# Performance Evaluation Of Stacked Ensemble Transferred Neural Network Model Using Adni And Oasis Dataset

**Archana W. Bhade[1,2]*, G. R. Bamnote[3]**

[1]Research Scholar Department of Computer Science & Engineering, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, 444701, India
[2]Assistant Professor, Information Technology Department, Government College of Engineering, Amravati, Maharashtra, 444701, India
[3]Professor & Head Department of Computer Science & Engineering, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati, 444701, India
Email: awbhade@gmail.com
*Corresponding Author

Received: 12.04.2024 Revised : 16.05.2024 Accepted: 24.05.2024

**ABSTRACT**
Alzheimer's disease (AD) is the most prevalent type of dementia of the nervous system that causes many brain functions to weaken e.g. memory loss. Early diagnosis of Alzheimer's disease has been shown to improve patient outcomes. Machine learning techniques that utilize magnetic resonance imaging (MRI) have been used for Alzheimer's disease diagnosis, but traditional methods require manual feature extraction by an expert, which can be complex. To address this problem, our study proposes a new ensemble learning approach using a stacked ensemble model of pre-trained convolutional neural networks as base learners and logistic regression as meta learner called Stacked Ensemble Transferred Neural Network (SETNN) model for classification of MRI images to identify Alzheimer's disease. The efficiency of the SETNN model, compared to conventional Softmax and support vector machine (SVM) methods, was evaluated using various metrics like confusion matrix, precision, accuracy and other. The suggested SETNN model performed better than other modern algorithms according to the results by achieving an accuracy of 96% when using the MRI images from OASIS dataset and achieved accuracy of 94% for ADNI dataset.

**Keywords:** Alzheimer disease, Machine learning, Stacked Ensemble Transferred Neural Network, ADNI, MRI

## 1. INTRODUCTION

Named after the German psychiatrist Alois Alzheimer, Alzheimer's disease (AD) is a neurological disorder that is the most common type of dementia. It is marked by a slowly progressive neurodegenerative illness. It impairs memory and cognitive judgement over time. A lot of the time, there are also mood swings, confusion, and finally delirium. Typically, it begins as mild cognitive impairment (MCI), which is an imperceptible memory loss that gradually gets worse until it becomes totally incapacitating [1]. It usually starts with a subtle and unnoticed memory failure (commonly referred to as moderate cognitive impairment or MCI) and progresses to become more severe and, finally, incapacitating. Hallucinations, poor judgement, language disorder, visual complaints, agitation, withdrawal, and confusion are a few other frequent results [2]. On rare occasions, mutism, myoclonus, increased muscular tone, Parkinsonian features, seizures, and Parkinsonian characteristics might all be present. Pneumonia, general malnutrition, and syphilis are the main causes of death. [2, 3].
The stages of Alzheimer's disease are divided into pre-clinical or pre-symptomatic, moderate or early stage and severe AD. The comparison of brain of normal and alzheimer patient is shown in figure 1. The first stage is marked by mild memory loss, early hippocampal and cortical degenerative changes, minimal functional impairment in day-to-day activities, and the lack of clinical AD signs and symptoms [4]. In the second stage patients experience a range of symptoms, such as difficulties doing everyday tasks due to memory and focus issues, confusion about where they are and when they are, mood swings, and the onset of depression [4]. When the disease has progressed to the entire cortex, it is referred to as, also known as late-stage AD. Patients eventually succumb to these complications, lose all ability to identify their family, become bedridden, experience difficulties swallowing and urinating, and experience gradual functional and cognitive impairment [4].

At the age of 45, one in five women and one in ten men will have Alzheimer's disease at some point in their lives [5]. India is home to 18% (1.37 billion) of the world's population. To find out how common dementia is there, a study led by professor of economics Jinkook Lee and funded in part by the National Institute on Ageing was conducted [6]. According to Lee's team's estimation, 8.8 million persons over the age of 60 in India have dementia, with the incidence rate being 7.4%. By 2050, India's share of seniors is projected to increase to nearly 20% of its population—319 million individuals, as shown in Figure 2 [6]. With age the strongest risk factor for Alzheimer's and related disorders; India faces an alarming potential increase in the number of people with dementia [6].

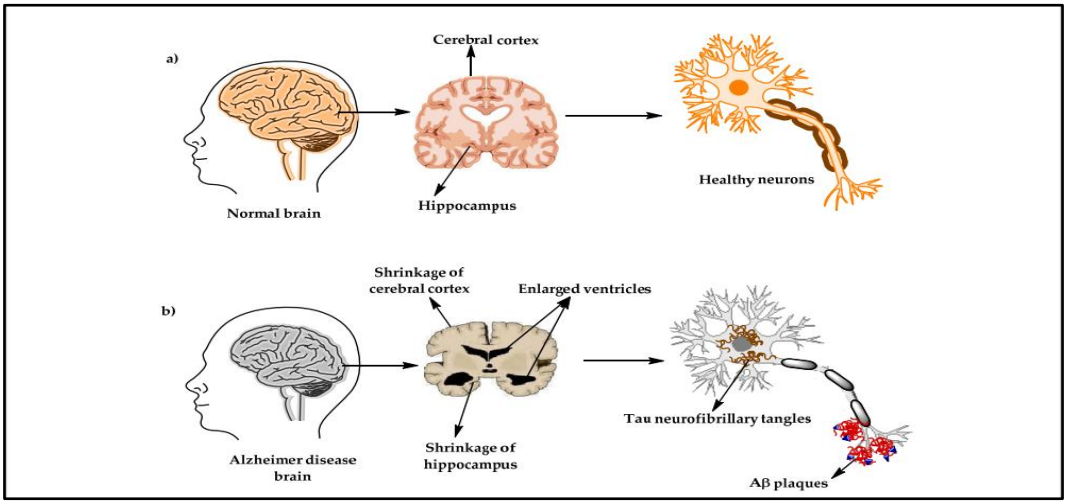


**Figure 1.** The physiological structure of the brain and neurons in (a) healthy brain and (b) Alzheimer's disease (AD) brain [4]

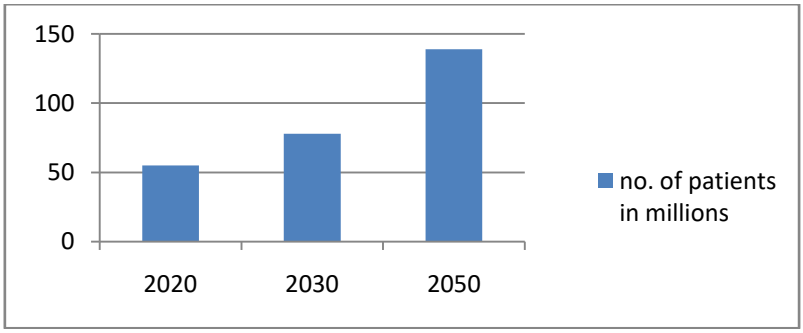


**Figure 2.** Worldwide dementia disease statistics [6]

Since disease-modifying therapies (DMTs) for Alzheimer's disease (AD) are becoming a reality, it is imperative to select affordable techniques that can accurately identify patients in the earliest stages of the disease. The bulk of medical scientists are therefore drawn to innovative methods for machine learning-based disease prediction [7, 8, 9, 10, 11].

Machine learning algorithms can anticipate a disease's early diagnosis and offer treatments by developing a model. The process of features engineering, or manual selection and extraction of features is the primary problem with traditional learning-based methods because it significantly affects the model's performance by reducing the robustness of the solution [12]. While several machine-learning models have been applied to automated prediction of neurological disorders, main area of advanced research is deep learning (DL) based diagnosis models. Using neural networks to learn directly from images, Deep Learning has automated the process without requiring human expertise in feature extraction, as opposed to manually extracting the features and performing them separately from the classifier [13]. Acquiring a substantial quantity of labelled data is a challenging task for deep learning-based methods. Lately, convolutional neural networks (CNNs) have achieved extraordinarily high precision and accuracy in image classification [14]. Numerous initiatives have been taken to develop an alternative for the fully automated detection of Alzheimer's disease that is both precise and reliable.

To address these problems, we proposed a workable alternative in this study (1) Robust and discriminative deep features are extracted from brain magnetic resonance imaging (MR) and used to

classify Alzheimer's disease using pre-trained models. A wide range of pre-trained deep convolutional neural networks (CNNs) are experimented with as feature extractors. (2) Also, extensive experimentation done by using hyperparameter tuning optimization on different pre-trained deep convolutional neural networks (CNNs) models to improve the performance. And (3) In order to provide accurate findings for MRI-based Alzheimer's disease detection, we also created a stacked ensemble of different pre-trained models as base models and logistic regression model as meta model and called SETNN (Stacked Ensemble Transfer Neural Network). This research introduces a novel method for automated classification of MRI images to find Alzheimer's disease employing Stacked Ensemble Transfer Neural Network (SETNN. The efficiency of the SETNN model, compared to conventional Softmax and support vector machine (SVM) methods, was evaluated using various metrics, including confusion matrix, precision, recall, F1score and accuracy. Because the differences between MCI groups are expected to be smaller than those between AD and controls, models trained with AD and control participants can be particularly helpful when attempting to distinguish between c-MCI and s-MCI patients [15]. Individuals with MCI who remain in the MCI stage are said to have stable MCI (s-MCI), whereas those who are moving towards AD have c-MCI [16].This study proposes a novel solution for early detection of Alzheimer's disease using MRI images, which offers high precision and high accuracy. The proposed model utilizes a technique that aims to achieve higher optimization during training compared to existing methods in the literature. This optimization would reduce the computational power required for training, making the model more practical for practitioners and researchers. The study model combines deep learning and transfer learning models, results in an excellent level of accuracy that outperforms the functionality of competing alternatives.

