People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University M'Hamed BOUGARA – Boumerdes

Université de Boumerdes
University of Boumerdes

Institute of Electrical and Electronic Engineering

Department of Electronics

Final Year Project Report Presented in Partial Fulfilment of
the Requirements for the Degree of

MASTER

In **Electrical and Electronic Engineering**

Option: **Telecommunications.**

Option: **Computer.**

Title:

Alzheimer Disease Classification Using Convolution Neural Networks and Transfer Learning.

Presented by:

- **BRAHIMI Kahina**
- **SLIMI Ounissa**

Supervisor:

Dr. CHERIFI Dalila

Registration Number:………/2023

Abstract

Alzheimer's Disease (AD) is a neurological disorder which causes brain cells to die, resulting in memory loss, language difficulties, and impulsive or erratic behavior. In recent years the number of individuals affected has seen a rapid increase, it is estimated that up to 107 million subjects will be affected by 2050 worldwide. Early diagnosis has become crucial to improve patients care and treatment. AD diagnosis is difficult due to the complexity of the brain structure and its pixel intensity similarity especially at its early stage. A comprehensive diagnosis must be led including clinical assessment and medical imaging, which is a process that requires the expertise of professionals including neurologists and radiologists. One of the drawbacks of medical imaging approach is the inability to detect changes in very mild impairment also known as mild cognitive impairment (MCI). Deep learning has inspired a lot of interest in recent years in tackling challenges in a variety of fields, including medical imaging and detecting abnormalities beyond human capabilities.

In this work we explored classification approaches of AD through two different datasets which are respectively MRI dataset and tabular dataset. First, we dealt with the tabular data for binary classification of AD into demented and non demented using classical machine learning algorithms namely Support Vector Machine, K Nearest Neighbor, Decision Tree, Naïve Bayes Gaussian, Random Forest and Logistic Regressor. The findings indicate that the models effectively utilize the Clinical Dementia Rating feature for AD classification. Second, we dealt with MRI dataset for multiclassification of AD into Non_Demented, Very_Mild_Demented, Mild_Demented, and Moderate_Demented using transfer learning models namely VGG19, ResNet50, Xception and MobileNet. The VGG19 model gave the best performance with 98.60% testing accuracy where the other models achieved 97.35%, 86.35% and 95.50% respectively. We also proposed a custom CNN model that outperformed the transfer learning models and achieved an accuracy of 99.00%.

Key words: Classification, Alzheimer's disease, machine learning, transfer learning, MRI, CNN.

Dedications

This study is dedicated wholeheartedly to my beloved parents Saddek and Djamila that have been a constant source of inspiration, providing me with hope and encouragement when I faced moments of doubt and despair. They have always been there, guiding and supporting me. Your sacrifices, which I could never fully repay, have shaped me into the person I am today. I am forever grateful to my dear brothers, Amine and Gaya, for their unwavering presence and continuous support throughout this entire process. I also extend my heartfelt gratitude to all my friends and family members who have stood by my side, offering their encouragement and assistance.

<div align="right">

"Kahina"

</div>

I dedicate this work to my beloved parents, Nora and Djaffer, whose love and support has been the foundation of my journey, who filled my life with ease and comfort. To my incredible aunts and sister in law, who have always been a source of joy in my life, I am grateful for your presence and guidance. A special dedication goes to my dearest Aunt Chafia, No words can express how lucky I am to have you and no matter what I do, I will never be able to fully repay your love. To my brothers and all family members with whom I share my best memories. Lastly, I dedicate this work to my dear grandma, who sadly passed away last year but whose memories continue to inspire me. This work is a tribute to her lasting impact on my life.

<div align="right">

"Ounissa"

</div>

Acknowledgments

All praise be to Allah Almighty who provided us with his blessing and the opportunity to successfully conclude our project.

*We would like to express our deepest gratitude to our supervisor, **Dr. CHERIFI Dalila**. We are immensely grateful for her valuable advices regarding our research. Her willingness to assist us, steadfast encouragement and continuous support were instrumental in making this project a reality. We deeply appreciate her patience, dedication, and unwavering belief in our abilities. Lastly, we are immensely thankful to our parents and friends for their unwavering support in completing this report. We would like to extend our gratitude to all the individuals who have supported us throughout this journey.*

List of Figures

List of Tables

List of Abbreviations

MRI	Magnetic Resonance Imaging
AI	Artificial Intelligence
ANNs	Artificial Neural Networks
CNNs	Convolutional Neural Networks
DL	Deep Learning
ReLu	Rectified Linear Unit
VGG	Visual Geometry Group
ResNet	Residual Network
DNN	Deep Neural Network
Adam	Adaptive Moment Estimation
OASIS	Open Access Series of Imaging Studies
CSV	Comma Separated Values
GPU	Graphics Processing Units

Contents

General introduction

Alzheimer's disease (AD) is a debilitating neurodegenerative disorder that affects millions of people worldwide. It is characterized by the progressive decline of cognitive functions, memory loss and behavioral changes. In recent years, the number of individuals affected has seen a rapid increase, posing significant concerns for healthcare systems and those impacted by the disease.

The accurate and timely diagnosis of early stage AD plays a crucial role in providing appropriate care and support to patients. Traditional diagnostic methods involve clinical assessments, cognitive tests, and neuroimaging techniques ,among them, MRI imaging, the most popular and efficient technique due to the rich information it provides. However, in the early stages of AD, an MRI scan of the brain may seem normal, not until later stages, where the MRI may show a decrease in the size of different areas of the brain. That is why another approach must be established to detect early stages.

Machine learning techniques have revolutionized numerous fields by facilitating the development of novel algorithms and inventions. Specifically, they have demonstrated tremendous potential in medical image classification, including the diagnosis of AD. Convolutional Neural Networks (CNNs) have emerged as powerful tools for image analysis, enabling precise and dependable AD diagnosis.

The objective of this work is to gain a comprehensive understanding of AD and identify its key associated features. Additionally, the main focus is on enhancing the performance of classification models for this disease using CNNs and Transfer Learning. A comparative study of existing research will be conducted to analyze different approaches used in the literature. Two datasets obtained from the Kaggle platform have been used. The first dataset comprises tabular data containing clinical assessments catagorized into Non Demeneted and Demented classes, while the second dataset consists of MRI scans categorized into Non_Demented, Very_Mild_Demented, Mild_Demented and Moderate_Demented classes. The aim is to use these datasets to improve the accuracy and effectiveness of AD classification models, ultimately contributing to the early diagnostic techniques and treatment strategies.

This report consists of three chapters. The first chapter serves as a comprehensive review about AD, exploring its background and significance, as well as medical imaging techniques employed in its diagnosis. In the second chapter, related work is presented. An overview of the classical machine learning algorithms used for tabular dataset classification and the utilized deep neural network for MRI dataset classification. The third chapter introduces the tools employed in this study, providing insights into the implementation details and experimental setup. The chapter also presents the results obtained from the conducted experiments and discusses these findings and ends with a general conclusion.

Chapter 1

Review About Alzheimer's Disease

1.1   Introduction:

Alzheimer's disease is a neurodegenerative disease, that is to say a progressive brain damage leading to neuronal death. It is characterized by a progressive loss of memory and certain cognitive intellectual functions, which in turn affects daily life activities. Examples of these activities include the ability to continue in a conversation and respond to the environment. The symptoms of AD change over time, and this progression can vary from one individual to another.

When considering the diagnosis of Alzheimer's disease, memory problems are typically the primary symptom. However, it is important to note that these memory issues must be accompanied by other cognitive function disorders for a confirmed diagnosis. These disorders can manifest in various ways, including language disorders (aphasia), difficulties in performing specific gestures (apraxia), or the loss of recognition of objects or people. Identifying these additional cognitive impairments is crucial in the diagnostic process of Alzheimer's disease [1].

Over time, a brain with Alzheimer's shrinks, neurons die, and plaques accumulate in areas in the brain including the hippocampus, which is the part of the brain that plays a role in memory and thinking where Alzheimer's disease usually begins. The brain may change trusted sources for up to a decade or longer, prior to symptoms of Alzheimer's manifesting [1].
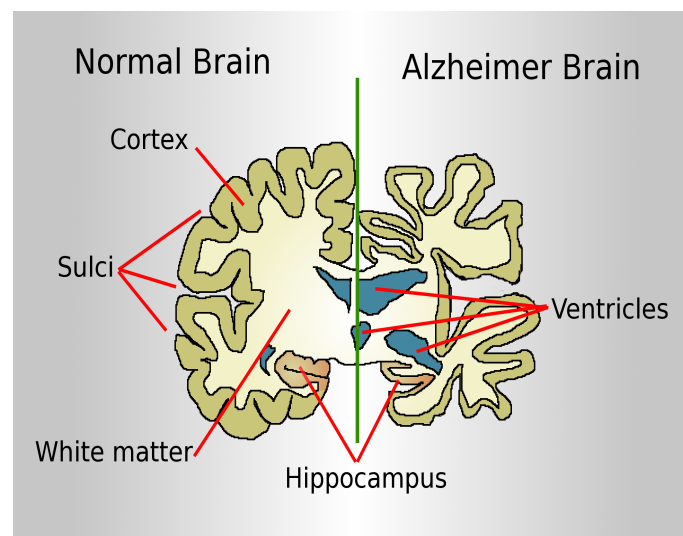
Figure 1.1: Comparison between normal brain and brain with Alzheimer
[2]

1.2   Causes of Alzheimer's disease:

The accumulation of abnormal proteins in and around brain cells is believed to be the cause of Alzheimer's disease. One of these proteins, called amyloid, forms plaques around brain cells, while another protein, called tau, forms tangles within brain cells. The exact cause of this process is not yet fully understood, but researchers know that it begins many years before the appearance of symptoms [1]. As the disease progresses, brain cells are progressively affected and various regions of the brain experience shrinkage, leading to a reduction in the levels of neurotransmitters, which are chemical messengers involved in transmitting signals between brain cells. While memory-related regions are commonly among the first to be affected, differ-

ent areas of the brain may also be impacted in typical forms of Alzheimer's disease. In some cases, individuals may initially experience symptoms that manifest as vision problems[1].

1.2.1 Non modifiable risk factors:

One well-known factor is aging, although it is not a direct cause of the disease. Genetic predisposing is another influential factor, as certain genes being associated with Alzheimer's disease. The ApoE gene, existing in different forms (ApoE2, ApoE3, and ApoE4), is a prominent genetic risk factor. In particular, the ApoE4 version significantly increases the risk of developing Alzheimer's disease, with approximately 33% of individuals with this form developing the disease by the age of 75. However, it's important to note that carrying the ApoE4 version does not guarantee the development of Alzheimer's disease. While these risk factors cannot be changed, individuals with a family history of dementia can take steps to reduce their risk by adopting a healthy lifestyle.

1.2.2 Modifiable risk factors:

Alzheimer's disease is influenced by certain environmental risk factors that have the potential to be managed with appropriate care. These factors include high blood pressure, diabetes, tobacco use, obesity, physical inactivity, educational level, alcohol consumption, depression, and head injuries. These risk factors can be reduced by making lifestyle changes such as exercising, eating well, and quitting smoking.

1.3 Alzheimer's disease statistics:

In 2020, there are more than 55 million people worldwide diagnosed with Alzheimer's disease. Alarmingly, this number is expected to double approximately every 20 years, with estimates suggesting that there will be 78 million people living with Alzheimer's disease by 2030 and 139 million by 2050. This significant growth is primarily anticipated in developing nations. Presently, around 60% of people affected reside in low and middle income countries. However, by 2050, this proportion is estimated to rise to 71% [3].
According to the latest WHO data published in 2020, Alzheimer's and dementia deaths in Algeria reached 5,635 or 3.04% of total deaths [3].
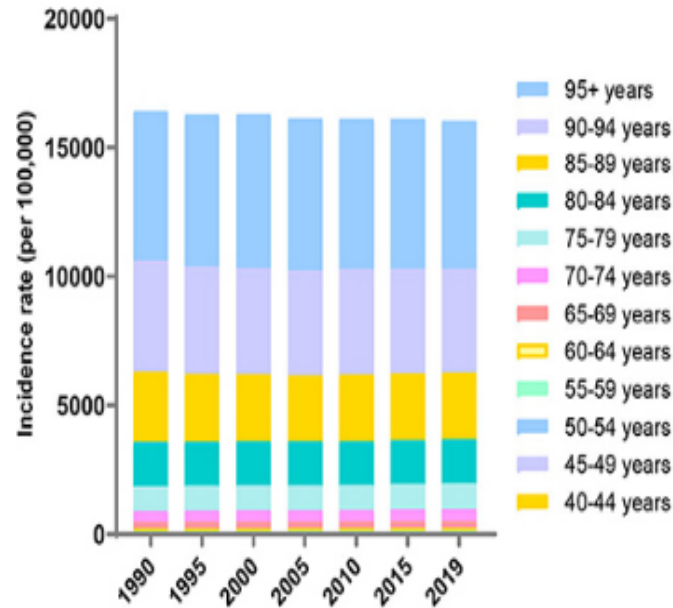
Figure 1.2: Incidence rate of Alzheimer in each age group
[4]

1.4 Diagnosis of Alzheimer's disease:

Doctors use several methods and tools to determine whether people with thinking or memory problems have Alzheimer's disease. To diagnose Alzheimer's disease they gather comprehensive information through interviews with the person experiencing symptoms, as well as their family and friends about their general health, prescription and over-the-counter drug use, diet, past medical conditions, ability to function in daily life, and behavioral or personality changes. Tests are conducted to evaluate memory, problem solving, attention, counting, and language skills. Standard medical tests and psychiatric evaluations are performed to identify any underlying conditions or mental illnesses. Brain scans such as CT or MRI help diagnose Alzheimer's by detecting brain changes and ruling out other causes [5].

1.5 Treatment of Alzheimer's disease:

Alzheimer's disease is a complex condition for which there is currently no cure known. However, there are treatments available that can help slow down the progression of the disease and alleviate symptoms. Cholinesterase inhibitors such as donepezil, rivastigmine, and galantamine are commonly prescribed medications for individuals with mild to moderate Alzheimer's disease. These medications work by increasing cell-to-cell communication and preserving a chemical messenger that is depleted in the brain by Alzheimer's disease. Another medication called memantine is used to treat moderate to severe Alzheimer's disease [6]. In addition to medication, certain lifestyle changes, such as regular exercise, a healthy diet, and social engagement activities, can also be beneficial in managing symptoms and improving the overall quality of life for people with Alzheimer's disease. However, it's important to note that while these treatments can be effective in managing symptoms, they do not stop or reverse the underlying progression of the disease.

1.6 Alzheimer's disease stages:

The terms "Non demented," "very mild demented," "mild demented," and "moderate demented" are used to describe different stages of cognitive impairment. Mild cognitive impairment (MCI) is an early stage of memory loss or other cognitive decline that may or may not progress to dementia. Primary care providers are often the first point of contact for patients with MCI or mild dementia. Moderate dementia is characterized by increased memory loss, confusion, a shortened attention span, inappropriate angry outbursts, and difficulty recognizing people and places [7].

- **Non demented:** Non-demented individuals usually have a normal MRI appearance of the brain, with no evidence of atrophy or shrinkage. In contrast, people with dementia usually show structural changes in the brain on MRI that become more pronounced as the disease progresses.

- **Very mild demented:** or early-stage dementia may show subtle changes in the brain structure, such as mild atrophy or shrinkage in certain areas of the brain that are involved in memory and cognitive functions.

- **Mild demented:** Mild dementia may show more pronounced atrophy or shrinkage in these areas, as well as changes in the brain's white matter, which is responsible for connecting different parts of the brain. In the mild dementia stage, Alzheimer's disease is frequently identified. It is at this point when family members and medical professionals realize the patient is experiencing considerable memory loss and cognitive impairment. The symptoms interfere with daily activities.

