tag{10}$$

In deep learning algorithms, an existing model can be repurposed for a new task through appropriate modifications, a process known as fine-tuning. The success of fine-tuning is significantly determined by the resemblance between the original and the new training datasets. Similarities are pivotal for achieving optimal fine-tuning results.

Transfer learning is one of the most widely employed strategies in various tasks. An essential facet of transfer learning involves pre-trained networks, which are constructed on the foundation of deep learning frameworks like CNNs. Various pre-trained networks exist, such as VGG-16, GoogLeNet, LeNet, VGG-19, Inception, AlexNet, DenseNet, SqueezeNet, ResNet, and more. The vast ImageNet dataset, which includes 1,000 different image classes, served as the initial training ground for these networks.

While the potential benefits of transfer learning seem evident, it requires rigorous testing and evaluation to ascertain whether fine-tuning a model would yield improved outcomes. The present work centers on classifying Alzheimer's Disease by harnessing pre-trained networks and transfer learning. The study involves modifying and training pre-trained networks using the ADNI MRI dataset. This research aims to establish the effectiveness of transfer learning techniques in improving the accuracy of AD classification.

3.4. Performance Analysis

In our study, we utilize accuracy, precision, error, false positivity rate, kappa, sensitivity/recall, specificity, and f-measure to evaluate the performance of the proposed model. The confusion matrix complements these metrics by visually summarizing the performance of a classification model across different classes. It is organized with actual class labels forming the rows and predicted class labels forming the columns. This is shown in Table 1.

Table 1. Confusion matrix

	Predicted Positive (P)	Predicted Negative (N)
Actual Positive (P)	True Positive (TP)	False Negative (FN)
Actual Negative (N)	False Positive (FP)	True Negative (TN)

Table Description

Positive (P): indicates the patient suffers from dementia.
Negative (N): indicates the patient is Cognitive Normal.
True Positive (TP): Instances correctly predicted as positive.
True Negative (TN): Instances correctly predicted as negative.
False Positive (FP): Instances incorrectly predicted as positive (Type I error).
False Negative (FN): Instances incorrectly predicted as negative (Type II error).

Accuracy: Accuracy measures the proportion of correctly predicted instances among all instances. It is calculated as:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{11}$$

Accuracy Error: The accuracy error, often called the misclassification error, calculates the proportion of instances a machine learning model has about the number of instances incorrectly classified. It is calculated as

$$Accuracy\ Error = 1 - Accuracy \tag{12}$$

Precision: Precision is the ratio of True Positives to all instances predicted as positive. It is calculated as:

$$Precision = TP/(TP + FP) \tag{13}$$

Sensitivity (Recall or True Positive Rate): Sensitivity measures the proportion of actual positive instances correctly predicted. It is calculated as:

$$Sensitivity\ (Recall) = True\ Positive\ Rate = Recall = TP/(TP + FN) \tag{14}$$

Specificity: Specificity, also referred to as the selectivity of a network, is a measure that indicates the ratio of accurately detected negative instances to those that are negative. The true negative rate represents this parameter and can be calculated using the following formula:

$$Specificity = True\ Negative\ Rate\ = \frac{TN}{TN+FP} \tag{15}$$

False positive rate: The false positive rate, also known as the fall-out, pertains to the frequency of positive predictions generated by a network that is not correct. It is computed as the ratio of wrongly predicted positives to the total count of actual negatives. This metric helps assess the model's tendency to classify negative instances as positive incorrectly.

$$FPR\ (1 - Specificity) = FP/(TN + FP) \tag{16}$$

F-Measure (F1-Score): The F1-Measure is the weighted harmonic average of precision and recall or mean. It is calculated as

$$F1 - score = (2\ \times\ Precision\ \times\ Recall)/(Precission\ +\ Recall) \tag{17}$$

The F1 score is particularly useful when the class distribution is imbalanced, as it equally considers false positives and negatives. It is widely used in machine learning to evaluate the overall quality of a classifier's predictions.

Kappa: The Kappa Coefficient, also known as Cohen's Kappa Score, is a statistical measure used to evaluate the performance and agreement of machine learning classification models. The Kappa Coefficient helps to assess the model's performance beyond simple accuracy, considering the possibility of agreement occurring by random chance. It is advantageous when dealing with imbalanced datasets or situations where the classes have different prior probabilities. A higher Kappa value suggests a better agreement between the model's predictions and the actual outcomes beyond what might be expected due to chance.

$$Kappa = (2 * (TP * TN - FN * FP))/((TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)) \tag{18}$$

4. Implementation and Results

This section of the paper encompasses the implementation details and outcomes of the study. The section delves into the practical execution of the proposed approach and presents the achieved results. It likely discusses the methods used to implement the deep learning models, the pre-processing steps, hyperparameter tuning, and any modifications made to pre-trained networks. The results obtained from the experiments are then presented, often in the form of tables, graphs, and performance metrics.

This study proposes a method for early Alzheimer’s detection on brain MRI images using transfer learning techniques for multi-class AD classification. To detect Alzheimer's disease using MRI scans, Convolutional Neural Networks were tailored by utilizing brain images. Instead of initiating the model training process anew, pre-trained deep learning models like VGG-19, ResNet-50, and Inception V3 were employed. Subsequently, a gradual transfer learning process was implemented on these base models for Alzheimer's disease detection. Fig. 5. presents an outline of the suggested approach, showcasing the steps involved in the methodology. The process involves the acquisition of MRI images, data pre-processing, employing transfer learning for training the deep learning models, and eventually performing a multi-class classification process.

