An Efficient Deep Learning Approach to Detect Neurodegenerative Diseases Using Retinal Images

by

Chowdhury Mohammad Irfanuddin
19101123
Wasique Islam Shafin
19101122
Koushik Ahmed
22241154
Md. Hasib Khan
19101127

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
January 2023

Declaration

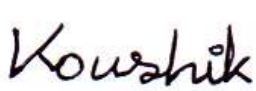
It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Chowdhury Mohammad Irfanuddin
19101123

Wasique Islam Shafin
19101122



Koushik Ahmed
22241154

Md. Hasib Khan
19101127

Approval

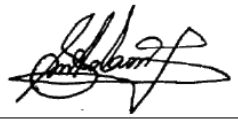
The thesis titled "An Efficient Deep Learning Approach to Detect Neurodegenerative Diseases Using Retinal Images" submitted by

1. Chowdhury Mohammad Irfanuddin (19101123)
2. Wasique Islam Shafin (19101122)
3. Koushik Ahmed (22241154)
4. Md. Hasib Khan (19101127)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 19, 2023.

Examining Committee:

Supervisor:
(Member)



Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Rafeed Rahman
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Neurodegenerative disorders are diagnosed through undergoing brain MRI, CT scans, genetic testing, and various laboratory screening tests which are often tedious, time-consuming and beyond the means of most people's financial capabilities and sometimes health unconducive. To remedy this, we proposed an efficient deep learning approach to detect neurodegenerative diseases, for instance, Multiple Sclerosis, Parkinson's disease, Amyotrophic Lateral Sclerosis, and Alzheimer's disease using retinal images. Efficient convolutional neural network-based architectures are used to classify brain diseases. The system enables the detection of brain diseases from retinal images rather than brain images effectively. Through the proposed system, we are able to proactively detect such disorders simply through retinal scans which are faster and simpler compared to the scanning of the brain itself which requires expensive and sophisticated equipment. We conducted our research on a dataset containing retinal cross-sectional images of 21 Multiple Sclerosis patients and 14 healthy individuals. Our model achieved 100% accuracy in classifying all healthy and diseased individuals from retinal scans.

Keywords: Neurodegenerative; Multiple Sclerosis; Retinal Images; Deep Learning; Convolutional Neural Network; Optical Coherence Tomography;

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AD Alzheimer's Disease

CNN Convolutional Neural Network

CT Computed Tomography

FNN Feed-Forward Neural Network

$GCIPL$ Ganglion Cell-Inner Plexiform Layer

HRT Heidelberg Retina Tomography

IRL Inner Retinal Layer

$mGCIPL$ Mean Macular Ganglion Cell-Inner Plexiform Layer

MRI Magnetic Resonance Imaging

MS Multiple Sclerosis

$MSON$ Multiple-Sclerosis-associated Optic Neuritis

OCT Optical Coherence Tomography

ON Optic Neuritis

PD Parkinson's Disease

PET Positron Emission Tomography

$pRNFL$ Peripapillary Retinal Nerve Fiber Layer

$RNFL$ Retinal Nerve Fiber Layer

Chapter 1

Introduction

1.1 Neurodegenerative Disease

The term "Neurodegenerative Disease" is used to collectively refer to the set of diseases that cause functional or structural damage to neurons located in the central or peripheral nervous system. Such degeneration of neurons eventually results in physical or cognitive impairment. Alzheimer's disease, Parkinson's disease, and Multiple Sclerosis are common types of neurodegenerative diseases that affect millions of people worldwide. Most of these diseases don't have any known cure, and so slowing down their progress is the only treatment available today. However, symptoms of these diseases don't become apparent until a considerable amount of neuron cells have deteriorated. Furthermore, the diagnosis of these diseases is typically done by Magnetic Resonance Imaging, Positron Emission Tomography, or Computed Tomography scan, all of which are time-consuming and expensive. As a result, researchers have been looking for alternatives to detect the presence of neurodegenerative diseases in a more accessible and cheaper way. One of them is to look at the brain through the eye via a technology known as Optical Coherence Tomography (OCT) which is used for cross-sectional ophthalmic imaging. It has revolutionized the field of eye evaluation, particularly retinal evaluation. It is a non-contact imaging technology that produces cross-sectional pictures of tissue at high-resolution [10]. As a result, it is particularly useful in organs where standard microscopic tissue identification by biopsy isn't possible, such as the human eye. OCT gives in vivo pictures without causing any damage to the tissue being examined, since it is fully noninvasive. Image visualization in real-time and at video rate is possible because of fast scanning rates and signal processing through OCT as well. Because of the noninvasive and more accessible nature of OCT, it is a perfect candidate to be used in the diagnosis of neurodegenerative diseases.