- **Moderate demented:** Moderate dementia often shows significant atrophy or shrinkage in multiple areas of the brain, including the hippocampus, which is important for memory, and the frontal lobes, which are responsible for planning, decision-making, and other executive functions. The ventricles, which are fluid-filled spaces in the brain, may also appear enlarged. patients at this stage start to require more assistance with self-care and daily tasks.
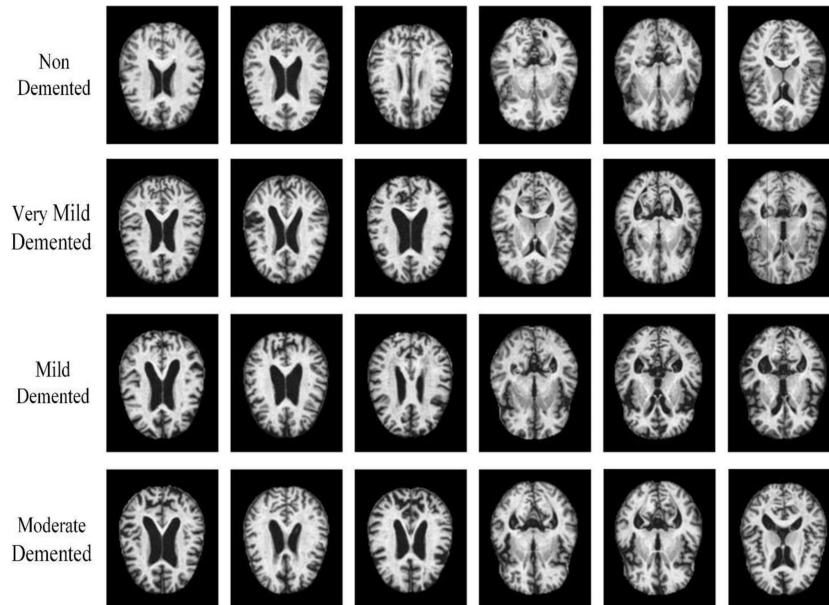
Figure 1.3: Alzheimer stages
[7]

1.7 Medical imaging:

Medical imaging, also referred to as radiology, is a specialized branch of medicine that involves creating images of various parts of the body for diagnostic or treatment purposes. These imaging procedures are non-invasive and allow doctors to diagnose injuries and diseases without the need for invasive procedures [8]. Medical imaging plays a central role in the improved outcomes of modern medicine, enabling doctors to accurately diagnose and treat a wide range of medical conditions. The frequently used ones are:

- Computed Tomography (CT).

- Magnetic Resonance Imaging (MRI).

- Magnetic Resonance Spectroscopy (MRS).

- Positron Emission Tomography (PET).

- Single-photon Emission Computed Tomography (SPECT).

X-rays and CT scans are powerful instruments, but they must be used with caution due to ionizing radiation. Ionizing radiation increases the risk of cancer, cataracts, cellular mutation, and improper fetal development.

1.7.1 Computed Tomography (CT):

A computed tomography (CT) scan is a medical imaging technique that involves taking a series of x-ray images from multiple angles around the body. These images are then processed by a computer to generate cross-sectional slices, providing detailed visualizations of bones, blood vessels, and soft tissues within the body.
CT scans are widely used and can be employed to examine nearly any body part. They are utilized in the diagnosis of various illnesses or injuries, as well as in the planning of medical, surgical, or radiation therapies [9].

1.7.2 Magnetic Resonance Imaging (MRI):

Magnetic resonance imaging (MRI) is a medical imaging method that employs magnetic fields and computer-generated radio waves to produce highly detailed images of organs and tissues within the body [10]. Typically, MRI machines consist of large tubular magnets. When a patient lies inside the MRI machine, the magnetic field temporarily reorients the water molecules in their body. Radio waves are then used to cause these aligned atoms to emit a weak signal, which is utilized to generate cross-sectional MRI images [9]. Additionally, MRI machines are capable of producing three-dimensional images that can be viewed from various angles.

1.7.3 Difference between MRI and CT in terms of dementia diagnosis:

Both CT and MRI scans can help diagnose Alzheimer, but they differ in the way they produce images of the brain. CT scans use X-rays to make images of brain structures, which can reveal indications of brain atrophy, strokes, ischemia, blood vessel abnormalities, and other factors that may indicate dementia. On the other hand, MRIs use magnetic fields and radio waves to create detailed images of the brain's soft tissues. They can detect abnormalities in white matter that may suggest dementia due to Alzheimer and show whether parts of the brain have atrophied (shrunk) [10]. Repeated MRI scans can also show how a person's brain changes over time. Therefore, both CT and MRI scans are useful for diagnosing dementia, but an MRI scan provides more detailed information about the structure of the brain than a CT scan. MRI scans are more commonly used than CT scans in diagnosing Alzheimer's disease because they can detect visible changes in the hippocampus, which is often affected first in Alzheimer's disease. MRI scans are generally more costly and less widely available than CT scans, but they are often a better choice for brain imaging due to their ability to provide more detailed and advanced images of the brain [10]. Therefore, both CT and MRI scans are useful for diagnosing dementia, but an MRI scan provides more detailed information about the structure of the brain than a CT scan.
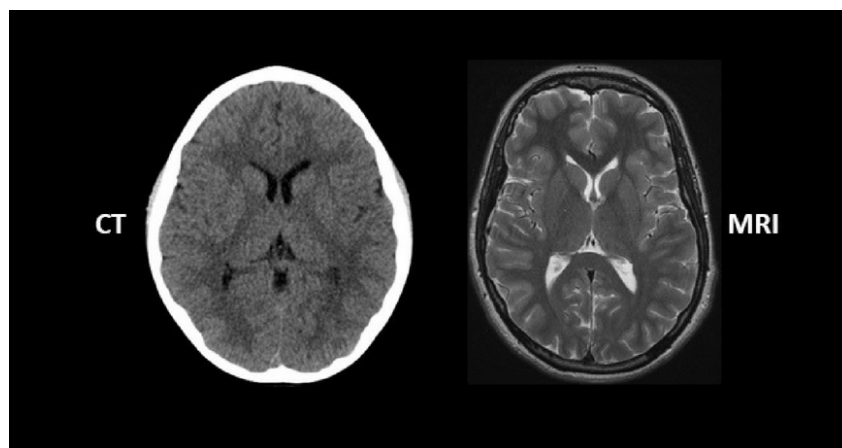
Figure 1.4: MRI VS CT scan of the brain
[9]

1.8   Summary:

This chapter provides a basic overview of Alzheimer's disease, including its origins and stages, as well as an examination of the many methods used to identify and treat it. It also includes medical imaging technology used in the diagnosis of this condition.

Chapter 2

Alzheimer's Disease Classification Using Machine Learning.

2.1   Introduction:

Artificial intelligence (AI) refers to the development of intelligent machines that can perform tasks that normally require human intelligence, such as visual recognition, speech recognition and decision making. Common AI techniques include machine learning, deep learning, natural language processing, computer vision, and robotics.

2.2   Related work:

In this section, information about the current studies in the literature closest to the method proposed in our article for the detection of AD is given. Note that all the following work used the same dataset that we used with different ways of dealing with it in terms of preprocessing and data augmentation techniques. In the most recent study, Feyza ALTUNBEY ÖZBAY and Erdal ÖZBAY [11], proposed a hybrid CNN method based on Neighborhood Component Analysis (NCA), they used DenseNet201, EfficientNet-B0, and AlexNet pre-trained CNN architectures as feature extractors, then the NCA method has been used to optimize all concatenated features. After this stage, the optimized features have been classified with KNN, they obtained an accuracy of 99.83%. Another study done by Madhusudan g lanewar et al.[12], combined a four block CNN model for feature extraction with KNN for classification, they obtained state of the art result of 99.58% accuracy. Furthermore, three deep learning CNNs ResNet50, VGG16 and MobileNetV2 were employed to compare their performance with the CNN-KNN model, the validation accuracies were 55.67%, 68.65% and 66.69% respectively. Gaurav kumar ameta, pushpendra sigh sisodia[13], reviewed deep transfer learning approaches they obtained validation accuracies of 96.56% for DenseNet201, 93.52 for ResNet50, 95.88% for VGG19, 89.77% for Xception and 83.20% for EfficientNet, they used data augmentation to balance the dataset. On the other hand, Mudiyala aparma, battula srinivasa[14], used cycle GAN for data augmentation by increasing 70% of dataset and employed DenseNet121 and MobileNetV2 the validation accuracies were 98.03% and 97.88% respectively. In addition, a study by Ms. Heta Acharya et al.[15], VGG19, ResNet50, AlexNet, and CNN were employed, resulting in accuracies of 85.07%, 75.25%, 95.70%, and 88.79% respectively. Marwa El-Geneedy et al.[16], employed CNN with five layers, VGG16, InceptionV3, and ResNet201, achieving respective accuracies of 96.84%, 96.39%, 87.71%, and 89.71%. Furthermore, Waleed Al Shehri [17] conducted a study , DenseNet169 and ResNet50 were used, yielding accuracies of 83.82% and 81.92% respectively. As we can observe from reviewing previous work Deep Learning architectures are increasingly being used in the multi-classification of Alzheimer disease.

2.3   Machine learning:

Machine learning is a branch of artificial intelligence and computer science that uses data and algorithms to learn in a manner similar to humans. It focuses on understanding and building methods that learn from data to improve performance with minimal human intervention.

- **Supervised learning:** It is defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately. Supervised learning helps organizations solve for a variety of real-world problems at scale [18].

- **Unsupervised learning:** Unsupervised learning analyzes and groups unlabeled datasets using machine learning techniques. Without the need for human intervention, these algorithms uncover hidden patterns and groups in data. Because of its capacity to detect similarities and contrasts in data, it is a perfect solution for exploratory data analysis, cross-selling techniques, consumer segmentation, and picture identification [18].

- **Semi-supervised learning:** Semi-supervised machine learning is a hybrid of supervised and unsupervised learning techniques. It employs a small number of labeled data and a large number of unlabeled data, providing the advantages of both unsupervised and supervised learning while avoiding the difficulties associated with locating a large amount of labeled data [18].
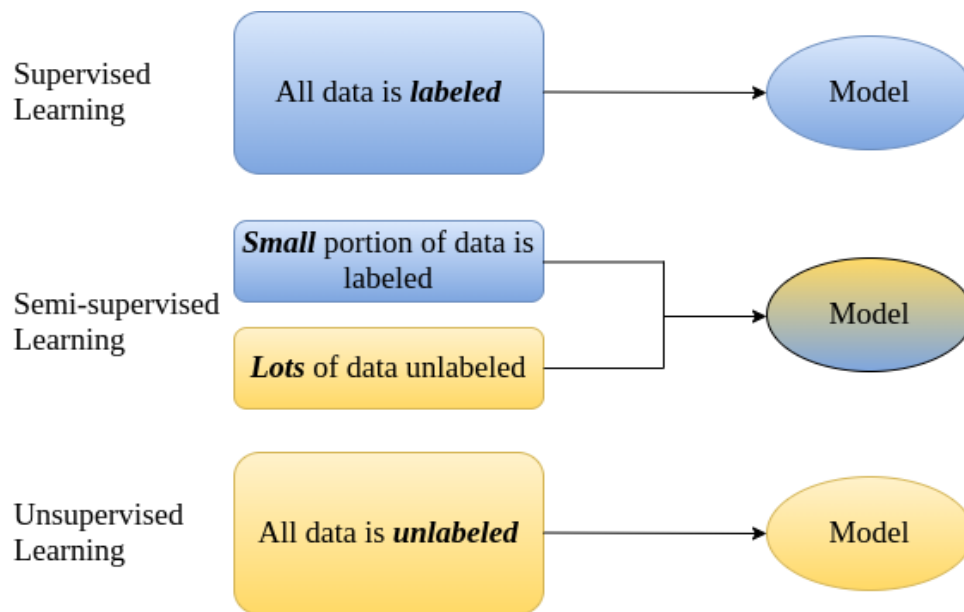
Figure 2.1: Supervised and unsupervised learning
[18]

2.3.1 Classical Machine Learning algorithms:

The algorithms described below are the classical machine learning algorithms employed in our study in order to classify the tabular dataset proposed in the third chapter.

2.3.1.1 Support Vector Machine:

The Support Vector Machine (SVM) is a well-known supervised machine learning technique that can be used for classification and regression applications [19]. SVMs are based on the idea of finding the super level that best separates the data into different classes.

2.3.1.2 K Nearest Neighbors:

The k-nearest neighbors or KNN algorithm is a simple, supervised machine learning algorithm that can be used for both classification and regression tasks. KNN works by finding the k closest data points in the training set to a new data point and then using the majority class of

those k neighbors to classify the new data point. The value of k must be chosen before training the model [19].

2.3.1.3 Logistic Regression:

Logistic regression is a statistical method used for predicting binary outcomes (i.e. outcomes that can take only two values). It is a type of regression analysis in which the dependent variable is binary and the independent variables can be continuous, discrete, or categorical. The logistic regression model uses the logistic function (also called the Sigmoid function) to model the relationship between the independent variables and the probability of the binary outcome [19].

2.3.1.4 Decision Tree:

A decision tree is a non-parametric supervised learning approach that can be used for classification as well as regression. It is a decision support tool that uses a tree-like model of decisions it maps out all possible courses of action and their potential outcomes [19].

2.3.1.5 Random Forest:

It consists of many decision trees that are built using feature randomness to create an uncorrelated forest of trees. The output of multiple decision trees is combined to make a final prediction that is more accurate than any individual tree [19].

2.3.1.6 Naïve Bayes Gaussian:

Gaussian Naive Bayes is a simple and efficient algorithm that can be used for both binary and multi-class classification problems based on the probabilistic approach and Gaussian distribution [19]. It assumes that all features are independent, which may not always be true in real-world scenarios. However, it still performs well in many cases.

2.4 Artificial Neural Networks:

Neural Networks (NNs) also known as Artificial Neural Networks (ANNs) are computing systems that take inspiration from the structure and function of biological neural networks of the human brains. These networks consist of interconnected units or nodes called artificial neurons that simulate the behavior of neurons in the brain. The connections between these neurons allow them to transmit signals, similar to how synapses work in the biological brain [20]. Artificial neurons can receive, analyze, and send impulses to neurons to which they are connected.
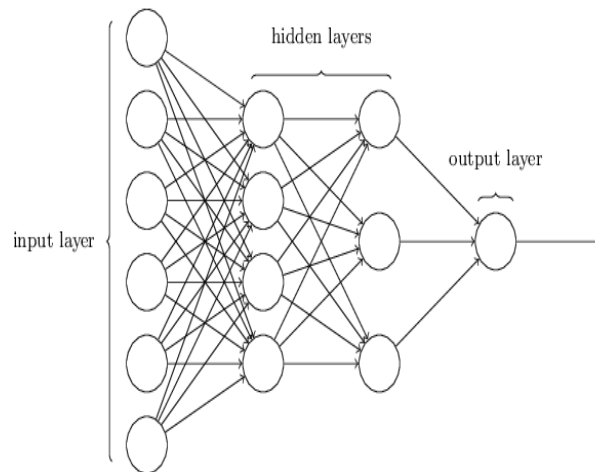
Figure 2.2: Topology of ANN
[20]

2.4.1   Convolutional Neural Network (CNN):

A convolutional neural network (CNN) is a type of artificial neural network commonly used for image or object detection and classification. Deep learning uses CNNs to recognize objects in images. CNNs play an important role in a variety of tasks, including image processing problems, computer vision tasks such as localization and segmentation, video analytics for detecting obstacles in self-driving cars, and speech recognition in natural language processing [21]. CNNs are very popular in deep learning because they play a key role in these rapidly growing new areas.
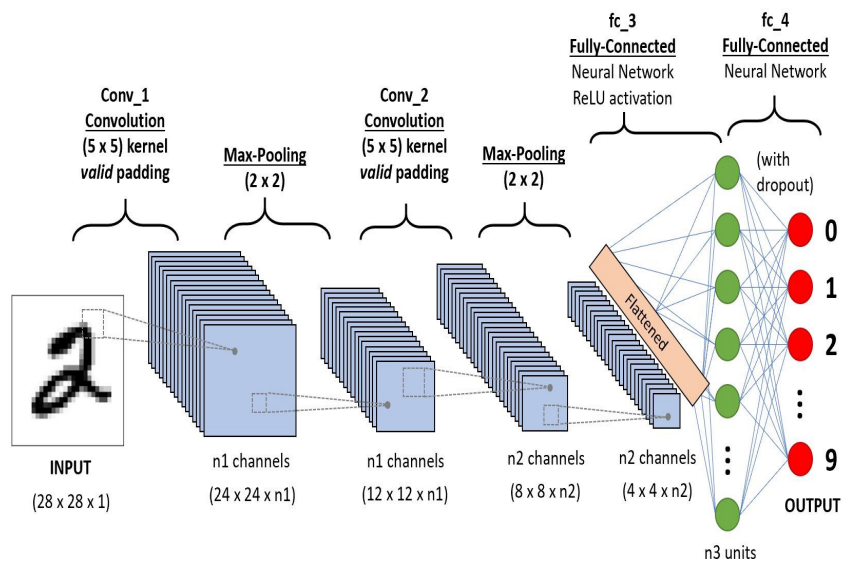
Figure 2.3: Convolution Neural Network model
[21]

2.4.2   Layers of Convolutional Neural Network:

2.4.2.1   Convolution layer:

This is the first layer utilized to extract the different features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a Kernel of a particular size M×M. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image mapped by the Kernel (M×M). The result is known as the feature map, and it contains information about the image such as the corners and edges.
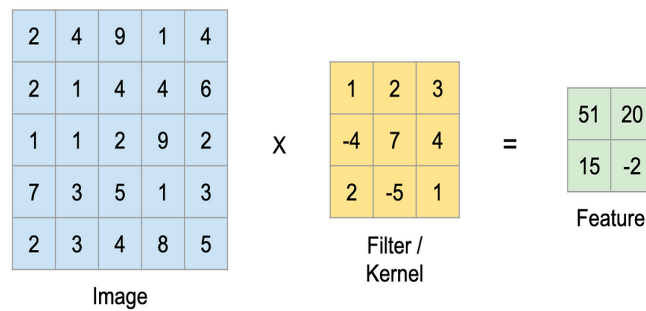
Figure 2.4: Convolution layer
[22]

Stride: Refers to the number of pixels that the filter also known as the kernel moves in each step while performing a convolution operation on the input image.
Padding: Used to describe how many pixels are added to an image when it is processed by the CNN kernel [23].