4.1. Dataset Acquisition

The data utilized for our study were sourced from the ADNI database, accessible at http://adni.loni.usc.edu/. The primary objective of ADNI is to develop more precise and sensitive methods for diagnosing Alzheimer's disease in its initial stages. Additionally, it seeks to track the advancement of Alzheimer's disease using biomarkers, providing insights into the disease's progression. The dataset used in this study is based on the T1-weighted MRI modality and includes a

total of 1230 patients categorized into five classes: Alzheimer's Disease (AD), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI), and Normal Control (NC). Each class consists of images from both male and female subjects. The distribution of images across the classes is as follows: AD class (1148 images), EMCI class (1084 images), MCI class (827 images), LMCI class (798 images), and NC class (661 images). The medical images have a 2D format of 256 × 256 pixels. Further demographic information about 1230 subjects from the ADNI dataset can be found in Table 2.

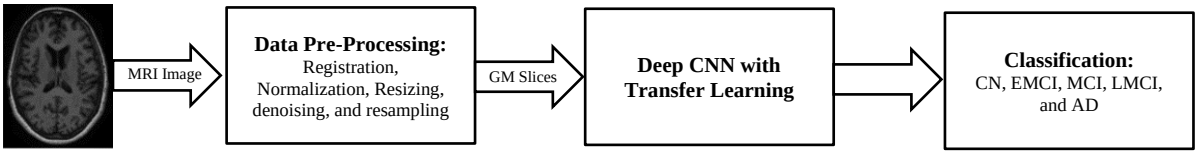


Fig.5. Workflow of the proposed model

Table 2. Demographic data

Alzheimer stages	NC	EMCI	MCI	LMCI	AD
Count	175	274	248	239	294
Male/female	77/98	85/189	112/136	108/131	114/180
Age (mean ± STD)	76.78±0.478942	75.84±0.864785	75.81±0.786485	76.89±1.215487	76.21±0.897835

The summary of the dataset includes various vital details. It provides the count of patients in each class, the distribution of male and female patients within each class, and the average ages along with their corresponding standard deviations (STD). Additionally, Fig. 6 displays representative MRI scans from all five classes, visually representing the data.

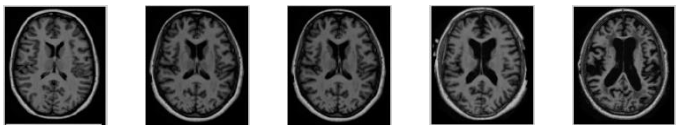


Fig.6. Stages of AD: (a) CN (b) EMCI (c) MCI (d) LMCI (e) AD

4.2. Data Pre-processing

The collected dataset exhibited class imbalances, prompting the implementation of two resampling techniques: oversampling and undersampling. In the oversampling approach, instances were duplicated for the underrepresented classes (CN, MCI, and LMCI), while in undersampling, instances were removed from the overrepresented classes (AD and EMCI). As a result of these methods, all AD classes were augmented to include 1000 MRI images, resulting in a final dataset of 5000 images. The dataset underwent various pre-processing steps, including normalization, standardization, resizing, denoising, and format conversion.

The MRI scans were subjected to spatial normalization into the MNI (Montreal Neurological Institute) space. This was achieved using the Statistical Parametric Mapping (SPM12) software (available at HTTP://WWW.FIL.ION.UCL.AC.UK/SPM/) and the Diffeomorphic Anatomical Registration Exponentiated Lie Algebra (DARTEL) registration method. The intensity values of the MRI scans were adjusted to fall within the range of [0, 1] through normalization. Additionally, denoising was performed on the images using a non-local means algorithm, which introduced image blurring to reduce the impact of noise in the images.

Due to the inherent challenge of limited medical datasets, we utilized conventional data augmentation techniques to expand the dataset. These techniques included image rotation and reflection (flipping), which encompass both horizontal and vertical flipping of images. As a result of these augmentation methods, the dataset's volume was substantially increased to a total of 10,000 images, equally distributed across all classes with 2,000 images per class. The implementation of data augmentation serves two primary purposes: firstly, it effectively enhances the dataset's size, and secondly, it serves as a strategy to alleviate the potential problem of overfitting in the model.

For training the Deep Neural Networks (DNNs), a set of 8,000 images was used, while a separate set of 2,000 images was allocated for testing. Ten thousand images were employed to train and evaluate the deep neural network classifiers. Table 3 describes the training and test datasets used in our study.

In deep learning, the images collected from each class serve as the image data repository. These images are incorporated into the data repository and resized to match the input dimensions of the neural network's input layer. The augmented image repository function is responsible for implementing the various pre-processing steps. Specifically, Inception V3 required input images to have dimensions of 299x299x3, whereas all other networks mandated input

dimensions of 224x224x3.

The pre-processing steps are applied to the target domain image samples before being utilized for training and testing. During the training process, there might be instances where MRI scanners introduce degradation due to factors like poor brightness or low contrast caused by optical devices. Image enhancement techniques such as linear contrast stretching are employed to rectify this issue and restore the visual characteristics of images. These techniques expand the distribution of pixels across a wide range of intensities, resulting in the recovery of image features.