The following research questions are addressed by the goals of the study. (1) Can Alzheimer's illnesses be identified in MRI brain pictures using the pre-train MobilenetV2 and Deep Learning CNN technique utilised in this study? (2) Softmax or SVM is the classifier that will perform better when applied with pre-trained SETNN in terms of classification.

## 2. LITERATURE REVIEW

Research is an on-going process in which researchers examine raw data from various angles. The growing interest in ML and DL for the identification of Alzheimer's disease led to the proposal of several methodology as well as common approaches for researchers. The detection is automatic. Because this study is more concerned with the DL methodology, we will we will limit our discussion to only DL models from the literature.

Loddo et al. [7], employed a deep ensemble strategy for classifying different MRI and fMRI image datasets through binary as well as multi-classes using AlexNet, ResNet-101 with Inception-ResNet-v2. Additionally, the study used data augmentation to enhance the amount of imagery before training the model. According to the stated findings, their suggested model achieves 98.67% accuracy for multiclass classification and 98.51% accuracy for binary classification.

Mahendran et al. [8], utilised a DL architecture with a feature selection technique incorporated for Alzheimer's disease detection at an early stage. Using this procedure, feature selection was carried out using typical machine learning techniques, and model evaluation was done using k-folds cross-validation. Recurrent neural network (RNN), CNN, deep RNN, and enhanced deep RNN—gives the best results when it came to feature selection. They found AdaBoost with 5-fold cross-validation was efficient for classification. EDRNN scored 88.7%, which was higher than the scores of the others with respect to classification accuracy.

Sava [9], on an ADNI dataset, the researchers employed various models using the architecture of CNN for classifying various phases of Alzheimer's disease. They tested the accuracy, precision, sensitivity, and specificity of the learning models using 29 distinct pre-trained models on MRI scans. EfficientNetB0, one of these models, attained a high accuracy of 92.98%, EfficientNetB3 precision was 89.78%, EfficientNetB2 sensitivity was 94.42%, and EfficientNetB3 specificity was 97.28%.

Murugan et al. [10], attempted to detect four stages of Alzheimer's disease using a CNN-based DEMNET (dementia network) model. The study's dataset, which included 6400 MRI images, was obtained from Kaggle. The SMOTE method was used for data augmentation. Results from the DEMNET model have an accuracy score of 95.23%.

Mohammed et al. [11], DL and hybrid DL methods were used to analyse MRI images for early detection of AD and dementia. They trained AlexNet and ResNet-50 models, as well as AlexNet+SVM and ResNet-50+SVM hybrid models, using OASIS and KaggleMRI datasets. The SMOTE approach was used for balancing the groups in an OASIS dataset, whereas data augmentation was employed in the Kaggle dataset. On the Kaggle MRI dataset, AlexNet+SVM beat other approaches after augmentation, obtaining an exceptional accuracy rate of 94.8%.

Gharaibeh et al. [12], used DL to detect Alzheimer's disease using a dataset of digital subtracted angiogram scans of Alzheimer's patients. Pre-trained models, such as InceptionV3 and DenseNet201, were used for feature extraction. To increase the number of scans in the dataset, data augmentation was used, and with variance ratio requirement of 0.99, features were chosen using principal component analysis (PCA). After training, the model has an accuracy rating of 99%.
Basher et al. [13], suggested a CNN and DNN-based volumetric feature-based AD diagnostic technique. The left and right hippocampi sMRI scans were used to extract volumetric features for the suggested AD detection approach. South Korea served as the collection site for the GARD MRI scan dataset. The suggested models' average accuracy on the pre-processed GARD dataset was 94.82% and 94.02% for left and right hippocampi respectively.
Islam and Zang [17] proposed a deep convolutional neural network for Alzheimer's disease diagnosis using brainMRI data analysis. Their model can identify different stages of Alzheimer's disease and obtains superior performance for early-stage diagnosis. The accuracy of the proposed model is 93.18% with 94% precision, 93% recall and 92% f1-score.

## 3. MATERIALS AND METHODOLOGY

### 3.1. Dataset

In this study we have used ADNI (Alzheimer's Disease Neuroimaging Initiative) and OASIS (Open Access Series of Imaging Studies) datasets. The overall goal of ADNI is to validate biomarkers for AD clinical trials. ADNI is a complex and unique collection of data, imaging and biospecimens gathered longitudinally from carefully phenotyped subjects, as shown in Table 1. It has massive potential for breakthrough discoveries in the field of Alzheimer's research [18].

**Table 1.** Details of ADNI1 dataset, such as the number of individuals, the number of images, descriptive age statistics, the percentage of women in the images compared to men, and the percentage of 1.5 T field strength images compared to 3.0 T images.

| Dataset | Subject | Groups | Images | Age(years) | | | | Female (%) | 1.5T (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Med | Avg ± Std | Min | Max | | |
| ADNI1 | 845 | All | 9149 | 76.6 | 76.3 ± 6.9 | 54.6 | 93 | 42.2 | 82.2 |
| | | CN | 2701 | 76.7 | 77.2 ± 5.1 | 60 | 92.8 | 50.2 | 80.5 |
| | | MCI | 4845 | 76.5 | 76.0 ± 7.4 | 54.6 | 90.9 | 35.3 | 83 |
| | | AD | 1603 | 76.5 | 76.1 ± 7.9 | 55.2 | 93 | 49.5 | 82.5 |

OASIS-3 and OASIS-4 are the latest releases in the Open Access Series of Imaging Studies (OASIS) that is aimed at making neuroimaging datasets freely available to the scientific community [19]. By compiling and freely distributing this multimodal dataset generated by the Knight ADRC and its affiliated studies, we hope to facilitate future discoveries in basic and clinical neuroscience. Previously released data for OASIS-Cross-sectional [20] and OASIS-Longitudinal [21] have been utilized for hypothesis driven data analyses, development of neuroanatomical atlases, and development of segmentation algorithms. OASIS-3 is a longitudinal multimodal neuroimaging, clinical, cognitive, and biomarker dataset for normal aging and Alzheimer's Disease. OASIS-4 contains MR, clinical, cognitive, and biomarker data for individuals that presented with memory complaints [21].
Used OASIS MRI dataset consists of around 86437 brain MRI images of 461 patients, as shown in Table 2. The dataset aims to provide a valuable resource for analyzing and detecting early signs of Alzheimer's disease. Original .img and .hdr files were converted into Nifti format (.nii) using FSL (FMRIB Software Library) to make the dataset accessible.

**Table 2.** Class wise distribution of OASIS dataset

| **Class** | **Class Name** | **No of images** |
|---|---|---|
| **0** | Non-Demented | 67222 |
| **1** | Very mild Dementia | 13725 |
| **2** | Mild Dementia | 5002 |
| **3** | Moderate Dementia | 488 |
| Total images | | 86437 |

For the neural network training, 2D images were used as input. The brain images were sliced along the z-axis into 256 pieces, and slices ranging from 100 to 160 were selected from each patient. Because of this

approach a comprehensive dataset was available for analysis. The proposed framework of model is shown in figure 3.

**Algorithm**
**ProposedStacked Ensemble Transfer Networking (SETNN) Model**
1. **Input:** Set of MRI images from ADNI dataset, **N:** No. of images
2. **Output:** Model, result, graphs
3. **Begin:**

//Image preprocessing
for i=1 to N
Resize image from (300x300) to (224x224)
Initialization
Divide data set 80% for training and 20% for testing
Set the epoch to 5
Set the batch size to 32
//initialize model 1 using mobilenetV2 pretrained model
Freeze feature layer
Establish and prepare learning model
For i= 1 to N*0.8
Normalize the image
Set up for data augmentation
Pretrain the model with MobilenetV2
Compile the model
Initialize ADAM optimizer and apply loss function (sparse cross entropy)
Train the model and save the model
Check the model
Get the results
//initialize model 2 using VGG16 pretrained model
Freeze feature layer
Establish and prepare learning model
For i= 1 to N*0.8
Normalize the image
Set up for data augmentation
Initialize ADAM optimizer
Pretrain the model with VGG16
Compile the model
Initialize ADAM optimizer and apply loss function (sparse cross entropy)
Train the model and save the model
Check the model
Get the results
//initialize model 3 using InceptionV3 pre-trained model
Establish and prepare learning model
Initialization
For i= 1 to N*0.8
Normalize the image
Set up for data augmentation
Initialize ADAM optimizer
Pretrain the model with InceptionV3
Compile the model and apply loss function (sparse cross entropy)
Train the model and save the model
Check the model
Get the results
//Ensemble Stacking model creation
Model= (model1, model2, model3)
Take model as base model and model 1 as meta model
Train the model and save the model
Check the model
Get the results
//testing phase

For i= 1 to N* 0.2
Load the model
Evaluate the model for testing the data
Measure confusion metric, precision, accuracy, specificity, loss, f1- score
Get results and graph

### 3.2. MRI Pre-processing

The initial stage of image processing, known as pre-processing, is typically applied to MRI images before they are further processed to detect Alzeimer's disease. Larger-than-needed MRI images will be downsized by a certain number of pixels in accordance with image processing requirements. Brightness, contrast, and other image quality adjustments will follow. Noise removal from MRI image is important, because in presence of noise processing the images and other required pre-processing will become complicated. The dataset has undergone initial pre-processing steps during acquisition, such as Grad warp, B1 non-uniformity, and N3 bias field correction.