2.4.2.2   Activation functions:

An activation function is a mathematical function used in deep learning to introduce non-linearity in a neural network. It helps the network learn complex patterns and relationships in data. There are different types of activation functions commonly used, in our study we used two activation functions ReLu and Softmax [24].

- **Rectified linear activation function (ReLU):** It is a function that will output the input directly if it is greater or equal to zero, otherwise, it will output zero. It has become the default activation function for many types of neural networks since it is quicker to train and frequently results in higher performance [25].
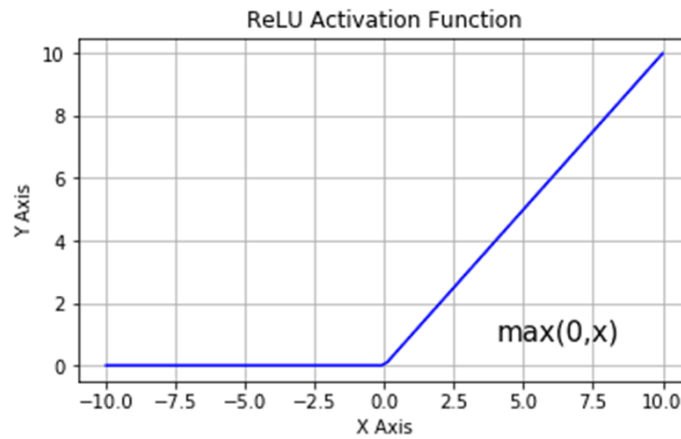
Figure 2.5: ReLu activation function
[25]

- **Softmax activation function:** Softmax activation function converts the neural network's raw outputs into a vector of probabilities. It assigns a value in the interval [0 1] to each class in a multi-class problem [26].

$$\delta(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^{N} e^{Z_j}} \tag{2.1}$$

Z : The vector of raw outputs from the neural network
i : The i-th entry in the softmax output vector softmax(z) can be thought of as the predicted probability of the test input belonging to class i.
N : The number of classes.
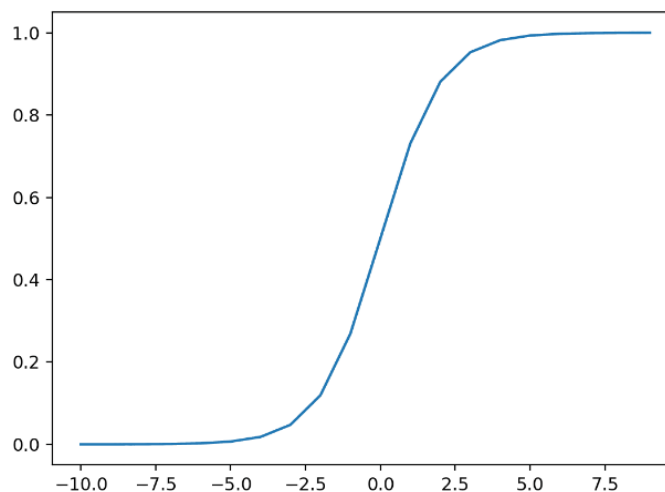
Figure 2.6: Softmax activation function
[27]

2.4.2.3 Pooling:

This layer's major goal is to lower the size of the convolved feature map in order to reduce computational expenses. This is accomplished by reducing the connections between layers and operating independently on each feature map. There are two types of pooling operations:

- **Max Pooling:** the largest element is taken from feature map [28].

- **Average Pooling:** calculates the average of the elements in a predefined sized image section [28].
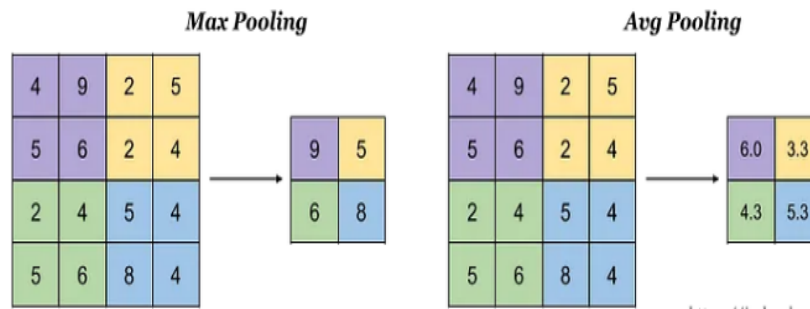
Figure 2.7: Max and average pooling
[28]

2.4.2.4 Fully Connected layer:

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers.
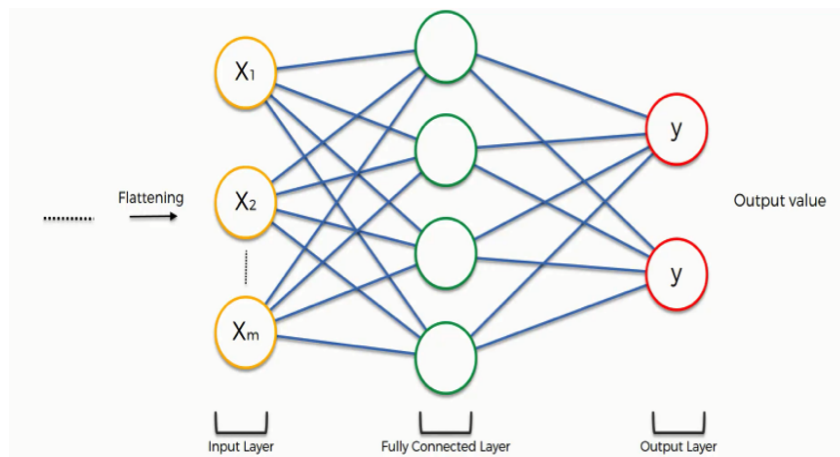
Figure 2.8: Fully connected layer
[29]

2.4.2.5 Dropout:

Dropout is a regularization technique used in the training of neural networks. Dropout aims to reduce overfitting in the network by randomly dropping out (i.e., setting to zero) part of the neurons during training. Each neuron has a probability of being removed from the network at each training step, and this probability is known as the dropout rate.

14
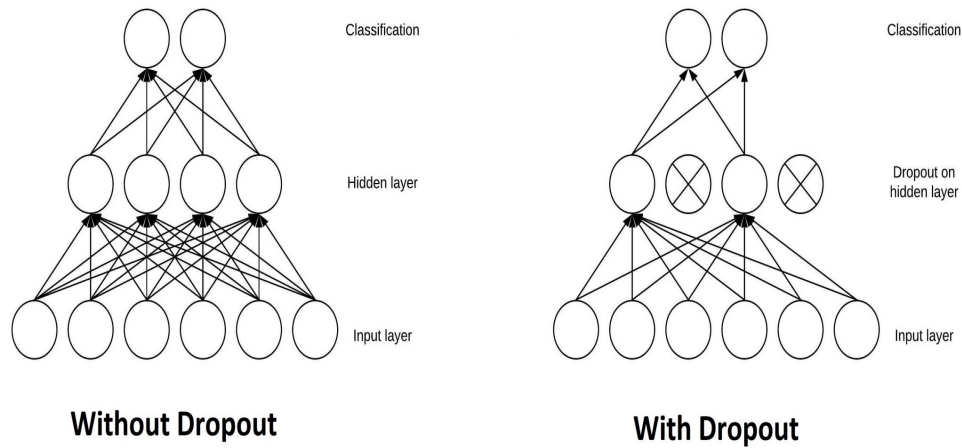
Figure 2.9: Neural network without and with dropout during training
[30]

2.4.2.6   Flattening layer:

Flattening is used to convert all the resultant arrays from pooled feature maps into a single long continuous linear vector (one dimensional vector).
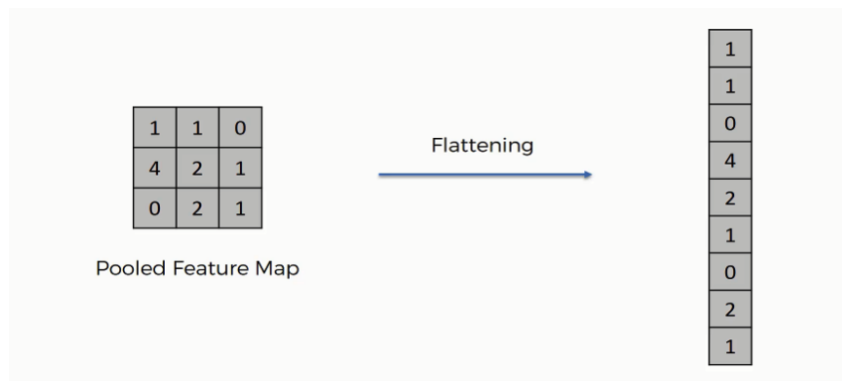
Figure 2.10: Flattening
[31]

2.5   Transfer learning:

Transfer learning is a machine learning technique that involves using a pre-existing model as a starting point for a new model instead of training it from scratch. The pre-existing model is usually pre-trained on a large and diverse dataset, such as ImageNet which is a large scale image database designed for visual object recognition research. It contains more than 14 million images covering over 20,000 categories organized and labeled, ranging from everyday objects to animals, plants, and scenes [32]. By doing that, the pre-existing model has learned to recognize basic patterns and features in the data that can be used to improve the performance of the new model [33]. We can use transfer learning by freezing pre-trained layers to prevent them from being updated or lose any knowledge during training. On the top we add new layers to the pre-trained model to adapt it to a specific task. These layers should be trainable and should be designed to perform the specific task. The model is then trained on the new data set.

2.5.1 Models of transfer learning:

2.5.1.1 VGG19:

VGG19 is a variant of the VGG model that supports 19 layers. The VGG19 model was introduced as an improvement over VGG16 but deeper and more complex by increasing the number of convolutional layers and adding more filters to each layer. The architecture of VGG19 consists of 19 layers, including 16 convolutional layers, 3 fully connected layers, and a softmax output layer. The input to the network is an image of dimensions (224, 224, 3). The first block consists of two convolutional layers having 64 channels of a 3*3 filter size, stride 1 and the same padding with ReLu activation function. Then it will pass thorough max pool layer of 2*2 filter size and stride 2, this results in output having shape of 112*112*64. Afterwards, it will pass through the second block which is similar to the first one but with 128 channels. The output will have the shape of 56*56*128. The third block consists of four convolutional layers having 256 channels of a 3*3 filter size, stride 1 and the same padding with ReLu activation function. Next it will pass through max pool layer of 2*2 filter size and stride 2, this results in output having with dimensions of 28*28*256. The fourth and fifth blocks are similar, they consist of four convolutional layers having 512 channels of a 3*3 filter size, stride 1 and the same padding with ReLu activation function. Then it will pass through max pool layer of 2*2 filter size and stride 2, the output of the fourth block 14*14*512 and for the fifth is 7*7*512. Finally, there are three fully connected layers with 4096 neurons each, followed by a softmax classifier with 1000 classes [34].
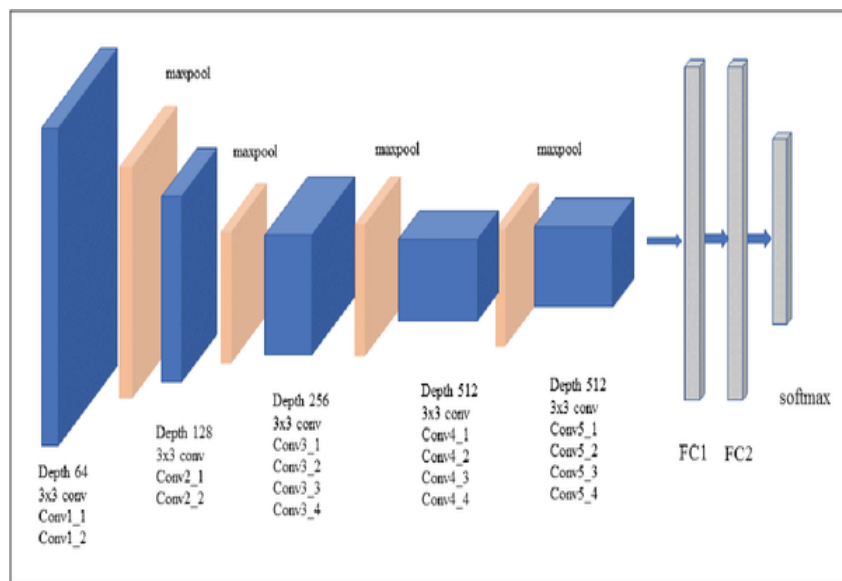
Fig. 3. VGG-19 network architecture

Figure 2.11: VGG19 architecture
[34]

2.5.1.2 ResNet 50:

ResNet stands for Residual Network which is designed to address the problem of vanishing gradients in very deep neural networks. It is a variant of the ResNet architecture consisting of 50 layers, developed by Microsoft Research in 2015 in research paper entitled "Deep Residual Learning for Image Recognition" by researchers Kaiming He, Xiangyu Zhang, Shaoqing Ren,

andJian Sunina. ResNet50 has a total of 50 layers, including 48 convolutional layers, one Max-Pool layer and one average pool layer. Its architecture is based on the residual learning concept to enable the training of very deep neural networks, which involves using skip connections to add the output of one layer to the input of another layer. The residual block contains skip connections in which the input to a layer is not directly passed to the next layer, but instead, it is passed through a shortcut connection that skips one or more layers. The output of this shortcut connection is then added to the output of the layers that it skiped, forming the final output of the block. This allows to the gradient to flow directly from one layer to another, bypassing the intervening layers enabling the model to learn from both the shallow and deep features of the data which leads to a better performance of the model.

ResNet 50 has 5 blocks each block contains convolution and identity block :
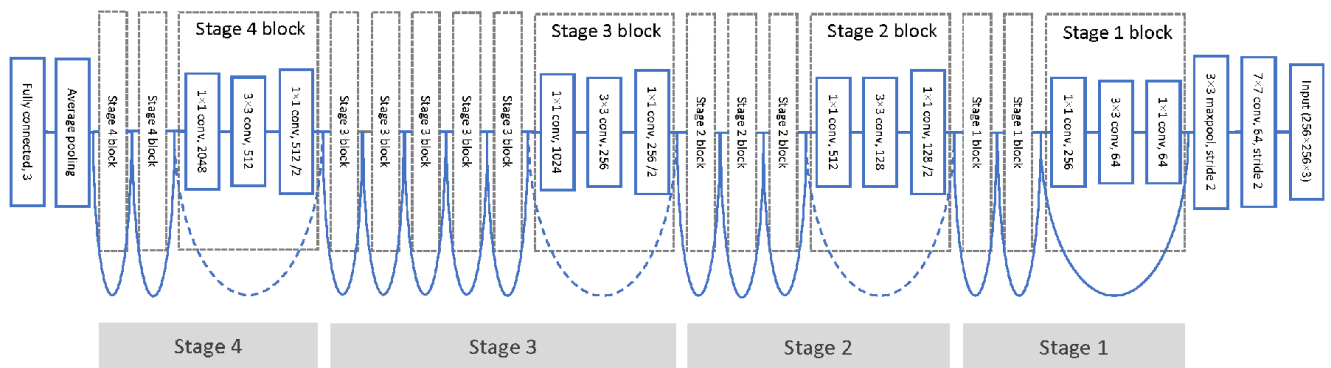
Figure 2.12: ResNet50 architecture
[35]

- **Convolution block:** The convolutional block is similar to the identity block, but it is used when the input and output dimensions are different, meaning that the shortcut connection needs to perform a convolution operation to match the dimensions. It has has three convolutional layers.