As part of the dataset preparation, the data repository is divided into training and testing subsets in an 80:20 ratio. This division ensures that the images in the testing set are distinct from those in the training set, which is crucial for effectively validating the model's classification performance.

Table 3. Description of training, validation, and test datasets

Class Label	Training set count	Test set count	Total Count
0-NC	1600	400	2000
1-EMCI	1600	400	2000
2-MCI	1600	400	2000
3-LMCI	1600	400	2000
4-AD	1600	400	2000
Total	8000	2000	10000

4.3. Deep Neural Network (DNN)

DNN architecture categorizes images from the ADNI database using transfer learning for binary and multi-class problems. The primary focus of medical sciences research is the early detection of Alzheimer's disease. However, it aids in their therapy during their phases' diagnosis, further increasing the significance of multi-class problems. The DNN model is trained to learn those specific task features using the GM, WM, CSF, and non-segmented images as four data sets. Every data set's training and testing procedures were time-stamped over 100 epochs, and their equivalent evaluation results arrived at using a confusion matrix. The suggested model has been evaluated for both binary and multi-class AD classification. The result illustrates the categorization outcomes for the separate NC, EMCI, MCI, LMCI, and AD segments.

This study used a total of 10,000 image scans for deep neural network classifiers, with 8,000 images used for training and 2,000 images used for testing.

Networks Employed

VGG-19, ResNet-50 and Inception V3 are used as pre-trained networks for transfer learning, and the effectiveness of each network is examined. It was discovered that the complexity of the networks utilized directly correlates with the training time needed. Using parameters derived from the confusion matrix, the study aims to assess their effectiveness in classifying Alzheimer's disease based on neuroimaging data.

4.4. Training Options

Table 4 in the study outlines the training choices and hyperparameters utilized for training the various network architectures. The purpose of this table is to facilitate a direct comparison of the performance of different network architectures under the same set of training conditions. This standardized approach enables researchers to evaluate and contrast the efficiency and effectiveness of the different architectures in a controlled manner.

The study employed the Stochastic Gradient Descent Method (SGDM) as the optimization algorithm for the training process. The momentum value for SGDM was set at 0.9. The networks underwent training for a total of 100 epochs, where each epoch represents a complete pass through the training dataset. The learning rate, a critical hyperparameter that influences the speed and direction of the optimization process, was set at 1e-4. During each epoch, the dataset was divided into mini batches of size 32 for more efficient computation, and the batches were shuffled to prevent any bias in the learning process.

The training procedure continued for 100 epochs, which represents 100 complete cycles of training through the entire dataset. The model's performance was monitored using a validation error metric, and the training process was stopped when a minimum validation error was reached. This approach prevents the model from being trained for an excessive number of epochs, which could lead to overfitting. Overfitting occurs when the model learns to perform exceptionally well on the training data but fails to generalize to new, unseen data.

Premature stopping, as described, is a common practice in deep learning to ensure that the model achieves optimal performance without overfitting. The graph of error rate (loss) versus the number of epochs often shows a trend where the error rate decreases initially but increases for subsequent epochs due to overfitting. The model's generalization performance is improved by stopping the training when the validation error increases.

Table 4. Hyperparameters

Parameter	Value
Optimization algorithm	SGDM
Momentum	0.9
Initial learning rate	1e-4
Maximum number of epochs	100
Mini-batch size	32
Validation frequency	3

4.5. Training Progress

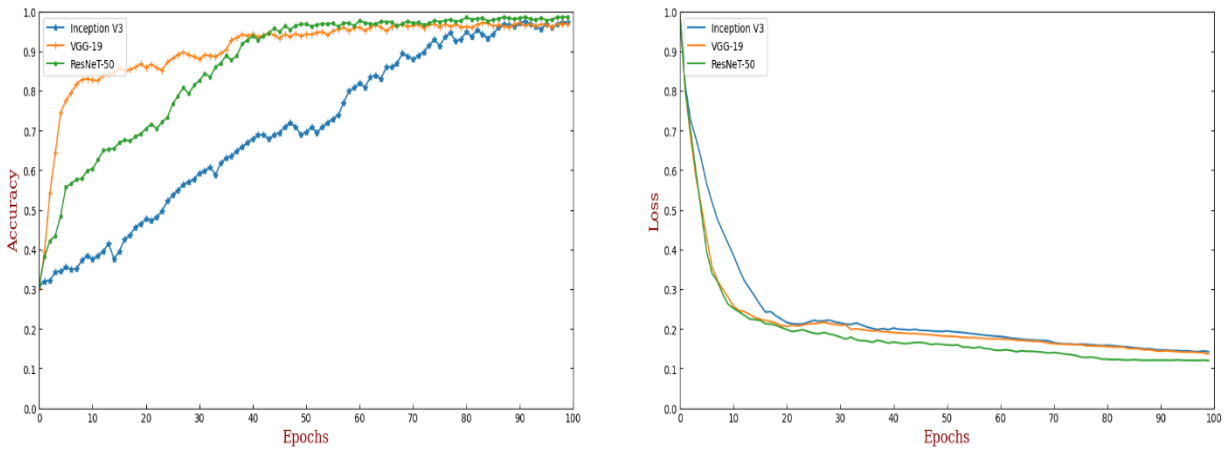
The study used the designated training dataset and established training choices to train the network architectures, including ResNet-50, Inception V3, and VGG-19. The goal was to compare the performance of these architectures in terms of accuracy and other relevant metrics. The training process involved optimizing the modified pre-trained networks using the same set of training parameters over a predetermined number of epochs.