1.2 Research Problem

The traditional methods of diagnosing most neurodegenerative diseases like Multiple Sclerosis and Alzheimer's disease require scanning of the central nervous system which consists of the brain and the spinal cord. These scans are typically done using MRI, CT scans, or PET which are complex and time-consuming. Moreover, these scans are quite expensive and their cost is beyond the reach of many patients. Since most of these diseases have no cure, getting them diagnosed as early as possible is

a top priority in order to slow down their further progression. So, the development of a cheaper and faster technique is necessary in order to provide the facility of neurodegenerative disease diagnosis to the masses. We propose Optical Coherence Tomography (OCT) as a potential diagnostic tool for the rapid diagnosis of neurodegenerative diseases as it is fast, cheap, and more accessible than the aforementioned methods. Here, the cross-sectional view of retinal layers is used as a biomarker for detecting these diseases. Several previous studies have also shown how people suffering from different neurodegenerative conditions have a reduction in the thickness of certain layers of the retina. As the retina is considered to be an extension of the central nervous system, any neurological change have also been reported to manifest in the retina as well [17]. So scanning the retina can be an easy and non-invasive way of detecting the presence of neurodegenerative diseases and this is the reason why we propose the method of using OCT to diagnose them.

1.3 Research Objectives

The main goal of our research is to develop a method of detecting neurodegenerative diseases using OCT scans of the patients, where we will use the thickness of retinal layers as a biomarker as suggested by various previous studies. This research also has the following objectives:

1. Put the use of different deep learning algorithms to predict the presence of a neurodegenerative condition.

2. Establish a comparison between the performance of various deep learning models in accomplishing the task.

3. Provide an appropriate explanation of the model's prediction by the means of explainable AI that will help in the process of verification of the model's accuracy.

Chapter 2

Related Work

Ever since the arrival of OCT, it has left many researchers fascinated with its ability to show the retinal tissue layer by layer accurately and wondered how it may contribute to the further development of neurological studies. Several past studies have indicated the link between the atrophy of retinal layers and neurodegenerative diseases. These researchers tested different methods under different scenarios to demonstrate how neurological functions in the brain also manifest in the retina. The following sections give an overview of a couple of such studies.

2.1 Cross-sectional / Thickness Observations

Most studies performed cross-sectional thickness observations. Through a cross-sectional approach, it is easier to differentiate various retinal layers compared to horizontal and longitudinal ones.

First of all, the research work [10] proposed that the inner plexiform layer and the ganglion cell layer thickness of the retina can be used to predict the nerves of patients suffering from MS. They used Spectralis OCT scan data from 204 MS patients and 138 control to perform logistic regression in order to correlate MS progression with retinal layer thickness. They found a noticeable amount of atrophy in the RNFL, ganglion cell layer, inner nuclear layer, and inner plexiform layer in MS patients.

Secondly, a study conducted in [11] assessed the thickness of the RNFL in individuals suffering from MS by using OCT scans. The study included 31 individuals with confirmed MS and 31 disease-free subjects. OCT parameters showed significant variations between the two groups, with the MS group having a reduced thickness of certain retinal layers. The researchers used Pearson's correlation coefficient to find out the relation between the thicknesses of RNFL, EDSS score, disease duration, and several other variables. The research team found out that MS had lower RNFL thickness compared to the control subjects.