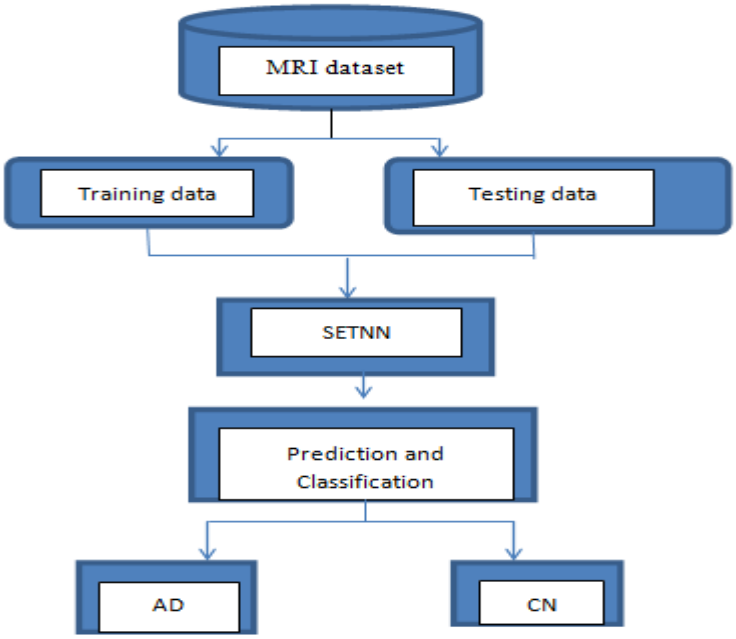


**Figure 3.** Flowchart of proposed framework

To prepare the data for our model, several pre-processing steps were applied:

- Spatial normalization was performed to validate the image position. In spatial normalization, regardless of differences in subject-to-subject variability in brain size, shape, and microarchitecture, images are processed to verify that the voxels being compa red represent the same brain regions [22]. The voxels in each brain picture are "registered" to represent the same region of the brain during this process. Typically, the pictures' voxels are registered to the voxels of an accepted "template" brain image. For image normalization, the pixel values are rescaled to [-1,1] using a pixel-wise multiplication factor as follows:

$$\widehat{IN} = \left(I - I_{MIN}^{\widehat{T}}\right) \frac{I_{MAX}^{\widehat{T}} - I_{MIN}^{\widehat{T}}}{MAX - MIN} + I_{MIN}^{\widehat{T}} \quad (1)$$

  where I and $\widehat{IN}$ represent input and normalized brain image, respectively, $I_{MIN}^{\widehat{T}}$ and $I_{MAX}^{\widehat{T}}$ are the normalised image's intensity range, and Min = 0; and Max = 255 represent the input brain image's pixel intensity range.
- Then we performed intensity normalisation, noise reduction, correction of bias, adjustments in contrast. The MANGO toolbox [23] was then used for same.
- Then rescaling is done. Each 3-dimensional volume into 2-dimensional layers with one channel for each region (axial, coronal, and sagittal) and a size of 300 x 300. The 256 x 256 x 166 slices that are contained in each 3D MRI volume make it impossible to directly feed them in 2D-CNN network.
- Then we performed cropping of images. To improve the classification performance of brain MRI images in our datasets, it is necessary to remove undesired spaces and areas through cropping.

Completely black portions of each of the images were removed, as shown in figure 4. To do this, extreme points (extreme top, extreme bottom, extreme right, and extreme left) were computed [18].

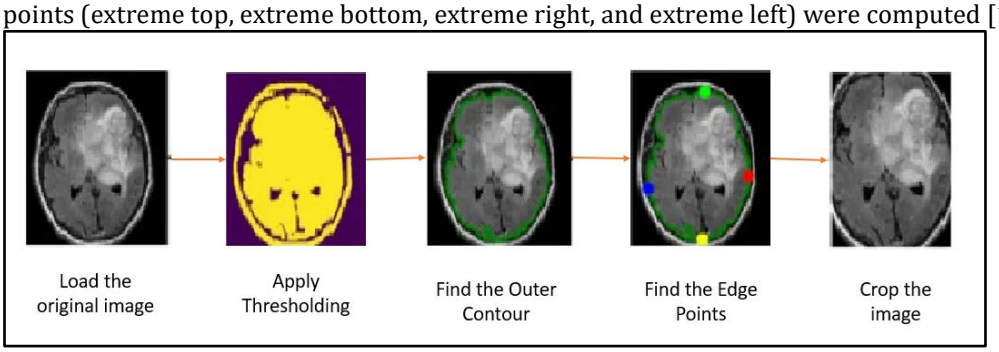


**Figure 4.** Steps to crop the magnetic resonance (MR) images

Cropped MR images then resized through bicubic interpolation [24]. Then, we extracted three crops from each image, each for one of the image planes: axial or horizontal plane, coronal or frontal plane, and sagittal or median plane. We used coronal planes for our study. All the pre-processing steps applied on MRI images is shown in figure 5.

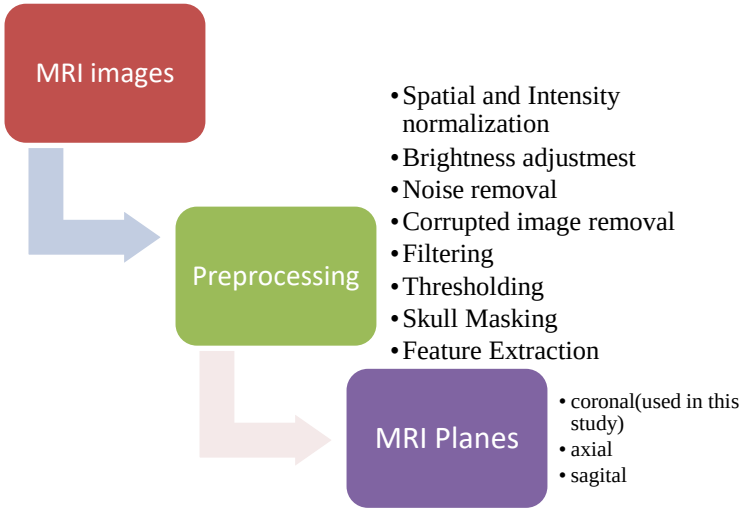


**Figure 5.** Basic Pre- processing steps

For AD, the availability of a large number of scans is a major problem in neuroimaging research because of patient's privacy issues. Additionally, a small imbalanced dataset can generate overfitting problems affecting the efficiency of the model. Thus, data augmentation is usually employed to overcome data unavailability and class imbalance. Image augmentation involves creating an artificial dataset by modifying the original images, which can include variations in scale, rotation, horizontal flipping, brightness, and other factors for generation of new training sets.

Then we splitted the data into training and testing part. For our work, we use 80% data from dataset as training set and 20% as test dataset. From the training dataset, a random selection of 20% images is used as validation dataset. To avoid data imbalance problem, we have used stratified sampling while performing train test split.

### 3.3. Convolutional Neural Networks (CNN)

CNNs have a hierarchical system with input layer, hidden layers consisting convolutional layers, dropout layers, and activation layers and finally output layer. In particular, two convolutional layers in two dimensions (2D) were employed in this study, each of which included a 2D max pooling. The convolutional layer is a crucial and fundamental building block of a Deep Learning Convolutional Neural Network (CNN) [25]. Convolution is a linear operation between the input and a kernel (or filter) that acts as a feature detector extracting low-level features such as colours, edges, blobs, and corners by convolving the input image with these filters during the training process. The output of convolution layer is known as feature maps. Number of filters are fixed in number. We have designed both convolutional layers with 256 filters. The size and type of activation methods used in CNNs can vary based on the

number of layers implemented and are determined empirically through trial and error. Mostly Relu (Rectified linear unit) [26] is used as an activation function because of its fast training speed. Immediately after each convolutional layer, it is convention to apply a nonlinear layer. This layer (ReLU in our experiment), which just changes all the negative activations to 0, increases the nonlinear properties of the model and the overall network without affecting the receptive fields of the convolutional layer.

CNNs have a hierarchical system with input layer, hidden layers, convolutional layers, batch normalization, and activation methods. The size and type of activation methods used in CNNs can vary based on the number of layers implemented and are determined empirically through trial and error. Mostly Relu (Rectified linear unit) [26] is used as an activation function because of its fast training speed. We also have used the same in our experiment as an activation function. The convolutional layer is a crucial and fundamental building block of a Deep Learning Convolutional Neural Network (CNN). The proposed CNN model consists of an input, an output, and multiple designed layers. In particular, two convolutional layers in two dimensions (2D) were employed in this study, each of which included a 2D max pooling layer. Convolution is a linear operation between the input and a kernel (or filter) that acts as a feature detector extracting low-level features such as colours, edges, blobs, and corners by convolving the input image with these filters during the training process. The output of convolution layer is known as feature maps. Number of filters are fixed in number.

Equation (2) shows design of convolution layer.

$$x_n^r = \alpha\left(\sum_{n-1}^{k} x_{n-1}^r * w_{mn}^r + b_m^r\right) \qquad (2)$$

where, $x_n^r$ represents $n^{th}$ activation map of the current $r^{th}$ layer, $x_{n-1}^r$ is the $(n\text{-}1)^{th}$ activation map of the previous $r^{th}$ layer, and k is the number of input activation maps. $w_{mn}^r$ and $b_m^r$ are weight and bias vectors. The * operator is used for convolution operation and $\alpha$ denotes the activation function.