Fig. 7 in the study presents a graphical representation of the training progress of these networks over the specified number of epochs. This plot provides insights into how well each network architecture optimized and improved its performance during training.

The plot reveals that VGG-19 and ResNet-50 exhibited more effective optimization than Inception V3. Additionally, ResNet-50 showed better optimization performance than VGG-19. This can be observed from the fact that the accuracy or performance metrics on the vertical axis increase more steeply for VGG-19 and ResNet-50 than Inception V3, indicating faster and more efficient learning.

Furthermore, it is evident from the plot that more than 50 epochs were needed for Inception V3 to achieve optimal performance. The graph for Inception V3 shows that the accuracy or performance metrics had not plateaued by the 50th epoch, suggesting that this architecture might require more epochs to converge to its best performance.

This observation aligns with the fact that larger and more complex network architectures, like Inception V3, might require additional training epochs to achieve their maximum potential in terms of accuracy and performance. It highlights the importance of tuning the number of training epochs based on the architecture's complexity and behavior during training.



(a) Accuracy (b) Loss

Fig.7. Accuracy and loss across different epochs

4.6. Pre-trained Networks for AD Classification

Table 5 in the study compares the classification performance of three pre-trained networks: ResNet-50, VGG-19, and Inception V3. The performance metrics evaluated include accuracy, error rate, recall, specificity, precision, false positive rate, F1-score, and kappa. These metrics are crucial in assessing the effectiveness of each network in accurately classifying the different classes in the dataset.

The results highlight that ResNet-50 achieved the highest overall accuracy of 98.7% for the classification of the AD class, surpassing the performance of both VGG-19 (97.16%) and Inception V3 (97.54%). This indicates that ResNet-50 has better discriminative power in distinguishing AD cases from other classes, making it more effective in this classification task.

Additionally, the study reveals that ResNet-50 also outperforms the other models regarding other performance metrics, including error rate, recall, specificity, precision, false positive rate, F1-score, and kappa. These metrics collectively indicate that ResNet-50 consistently provides superior results compared to the other two pre-trained networks.

To further illustrate the comparative performance of VGG-19, ResNet-50, and Inception V3, Fig. 8 in the study presents a graphical representation. This visualization showcases the varying performance of these networks across different classes, helping to visually compare their classification outcomes and understand their strengths and weaknesses in classifying AD and other classes.

In summary, ResNet-50 is the most effective pre-trained network for classifying AD cases in this study, demonstrating superior accuracy and performance across various evaluation metrics compared to VGG-19 and Inception V3.

4.7. Class-specific Evaluation of Performance

The process we describe involves evaluating the performance of three different networks (ResNet-50, VGG-19, and Inception V3) for classifying different classes of AD. By analyzing the combined confusion matrix of all three networks, we aim to assess the consistency of their performance across various AD classes. This evaluation helps identify whether the networks consistently perform well across all classes and aids in selecting optimal training parameters for classifying different AD classes effectively.

A. Confusion Matrix

Table 6 depicts the VGG-19, ResNet-50, and Inception V3 confusion matrix. The AD class contains more accurately predicted images, suggesting that the networks perform well in this particular class.

Table 5. Overall performance evaluation of pre-trained networks

Performance	Inception V3	VGG-19	Resnet-50
Accuracy	97.54	97.16	98.70
Accuracy Error	2.46	2.84	1.30
Sensitivity/Recall	93.80	92.90	96.90
Specificity	98.48	98.23	99.15
Precision	93.93	92.96	96.65
False Positive Rate	1.53	1.78	0.85
F1-Score	93.84	92.89	96.73
Kappa	92.31	91.12	95.92

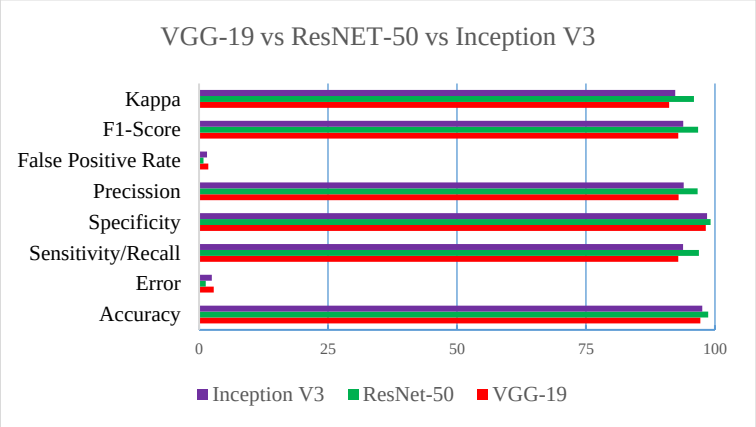


Fig.8. Performance comparison of VGG-19, ResNet-50, and Inception V3 networks

Table 6. The VGG-19, resnet-50, and inception V3 confusion matrix with five instances

	VGG-19				ResNet-50				Inception V3			
	TP	FP	TN	FN	TP	FP	TN	FN	TP	FP	TN	FN
CN	389	42	1558	11	394	17	1583	6	382	40	1560	18
EMCI	373	22	1578	27	397	14	1586	3	384	20	1580	16
MCI	370	21	1579	30	389	15	1585	11	383	15	1585	17
LMCI	367	22	1578	33	391	13	1587	9	369	18	1582	31
AD	359	35	1565	41	363	6	1594	37	358	31	1569	42