  - Convolution layer with kernel size 1x1 and stride 2 to reduce the spatial dimensions and increase the number of channels.

  - Convolution layer with kernel size 3x3 and padding to maintain the spatial dimensions.

  - Convolution layer with kernel size 1x1 to reduce the number of channels.

  - Shortcut connection performing a convolution operation to match the dimensions.

  - Addition of the output of the third convolution layer and the shortcut connection.

  - ReLU activation function applied to the output.

- **Identity block:** An identity block is used when the input and output dimensions are the same, meaning that the shortcut connection can be directly added to the main path. An identity block has three convolutional layers with Batch Norm and a ReLU activation function.

  - Convolution layer with kernel size 1x1 to reduce the number of channels.

17

–   Convolution layer with kernel size 3x3.

–   Convolution layer with kernel size 1x1 to increase the number of channels.

–   Addition of the original input to the output of the third convolution layer.

–   ReLU activation function applied to the output.
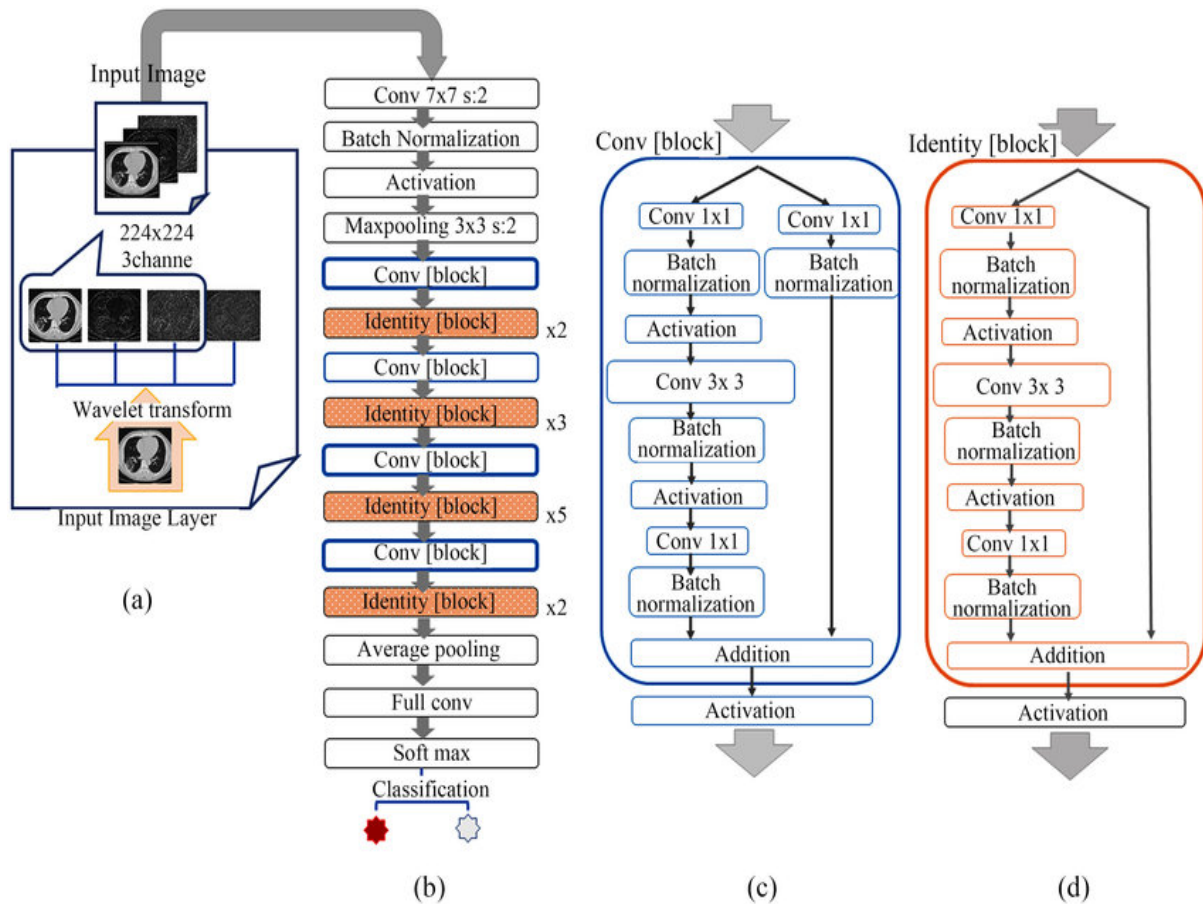
Figure 2.13: ResNet50 architecture showing convolutional and identity blocks
[36]

2.5.1.3   MobileNet:

MobileNet is an efficient convolutional neural network architecture designed by Google for efficient image classification tasks. It is composed of various layers, including depthwise separable convolutions, pointwise convolutions and global average pooling. The core idea behind MobileNet is to reduce the computational complexity and model size while still maintaining reasonable accuracy. There are two versions of MobileNet: MobileNet V1 and MobileNet V2, both of which are utilized for object recognition and classification. MobileNet V2 is an enhanced iteration of MobileNet V1, offering improved performance, efficiency and power. The architecture of MobileNet versions 1 and 2 is illustrated in Figure 2.14.
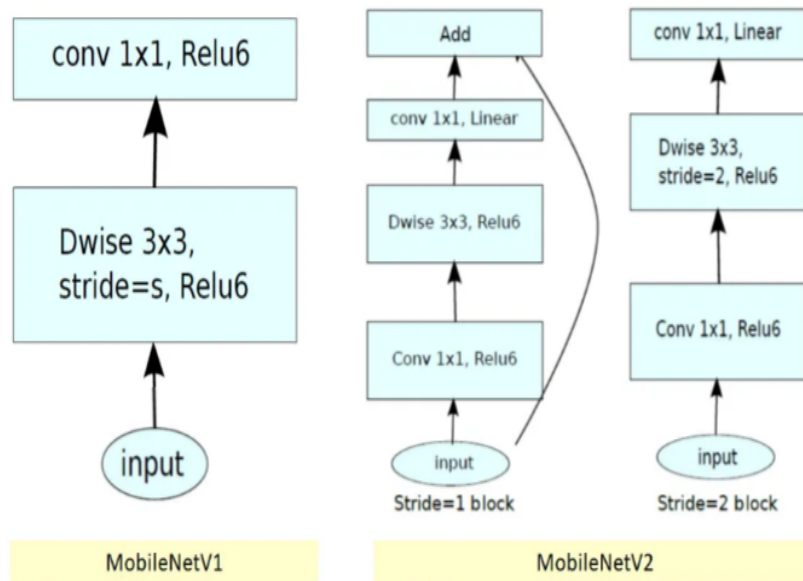
Figure 2.14: MobileNet architecture
[37]

2.5.1.4 Xception:

Xception is a convolutional neural network (CNN) architecture developed by François Chollet, Google researcher, in a 2016 paper called "Xception: Deep Learning with Depthwise Separable Convolutions", it was designed for image classification tasks and it is often used in transfer learning. The Xception architecture is composed of three parts: the entry flow, the middle flow, and the exit flow.

- **The entry flow:** It is the first part of the architecture of Xception.

    - Input image: 299 x 299 x 3 image.
    - Convolutional layer with 32 (3 x 3) filters, stride 2, and "same" padding.
    - Batch normalization.
    - ReLU activation function.
    - Convolutional layer with 64 (3 x 3) filters.
    - Batch normalization.
    - ReLU activation.
    - Depth-wise separable convolution with 128 (3 x 3) filters, "same" padding, and stride 2.
    - Batch normalization.
    - ReLU activation.
    - Depth-wise separable convolution with 128 (3 x 3) filters.
    - Batch normalization.
    - ReLU activation.

– Max pooling with (3x3) size and stride 2.

- **The middle flow:** It is repeated height times.

  – Depth-wise separable convolution with 128 (3 x 3) filters, "same" padding.

  – Batch normalization.

  – ReLU activation.

  – Depth-wise separable convolution with 128 (3 x 3) filters, "same" padding.

  – Batch normalization.

  – Residual connection: add input to output.

  – ReLU activation.

- **The exit flow:** The exit flow is designed to produce the final output of the model.

  – Depthwise separable convolution with 256 (3 x 3) filters, "same" padding, and stride 2.

  – Batch normalization.

  – ReLU activation.

  – Depth-wise separable convolution with 256 (3 x 3) filters, "same" padding.

  – Batch normalization.

  – ReLU activation.

  – Depth-wise separable convolution with 728 (3 x 3) filters, "same" padding.

  – Batch normalization.

  – ReLU activation.

  – Max pooling with (3 x 3) size and stride 2.

  – Depth-wise separable convolution with 728 (3 x 3) filters, "same" padding.

  – Batch normalization.

  – ReLU activation.

  – Depth-wise separable convolution with 1024 (3 x 3) filters, "same" padding.

  – Batch normalization.

  – ReLU activation.

  – Global average pooling.

Depth-wise convolution is a type of convolutional operation often used in deep learning models that require a large number of convolutional layers. It does not perform the convolution computation over all channels, but rather one at a time. Point-wise is a 1x1 convolution, used to reduce the dimensions.
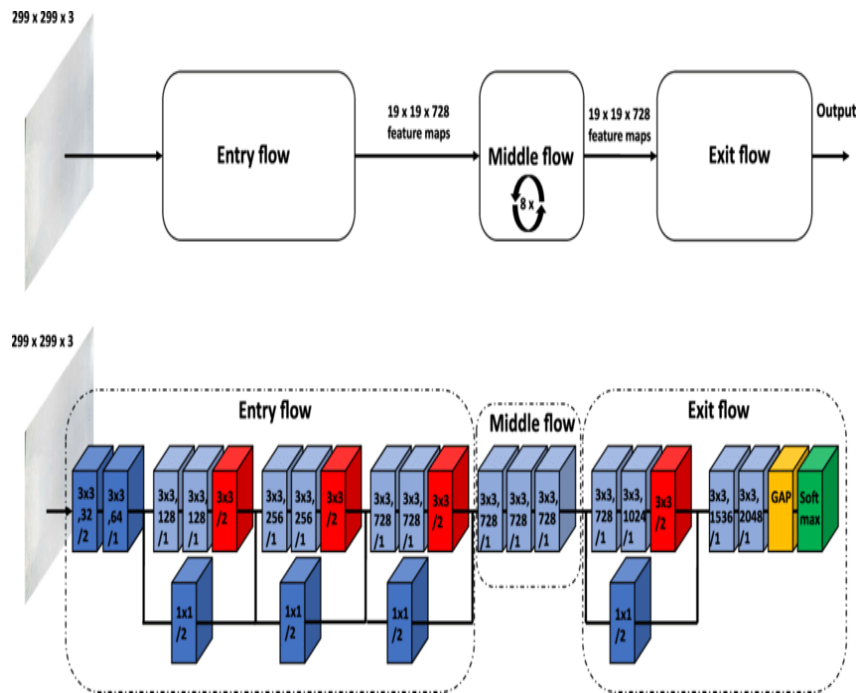
Figure 2.15: Xception architecture
[38]

2.5.2 Fine tuning:

Fine-tuning is a technique used in transfer learning to improve the performance of a pre-trained model on a new task. Once the model has been trained on the new data, all or part of the base model can be unfrozen and retrained end-to-end with a very low learning rate. This optional step can potentially give incremental improvements to the model, but it could also lead to overfitting [33].

2.5.3 Advantages of transfer learning:

Transfer learning has several advantages, such as faster training times, improved model performance, and the ability to train models with limited data. It can reduce the time and computational resources required to train a model from scratch. Furthermore, transfer learning can enhance model performance by leveraging the knowledge gained from a pre-trained model on a related task. This is especially beneficial when data is limited for the new task, as the pre-trained model can provide a strong starting point for the new model. Ultimately, transfer learning is a powerful technique that can enhance the efficiency and effectiveness of the models [33].

2.6 Overfitting and Underfitting:

2.6.1 Overfitting:

Overfitting happens when a machine learning model fits the training data too well, to the extent that it begins to memorize the data rather than learn the patterns. Consequently, the model becomes skilled at classifying or predicting the training data, but performs poorly on

new, unseen data [38]. The overfitting problem can be detected by comparing the validation metrics to the training metrics. If the validation metrics are significantly worse than the training metrics, it suggests that the model has overfit the training data [39].

2.6.2   Underfitting:

Underfitting occurs when a machine learning model is too simple to capture the underlying patterns in the data, leading to poor performance on both the training and testing data. This is often indicated by a large gap between the training and testing performance. The model may need to be made more complex or the training process may need to be improved in order to better capture the patterns in the data [39].
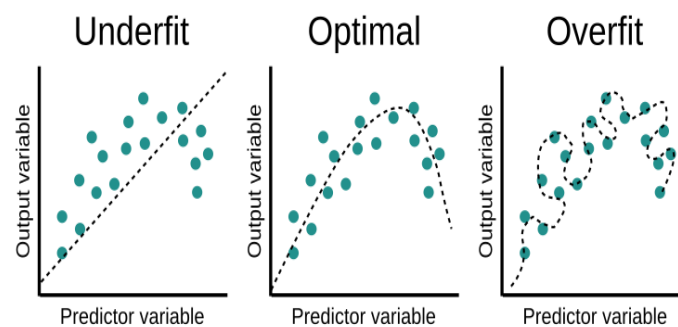
Figure 2.16: Overfitting underfitting and optimal-fitting
[39]

2.7   Summary:

This chapter provides an introduction to classical machine learning algorithms and deep learning. The chapter begins by presenting the algorithms of machine learning used for Alzheimer disease classification using the first dataset provided in the third chapter, and then delves into convolution neural networks (CNNs) by explaining its different layers. Transfer learning is also introduced, followed by an overview of the commonly used deep learning models and their architectures, which are VGG19, ResNet 50, MbileNet and Xception.

Chapter 3

Experiments & Results

3.1   Introduction:

In this section, we will discuss tha various performance metrics and tools used for the implementation and evaluation of the Alzheimer classification. Our approach involves utilizing two different datasets: a tabular dataset comprising demented and non-demented subjects for binary classification, and an MRI dataset encompassing four different stages of Alzheimer for multi-classification. We employ machine learning algorithms for the binary classification task and a Deep Neural Network (DNN) with Transfer Learning for the multi-classification task. The obtained results for both the binary and multi-classification tasks are thoroughly discussed in this chapter.

3.2   Tools:

In our study we used Python programming language and its different packages. For training and testing our models, we used Tensorflow and Keras framework as well as GPU which is provided by Kaggle.

3.2.1   Kaggle:

Kaggle is a platform for data scientists and machine learning practitioners that provides access to public datasets, code snippets, and notebooks. It has over 1 million registered users and is used by some of the world's best data scientists [40].

3.2.2   Python:

Python is an open-source, high-level, interpreted, general-purpose programming language and has efficient high-level data structures. It supports multiple programming paradigms, including structured, object-oriented, and functional programming. It was first released in 1991 and has since become one of the most used programming language in the world. Python has a clear and concise syntax, making it easy to learn and read, even for beginners [41].

3.2.3   TensorFlow:

TensorFlow is an end-to-end open-source platform for building and deploying machine learning models. It provides a comprehensive ecosystem of tools, libraries, and community resources that lets researchers and developers easily build and deploy machine learning-powered applications. TensorFlow was developed by Google and is widely used in industry and academia for a variety of machine learning tasks [42].