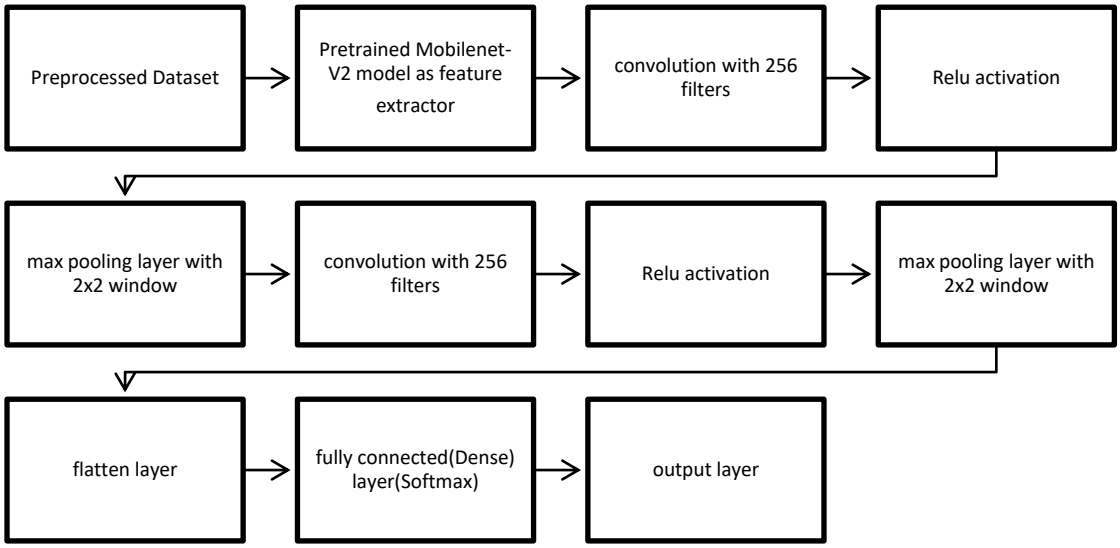


**Figure 6.** Basic Transfer learning model

Pooling layers are typically placed after the convolutional layers. After applying the activation function to each activation map, they are then sent to the pooling layer, as shown in figure 7. By lowering the resolution of the activation maps, the pooling layer offers translation invariance [27]. The d × d (e.g. d = 2) window of activation maps in convolution layer generates pooling layer activations value. Max pooling is the most commonly used pooling method which reduces the size of the feature maps by selecting the maximum value from non-overlapping regions of the image. Max pooling helps in avoiding overfitting by providing an abstract representation of the image and reduces computational cost by reducing the number of parameters [27].

The optimizer plays a significant role in training the deep CNN model by iteratively changing the parameters of all the layers in the network. After every iteration, the desired output and predicted output are compared, and the error is back-propagated. One of the most popular performance measurement metrics is categorical cross-entropy when there are more than two classes. The categorical cross-entropy value is near to zero when the desired output and predicted output are exactly the same, and this is the main aim of any optimization technique [28].

The dropout layer is used to address the issue of overfitting in the network. During training, neurons are randomly dropped out based on a dropout rate parameter, which determines the likelihood of neuron removal. As CNN's run internally with convolutions in multiple sliding windows, these models will locally distinguish patterns and thus allow a stronger distinction between which each class is represented [26].
The fully connected layer creates a categorization map using the information from all of the previous layer's activation maps.
Both Conv + ReLU blocks shown in figure 6 have 256 filters following the last fully connected layer, a dropout(MaxPool) layer (p = 0.2) was added after each convolution before being linked to next layer with ReLU activation. The model can extract patterns from the input data, and deliver it to the next layers. Finally, softmax activation with 3 neurons supplied the model output [29]. This layer of the network is in charge of categorization by calculating the likelihood that the input supplied belongs to a specific label. The training was carried out for a maximum of 4 epochs at a learning rate of 0.0001, using the Adam optimizer to estimate model parameters and a batch size of 32 and utilising the categorical cross-entropy as a loss function computed on the classifier output.
The categorical-cross entropy is defined as

$$H(p, q) = -\sum_x p(x)\log(q(x)) \quad (3)$$

where p is the true distribution and q is the computed distribution.
The only difference in categorical cross entropy loss and sparse cross entropy loss is we use one hot encoded label in categorical cross entropy while integer labels in case of sparse entropy loss. Loss function for both losses is same. This technique helps in regularizing the network and prevents over-reliance on specific neurons during training [30]. The choice of hyperparameters such as number of filters in the convolution layers and the depth of the CNN is of high importance to ensure that our CNN model generalizes well. These are chosen carefully through cross validation. Basic Transfer learned model with either pre-trained model mobilenet V2, VGG16 and InceptionV3 is as shown in figure 6

### 3.4. Pre-trained models

A machine learning (ML) model that has been trained on a large dataset and is ready to be adjusted for a particular task is known as a pre-trained model. ML models are frequently developed using pre-trained models as a foundation since they offer a baseline set of weights and biases that may be adjusted for a particular task [31].
CNN may be trained from scratch or using pre-trained models [31], such as DenseNet [32], MobileNet [33], and InceptionV3[34]. Pretrained models are the most successful strategy for medical imagines categorization due to restricted training data. Pre-trained models are neural network models that have been trained on big benchmark datasets such as ImageNet [31, 34]. The Deep Learning community has tremendously benefited from these open-source models. Stanford University maintains the ImageNet collection [34], which includes approximately 14 million images which fall under several categories. It is frequently utilised for a wide range of image-related deep learning tasks.

#### 3.4.1. Pre-trained InceptionV3 model

In 2019, the Google team presented the InceptionV3 CNN [34]. InceptionV3's architecture was changed based on the InceptionV1 model. It solved several of the concerns identified in the preceding inceptionV1, including auxiliary classifiers with batch normalisation and representation bottlenecks with kernel factorization. The InceptionV3 architecture enables a wide range of kernels (i.e. kernel sizes) on the same level. This structure seeks to address the issue of high variability in the location of notable sections in the input pictures under investigation. The InceptionV3 use lower filter sizes (1 × 7 and 1 × 5) rather than bigger filters (7 × 7 and 5 x 5).

#### 3.4.2. Pre-trained VGG16 model

The VGG-16[35] networks were featured at the ILSVRC 2014 conference since they are among the most popular pre-trained models. The University of Oxford's Visual Graphics Group invented it. VGG16 is a more detailed convolutional neural network model. The underlying principle of this model is to increase the CNN model's depth by replacing big kernels with smaller kernels. As a result, the VGG16 may become more reliable in completing classification tasks. It consists of five 41-layer blocks: 16 with learnable weights, 13 convolutional layers, and three FCC layers. There are two convolutional layers in the first two blocks and three convolutional layers in the next three blocks. 1-pixel padding is used to 3 × 3 kernels in convolutional layers. Having a filter size of 2 × 2 and padding of 1, max-pooling layers are used to split convolutional layers. The last convolutional layer's output is 4096, hence the FCC contains 4,096 neurons [35].

### 3.4.3. MobileNetV2 Architecture

MobileNetV2 [33] is a powerful image classification tool. TensorFlow provides the image weights in MobileNetV2, a lightweight CNN-based deep learning model. It is based on an inverted residual structure where the residual connections are between the bottleneck layers. The intermediate expansion layer uses lightweight depth wise convolutions to filter features as a source of non-linearity. As a whole, the architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. First, the MobileNetV2 base layer is removed, and a new trainable layer is added. The model analyzes the data and extracts the most relevant features from our images.

### 3.5. Transfer learning

Transfer learning involves transferring weights to a corresponding model from a previously learned network. This approach is crucial for problems where there is a dearth of training data. The network could overfit with insufficient data, which would impede generalisation. The parameters of the transferred network offer proper categorization of moderate quantities of input when the pre-trained model's training dataset is large enough. In the last step, the classifiers of the new model were trained with the expected weights from the pre-trained model [36].

CNN models require a significant amount of memory and processing power to train, and over-fitting issues are typically a hindrance. Additionally, a substantial quantity of training dataset is needed. Recent research in this area has shown that sufficiently adjusted pre-trained CNN models function far more reliably than those that are either taught from scratch or, in the worst-case scenario, perform the same. This method is applied to deep learning techniques when the CNN is trained on the large-scale base dataset (source domain), such as ImageNet. Subsequently, the convolutional layer weights are moved to the newly created tiny dataset (target domain) [36].

There are two primary approaches while using pre-trained models for classification tasks: fine-tuning the models and freezing the layers of the pre-trained model [35]. In the first case, a deep CNN model's convolutional layers are frozen, and the last FCC is skipped over when extracting features. Following that, a particular classifier receives these features. In contrast, in the latter scenario, some hyper-parameters are changed and the layers are fine-tuned to tackle a new task. Additionally, the completely linked layer at the top is modified for the target domain. The number of classes in the ADNI dataset is taken into consideration while configuring the number of neurons in layer, three in this study.

By making the most of the source field (i.e., ImageNet), TL seeks to increase the accuracy of the target field. As part of our study to improve the diagnostic performance of Alzheimer's disease image categorization, we adjusted the weights of three powerful pre-trained CNN models - MobileNetV2, VGG16, and InceptionV3. Basic Transfer learned model with either pretrained model mobilenet V2, VGG16 and InceptionV3 is as shown in figure 6.