B. Performance Parameters of Every Class

This subsection describes the outcomes of the class-specific performance evaluation for the pre-trained networks: VGG-19, ResNet-50, and Inception V3. The networks' performance is evaluated using a separate testing dataset that contains 20% of the data acquired from the ADNI dataset. This dataset assesses how well the trained models generalize to new, unseen data. The pre-trained networks' performance evaluation involves using seven performance metrics: accuracy, error, recall, specificity, precision, false-positive rate, and F1-score.

a. Class-based Performance Metrics for VGG-19

Table 7 presents the class-specific results obtained from the VGG-19 model. These results offer insights into the model's performance for each class within the ADNI dataset. The findings reveal that the VGG-19 model excels in classifying the "CN" category with a high recall rate of 97.25%. On the other hand, its performance is comparatively less effective for the "AD" class, with a recall of 89.75%. The recall metric for the other classes is as follows: "EMCI" – 93.25%, "MCI" – 92.5%, and "LMCI" – 91.75%.

The provided accuracy rates demonstrate how well the VGG-19 model can correctly classify instances from each class. While it achieves high accuracy for some classes, its performance is relatively weaker for others. This variance in performance highlights the model's ability to distinguish between different cognitive impairment stages and typical cases.

The graphical representation in Fig. 9 visually represents the VGG-19 model's performance across the various classes.

b. Class-based Performance Metrics for Resnet-50

Table 8 presents the class-wise performance outcomes of the ResNet-50 model. This table provides insights into how well the ResNet-50 model can classify brain MRI images for each class within the dataset.

The results indicate that the ResNet-50 model achieves its highest accuracy in classifying the "EMCI" category, with a recall rate of 99.25%. This implies that the model is particularly effective at identifying instances within this class. On the other hand, the recall is relatively lower for the "AD" class, which is classified with a recall of 90.75%. The recall rates for the other classes are as follows: "CN" - 98.50%, "MCI" - 98.25%, and "LMCI" – 97.75%.

The graphical representation in Fig. 10 visually depicts the ResNet-50 model's performance across the various classes. This visualization enhances understanding of the model's strengths and weaknesses when classifying different cognitive impairment stages and typical cases.

c. Class-based Performance Metrics for Inception V3

Table 9 presents the class-wise performance metrics for the Inception V3 model. This table shows how effectively the Inception V3 model can classify brain MRI images for each class within the dataset.

The results indicate that the Inception V3 model achieves its highest accuracy in classifying the "EMCI" category, with a recall of 96.00%. This suggests that the model excels at identifying instances within this class. Conversely, the accuracy is relatively lower for the "AD" class, classified with a recall of 89.50%. The recall rates for the other classes are as follows: "CN" - 95.50%, "MCI" – 95.75%, and "LMCI" – 92.25%.

The graphical representation in Fig. 11 visually summarizes the Inception V3 model's performance across the different classes. This visualization clearly explains the model's performance trends and helps assess its capabilities in classifying various cognitive impairment stages and typical cases.

The analysis shows that when the ResNet-50 model is used, it achieves a recall of 99.25% (EMCI class) in detecting early stages of cognitive development. This rate is higher than the results obtained from the other two pre-trained networks, namely VGG-19 and Inception V3 (93.25% and 96.00%, respectively).

This suggests that the ResNet-50 model is particularly effective in identifying early stages of cognitive impairment. The deeper architecture of ResNet-50 might contribute to this improved accuracy, as it can capture more intricate and complex features in brain MRI scans. Detecting cognitive decline at an early stage is crucial for timely intervention and treatment, and the higher accuracy achieved by ResNet-50 could have significant implications for accurate and early diagnosis.

These observations indicate that the choice of pre-trained network architecture can notably impact the model's performance, especially in differentiating between different stages of cognitive impairment and AD.

Table 7. Class-based performance metrics for VGG-19

Performance Metric	CN	EMCI	MCI	LMCI	AD
Accuracy	97.35	97.55	97.45	97.25	96.20
Accuracy Error	2.65	2.45	2.55	2.75	3.80
Recall	97.25	93.25	92.50	91.75	89.75
Specificity	97.38	98.63	98.69	98.63	97.81
Precision	90.26	94.43	94.63	94.34	91.12
False Positive Rate	2.63	1.38	1.31	1.38	2.19
F1 Score	93.62	93.84	93.55	93.03	90.43