3.2.4   Keras:

Keras is an open-source software library that offers a Python interface for building artificial neural networks. It is designed to be user-friendly, modular, and easy to extend, and it follows best practices for reducing cognitive load. Keras acts as an interface for the TensorFlow library, which is used as the back-end for computation.

3.3   Performance Metrics:

Performance metrics are used to evaluate the performance of a machine learning models. There are a wide variety of metrics that can be used to evaluate machine learning models, but we will focus on the four most commonly used performance metrics.

3.3.1   Confusion Matrix:

The confusion matrix is a a square performance matrix used to visualize the performance of the model. It provides a detailed breakdown of the model's predictions and the actual labels, help in understanding the strengths and weaknesses of the classifier. It can be used for both binary and multi-classifications [43].

In case of binary classification the matrix is organized as follows:

• **True Positive (TP):** The number of instances that are actually positive and are correctly predicted as positive by the model.

• **False Negative (FN):** The number of instances that are actually positive but are incorrectly predicted as negative by the model.

• **False Positive (FP):** The number of instances that are actually negative but are incorrectly predicted as positive by the model.

• **True Negative (TN):** The number of instances that are actually negative and are correctly predicted as negative by the model.

Table 3.1: Binary confusion matrix structure

Classes	Class 1	Class 2
Class 1	TP	FP
Class 2	FN	TN

In case of multi-classification the confusion matrix is organized differently. Here is the organization for four classes classification.

Table 3.2: 4×4 confusion matrix structure

		Predicted class			
	Classes	ND	VMD	MD	MOD
Actual Class	ND	Number of samples predicted as ND and labeled as ND.	Number of samples predicted as VMD and labeled as ND.	Number of samples predicted as MD and labeled as ND.	Number of samples predicted as MOD and labeled as ND.
	VMD	Number of samples predicted as ND and labeled as VMD.	Number of samples predicted as VMD and labeled as VMD.	Number of samples predicted as MD and labeled as VMD.	Number of samples predicted as MOD and labeled as VMD.
	MD	Number of samples predicted as ND and labeled as MD.	Number of samples predicted as VMD and labeled as MD.	Number of samples predicted as MD and labeled as MD.	Number of samples predicted as MOD and labeled as MD.
	MOD	Number of samples predicted as ND and labeled as MOD.	Number of samples predicted as VMD and labeled as MOD.	Number of samples predicted as MD and labeled as MOD.	Number of samples predicted as MOD and labeled as MOD.

Where:
- **ND:** Non Demented.
- **VMD:** Very Mild Demented.
- **MD:** Mild Demented.
- **MOD:** Moderate Demenetd.

3.3.2 Accuracy:

Accuracy is a commonly used performance metric for classification tasks. It represents the or the ratio of the number of correct predictions to the total number of predictions [44].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

3.3.3 Recall:

Recall is a performance metric that measures the ability of a model to correctly identify positive instances for each class. Recall is calculated as the ratio of true positive instances to the sum of true positive and false negative instances for a particular class [44].

$$Recall = \frac{TP}{TP + FN} \tag{3.2}$$

3.3.4 Precision:

Precision measures the proportion of true positives among all positive predictions. A higher precision value indicates that the model has a lower rate of falsely labeling negative instances

as positive [44].

$$Precision = \frac{TP}{TP + FP} \qquad (3.3)$$

3.3.5 F1_score:

The F1_score combines precision and recall into a single value. It provides a balanced measure of a model's accuracy by taking into account both the ability to correctly identify positive instances (Precision) and the ability to capture all positive instances (Recall). It is particularly useful when the dataset is imbalanced[44].

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (3.4)$$

3.3.6 Macro average:

Macro accuracy is way to handle class imbalances by calculating the accuracy for each class independently and then taking the average across all classes. It treats each class equally, regardless of the number of instances it contains.

3.3.7 Weighted average:

Weighted accuracy is an extension of accuracy that considers the imbalance in class distribution within the dataset. It calculates the accuracy for each class individually and then computes a weighted average of these accuracies based on the proportion of instances in each class.

3.4 Hyperparameters:

Hyperparameters are model learning parameters that are defined by the user before training the neural networks and cannot be learnt from the data.

3.4.1 Epochs:

An epoch is one full pass through the algorithm of the training dataset. During each epoch, the model is presented with the training data, and the model's weights and biases are updated based on the observed errors or loss [45].

3.4.2 Batch size:

Batch size refers to the number of training samples used during training. The batch size is a hyperparameter that can be tuned during the training process.

3.4.3 Learning rate:

In machine learning, the learning rate is a hyperparameter that controls the step size or the rate at which a model's parameters are updated during the training process. Finding the optimal learning rate is challenging because a too high learning rate can cause the model to jump over minima, while a too low learning rate can either take too long to converge[46].

3.4.4 Optimizer:

In machine learning, an optimizer is an algorithm used to adjust the weights of a neural network in order to minimize the losses.

- **Adaptive Moment Estimation (Adam):** It maintains adaptive learning rates for each parameter and adapts the learning rate based on the moving average of gradient and gradient squared moments.

3.5 Tabular dataset classification:

3.5.1 Data description:

This data was obtained from Kaggle under the name "Alzheimer feature", which originates from the Open Access Series of Imaging Studies (OASIS). This dataset aims to classify subjects into two categories: demented and non demented. It is provided as a CSV (comma-separated values) file consists of a collection of 373 observations for 150 subjects, aged between 60 and 98 years. The dataset includes multiple observations for some patients, recorded at different time points separated by at least one year. All subjects included in the dataset are right-handed. Out of the total subjects, 64 individuals were characterized as demented during their initial visits and remained so for subsequent visits. Among these, 51 individuals had mild to moderate Alzheimer's disease. Additionally, 14 subjects initially characterized as non demented were later classified as demented during subsequent visits. The dataset is structured with 373 rows representing the subjects and nine columns representing the following features: 'M/F' (gender), 'Age', 'EDUC' (years of education), 'SES' (socioeconomic status), 'MMSE' (mini mental state examination), 'CDR' (clinical dementia rating), 'eTIV' (estimated total intracranial volume), 'nWBV' (normalize whole brain volume), 'ASF' (atlas scaling factor) and the target variable 'Group'. The table 3.3 given below provides the dataset features description, while figure 3.1 shows the five first samples from the dataset.

Table 3.3: Tabular dataset description

Features	Discription	Range
Group	Demented or Non_Demented	
Age	Age of a subject at the time of test observation	60 − 98
EDUC	Years of education	6 - 23
SES	Socioeconomic status	1 - 5
MMSE	Mini mental state examination value assessed through questionnaire	4 - 30
CDR	Clinical dimentia rating	0 - 2
eTIV	Estimated total intracranial volume	1106 - 2004
nWBV	Normalized whole brain volume: expressed as a percent of all voxels ("constant" for any value of estimated total intracranial volume)	0.644 - 0.837
ASF	Atlas scale factor: volume scaling factor for brain size ("constant" for any value of estimated total intracranial volume)	0.87 − 1.58

```
SHAPE:
NCOLS: 10  NROWS: 373
```

	Group	M/F	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
0	Nondemented	M	87	14	2.0	27.0	0.0	1987	0.696	0.883
1	Nondemented	M	88	14	2.0	30.0	0.0	2004	0.681	0.876
2	Demented	M	75	12	NaN	23.0	0.5	1678	0.736	1.046
3	Demented	M	76	12	NaN	28.0	0.5	1738	0.713	1.010
4	Demented	M	80	12	NaN	22.0	0.5	1698	0.701	1.034

Figure 3.1: Five first samples from the tabular dataset

3.5.2   Data preparation:

In this phase, several techniques were employed to clean and preprocess the data. The dataset initially consisted of three categories: 51% non demented, 39% demented, and 10% converted. However, since our goal is to classify subjects as either demented or non demented for Alzheimer classification, we needed to remove the converted group class. Additionally, categorical variables like gender and group were encoded into numerical values. During the data cleaning process, missing values were identified in two columns. Specifically, the SES column contained 19 missing values (5.09% of the data), and the MMSE column contained 2 missing values (0.54% of the data). To address this issue, two approaches were adopted :

- **Filtered data:** Is a process of selecting or cleaning the data to eliminate irrelevant, redundant, or inaccurate information.Filtering reduces the volume of data, improves the overall data quality and makes it more usable for analysis and decision making.

- **Imputed data:** Imputation is the process of replacing the missing values with estimated or predicted values based on the available data. The purpose is to preserve as much information as possible in the dataset, even if it leads to a slight loss of accuracy.

The dataset was divided into 70% of the total data for training and the remaining 30% for testing as shown in figure 3.2.

3.5.3   Features selection:

Feature selection is the process of reducing the number of the input variables when creating classification model. Its purpose is to identify most relevant informative features and eliminating irrelevant and redundant once. The feature selection methods used in our work are:

- **Correlation matrix:** A correlation matrix is employed to assess the linear association among multiple variables, enabling the identification of positive or negative correlations and the quantification of their strength. It proves beneficial in analyzing the relationships between variables and it is useful for:

Figure 3.2: Tabular dataset spliting

1. Identifying multicollinearity in a dataset before building a regression or classification model.

2. Understanding the inter-dependency between variables.

3. Data interpretation and visualization.

4. Feature selection for machine learning models.

In order to decide which features are related to our target 'Group', we plotted the correlation matrix as shown in figure 3.3.

The correlation matrix provides valuable insights about the relationship between Group variable and other features. The group variable exhibits a correlation of -0.22 with the EDUC feature, indicating a negative association. Conversely, there is a positive correlation of 0.16 between the 'Group' variable and the SES feature. Furthermore, the correlation analysis reveals that lower scores on the MMSE feature are associated with a higher likelihood of being in a certain group, with a negative correlation of -0.62. Similarly, the CDR exhibits a strong positive correlation of 0.86, indicating that as the CDR score increases, the likelihood of being in a certain group also significantly increases. Regarding the features eTIV and Age, they demonstrate weak negative correlations of -0.013 and -0.054, respectively, implying a minimal association. The nWBV exhibits a negative correlation of -0.33. Lastly, the correlation between the ASF and 'Group' is weakly positive, with a correlation coefficient of 0.0048, indicating that the ASF has a minimal impact on the likelihood of belonging to a certain group.

Based on the correlation matrix results, it is observed that gender, being a categorical variable, was not included. Therefore, conducting a hypothesis test is necessary to examine the its relationship with the target value.

- **Hypothesis independence test:** The hypothesis test for independence, also known as the chi-square test of association, is employed to examine the relationship between two variables. This test involves formulating null and alternative hypotheses to address the
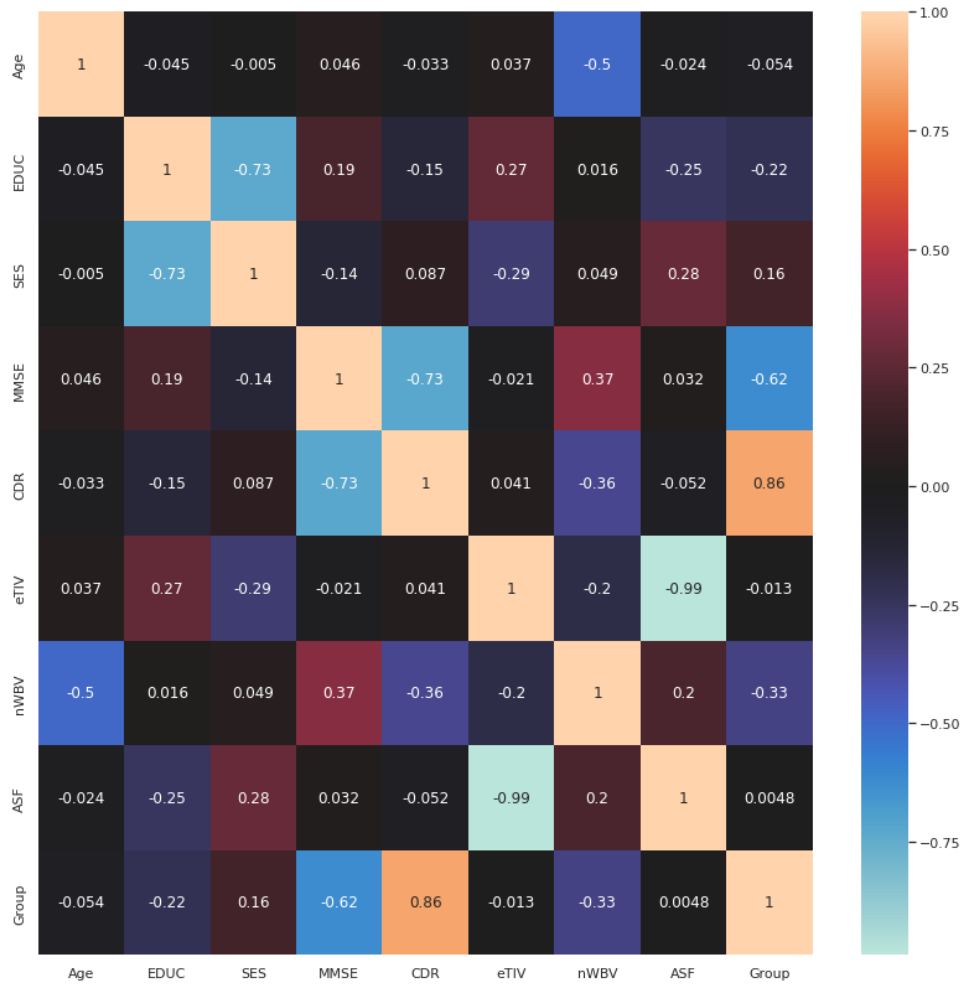
Figure 3.3: Correlation matrix plot

question of whether the two variables are related or not. The null hypothesis (H0) states that there is no relationship between variable 1 and variable 2 in the population. It suggests that the proportions or distributions of variable 1 are the same across different values of variable 2. On the other hand, the alternative hypothesis (Ha) proposes that there is a relationship between variable 1 and variable 2 in the population. It posits that the proportions or distributions of variable 1 differ across different values of variable 2.

To investigate the potential relationships between gender and the target variable group we conducted hypothesis test.We set the significance threshold to 0.05, indicating a level at which we would suspect a potential relationship, and the acceptance threshold to 0.2, which represents a higher level of acceptance for a relationship. The results obtained from these tests are presented below:

Since the p-value is less than the significance level (0.05), it provides evidence to reject the null hypothesis. In other words, there is strong evidence to suggest that the gender distribution is significantly different between the demented and non demented groups. The NON-DENIAL ZONE and ACCEPT ZONE are indicators of whether the null hypothesis is rejected or not based on the p-value and predetermined thresholds. In this case, both of them are labeled as False indicating that the null hypothesis is rejected and the alternative hypothesis is supported. To decide between the variables gender and CDR, we built a scatter plot to visually analyze relationship between the two features and the target.

```
--------------------------------------------------
HYPHOTESIS: Independece Test
--------------------------------------------------
H0: P(M/F | Dement) == P(M/F | Non-Dement)
H1: P(M/F | Dement) != P(M/F | Non-Dement)


--------------------------------------------------
TEST RESULT
--------------------------------------------------
T      : 24.187
P-VALUE: 7.326e-05

NON-DENIAL ZONE: False
ACCEPT     ZONE: False
--------------------------------------------------
```

Figure 3.4: Hypothesis test result
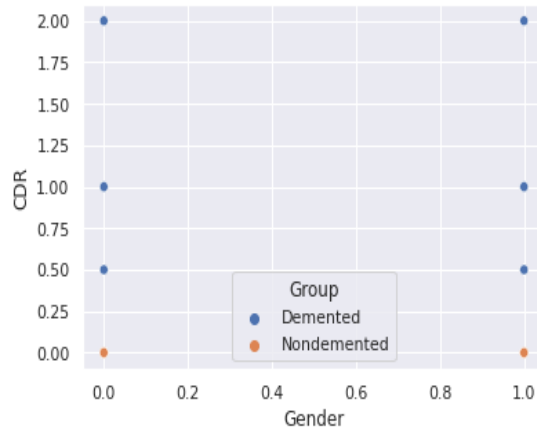
Figure 3.5: Scatter plot

Based on the results of the hypothesis test for independence, and correlation matrix and upon examining the scatter plot, a clear pattern emerges that when the CDR is greater than or equal to 0.5, the corresponding group is classified as demented. As a result of these findings, we choosed the CDR feature over gender for the rest of the experiment.