Third, the purpose of the research done on paper [1] was to examine the relationship between the RNFL thickness and visual function of the eyes with the help of a functional biomarker for nerve cell loss in MS. The investigation was designed in a cross-sectional manner. To conduct this research, they took data from 90 patients

suffering from MS and 36 healthy individuals. The authors used OCT to determine the density of RNFL. They found that MS patients had a considerable amount of reduction in RNFL. Also, MS with ON patients showed more reduction of RNFL compared to non-ON MS patients.

Fourth, regarding the [15] paper study, the objective of the authors was to determine whether OCT-measured INL density changes with age and disease symptoms in individuals with progressive MS. The authors tested lots of variations in the pRNFL, GCIPL, and INL retinal layers among 84 individuals suffering from P-MS and 36 healthy individuals. All patients were segmented as per age. The researchers found that P-MS patients had lower pRNFL as well as GCIPL thickness. Moreover, older P-MS patients aging more than 51 had shown reduced thickness of the INL compared to younger patients.

Fifth, in [17], the researchers implemented a system to diagnose MS in the early stages using a CNN to classify swept-source OCT imaging data from 48 control and MS patients each. These images contained complete retina, retinal nerve fiber layer, two ganglion cell layers and choroid. They made use of a deep convolutional generative adversarial network to grow their dataset and selected features before training the CNN model. Their method represents an improvement in the detection of early state MS by correlating the retinal thickness through the use of CNNs.

Finally, again, the research team from paper [4] used OCT to assess axonal loss in the RNFL in individuals with MS, both with and without a history of ON. This study included 50 MS patients and 25 healthy individuals who were age and sex-matched. All RNFL parameters assessed by OCT showed substantial differences between MS and healthy eyes. According to their research, the metric with the biggest variations across different groups was the thickness of RNFL in the temporal quadrant. They found that people suffering from MS had an overall lesser thickness of RNFL than healthy individuals and this trend was prevalent in both MS with ON and MS without ON patients.

2.2 Horizontal / Longitudinal Observations

Although not as common as the cross-sectional thickness approach, it provides a fresh perspective.

Firstly, the objective of the study [7] was to demonstrate the relationship between RNFL thinning and visual loss by longitudinal examination of MS patients. Using OCT, approximately 1005 patients were measured as baseline and 299 patients were followed in 6-month increments. The longitudinal study that was performed showed not only a steady decrease in RNFL thickness, but also an increase in visual acuity over time. The study also showed a significant correlation between RNFL thinning and decreased visual acuity.

Secondly, the research paper [9] worked on evaluating relationships of MS with all gathered retinal layers from OCT scans. They extracted usable horizontal scan data by performing surface & deeper paramacular retinal and peripapillary RNFL

scans, and then classifying individual layers. This paper analyzed data gathered from 95 MS patients and 91 healthy individuals of similar age and sex. They found notable atrophy in the INL and similar thinning in the pRNFL, retinal ganglion cell layer, and inner plexiform layer. Their results affirmed the existence of correlation between the disease severity and the thickness of retinal layers.

Lastly, the objective of the study conducted on [8] was to examine various retinal layer thicknesses of MS patients having no history of optical neuritis using Cirrus HD-OCT scans. The study gathered all pRNFL, central and average macular thickness data from 60 eyes of affected MS patients and 32 eyes of healthy individuals. They found a major reduction in average macular thickness and a meaningful correlation between average macular and RNFL thickness of MS patients. The findings are consistent with already existing studies, but also dictate the reduction of thickness in the RNFL in patients with multiple sclerosis even without optic neuritis presence. This study experimented only on horizontal scan data excluding longitudinal features.

2.3 OCT compared with Traditional Apparatus

Several studies have compared OCT with MRI, HRT, and other conventional devices and methods.