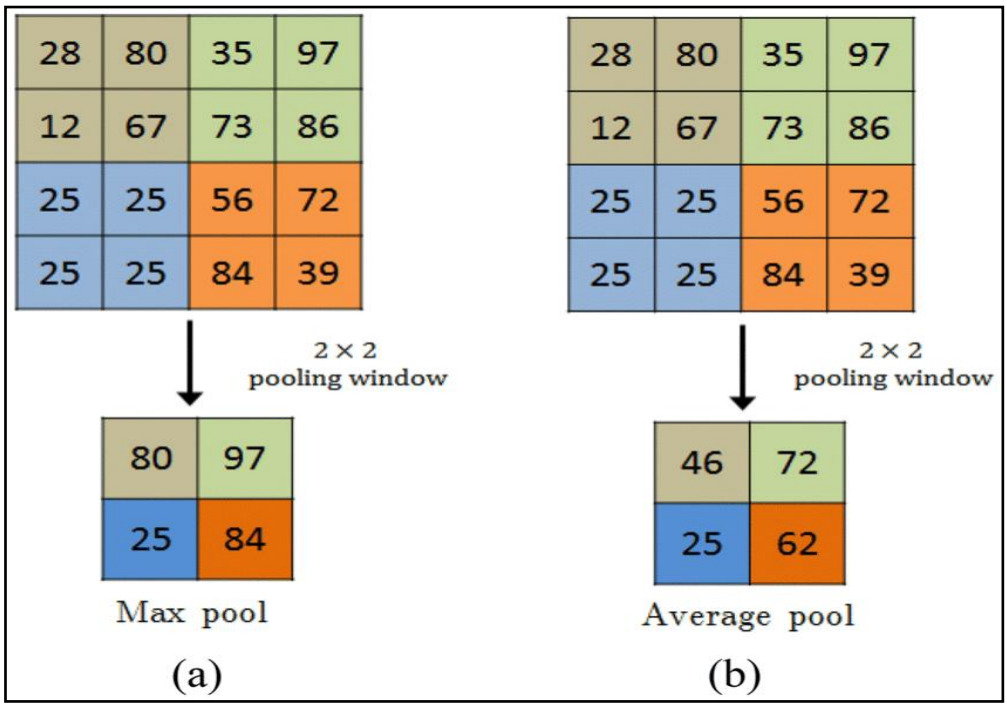


**Figure 7.** Max pool vs Average pool

### 3.6. Ensemble Learning

Ensemble learning models works just like a group of diverse experts teaming up to decide. Different models often of the same or different types team up to enhance predictive performance. It's all about leveraging the collective wisdom to overcome individual limitations and make more informed decisions in ML tasks. When used for binary or multiclass image classification tasks, a single deep CNN (also known as a weak learner) model's limited ability to extract discriminative features from images leads to subpar classification performance. The goal of ensemble learning for deep pre-trained models is to improve classification performance by combining the judgements of several weak learners (individuals). When pre-trained models are learned together, they perform well in a variety of image classification applications. Mainly four types of ensemble learning techniques are there: Bagging, Boosting, Augmenting and stacking. In our research we are using stacking method.

Thus, by incorporating a series of CNN-based transfer learning models, this research closes the gap left by other studies on the categorization of ADNI/OASIS images for Alzhiemer's diseases. After then, embrace the advantages of ensemble learning, we will use outputs of base models to train meta model to take accurate decision. An ensemble of transfer learning networks can be a reliable method for decreasing mistakes. It generates optimal outputs from the integrated networks with the fewest potential faults. Following data preparation, pre-trained models are used to build the convolutional neural network architecture. The previous subsections outline the key components of the employed models.

### 3.7. MRI Image Classification

The fully connected (FC) layers of the Convolutional Neural Network (CNN) can be substituted with other classifiers, such as logistic regression or Support Vector Machines (SVMs), which are specifically optimized for classification tasks [37]. In this project, we will evaluate the performance of SETNN and SVM classifiers.

**Softmax Classification Layer**

Typically, the Softmax function is used in the last layer of the CNN architecture for classifying labeled data. It calculates the probability of each ground-truth label for the outputs, ranging between 0 and 1, and converts the output values into interpretable values [37].

The predicted probabilities of the classes via softmax function is given by following equation:

$$P_i^j = \text{softmax}^j(O_i) = \frac{O_i^j}{\sum_{k=1}^{K} \exp(O_k^j)} \qquad (4)$$

**SVM Classification**

SVMs are commonly used for binary image classification tasks, such as AD vs. NC, and have shown promising results in real-world problems. By using the Radial Basis Function (RBF) kernel shown in equation (5), the SVM classifier can generate a nonlinear classifier that maps the original dataset to a higher-dimensional space by creating linearly separable data [37].

$$\mathbf{K(x, x_i) = \exp(-gamma * ||x - x_i||^2)} \qquad (5)$$

### 3.8. Proposed Ensemble Model

As show in figure 8, in our proposed Stacked Ensemble Transfer Neural Networks (SETNN) model, Transfer learned models with InceptionV3, VGG16, and MobilenetV2 are used to make first-level predictions from the dataset, divided into training and test sets. The three models are each trained from scratch on the dataset for classification of classes of MRI images. Outputs from the first-level prediction models are combined and sent as input to the logistic regression for second-level prediction. The new test data is used to evaluate the performance and progress of the algorithm training and adjust or optimize it for improved results. Logistic regression model will make the final prediction and classifications.

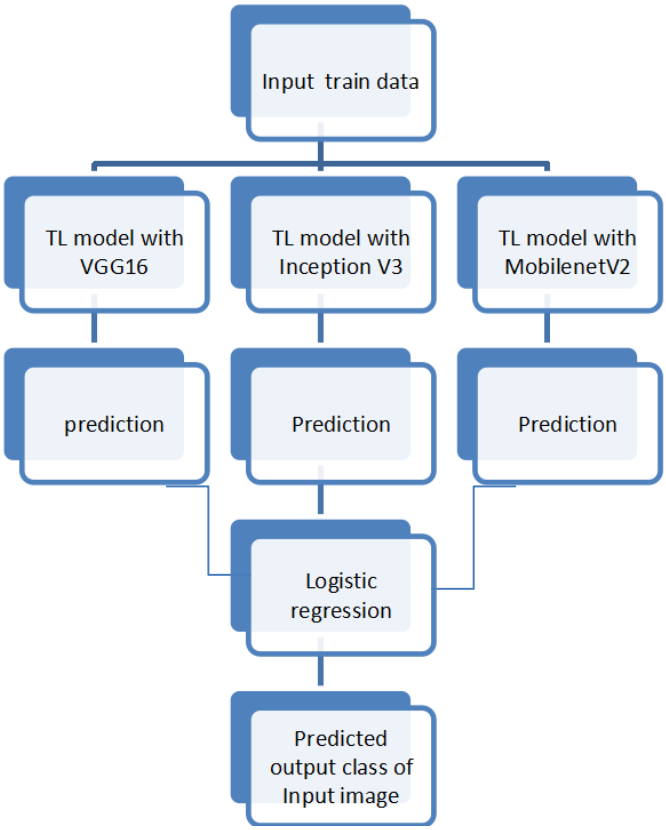


**Figure 8.** Internal Architecture of SETNN model

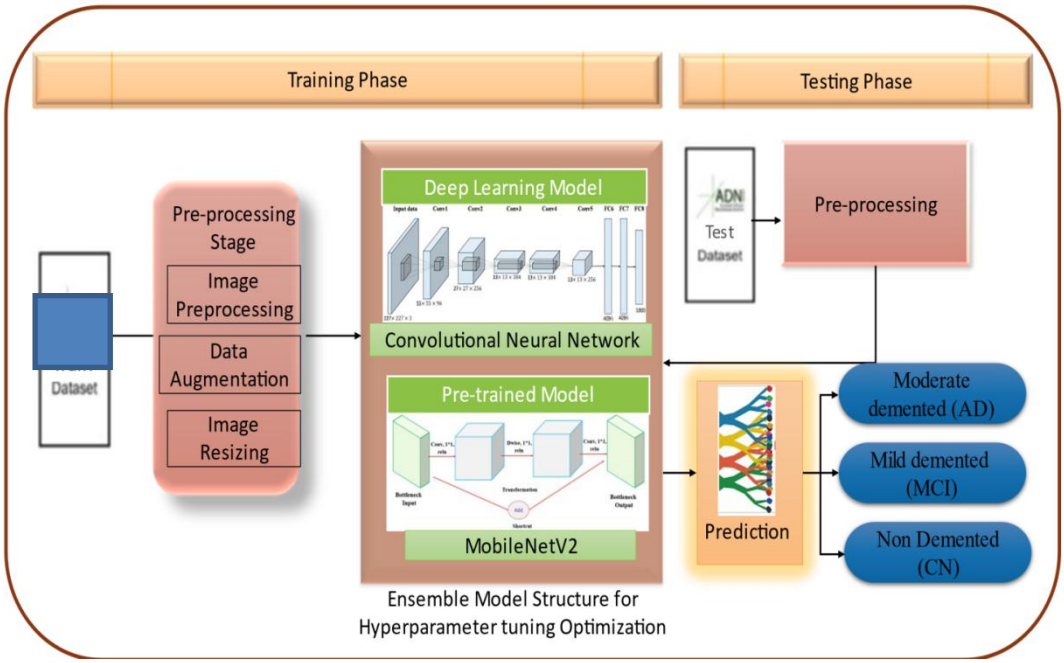


**Figure 9.** Single Transfer Learning Model Architecture using MobileNetV2

The ensemble learning in our approach involves data ensembles and classifiers, with feature set incorporating more data from MRI images to improve classification performance. The internal architecture of our proposed SETNN model is depicted in figure 9, showing the arrangement of the ensemble models and their outputs. Again, experimentation is done with proposed model SETNN versus SVM with three different kernels viz., linear, sigmoid (non-linear), RBF. Their comparisons are shown in results section.

In figure 9, we can see single transfer learned model. The convolutional base and the classifier head are the two components of the method. The pre-trained base model, which transforms the pixels of the input image into features, is contained in the convolutional base. The classifier head then receives these features to make a prediction about the likelihood of each class. The classifier head's fully connected layers link all of the neurons together and incorporate all of the features to provide the most accurate prediction. For classification, the activation values from the global average pooling layer are sent into the Softmax layer, where they are added together and divided by the total to produce probability values.
We conducted thorough experiments using different pre-trained deep convolutional neural networks (CNNs) as feature extractors to obtain meaningful and discriminative features from brain magnetic resonance (MR) images. Additionally, we performed extensive experiments with hyperparameter tuning optimization on various pre-trained CNN models to enhance the performance of our approach. The weight of the model is initialized with the ImageNet weight used in the training of the original model, and we use sparse cross-entropy loss function due to the multiclass classification task while using ADNI dataset and binary cross entropy for OASIS dataset. Details of hyperparameters are provided in Table 3.