3.5.4 Obtained results using classical machine learning algorithms:

In this section, we employed two different methods that differ in their approach to handling missing values. The first method involved filtering the data, resulting in a reduced dataset of 317 rows. The second method involved imputing the missing values with the median value, resulting in a dataset of 336 rows. Table 3.4 provides an analysis of the resulting dataset, which was used for subsequent analyses.

We have conducted an evaluation of the performance using a range of different models, including Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Naive Bayes Gaussian and K Nearest Neighbor, utilizing distance-based weighting and the Manhat-

Table 3.4: Tabular dataset analysis

	Data	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
count	Filtered	317.00	317.00	317.00	317.00	317.00	317.00	317.00	317.00
	Imputed	336.00	336.00	336.00	336.00	336.00	336.00	336.00	336.00
mean	Filtered	76.71	14.62	2.55	27.26	0.27	1493.58	0.73	1.19
	Imputed	76.71	14.50	2.51	27.20	0.29	1491.30	0.73	1.19
std	Filtered	7.81	2.93	1.12	3.86	0.38	179.72	0.04	0.14
	Imputed	7.61	2.90	1.098	3.81	0.38	179.94	0.03	0.14
min	Filtered	60.00	6.00	1.00	4.00	0.00	1106.00	0.64	0.88
	Imputed	60.00	6.00	1.00	4.00	0.00	1106.00	0.64	0.88
25%	Filtered	71.00	12.00	2.00	27.00	0.00	1358.00	0.70	1.10
	Imputed	71.00	12.00	2.00	26.00	0.00	1357.00	0.70	1.09
50%	Filtered	76.00	15.00	2.00	29.00	0.00	1476.00	0.73	1.19
	Imputed	76.00	14.00	2.00	26.00	0.00	1475.00	0.73	1.19
75%	Filtered	82.00	16.00	3.00	30.00	0.50	1599.00	0.76	1.29
	Imputed	82.00	16.00	3.00	30.00	0.50	1599.75	0.76	1.29
max	Filtered	98.00	23.00	5.00	30.00	2.00	2004.00	0.84	1.59
	Imputed	98.00	23.00	5.00	30.00	2.00	2004.00	0.84	1.59

tan distance metric. The code incorporates several metrics to assess the performance of these classification models. The binary F1_score, accuracy and recall score or sensitivity. These metrics provide valuable insights into various aspects of model performance, encompassing overall accuracy, class-specific performance, and generalization capabilities. The results are shown in table 3.5.

Table 3.5: Tabular dataset classification results after feature selection

Model	Data	Accuracy	Recall	F1_score
Logistic Regressor	Filtered	1.0000	1.0000	1.0000
	Imputed	0.9881	1.0000	0.9863
Decision Tree	Filtered	1.0000	1.0000	1.0000
	Imputed	0.9881	1.0000	0.9863
Random Forest	Filtered	1.0000	1.0000	1.0000
	Imputed	0.9881	1.0000	0.9863
Naive Bayes Gaussian	Filtered	1.0000	1.0000	1.0000
	Imputed	0.9881	1.0000	0.9863
Support Vector Machine	Filtered	1.0000	1.0000	1.0000
	Imputed	0.9881	1.0000	0.9863
K Nearest Neighbors	Filtered	1.0000	1.0000	1.0000
	Imputed	0.9881	1.0000	0.9863

3.5.5 Discussion:

The results show the performance metrics of different models on both the filtered and imputed datasets. For all the models, including Logistic Regression, Decision Tree, Random Forest, Naive Bayes Gaussian, Support Vector Machine, and K Nearest Neighbors, the accuracy, recall and F1_score are consistently high for both the filtered and imputed datasets. In the filtered dataset, all models achieved impeccable accuracy with a score of 100%, signifying

accurate classification without any misclassifications. The recall, which evaluates the models' ability to identify positive instances, also attained a flawless score of 1.0 for both the filtered and imputed datasets, meaning there were no false negatives. Additionally, the F1_score, achieved a perfect score of 1.0 for all models on the filtered dataset, indicating an ideal balance between these two metrics. On the other hand the imputed data shows a slightly lower accuracy of 0.9881. While this indicates a small number of misclassifications, the overall performance is still quite high. The recall remains perfect at 1.00, suggesting that there were no false negatives in the imputed dataset. The F1_score is also high at 98.63%. Overall, the models perform well on both data.

3.6 MRI dataset classification:

3.6.1 Data description:

We performed a set of experiments on MRI dataset which is publicly available for the tasks of Alzheimer multi-stage classification on the KAGGLE platform. This dataset contains 6400 images that belong to four classes entitled Non Demented, Mild_Demented, Very_Mild_Demented and Moderate_Demented with each containing 3200, 896, 2240 and 64 images respectively. The size of the images is 128x128 pixels. Figure 3.6 represents the disribution of MRI scans dataset in each class while figure 3.7 represents samples from each.
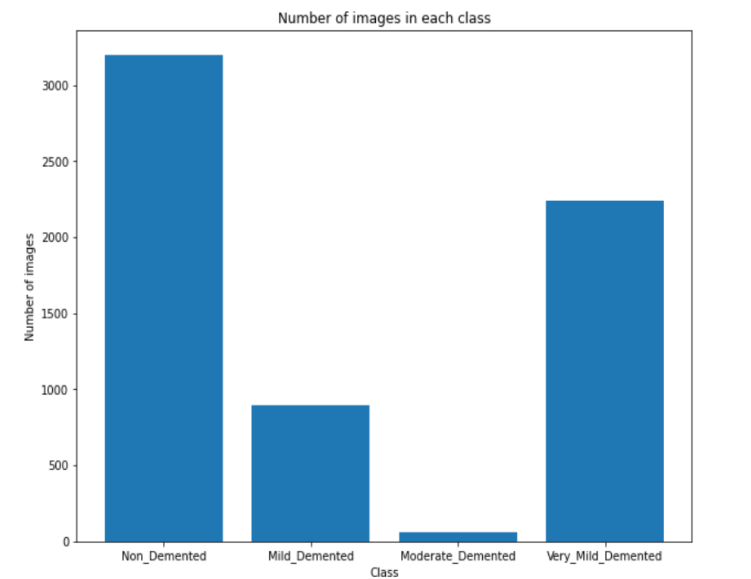
Figure 3.6: Distribution of MRI images in each class

3.6.2 Obtained results using Transfer learning and CNN:

This section will cover the training and evaluation process of four pre-trained models using the architectures presented in section 2.6 in addition to proposed CNN model. These models are evaluated using the metrics defined above. The dataset used for training was split into 70% for training, 10% for validation and 20% for testing. The training data was augmented using the ImageDataGenerator class from the Keras library.

Figure 3.7: Samples from MRI dataset

- Batch size: 32

- Optimizer: Adam

- Initial learning rate: 0.001

- Epochs: we changed depending on when the training loss and accuracy stopped improving (early stopping).

We used four different pre-trained deep convolutional neural networks: ResNet-50, VGG-19, Xception and MobileNet. We froze the weights of layers pre-trained on the ImageNet dataset. In addition to that, we have converted the images to the size 224x224 or 299x299 pixels in accordance with the specifications of the pre-trained networks we utilized in our trials.
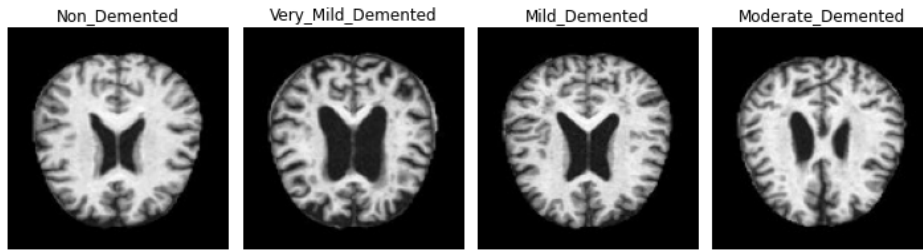
3.6.2.1 VGG19:

The VGG19 model has 28579268 total parameters (8488580 trainable and 20090688 non-trainable). It resulted in the learning curves of the accuracy and loss over 20 epochs, as shown in figure 3.8.

Figure 3.8: VGG19 model accuracy and loss curves of taining and validation sets

It can be seen from the accuracy and loss curves that the model is performing well on both training and validation data. Furthermore, there is no gap between the validation and training loss meaning that the model is generalizing well on unseen data and that model is not overfitting. The absence of fluctuations means that the model predictions are accurate. The confusion matrix is plotted to verify and compare the misclassified labels and the correctly classified labels as can be seen in figure 3.9. The resulted evaluations metrics are given in table 3.6.

Figure 3.9: Confusion matrix of VGG19 model

Table 3.6: Classification report of VGG19 model

	Precision	Recall	F1_score	Support
Mild_Demented	0.99	0.99	0.99	180
Moderate_Demented	0.86	1.0	0.92	14
Non_Demented	1.0	0.98	0.99	640
Very_Mild_Demented	0.98	0.99	0.98	448
Accuracy			98.60%	1282
Macro avg	0.96	0.99	0.97	1282
Weighted avg	0.99	0.99	0.99	1282

3.6.2.2 ResNet 50:

The Resnet 50 model has 57701636 total parameters (33851012 trainable and 23850626 non- trainable). It resulted in the learning curves of the accuracy and loss over 30 epochs, as shown in figure 3.10.
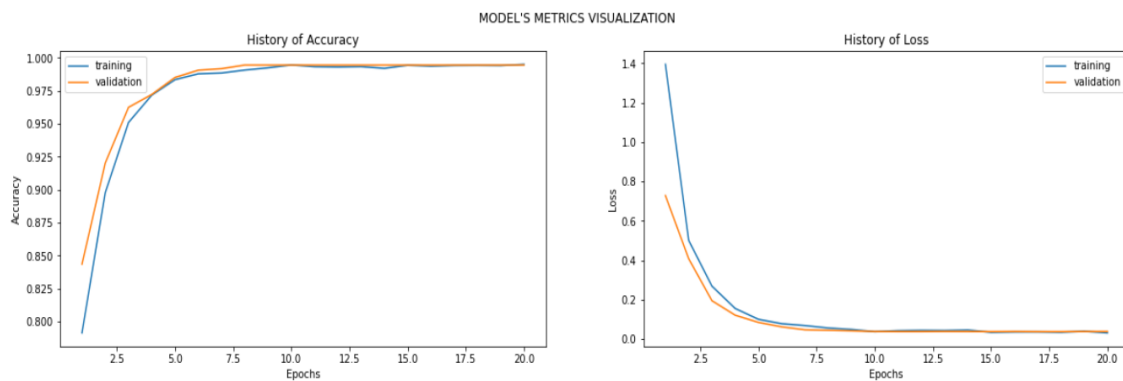

Figure 3.10: ResNet50 model accuracy and loss curves of taining and validation sets

It can be seen from the above figure that the validation accuracy is higher than the training accuracy and that the model is performing well on both training and validation data, meaning that the model is generalizing well on unseen data. The confusion matrix is plotted to verify and compare the misclassified labels and the correctly classified labels on the testing set as can be seen in figure 3.11. The resulted evaluations metrics are given in table 3.7.
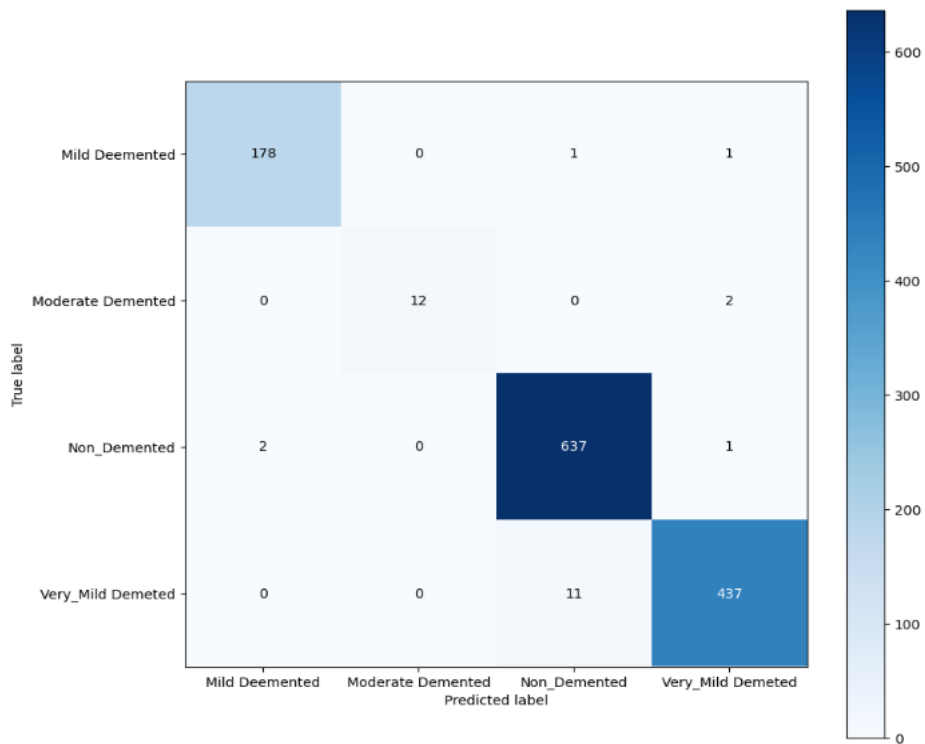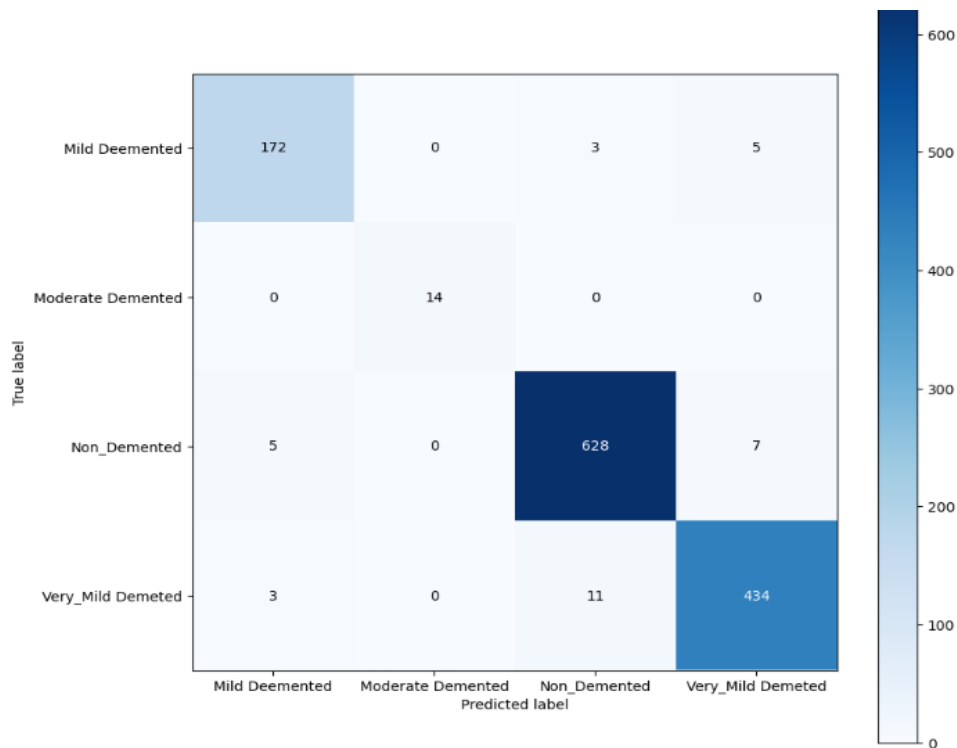
Figure 3.11: Confusion matrix of ResNet50 model

Table 3.7: Classification report of ResNet 50 model

	Precision	Recall	F1_score	Support
Mild_Demented	0.96	0.96	0.96	180
Moderate_Demented	1.0	1.0	1.0	14
Non_Demented	0.98	0.98	0.98	640
Very_Mild_Demented	0.97	0.97	0.97	448
Accuracy			97.35%	1282
Macro avg	0.98	0.98	0.98	1282
Weighted avg	0.98	0.98	0.98	1282

3.6.2.3 Xception:

The Xception model has 54975404 total parameters (33851012 trainable and 21124392 non-trainable). It resulted in the learning curves of the accuracy and loss over 25 epochs, as shown in figure 3.12.
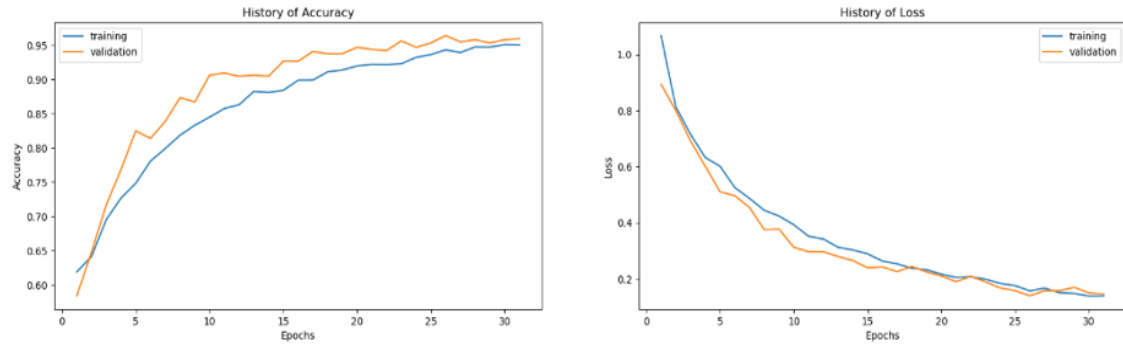It can be seen from the accuracy and loss curves that the model is performing well on training data but less performing on validation data. Meaning that model in unable to generalize well on unseen data. The confusion matrix is plotted to verify and compare the miss classified labels and the correctly classified labels as can be seen figure 3.13.