Firstly, the paper [3] used not only retinal OCT images but also MRI brain imaging to associate MS and RNFL thickness. MRI was used to detect the inflammation part and OCT was used to detect the neurodegeneration part of MS. 30 MS patients were tested using OCT and of that, 18 patients again were tested using MRI. The paper found a strong relation between RNFL thickness measured by OCT and that of brain imaging characteristics found by MRI scans in MS patients. Moreover, comparing the results, they pointed out the significant potential of OCT in terms of detecting MS.

Second, the research group from [5] compares the results obtained from both OCT and HRT scanning technologies in order to measure the RNFL thickness in patients suffering from MS as well as to figure out the relationship between thickness change in the RNFL with cognitive alongside physical disability. The BRB-N test was used for evaluating the cognitive state, and OCT alongside HRT was employed for measuring the thickness of RNFL on 52 MS patients and 18 healthy individuals. They found that the atrophy of RNFL was associated with physical, as well as cognitive disability in patients suffering from MS. The researchers also compared OCT and HRT results of the retinal nerve fiber layer in order to see if these approaches are equal for monitoring MS patients or if they provide supplementary information regarding disease activity. It was found that only OCT image data had a decent correlation between RNFL thickness and physical disability. They also observed that patients suffering from cognitive impairment had lower RNFL thickness than non-cognitively impaired MS patients.

Third, the aim of the study conducted in [2] was to use OCT to test whether RNFL thickness has any relationship with MS. The researchers followed 61 MS pa-

tients and 29 matched controls with neurological assessments such as clockwork and ophthalmologic assessments including OCT controls on a regular basis over time. The research team performed multiple linear and logistic regression using SPSS 13 software. A reduction in the thickness of RNFL thickness was observed in patients MS patients. In addition, they also found that gray and white matter volumes determined by MRI correlated with the baseline thickness of RNFL.

Fourth, in [6], the researchers examined how disease duration effects RNFL in people not receiving disease-modifying treatment. The authors examined the link between RNFL loss and illness time in unmanaged MS patients to see if it was related to corticospinal pathway malfunction. Vision tests were performed on 52 participants using OCT and 60 patients either went through vision testing or took part in a verified telephone interview so that their EDSS scores could be determined. The researchers used Spearman's correlation coefficient in order to determine the correlations between OCT parameters, EDSS scores, and several other variables. It has been found that in the patients with symptomatic MS, RNFL loss is linked to illness time and EDSS scores and authors suggest that OCT might be a valuable method for obtaining the impact of disease and neurodegeneration on the CNS in clinical trials and application.

Finally, the purpose of the study conducted in [13] was to examine the link between IRL loss and cognitive disability in MS patients. This research was conducted on 217 sufferers and 59 healthy controls. Besides, medical examinations, retinal OCT, and memory tests were all performed on the participants. The authors claim that mentally curtailed individuals showed a substantially lower mean pRNFL and mGCIPL density than MSON. The significant correlation between atrophy of the retinal layers and the cognitive disability across multiple cognitive domains in MS-ON patients suggests the usefulness of OCT in diagnosing MS.

2.4 Other than MS

Some studies also found association of RNFL loss with Parkinson's disease & Alzheimer's disease along with MS.

For example, the research team working on [12] conducted an experiment on 150 individuals who were suffering from Alzheimer's disease with 75 control subjects and used Spectralis OCT in order to segment different retinal layers. In order to determine the correlation between Alzheimer's disease progression and retinal atrophy, they used a test called MMSE, which is an exam of 30 points that reflects the status of overall cognitive decline in AD suffering patients. In this research, they found noteworthy reductions in RNFL, GCL, outer nuclear layer, and inner plexiform layer thickness in AD patients.