**Table 3.** Hyper-parameters values

| **Parameter** | **Value for SETNN** |
|---|---|
| Input Shape | (32,32,3) |
| Weight | Initialized to ImageNet |
| Optimizer | ADAM |
| Learning Rate | 1e-1 |
| Loss Function | Categorical Cross Entropy |
| Classifier | Softmax |
| Epochs | 2 for OASIS dataset /5 for ADNIdataset |
| Batch Size | 32 |
| Dropout Rate | 0.2 |

The table 4 displays the average time per epoch (in seconds) for training different pre-trained deep convolutional neural network (CNN) models on non-demented data. The models include popular architectures such as VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet201, NASNetM, EfficientNetB0, and a custom CNN+MobileNet. The results show that InceptionResNetV2 has the highest average time per epoch at 1102 seconds, while InceptionV3 has the lowest at 298 seconds. Other models such as DenseNet201, NASNetM, and EfficientNetB0 also require a considerable amount of time per epoch for training, ranging from 868 to 1022 seconds. On the other hand, VGG16, VGG19, and MobileNet show moderate training times, while MobileNetV2 and CNN+MobileNet exhibit similar times around 580 to 590 seconds per epoch, as shown in figure 10. These results provide insights into the computational costs associated with training different CNN models on non-demented data, which can be valuable for choosing an appropriate architecture based on time constraints in a specific research or application context.

**Table 4.** Time required for execution

| Non-Demented | Average time per epoch (s) |
|---|---|
| VGG16 | 302 |
| VGG19 | 348 |
| ResNet50 | 648 |
| InceptionV3 | 298 |
| InceptionResNetV2 | 1102 |
| MobileNet | 578 |
| MobileNetV2 | 582 |
| DenseNet201 | 894 |
| NASNetM | 868 |
| EfficientNetB0 | 1022 |
| CNN+MobileNet | 590 |

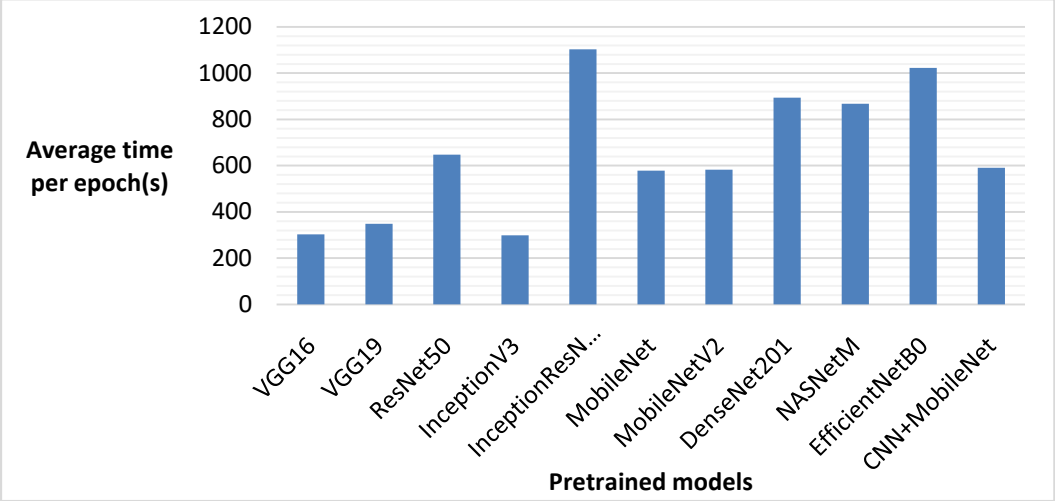


**Figure 10.** Time required for execution

### 3.9. Performance Evaluation

Performance evaluation of ensemble modelling involves assessing the effectiveness and efficiency of combining multiple individual models to improve predictive accuracy and robustness. Ensembles are known for reducing over fitting, increasing generalization, and enhancing overall model performance. Here's a comprehensive guide on how to evaluate ensemble models:

1. Accuracy

Performance models are evaluated using the accuracy (ACC) as the most important performance indicator for AD diagnosis. Additionally, sensitivity (SEN) and specificity (SPE) are also used as performance indicators. True positives (TP) are the positive tuples correctly labeled by the classifier, while false positives (FP) are the negative tuples incorrectly labeled as positive. True negatives (TN) are the negative tuples correctly labeled, and false negatives (FN) are the positive tuples mislabeled as negative.

- Accuracy: The proportion of correct predictions among all predictions. It's suitable when classes are balanced.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

- Precision: The ratio of true positive predictions to the total predicted positives. It emphasizes the correctness of positive predictions.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

- Recall (Sensitivity): The ratio of true positive predictions to the total actual positives. It focuses on capturing all positive cases.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (8)$$

- F1-Score: It is the harmonic mean of precision and recall. It is useful for imbalanced classes.

$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

- Specificity: It is the ratio of true negative predictions to the total actual negatives. It is the relevant when false negatives are critical.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (10)$$

- ROC Curve and AUC: It is the Receiver Operating Characteristic curve plots the true positive rate against the false positive rate at various thresholds. Area Under the Curve (AUC) tells the ROC curve's performance in a single value.

2. Confusion Matrix

A tabular representation of predicted versus actual class labels. It helps visualize the performance of the ensemble across different classes, aiding in error analysis.

3. Cross-Validation

To avoid overfitting during training, perform k-fold cross-validation. Divide the dataset into k subsets, train on k-1 subsets, and validate on the remaining one. Repeat k times, ensuring each subset is used for validation once.

4. Hyperparameter Tuning
Optimize the hyperparameters of the individual models and the ensemble itself. Techniques like grid search or random search can be used.

## 4. EXPERIMENTS, RESULTS AND DISCUSSION

In this section, we provide a comprehensive overview of the experiment conducted, including its setup and results. We start by describing the experiment's setup, which encompasses the software and hardware settings employed. The results of training and validation of the model procedure are then presented. The findings from combining three classifiers—Softmax, SVM, and the CNN model for feature extraction—are discussed in the fourth subsection. Finally, we compare the outcomes of our suggested methodology with those of alternative approaches.
The trials were carried out using the Python programming environment provided by the Google Colaboratory Pro platform. Google's cloud offering Colab Pro enables customers to create and run Python programmes on a hosted GPU. We used a variety of DL Python modules, including TensorFlow, Keras, Scikit-learn, Numpy, and OpenCV, to create our suggested solution. Additionally, we used the Python modules Nibabel, Nilearn and DeepBrain to analyse MRI neuroimaging data. This work focused on coronal plane visualisation of brain anatomy using the ADNI dataset of MRI images in NIFTI format. The coronal plane, an x-z plane that divides the anterior from the posterior, is perpendicular to the ground. According to studies, employing the coronal plane is more efficient [38]
For our investigation, the dataset was randomly partitioned into two sets: a training set including 80% of the data and a testing set containing 20% of the data. Table 5 displays the performance of 10 pre-trained models as well as the suggested ensemble model SETNN with the two classifiers.

**Table 5.** Performance of ten pre-trained model along with proposed ensemble model SETNN with the two classifiers

| **Pre-trained model Deep Features** | **SVM (Linear)** | **Classifier – Accuracy (%) SVM (Non-linear/Sigmoid)** | **SVM(RBF)** | **Softmax (FC)** |
|---|---|---|---|---|
| VGG16 | 86.27 | 86.27 | 80.39 | 90.1 |
| VGG19 | 82.35 | 82.35 | 83.78 | 90 |
| ResNet50 | 82.35 | 88.24 | 90.2 | 92.1 |
| InceptionV3 | 90.2 | 90.2 | 90.2 | 93.7 |
| InceptionResNetV2 | 92.16 | 92.16 | 92.16 | 95.3 |
| MobileNet | 86.27 | 88.24 | 88.24 | 89.5 |
| MobileNetV2 | 87.32 | 88.69 | 89.88 | 90.1 |
| DenseNet201 | 84.31 | 88.24 | 86.27 | 93.6 |
| NASNetM | 84.31 | 86.27 | 86.45 | 91.1 |
| EfficientNetB0 | 86.23 | 90.2 | 92.16 | 93.3 |
| SETNN | 87.2 | 88.79 | 90 | 91.9 |

The proposed SETNN model structure is based on the transfer learning with pre-trained models like MobilenetV2, Inception V3 and VGG16 as base models with some modifications to prevent overfitting and improve model performance. To reduce overfitting, a dropout layer with a rate of 0.2 was inserted before the classifier and after the final fully connected layer. With a learning rate of 0.0001, the ADAM optimizer was used to train the model. The batch size was set to 32 for the training and validation sets. The predetermined hyperparameter for model training, known as the epoch, was set at 4 empirically. For distinguishing AD and normal MRI imagery, model assessment depends on precision, recall, F1 score, top 5 accuracy and categorical cross-entropy (loss). The amount of data which an algorithm should minimise during training is determined by loss functions. The training and validation curves of SETNN model is shown in Figure 11 for ADNI data and in figure 12 for OASIS data.
For distinguishing AD and normal MRI imagery, model assessment depends on accuracy and categorical cross-entropy (loss). The amount of data which an algorithm should minimise during training is determined by loss functions. The training and validation curves of SETNN model is shown in Figure 11 for ADNI data and in figure 12 for OASIS data. The right graphs, over 5 epochs, show accuracy vs. epochs whereas the left graphs indicate loss vs. epochs. The training results are shown in red and the validation results are shown in orange.