Figure 3.12: Xception model accuracy and loss curves of taining and validation sets
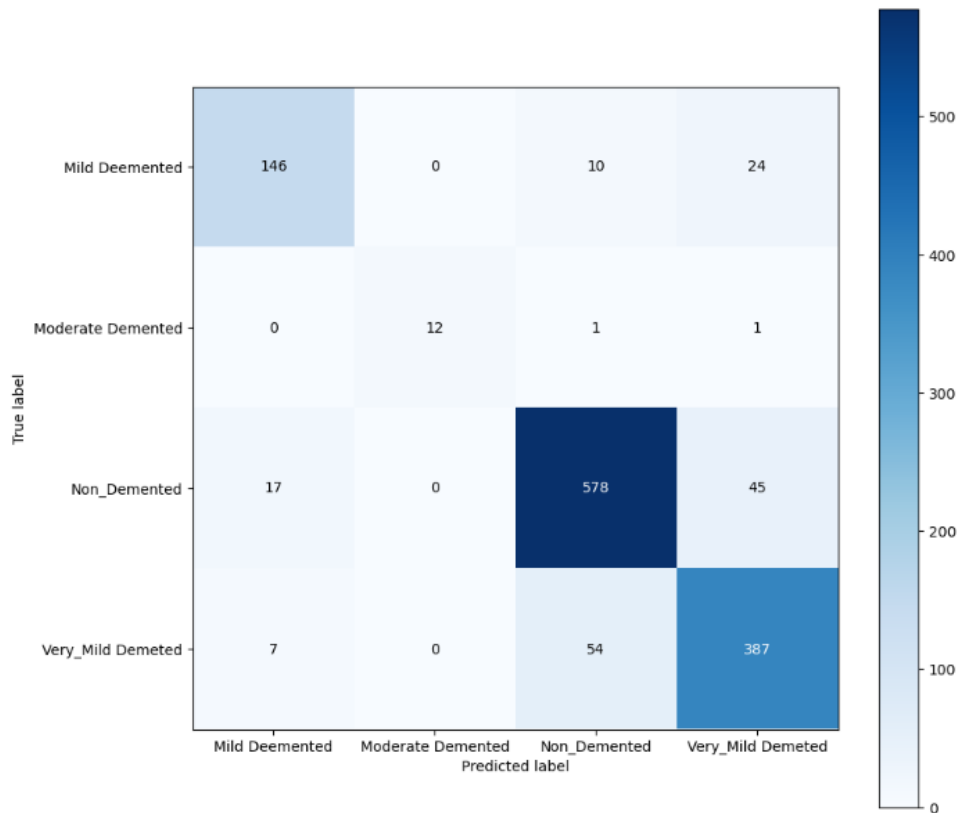
Figure 3.13: Confusion matrix of Xception model

We noticed an increase in misclassification between the classes Non_Demented and Very_Mild_Demented due the difficulty of detecting early stages of Alzheimer through MRI scans. The resulted evaluations metrics are given in table 3.8.

3.6.2.4 MobileNet:

The MobileNet model has 20303428 total parameters (16942724 trainable and 3360704 non- trainable). It resulted in the learning curves of the accuracy and loss over 14 epochs, as shown in figure 3.14.

Table 3.8: Classification report of Xception model

	Precision	Recall	F1_score	Support
Mild_Demented	0.81	0.86	0.83	180
Moderate_Demented	0.86	1.0	0.92	14
Non_Demented	0.90	0.90	0.90	640
Very_Mild_Demented	0.86	0.85	0.86	448
Accuracy			86.35%	1282
Macro avg	0.86	0.88	0.89	1282
Weighted avg	0.89	0.90	0.98	1282

Figure 3.14: MobileNet model accuracy and loss curves of taining and validation sets

It can be seen from the accuracy and loss curves that the model is doing better than Xception model but overfitting is still present. The confusion matrix is plotted to verify and compare the misclassified labels and the correctly classified labels as can be seen in figure 3.15.
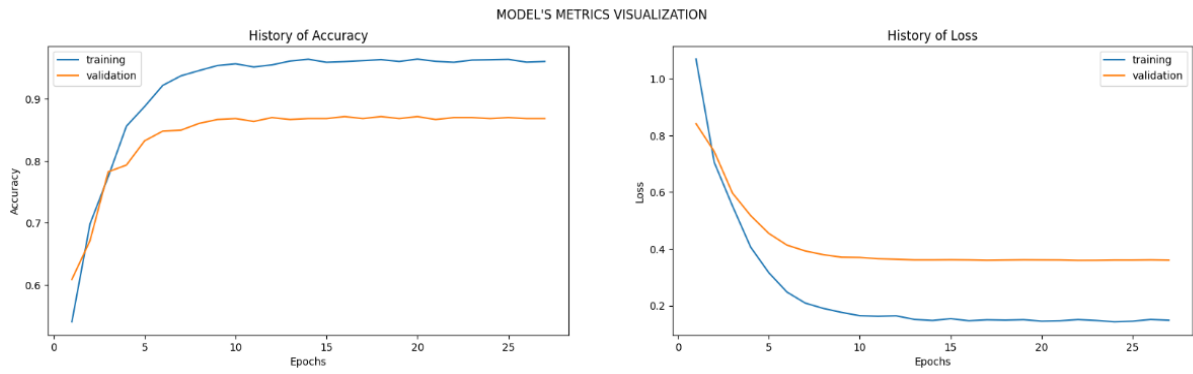We noticed that the misclassification between the classes Non_Demented and Very_Mild_Demented are reduced compared to the Xception model. The resulted evaluations metrics are given in table 3.9.

Table 3.9: Classification report of MobileNet model

	Precision	Recall	F1_score	Support
Mild_Demented	0.94	0.94	0.94	180
Moderate_Demented	1.0	1.0	1.0	14
Non_Demented	0.97	0.96	0.97	640
Very_Mild_Demented	0.94	0.94	0.94	448
Accuracy			95.40%	1282
Macro avg	0.96	0.96	0.96	1282
Weighted avg	0.96	0.96	0.96	1282
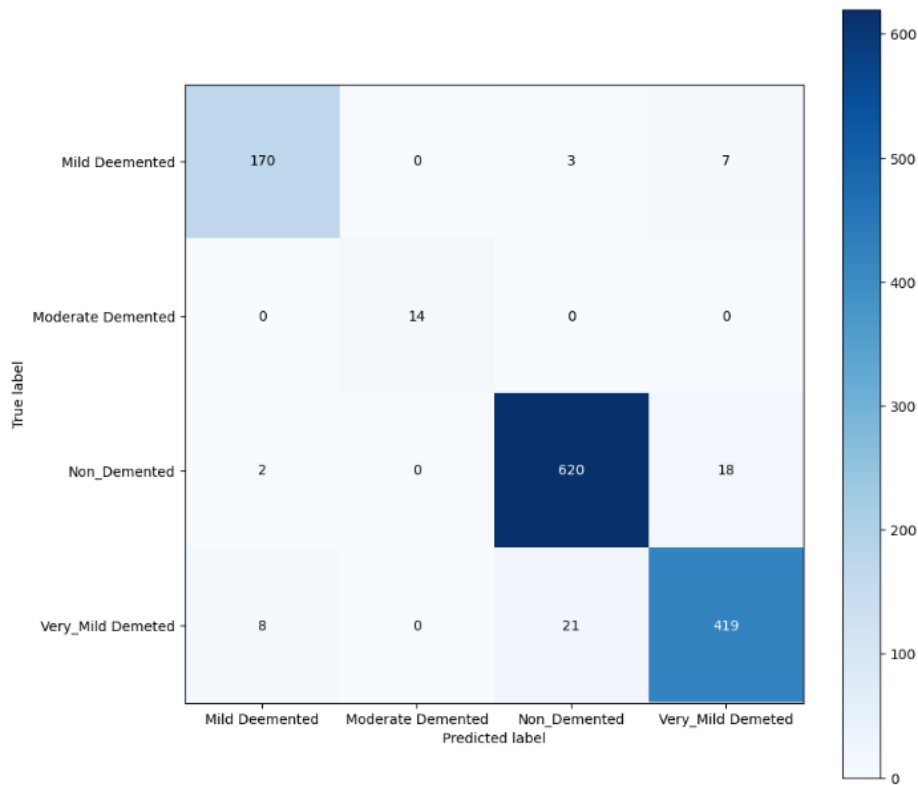
Figure 3.15: Confusion matrix of MobileNet model

3.6.3 Implementing custom CNN model:

In this experiment, we have designed a CNN of 4 convolutional layers. Every convolutional layers is followed by one max pooling layer and a dropout. The architecture ends with two fully connected layers and finally an output layer with four units representing our four classes. The batch size is set to 32 with 50 epochs, Adam as optimizer with a learning rate of 0.001. The model has 593333 total parameters all trainable. The accuracy and loss curves are plotted with respect to the number of epochs as shown in figure 3.16.

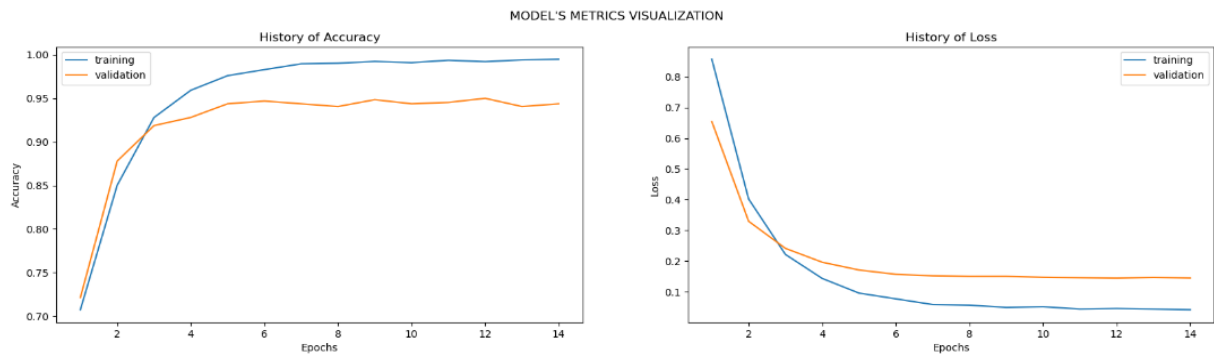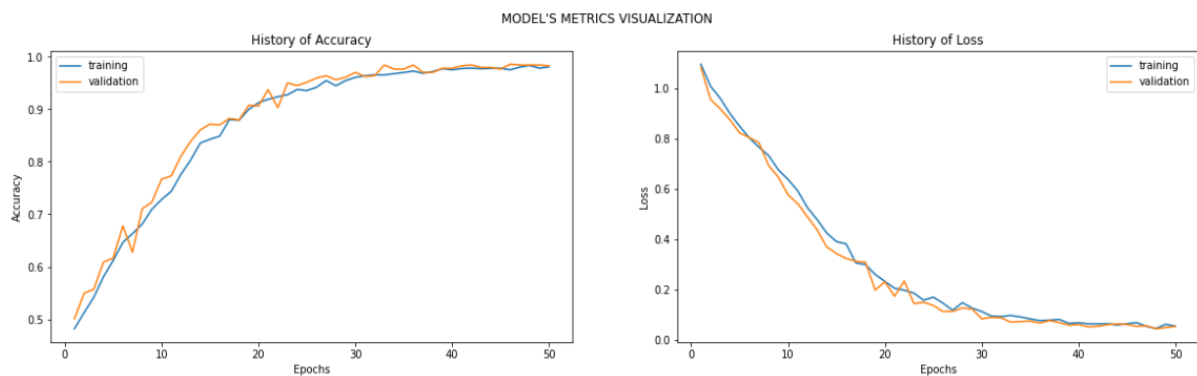

Figure 3.16: Proposed CNN model accuracy and loss curves of taining and validation sets

It can be seen from the accuracy and loss curves that there is no overfitting, some fluctations are seen due the noise present in some batches. The confusion matrix is plotted to verify and compare the misclassified labels and the correctly classified labels as can be seen the figure

40

Figure 3.17: Confusion matrix of the proposed CNN model

below.

We noticed from the confusion matrix that the model is doing well on distinguishing between the classes Non_Demented and Very_Mild_Demented. The resulted evaluations metrics are given in table 3.10.

Table 3.10: Classification report of the proposed CNN model

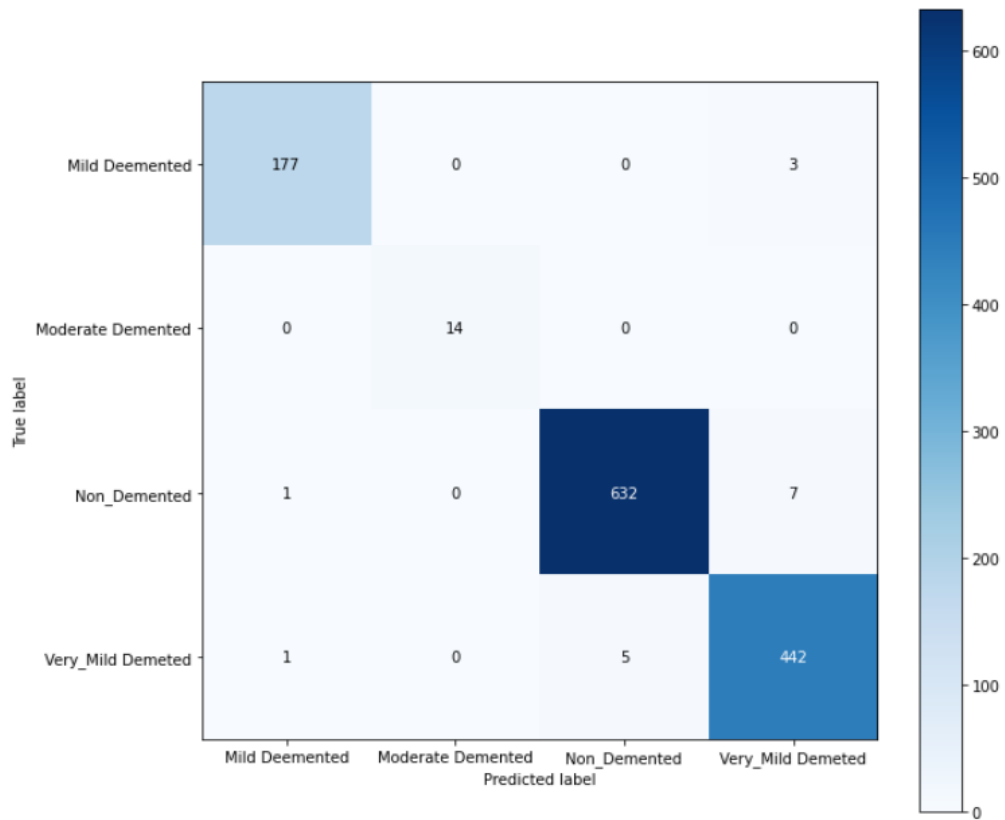|	Precision	Recall	F1_score	Support
Mild_Demented	0.99	0.99	0.99	180
Moderate_Demented	1.0	1.0	1.0	14
Non_Demented	1.0	0.98	0.99	640
Very_Mild_Demented	0.98	1.0	0.99	448
Accuracy			99.00%	1282
Macro avg	0.99	0.99	0.99	1282
Weighted avg	0.99	0.99	0.99	1282

3.6.4 Discussion:

In this study four transfer learning experiments (VGG19, ResNet50, Xception, and MobileNet) were implemented to classify brain MRI scans into four classes. The testing accuracies obtained were 98.60% for VGG19, 97.35% for ResNet50, 86.35% for Xception, and 95.40% for MobileNet. The VGG19 model outperformed other models by producing the most accurate predictions on the testing dataset. Not only it did produce better accuracy, but it also demonstrated shorter training and execution times, as well as fewer trainable parameters. The ResNet50 model achieved the next best accuracy; however, it had the longest training and execution times and involved the highest number of trainable parameters. On the other hand, the Xception model yielded the lowest accuracy. Although this architectural design is known for efficient learning, it does not guarantee improved performance. In comparison, the MobileNet model, being simpler in architecture, exhibited improved performance compared to Xception and showcased the least training and execution time. These variations in performance can be attributed to several factors. Firstly, the architecture of the models that played a role allowing VGG19 and ResNet50 to effectively capture and generalize patterns enabling them to capture finer details and subtle patterns in the MRI scans, leading to higher accuracy while Xception and mobileNet may have struggled to generalize well on this specific dataset. Additionally, the number of parameters in the models influenced their capacity to learn from the data. Moreover, the complexity of the task and the limited size of the dataset could have affected the performance. The process of fine-tuning and hyperparameter selection also played a role, as VGG19 and ResNet50 might have been better aligned with the dataset during the fine-tuning process. Despite the satisfying results obtained from the transfer learning models, we proposed a customized CNN model made specifically for our dataset. This model achieved the best accuracy with reduced training and execution times, along with a lower parameter count.