Chapter 3

Research Methodology

3.1 Work Plan

In order to detect neurodegenerative diseases from OCT Scans, a decent dataset containing images of both healthy people and neurodegenerative disease-affected individuals is needed. The dataset has to be labeled. Since we are dealing with a binary classification problem of distinguishing between the OCT images of healthy and diseased patients, the images of the dataset can simply be organized in two folders representing these two categories. Then, the automatic labeling tools found in popular libraries like TensorFlow can be used to easily label our dataset. After the dataset acquisition, necessary pre-processing of the images must be done which may include resizing, segmentation, enhancement, etc. Then, the dataset will be split for training and testing purposes. In our case, we have used 8:1:1 split for dividing our dataset in training, testing and validation sets. Then, the training data was fed into different deep learning models.

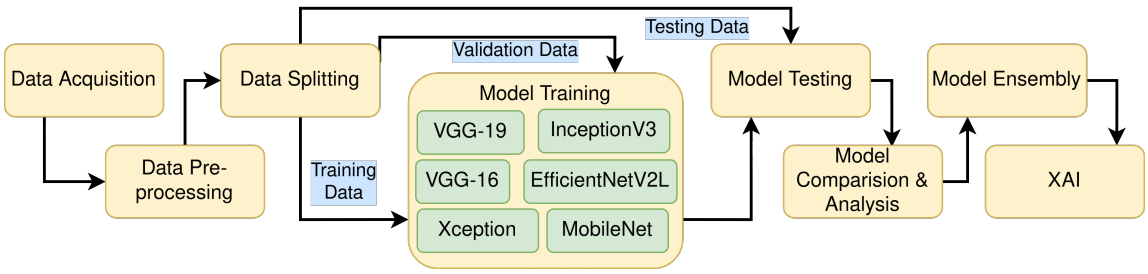


Figure 3.1: Work Plan

We used VGG-16, VGG-19, InceptionV3, Xception, EfficientNetV2L and MobileNet architecture to train 6 different models using our training dataset. Transfer learning was employed to optimize the training process, where we fetched pretrained models of aforementioned architectures that contained ImageNet weights and then further trained and fine-tuned these models using our own data. After training all the six models, we selected top 3 best performing models among them and ensembled their outputs to make sure that any type of bias from any particular model has been eradicated. After the ensembling was done, we also developed an explainable model that shows which part of the input images are drawing the neural network's attention most. As neural network-based deep learning models are quite complicated, the reasoning behind the output generated by them may not be apparent to the users

and so verification and trust in the model will be quite hard. We solve this problem by including explainable AI in our approach that will help the users to further analyze and verify the prediction generated by the other three models. The process can be represented by figure 3.1.

3.2 Used Architecture

We employed 6 architectures in our research, VGG-16, VGG-19, InceptionV3, Xception, EfficientNetV2L and MobileNet.

3.2.1 VGG-16

VGG-16 is an altered version of Visual Geometry Group (VGG) model having 16 layers. It consists of 13 convolution layers having 3x3 kernels and 3 fully connected layers. It makes use of 5 2x2 max-pooling and a SoftMax function. The hidden layers make use of ReLu as an activation function because of its efficiently in computationally demanding tasks and lower probability of vanishing gradient issues. Although VGG-16 is very performant, it is significantly slower to train and its larger parameters lead to problems of exploding gradient

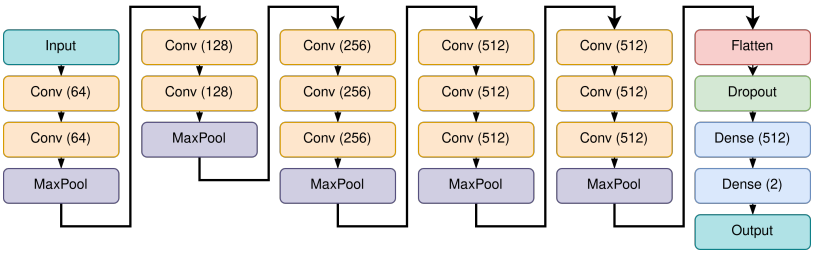


Figure 3.2: VGG-16 Internal Architecture and Layers

3.2.2 VGG-19

VGG-19 is yet another modified version of Visual Geometry Group (VGG) model having 19 layers. It consists of 16 convolution layers having 3x3 kernels and 3 fully connected layers. It makes use of 5 2x2 max-pooling and a SoftMax function. VGG achieves more in less time having less parameters by using 3x3 filters instead of conventional 7x7 convolutional filters.