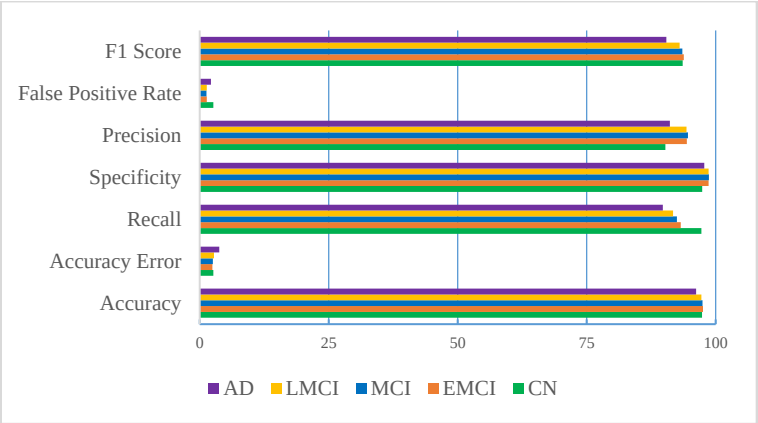


Fig.9. Graphical representation of VGG-19 model performance

Table 8. Class-based performance metrics for ResNet-50

Performance Metric	CN	EMCI	MCI	LMCI	AD
Accuracy	98.65	99.15	98.95	98.90	97.85
Accuracy Error	1.35	0.85	1.05	1.10	2.15
Recall	98.50	99.25	98.25	97.75	90.75
Specificity	98.69	99.13	99.13	99.19	99.63
Precision	94.94	96.59	96.56	96.78	98.37
False Positive Rate	1.31	0.88	0.88	0.81	0.38
F1 Score	96.69	97.90	97.40	97.26	94.41

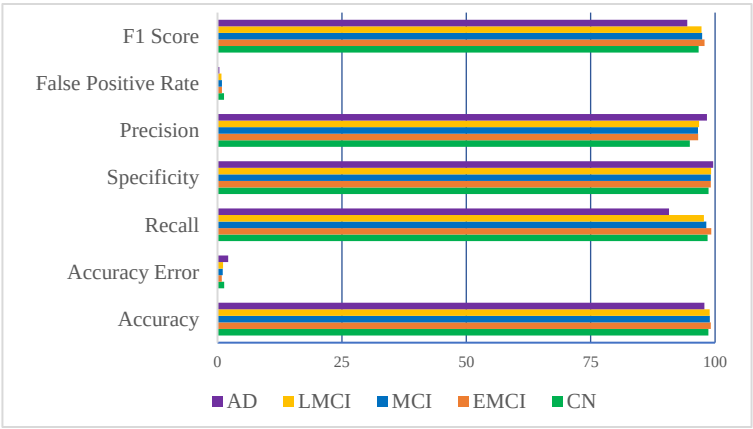


Fig.10. Graphical representation of ResNet-50 model performance

Table 9. Class-based performance metrics for inception V3

Performance Metric	CN	EMCI	MCI	LMCI	AD
Accuracy	97.15	98.25	98.40	97.55	96.35
Accuracy Error	2.85	1.75	1.60	2.45	3.65
Recall	95.50	96.00	95.75	92.25	89.50
Specificity	97.56	98.81	99.06	98.88	98.06
Precision	90.74	95.29	96.23	95.35	92.03
False Positive Rate	2.44	1.19	0.94	1.13	1.94
F1 Score	93.06	95.64	95.99	93.77	90.75

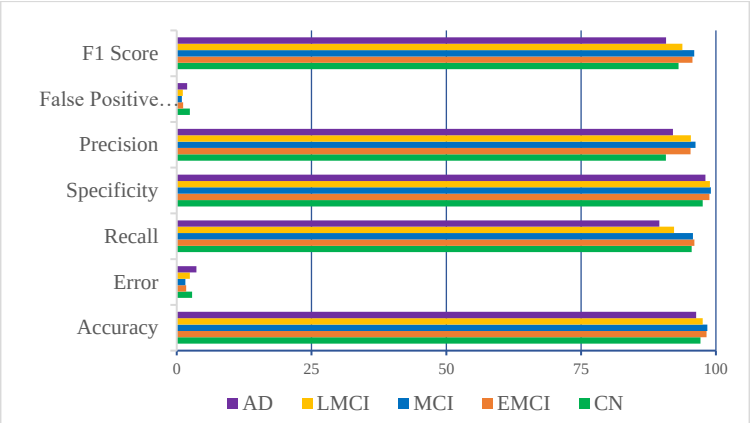


Fig.11. Graphical representation of inception V3 model performance

5. Discussion

The research presented in this study addresses the challenges of early Alzheimer's disease diagnosis using deep learning techniques, specifically Convolutional Neural Networks. The study aims to overcome the limitations of training deep learning models from scratch by introducing a transfer learning-based fine-tuning method for detecting different stages of Alzheimer's disease. The effectiveness of three pre-trained networks, VGG-19, ResNet-50, and Inception V3, is evaluated for this task.

To address the class imbalance in the dataset, the study employs resampling techniques such as oversampling and undersampling. Data augmentation techniques are also utilized to mitigate overfitting and enhance the dataset size, a common challenge in medical image classification due to the limited availability of medical data. The research focuses on learning representations from a more significant number of images to improve the reliability and precision of the results.

One significant aspect of the study is its emphasis on conducting genuine predictions. To ensure this, the model is trained without using the test data, preventing potential biases and overfitting that could compromise the validity of the results.

The evolving landscape of deep learning research highlights the importance of transfer learning. In biomedical image analysis and classification, it becomes essential to identify the most influential and effective models from a diverse range of successful models used in various image classification tasks. The study proposes shifting the focus from creating individualized models for specific subjects or scenarios to developing universally applicable models to promote consistency and reliability across different research studies. This is particularly relevant considering the challenges associated with acquiring and processing medical images and the scarcity of available data. Identifying models that consistently produce dependable results across different imaging techniques is crucial for advancing the field and improving the diagnosis and treatment of medical conditions like Alzheimer’s.