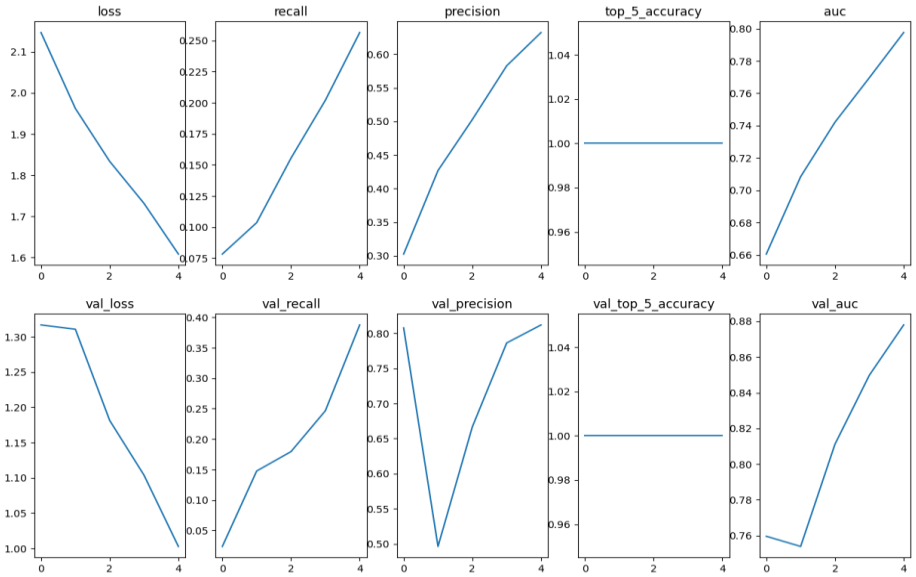


**Figure 11.** Training and validation performance of SETNN model for ADNI dataset

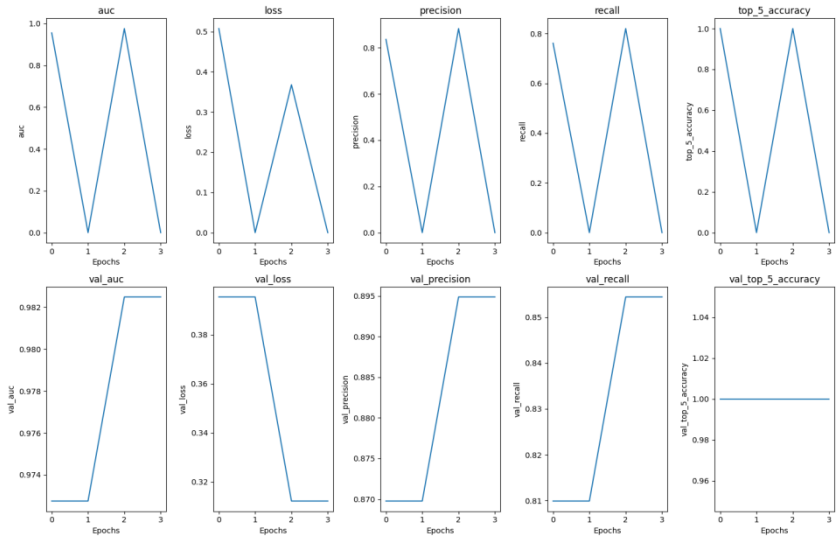


**Figure 12.** Training and validation performance of SETNN model for OASIS dataset

The confusion matrix of the Softmax classifier is shown in Figure 13 for ADNI dataset classification while figure 14 shows OASIS dataset classification. The prediction results of SETNN are displayed in Table 6 and table 7 while classifying ADNI dataset and OASIS dataset in terms of accuracy, recall, f1-measure, and support, while support is the number of samples.

To address our research questions, we conducted various experiments using two different classifiers, Softmax and SVM. We evaluated the classification performance of the proposed model using Softmax and SVM classifiers with the ADNI dataset and OASIS dataset. These tests aimed to identify the best precise method for pre-trained AD diagnosis model. In the classifier layer, we initially used Softmax to apply transfer learning to SETNN. Then, using the both datasets, we evaluated the suggested methods (SETNN-Softmax and SETNN-SVM). The outcomes demonstrated that, across all performance metrics, the model using the Softmax classifier beat SVM. The confusion matrix of the Softmax classifier is shown in Figure 13 for ADNI dataset classification while figure 14 shows OASIS dataset classification. The prediction results of Softmax are displayed in Table 5 in terms of accuracy, recall, f1-measure, and support, while support is the number of samples.

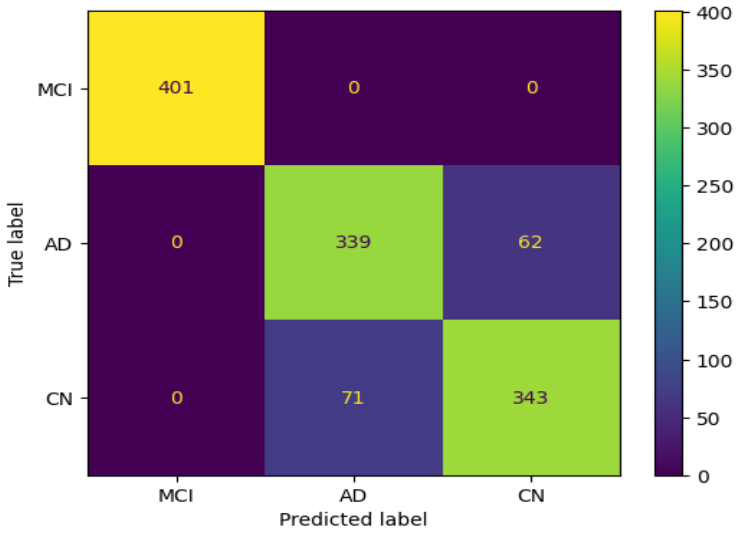


**Figure 13.** Confusion matrix during testing the SETNN model with ADNI data

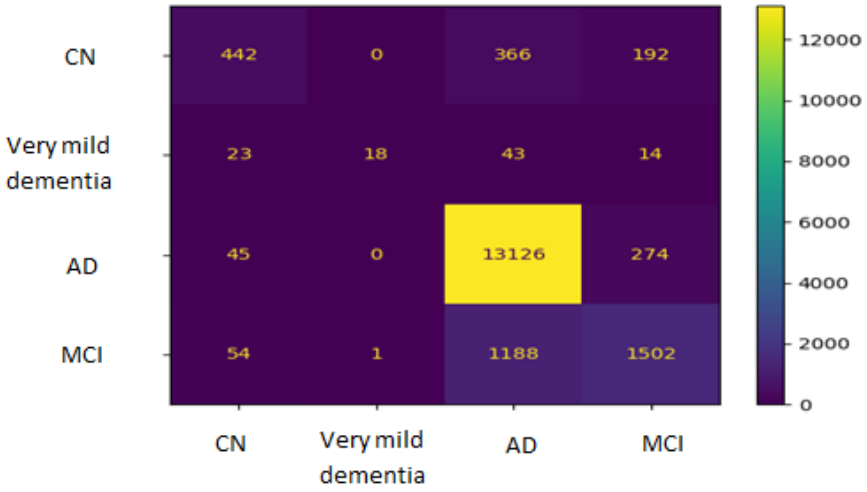


**Figure14.** Confusion matrix during testing the SETNN model with OASIS data

**Table 6.** SETNN model experiment results with ADNI dataset

| Data | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| AD | 0.92 | 0.86 | 0.89 | 401 |
| CN | 0.87 | 0.90 | 0.89 | 414 |
| MCI | 0.98 | 1.00 | 0.99 | 401 |
| Accuracy | 0.92 | | | 1216 |
| Loss | 0.2 | | | |
| Area under the Curve (ROC) | 0.92 | | | |
| Precision | 0.96 | | | |
| Recall | 0.99 | | | |

**Table 7.** SETNN model experiment results with OASIS dataset

| Data | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Non Demented (CN) | 0.78 | 0.44 | 0.57 | 1000 |
| Very mild Dementia | 0.95 | 0.18 | 0.31 | 98 |

| | | | | |
|---|---|---|---|---|
| Moderate Dementia (AD) | 0.89 | 0.98 | 0.93 | 13445 |
| Mild Dementia (MCI) | 0.76 | 0.55 | 064 | 2745 |
| Accuracy | 0.88 | | | 17288 |
| Loss | 0.31 | | | |
| Area under the Curve (ROC) | 0.89 | | | |
| Precision | 1.00 | | | |
| Recall | 0.98 | | | |

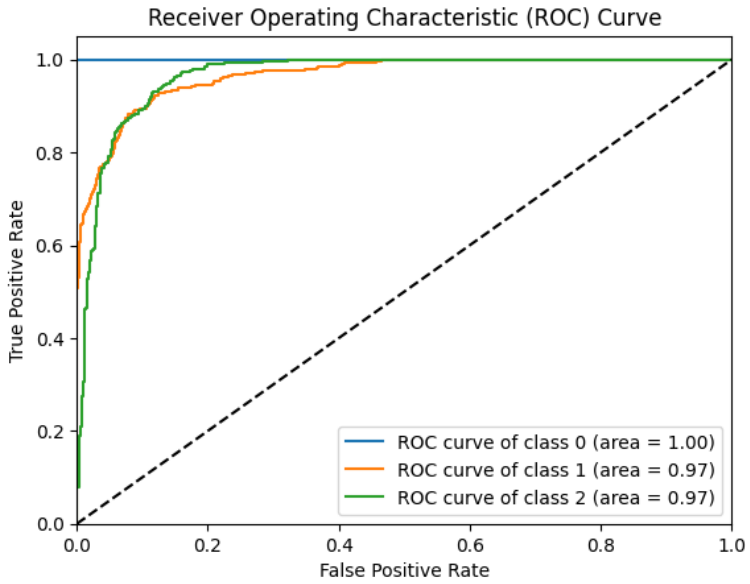


**Figure 3.** ROC-AUC curves for ADNI dataset

**Table 8.** Discussion for proposed model test performance with state-of-the-art existing methods.

| **References** | **Feature Extraction** | **Accuracy** |
|---|---|---|
| Loddo et al. [7] | Inception-ResNet-v2 | 98.51% in the binary case, and 98.67% in the multiclass case |
| Mahendran et al. [8] | EDRNN | 88.7% |
| Sava [9] | EfficientNetB0 | 92.98% |
| Murugan et al. [10] | DEMNET | 95.23%. |
| Mohammed et al. [11] | AlexNet+SVM | 98.3% |
| Gharaibeh et al. [12] | InceptionV3 and DenseNet201 | 99% |
| Basher et al. [13] | CNN and DNN | 94.82% and 94.02% for left and right hippocampi respectively |
| Yigit et al. [38] | CNN | 83% on the whole brain images |
| Proposed Model | SETNN | 92% with ADNI dataset and 88% with OASIS dataset |

The ROC (Receiver Operating Characteristic) curve compares the true positive rate to the false positive rate at various thresholds. The AUC (Area Under the Curve) measures the model's ability to differentiate across classes. We can see in figure 15 that curve is hugging the upper left corner of the plot. This means that our SETNN model is functioning well because it has high sensitivity and specificity. Also, high AUC

(near to 1) indicates that the model can effectively distinguish between patients with and without the Alzheimer's condition. This is suitable for medical diagnostics in which both sensitivity (recall) and specificity are required.