When building the CNN model, we experimented various experiments to assess the impact of adding or removing the convolutional layer, modifying the number of kernels and kernel sizes as well as different batch size and the initial learning rate.

Table 3.11 provides a comparison of accuracy, macro average, weighted average, training time, and execution time for different models: VGG19, RESNET50, Xception MobileNet, and CNN model.

Table 3.11: Summary of the obtained results

	Accuracy	Macro avg	Weighted avg	Training time	Execution time
VGG19	98.60%	95.41%	98.59%	143s309ms	550ms
ResNet50	97.35%	96.25%	97.43%	1327s488ms	1146ms
Xception	86.35%	85.88%	87.59%	610s352ms	902ms
MobileNet	95.50%	96.21%	95.39%	176s786ms	577ms
Proposed CNN	99.00%	99.35%	99.06%	205s358ms	47ms

Table 3.12 presents the accuracy achieved by each model under different conditions, considering data splitting ratios and the presence or absence of dropout layer. It showcases the impact of these factors on the performance of the models. The table showcases the stability of the model, as evidenced by its consistent performance across different data split ratios.

Table 3.12: Results under different conditions

Model	Dropout	70-10-20	60-10-30
VGG19	With	98.60%	95.00%
	Without	98.05%	97.00%
ResNet50	With	97.35%	93.00%
	Without	96.96%	95.00%
Xception	With	86.35%	87.00%
	Without	90.41%	84.00%
MobileNet	With	94.40%	93.00%
	Without	94.53%	92.00%

Table 3.13 provides a summary of different studies and their reported accuracies using various methods and models and splitting along with our study. It's important to note that the reported accuracies are specific to each study's methodology, and experimental setup. Comparisons between studies should be made with caution, as experimental conditions can vary significantly. However, these results provide valuable insights into the performance of different models across multiple studies, showcasing the capabilities and limitations of each approach. By considering the data splitting as our point of comparison, we observed significant improvements in the accuracy of our ResNet50 and VGG19 models. Specifically, we achieved an accuracy of 97.45% using the same data splitting ratio as in previous work. Moreover, even when applying a 70-30 data split ratio, our models consistently outperformed previous results, demonstrating their superior performance.

Table 3.13: Comparison of obtained results with related work

Study	Method	Splitting	Accuracy
Ms.Heta Acharya et al. [15]	VGG19	80-20	85.07%
	ResNet50		75.25%
	CNN		88.79%
	AlexNet		95.70%
Gaurav kumar ameta et al. [13]	VGG19	80-20	95.88%
	ResNet50		93.52%
	Xception		89.77%
	DenseNet201		96.59%
	EfficentNet		83.20%
Madhusudn G et al. [12]	CNN with KNN	80-20	99.92%
Marwa El-Geneedy et al. [16]	CNN	80-20	96.84%
	VGG16		96.39%
	InceptionV3		87.71%
	ResNet201		89.71%
Waleed Al Shehri [17]	DenseNet169	70-30	83.82%
	ResNet50		81.92%
Our study	VGG19	80-20	98.60%
	ResNet50		97.35%
	Xception		86.35%
	MobileNet		95.50%
	CNN		99.00%

3.7 Summary:

In this chapter, we described the performance metrics used to implement and evaluate transfer learning models and CNN model for MRI dataset along with classical machine learning models for the tabular dataset. Then, we separately plotted and discussed for each experiment the curves, confusion matrices, and classification reports. Next, we drew a summary table to resume all the models' findings. Additionally, we interpreted compared and discussed the results for all experiments.

General conclusion

    The objective of this project was to classify and categorize the various stages of Alzheimer's disease mainly, the early stage identification and diagnosis of the disease in order to offer early intervention, better care and provide preventive measures to combat it for those affected, thereby preventing further deterioration that eventually results in death.

In this project, we used tabular data containing nine features and determined that the Clinical Dementia Rating (CDR) is the key feature for classifying Alzheimer's disease. Additionally, the mini mental state examination (MMSE) emerged as another important feature. Since both these features are highly correlated, it is important to choose only one to avoid redundancy in the model.

In this study, two approaches were proposed and develloped to tackle the multi-classification task of MRI dataset: a CNN model and transfer learning. The project started by presenting the implementation of four different DL architectures (VGG19, Xception, MobileNet and ResNet50). These models were trained using transfer learning techniques on the dataset consisting of 6400 pictures, categorized into four classes that represent different stages of dementia in Alzheimer's disease. After evaluating the models, we found promising results. the VGG19 model outperformed the transfer learning models followed by ResNet50 model which had the slower execution time. Next the mobileNet with the fastest execution time and finally Xception model. Secondly we moved into the development of a CNN model to successfully identify Alzheimer's disease stages with 99.00% accuracy.

In the transfer learning section, we deduced that VGG19 Model was best suited for our specific task. Furthermore, fine tuning the model is crucial to obtain best results, adding dropout and batch normalization layers had a noticeable positive impact on performance. We also found that using the learning rate scheduler callback helped the model converging more rapidly than when we used only the optimizer, the use of early stopping callback also helped saving time in the training process.

The studies lead us to the conclusion that the depth and complexity of the architectures is critical to achieve desired results, and that the optimal depth is directly linked to the complexity and size of the dataset. Additionally, we concluded that a CNN model that is build and trained on our dataset gives the best results since pre-trained models are built for more general images whereas CNN is more specialized for our specific task. On the other hand, one notable limitation of this work is the limited availability of data and the associated privacy concerns.

Our contributions in this work were firstly, successfully implementing and enhancing the accuracy of the VGG19 and ResNet50 models. Additionally, We thoroughly compared our work with recent published work that utilized the same dataset, ensuring the validity of our results. We compared four transfer learning models to assess their performance and detailed comparison between transfer learning and CNN model. Furthermore, we conducted experiments involving the fine-tuning as well as experimenting with diffrent splitting without degrading the performance of the models.

Looking ahead, there are several ways for future research and development. One direction

involves integrating both tabular and MRI data into a unified model using fusion methods algorithms. This integration would provide a more comprehensive understanding of the disease by incorporating different types of information. Furthermore, we plan to explore MRIs with improved resolution to handle more intricate features and details.

Bibliography

[1] *La maladie d'Alzheimer.* Assessed on March 2023. URL: https://sante.gouv.fr/soins-et-maladies/maladies/maladies-neurodegeneratives/article/la-maladie-d-alzheimer.

[2] Wikipedia contributors. *Alzheimer's disease.* Assessed on March 2023. URL: https://en.wikipedia.org/w/index.php?title=Special:CiteThisPage&page=Alzheimer%5C%27s_disease&id=1155290865&wpFormIdentifier=titleform.

[3] *Dimentia.* Assessed on March 2023. URL: https://www.who.int/news-room/fact-sheets/detail/dementia#.

[4] Li et al. "Global, regional, and national burden of Alzheimer's disease and other dementias". In: *The alzheimer's disease challenge, volume II* 16648714 (2023), p. 194.

[5] *Diagnosing Alzheimer's: How Alzheimer's is diagnosed.* Assessed on March 2023. URL: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers/art-20048075.

[6] *Treateement.* Assessed on March 2023. URL: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/diagnosis-treatment/drc-20350453.

[7] Liu Y el al. "MPC-STANet: Alzheimer's disease recognition method based on multiple phantom convolution and spatial transformation attention mechanism". In: *Frontiers in Aging Neuroscience* 14 (2022), p. 530.

[8] *Medical imagind.* Assessed on March 2023. URL: https://en.wikipedia.org/wiki/Medical_imaging.

[9] *Comparing MRI and CT scan.* Assessed on April 2023. URL: https://www.omegapds.com/comparing-mri-and-ct-scan/.

[10] *Brain Scans and Dementia.* Assessed on April 2023. URL: https://stanfordhealthcare.org/medical-conditions/brain-and-nerves/dementia/diagnosis/brain-scans.html.

[11] Feyza ALTUNBEY ÖZBAY and Erdal ÖZBAY. "An NCA-based Hybrid CNN Model for Classification of Alzheimer's Disease on Grad-CAM-enhanced Brain MRI Images". In: *Turkish Journal of Science and Technology* 18.1 (2023), pp. 139–155.

[12] Madhusudan G and Parab et al. "Development of framework by combining CNN with KNN to detect Alzheimer's disease using MRI images". In: *Multimedia Tools and Applications* 82.8 (2023), pp. 12699–12717.

[13] Sisodia et al. "A review of deep transfer learning approaches for class-wise prediction of Alzheimer's disease using MRI images." In: *Archives of Computational Methods in Engineering.* 30.(4) (May 2023), pp. 2409–2429.

[14] Mudiyala et al. "A novel automated deep learning approach for Alzheimer's disease classification". In: *IAES International Journal of Artificial Intelligence* 12.1 (2023), p. 451.

[15] Heta Acharya, Rutvik Mehta, and Dheeraj Kumar Singh. "Alzheimer disease classification using transfer learning". In: (2021), pp. 1503–1508.

[16] Marwa EL-Geneedy et al. "An MRI-based deep learning approach for accurate detection of Alzheimer's disease". In: *Alexandria Engineering Journal* 63 (2023), pp. 211–221.

[17] Waleed Al Shehri. "Alzheimer's disease diagnosis and classification using deep learning techniques". In: *PeerJ Computer Science* 8 (2022), e1177.

[18] Petru Potrimba. *What is Semi-Supervised Learning? A Guide for Beginners.* Assessed on April 2023. URL: `https://blog.roboflow.com/what-is-semi-supervised-learning/`.

[19] Simon Tavasoli. *Simplilearn.* assessed on April 2023. URL: `https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article`.

[20] *How Neural Networks Work.* Assessed on April 2023. URL: `https://chatbotslife.com/how-neural-networks-work-ff4c7ad371f7`.

[21] Shivam Bhardwaj. *HISTORY & Background of Convolutional Neural Networks.* Assessed on April 2023. URL: `https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-understand-the-basics/`.

[22] Krut Patel. *Convolutional Neural Networks — A Beginner's Guide.* Assessed on April 2023. URL: `https://towardsdatascience.com/convolution-neural-networks-a-beginners-guide-implementing-a-mnist-hand-written-digit-8aa60330d022`.

[23] Marwan Adnan Jasim and Jamal Mustafa Al-Tuwaijari. "Plant leaf diseases detection and classification using image processing and deep learning techniques". In: *2020 International Conference on Computer Science and Software Engineering (CSASE).* IEEE. 2020, pp. 259–265.

[24] Yingying Wang et al. "The influence of the activation function in a convolution neural network model of facial expression recognition". In: *Applied Sciences* 10.5 (2020), p. 1897.

[25] *What is ReLU and Sigmoid activation function?* Assessed on April 2023. URL: `https://www.nomidl.com/deep-learning/what-is-relu-and-sigmoid-activation-function/`.

[26] Thomas Wood. *What is the Softmax Function?* assessed on April 2023. URL: `https://deepai.org/machine-learning-glossary-and-terms/softmax-layer`.

[27] Jason Brownlee. *How to Choose an Activation Function for Deep Learning.* Assessed on April 2023. URL: `https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/`.

[28] Sabina Pokhrel. *Beginners Guide to Convolutional Neural Networks*. Assesed on April 2023. URL: https://towardsdatascience.com/beginners-guide-to-understanding-convolutional-neural-networks-ae9ed58bb17d.

[29] SuperDataScience Team. *Convolutional Neural Networks (CNN): Step 4 - Full Connection*. assessed on April 2023. URL: https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-4-full-connection.

[30] baeldung. *How ReLU and Dropout Layers Work in CNNs*. Assessed on April 2023. URL: https://www.baeldung.com/cs/ml-relu-dropout-layers.

[31] SuperDataScience Team. *Convolutional Neural Networks (CNN): Step 3 - Flattening*. Assessed on April 2023. URL: https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-3-flattening.

[32] *ImageNet*. Assessed on Juin 2023. URL: https://devopedia.org/imagenet.

[33] M. Tim Jones. *Transfer learning for deep learning*. Assessed on April 2023. URL: https://developer.ibm.com/articles/transfer-learning-for-deep-learning/.

[34] *VGG-19 Convolutional Neural Network*. Assessed on Juin 2023. URL: https://blog.techcraft.org/vgg-19-convolutional-neural-network/.

[35] Konovalenko Ihor et al. "Steel surface defect classification using deep residual neural network". In: *Metals* 10.6 (2020), p. 846.

[36] Matsuyama Eri et al. "A deep learning interpretable model for novel coronavirus disease (COVID-19) screening with chest CT images". In: *Journal of Biomedical Science and Engineering* 13.07 (2020), p. 140.

[37] Sik-Ho Tsang. *Review: MobileNetV2 — Light Weight Model (Image Classification)*. Assessed on Juin 2023. URL: https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c.

[38] Erik Westphal and Hermann Seitz. "A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks". In: *Additive Manufacturing* 41 (2021), p. 101965.

[39] *Overfitting and Underfitting*. Assessed on April 2023. URL: https://www.fastaireference.com/overfitting.

[40] Dabic Ozren et al. "Sampling projects in github for MSR studies". In: *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. IEEE. 2021, pp. 560–564.

[41] *What is Python? Executive Summaryx*. Assessed on may 2023. URL: https://www.python.org/doc/essays/blurb/.

[42] Wikipedia contributors. *TensorFlow*. Assessed on May 2023. URL: https://en.wikipedia.org/wiki/TensorFlow.

[43] Sarang Narkhede. *Understanding Confusion Matrix*. Assessed on Juin 2023. URL: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.

[44] Sahoo Karam Kumarel al. "Wrapper-based deep feature optimization for activity recognition in the wearable sensor networks of healthcare systems". In: *Scientific Reports* 13.1 (2023), p. 965.

[45] *Epoch in Machine Learning*. Assessed on Juin 2023. URL: `https://www.javatpoint.com/epoch-in-machine-learning`.

[46] Hafidz Zulkifli. *Understanding Learning Rates and How It Improves Performance in Deep Learning*. Assessed on Juin 2023. URL: `https://towardsdatascience.com/understanding-learning-rates-and-how-it-improves-performance-in-deep-learning-d0d4059c1c10`.