6. Conclusions

Detecting and classifying Alzheimer's Disease in a multi-class setting remains a formidable challenge, necessitating classification frameworks and automated systems for effective disease management. This study introduces a system based on a transfer learning classification model to address the multi-stage detection of AD. Early detection of Alzheimer's disease can play a crucial role in reducing the risk of neuron disorders so that the disease progression can be retarded or even prevented. MCI is considered one of the early stages of AD and can result in a mild yet observable and measurable decline in cognitive abilities. The proposed algorithm leverages pre-trained networks, namely VGG-19, ResNet-50, and Inception V3, and re-trains the convolutional neural network for the multi-class problem. Using MRI images trained on the ADNI database, these models can categorize five NC, EMCI, MCI, LMCI, and AD classes. The deep learning networks VGG-19, ResNet-50, and Inception V3 were built to classify 1000 classes from an image net database. Transfer learning is used to change the networks' architecture to classify the five different AD classes found in the ADNI database. The confusion matrix and its parameters are used to examine the classification performance of these three networks. ResNet-50, Inception V3, and VGG-19 have been found to have 97.85%, 96.35%, and 96.20% overall accuracy in detecting AD, respectively. From the observations, the ResNet-50 outperforms the other two models with a recall rate of 99.25% in detecting early development. Notably, the proposed approach relies on something other than manually crafted features, is swift, and is adept at handling small image datasets.

In future research, there is potential to enhance model accuracy by fine-tuning all convolutional layers. Additionally, exploration of other pre-trained networks or CNN architectures could be beneficial. The utilization of other datasets could

yield comparable or superior results. Lastly, supervised and unsupervised deep learning methods can be explored for the multi-class Alzheimer's disease detection task.

Acknowledgements

The author(s) received no specific funding for this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Srivastava, R. Ahmad, and S. K. Khare, “Alzheimer’s disease and its treatment by different approaches: A review,” Eur. J. Med. Chem., vol. 216, Apr. 2021, doi: 10.1016/J.EJMECH.2021.113320.

[2] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, “Forecasting the global burden of alzheimer’s disease,” Alzheimer’s & Dementia, vol. 3, no. 3, pp. 186–191, 2007. doi:10.1016/j.jalz.2007.04.381.

[3] G. Livingston, J. Huntley, A. Sommerlad, D. Ames, C. Ballard, S. Banerjee, et al., “Dementia prevention, intervention, and care: 2020 report of The Lancet Commission,” The Lancet, vol. 396, no. 10248, pp. 413–446, 2020. doi:10.1016/s0140-6736(20)30367-6.

[4] M. Dadar, A. L. Manera, S. Ducharme, and D. L. Collins, “White matter hyperintensities are associated with grey matter atrophy and cognitive decline in Alzheimer’s disease and frontotemporal dementia,” Neurobiol. Aging, vol. 111, pp. 54–63, Mar. 2022, doi: 10.1016/J.NEUROBIOLAGING.2021.11.007.

[5] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, “Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning,” Proc. - IEEE Symp. Comput. Med. Syst., vol. 2018-June, pp. 345–350, Jul. 2018, doi: 10.1109/CBMS.2018.00067.

[6] R. Petersen, “Early diagnosis of alzheimers disease: Is MCI too late?” Current Alzheimer Research, vol. 6, no. 4, pp. 324–330, 2009. doi:10.2174/156720509788929237.

[7] A. Tahami Monfared, M. J. Byrnes, L. A. White, and Q. Zhang, “Alzheimer’s disease: Epidemiology and clinical progression,” Neurology and Therapy, vol. 11, no. 2, pp. 553–569, 2022. doi:10.1007/s40120-022-00338-8.

[8] D.P. Veitch, M.W. Weiner, P.S. Aisen, L.A. Beckett, N.J. Cairns, R.C. Green, et al., “Understanding disease progression and improving Alzheimer’s disease clinical trials: Recent highlights from the Alzheimer’s Disease Neuroimaging Initiative,” Alzheimers. Dement., vol. 15, no. 1, pp. 106–152, Jan. 2019, doi: 10.1016/J.JALZ.2018.08.005.

[9] C. Plant, S.J. Teipel, A. Oswald, C. Böhm, T. Meindl, J. Mourao-Miranda, et al., “Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer’s disease,” Neuroimage, vol. 50, no. 1, pp. 162–174, Mar. 2010, doi: 10.1016/J.NEUROIMAGE.2009.11.046.

[10] S. S. Kundaram and K. C. Pathak, “Deep Learning-Based Alzheimer Disease Detection,” Lect. Notes Electr. Eng., vol. 673, pp. 587–597, 2021, Accessed: Apr. 05, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-5546-6_50

[11] M. Tamoor and I. Younas, “Automatic segmentation of medical images using a novel Harris Hawk optimization method and an active contour model,” J. Xray. Sci. Technol., vol. 29, no. 4, pp. 721–739, 2021, doi: 10.3233/XST-210879.

[12] X. Liu, C. Wang, J. Bai, and G. Liao, “Fine-tuning Pre-trained Convolutional Neural Networks for Gastric Precancerous Disease Classification on Magnification Narrow-band Imaging Images,” Neurocomputing, vol. 392, pp. 253–267, Jun. 2020, doi: 10.1016/J.NEUCOM.2018.10.100.

[13] M. Hon and N. M. Khan, “Towards Alzheimer’s Disease Classification through Transfer Learning,” Proc. - 2017 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2017, vol. 2017-Janua, pp. 1166–1169, Nov. 2017, doi: 10.1109/BIBM.2017.8217822.