The suggested AD diagnostic model has been demonstrated to be successful, with a good AD classification accuracy (92%) with ADNI dataset while it has achieved accuracy of 88%, is shown in Tables 5, 6 and 7. Test loss with OASIS dataset was 0.3. The results indicate that the Softmax classifiers exhibit the highest accuracy among them, as shown in table 8 compared to existing studies. Additionally, the SVM with RBF kernel ranks as the second-best classifier in a comparable manner.

## 5. CONCLUSION AND FUTURE SCOPE

To improve early diagnosis of Alzheimer's disease, a deep learning-based classification model with an ensemble approach was utilized to classify patients with Alzheimer's disease. The analysis was performed using an ADNI dataset and OASIS dataset from the Alzheimer's disease Neuroimaging Initiative database. The implementation of a Stacked Ensemble Transfer Learned neural network (SETNN) and comparisons with other current classification models were conducted. Stacked ensemble models outperform individual models by combining the predictions of multiple networks to reduce variance and bias in the overall prediction. The findings showed that, when compared to previous approaches, the suggested model's classification accuracy, precision and recall as well as AUC had significantly improved.

Deep learning-based AD research is continuously being developed for improved performance and transparency. Research on the detection of AD by utilising deep learning is moving away from hybrid approaches and towards a model that uses only deep learning algorithms. However, methods must still be evolved to combine totally distinct types of data in a network based on deep learning. Future research must prioritize optimizing these models to achieve a balance between performance and efficiency, by incorporating advanced techniques such as automated machine learning (AutoML) to streamline the model selection and tuning process.

## REFERENCES

[1] Snyder, P. J., Jackson, C. E., Petersen, R. C., Khachaturian, A. S., Kaye, J., Albert, M. S., & Weintraub, S. (2011). Assessment of cognition in mild cognitive impairment: a comparative study. Alzheimer's & Dementia, 7(3), 338-355.

[2] McCollum, L., & Karlawish, J. (2020). Cognitive impairment evaluation and management. Medical Clinics, 104(5), 807-825.

[3] Mukherjee, S., Ali, S., Hashmi, S., & Jahan, S. (2023). History, Origin and Types of Neurological Disorders. In Applications of Stem Cells and derived Exosomes in Neurodegenerative Disorders (pp. 1-32). Singapore: Springer Nature Singapore.

[4] Breijyeh, Z., & Karaman, R. (2020). Comprehensive review on Alzheimer's disease: causes and treatment. Molecules, 25(24), 5789.

[5] Alzheimer's Association. (2017). 2017 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 13(4), 325-373.

[6] Lee, J., Phillips, D., & Wilkens, J. (2022). Gateway to global aging data. In Encyclopedia of gerontology and population aging (pp. 1951-1959). Cham: Springer International Publishing.

[7] Loddo, A., Buttau, S., & Di Ruberto, C. (2022). Deep learning based pipelines for Alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method. Computers in Biology and Medicine, 141, 1-15.

[8] Mahendran, N., & PM, D. R. V. (2022). A deep learning framework with an embedded-based feature selection approach for the early detection of Alzheimer's disease. Computers in Biology and Medicine, 141(September 2021).

[9] Sava¸s, S. (2022). Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. Arabian Journal for Science and Engineering, 47(2), 2201-2218.

[10] Murugan, S., Venkatesan, C., Sumithra, M. G., Gao, X. Z., Elakkiya, B., et al. (2021). DEMNET: A deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. IEEE Access, 9, 90319-90329.

[11] Mohammed, B. A., Senan, E. M., Rassem, T. H., Makbol, N. M., Alanazi, A. A., et al. (2021). Multi-method analysis of medical records and MRI images for early diagnosis of dementia and Alzheimer's disease based on deep learning and hybrid methods. Electronics, 10, 1-20.

[12] Gharaibeh, M., Almahmoud, M., Ali, M. Z., Al-Badarneh, A., El-Heis, M., et al. (2022). Early diagnosis of Alzheimer's disease using cerebral catheter angiogram neuroimaging: A novel model based on deep learning approaches. Big Data and Cognitive Computing, 6(2), 1-23.

[13] Basher, A., Kim, B. C., Lee, K. H., & Jung, H. Y. (2021). Volumetric feature-based Alzheimer's disease diagnosis from sMRI data using a convolutional neural network and a deep neural network. IEEE Access, 9, 29870-29882.
[14] Mirzaei, G., & Adeli, H. (2022). Machine learning techniques for diagnosis of Alzheimer disease, mild cognitive disorder, and other types of dementia. Biomedical Signal Processing and Control, 72.
[15] Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., ... & Alzheimer's Disease Neuroimaging Initiative. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage: Clinical, 21, 101645.
[16] Hsiao, J. J., Lu, P. H., Grill, J. D., & Teng, E. (2015). Longitudinal declines in instrumental activities of daily living in stable and progressive mild cognitive impairment. Dementia and geriatric cognitive disorders, 39(1-2), 12-24.
[17] Islam, J., & Zhang, Y. (2018). Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures. In IEEE Conf. on Computer Vision and Pattern Recognition Workshops, Salt Lake City, USA, pp. 1994-1996.
[18] Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., DeCarli, C., Green, R. C., ... & Alzheimer's Disease Neuroimaging Initiative. (2022). Using the Alzheimer's Disease Neuroimaging Initiative to improve early detection, diagnosis, and treatment of Alzheimer's disease. Alzheimer's & Dementia, 18(4), 824-857.
[19] Nishimaki, K., Onda, K., Ikuta, K., Uchida, Y., Mori, S., Iyatomi, H., & Oishi, K. (2024). OpenMAP-T1: A Rapid Deep Learning Approach to Parcellate 280 Anatomical Regions to Cover the Whole Brain. medRxiv, 2024-01.
[20] Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience, 19(9), 1498-1507.
[21] LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., ... & Marcus, D. (2019). OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medrxiv, 2019-12.
[22] Morra, J. H., Tu, Z., Apostolova, L. G., Green, A. E., Avedissian, C., Madsen, S. K., ... & Alzheimer's Disease Neuroimaging Initiative. (2008). Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. Neuroimage, 43(1), 59-68.
[23] Nawaz, A., Anwar, S. M., Liaqat, R., Iqbal, J., Bagci, U., & Majid, M. (2020, November). Deep convolutional neural network based classification of Alzheimer's disease using MRI data. In 2020 IEEE 23rd International Multitopic Conference (INMIC) (pp. 1-6). IEEE.
[24] Venkatesan, R., & Haacke, E. M. (1997). Role of high resolution in magnetic resonance (MR) imaging: Applications to MR angiography, intracranial T1-weighted imaging, and image interpolation. International journal of imaging systems and technology, 8(6), 529-543.
[25] Gao, J., Jiang, Q., Zhou, B., & Chen, D. "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview." Mathematical Biosciences and Engineering, vol. 16, no. 6, 2019, pp. 6536-6561.
[26] Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. Neurocomputing, 503, 92-108.
[27] Shen, X., Tian, X., He, A., Sun, S., & Tao, D. (2016, October). Transform-invariant convolutional neural networks for image classification and search. In Proceedings of the 24th ACM international conference on Multimedia (pp. 1345-1354).
[28] Zhang, Z., & Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. Advances in neural information processing systems, 31.
[29] Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, 9(2), 85-112.
[30] Serrà, J., & Karatzoglou, A. (2017, August). Getting deep recommenders fit: Bloom embeddings for sparse binary input/output networks. In Proceedings of the Eleventh ACM Conference on Recommender Systems (pp. 279-287).
[31] Kolides, A., Nawaz, A., Rathor, A., Beeman, D., Hashmi, M., Fatima, S., ... & Jararweh, Y. (2023). Artificial intelligence foundation and pre-trained models: Fundamentals, applications, opportunities, and social impacts. Simulation Modelling Practice and Theory, 126, 102754.
[32] Agarap, A. F. (2018). Deep Learning Using Rectified Linear Units (ReLU). arXiv preprint arXiv:1803.08375.

[33] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).
[34] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). IEEE.
[35] Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP), 9(10), 143-150.
[36] Hussain, M., Bird, J. J., & Faria, D. R. (2019). A study on CNN transfer learning for image classification. In Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK (pp. 191-202). Springer International Publishing.
[37] Huang, Z., Sun, M., & Guo, C. (2021). Automatic Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment Based on CNN+ SVM Networks with End-to-End Training. Computational Intelligence and Neuroscience, 2021(1), 9121770.
[38] Yiğit, A., & Işik, Z. (2020). Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease. Turkish Journal of Electrical Engineering and Computer Sciences, 28(1), 196-210.