[14] S. Sarraf and G. Tofighi, “Deep learning-based pipeline to recognize Alzheimer’s disease using fMRI data,” FTC 2016 - Proc. Futur. Technol. Conf., pp. 816–820, Jan. 2017, doi: 10.1109/FTC.2016.7821697.

[15] J. Islam and Y. Zhang, “Brain MRI analysis for Alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks,” Brain Informatics, vol. 5, no. 2, pp. 1–14, Dec. 2018, doi: 10.1186/S40708-018-0080-3/FIGURES/10.

[16] K. Oh, Y. C. Chung, K. W. Kim, W. S. Kim, and I. S. Oh, “Classification and Visualization of Alzheimer’s Disease using Volumetric Convolutional Neural Network and Transfer Learning,” Sci. Reports 2019 91, vol. 9, no. 1, pp. 1–16, Dec. 2019, doi: 10.1038/s41598-019-54548-6.

[17] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, “Convolutional neural network based Alzheimer’s disease classification from magnetic resonance brain images,” Cogn. Syst. Res., vol. 57, pp. 147–159, Oct. 2019, doi: 10.1016/J.COGSYS.2018.12.015.

[18] M. Liu, D. Zhang, and D. Shen, “Identifying informative imaging biomarkers via tree structured sparse learning for AD diagnosis,” Neuroinformatics, vol. 12, no. 3, pp. 381–394, Dec. 2014, doi: 10.1007/S12021-013-9218-X/METRICS.

[19] X. Zhao and X. M. Zhao, “Deep learning of brain magnetic resonance images: A brief review,” Methods, vol. 192, pp. 131–140, Aug. 2021, doi: 10.1016/J.YMETH.2020.09.007.

[20] Ebrahimi, S. Luo, and for the A. Disease Neuroimaging Initiative, “Convolutional neural networks for Alzheimer’s disease detection on MRI images,” J. Med. imaging (Bellingham, Wash.), vol. 8, no. 2, Apr. 2021, doi: 10.1117/1.JMI.8.2.024503.

[21] Khan and S. Zubair, “Development of a three tiered cognitive hybrid machine learning algorithm for effective diagnosis of Alzheimer’s disease,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 10, pp. 8000–8018, Nov. 2022, doi: 10.1016/J.JKSUCI.2022.07.016.

[22] B. Lei, E. Liang, M. Yang, P. Yang, F. Zhou, E.-L. Tan, et al., “Predicting clinical scores for Alzheimer’s disease based on joint and deep learning,” Expert Syst. Appl., vol. 187, p. 115966, Jan. 2022, doi: 10.1016/J.ESWA.2021.115966.

[23] Revathi, R. Kaladevi, K. Ramana, R. H. Jhaveri, M. Rudra Kumar, and M. Sankara Prasanna Kumar, "Early Detection of Cognitive Decline Using Machine Learning Algorithm and Cognitive Ability Test," Secur. Commun. Networks, vol. 2022, 2022, doi: 10.1155/2022/4190023.
[24] N. Mahendran and D. R. V. P M, "A deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease," Comput. Biol. Med., vol. 141, Feb. 2022, doi: 10.1016/J.COMPBIOMED.2021.105056.
[25] T. M. Ghazal et al., "Alzheimer Disease Detection Empowered with Transfer Learning," Comput. Mater. Contin., vol. 70, no. 3, pp. 5005–5019, Oct. 2021, doi: 10.32604/CMC.2022.020866.
[26] M. Odusami, R. Maskeliūnas, and R. Damaševičius, "An Intelligent System for Early Recognition of Alzheimer’s Disease Using Neuroimaging," Sensors 2022, Vol. 22, Page 740, vol. 22, no. 3, p. 740, Jan. 2022, doi: 10.3390/S22030740.

Authors' Profiles

Naveen. N. is currently working as an Assistant Professor in the department of Computer Science at Ramaiah University of Applied Sciences, Bengaluru, India. He has completed his bachelor's in Information Science and Engineering, from Visvesvaraya Technological University, Belagavi, India, and Master's degree in Computer Network engineering, from Visvesvaraya technological University, Belagavi, India. He is currently pursuing his Ph. D in the field of Artificial Intelligence under the guidance of Dr. Nagaraj. G. Cholli from Visvesvaraya Technological University, Belagavi. He is interested in research areas including Artificial Intelligence, Machine Learning, Deep Learning, and Image Processing.

Dr. Nagaraj. G. Cholli is currently working as Registrar, Mandya University, Karnataka, India. He has completed his bachelor's in computer science and engineering, from Visvesvaraya Technological University, Belagavi, India, and master's degree in computer science, IIT, Roorkee, India. He holds Doctorate degree from Visvesvaraya Technological University, Belagavi, India. He has a total of 17 years of experience in teaching, research, industry in India and abroad. He has published several research articles in international journals and presented papers at International Conferences. He is active in research, has filed patents and guiding several Ph. D. scholars. He is also a life member of ISTE and CSI society.

How to cite this paper: Naveen. N., Nagaraj. G. Cholli, "Enhancing Early Alzheimer's Disease Detection: Leveraging Pre-trained Networks and Transfer Learning", International Journal of Intelligent Systems and Applications(IJISA), Vol.16, No.1, pp.52-69, 2024. DOI:10.5815/ijisa.2024.01.05