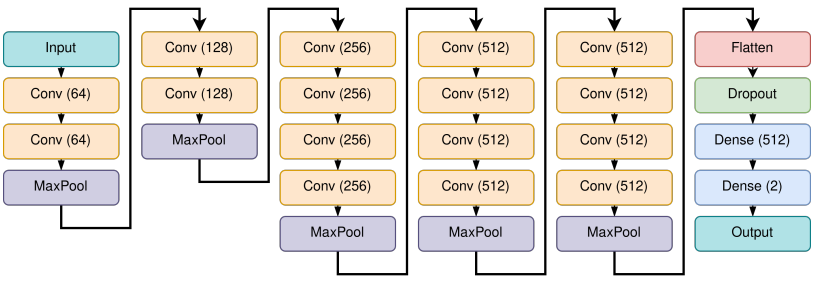


Figure 3.3: VGG-19 Internal Architecture and Layers

3.2.3 InceptionV3

InceptionV3 is the superior and optimal version of the inceptionV1 model developed by Google. Its architecture is resistant to overfitting of the data and achieves higher accuracy with better computational efficiency due to having fewer parameters. It has 42 layers in total. While having a deeper network and more layers compared to its previous versions, it doesn't lose its speed and it boasts a very low error rate.

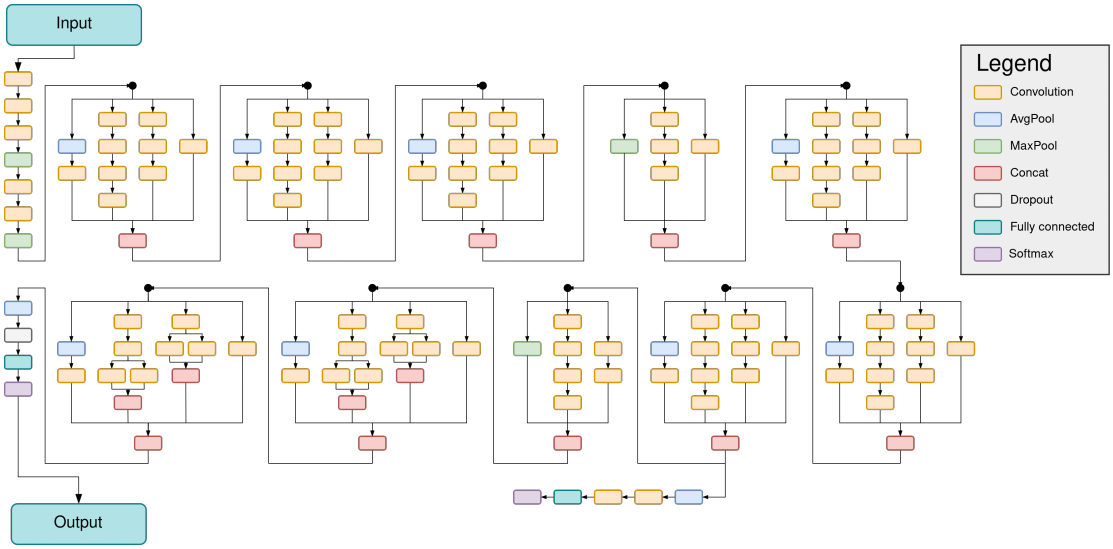


Figure 3.4: InceptionV3 Internal Architecture and Layers

3.2.4 Xception

Xception is heavily inspired from Inception model and replaces Inception modules with depthwise separable convolutions. This architecture includes the same parameter count and achieves more performance by properly utilizing them. While the Inception model makes use of ReLU non-linearity, Xception refrains from adding any non-linearity and outperforms every contemporary model.

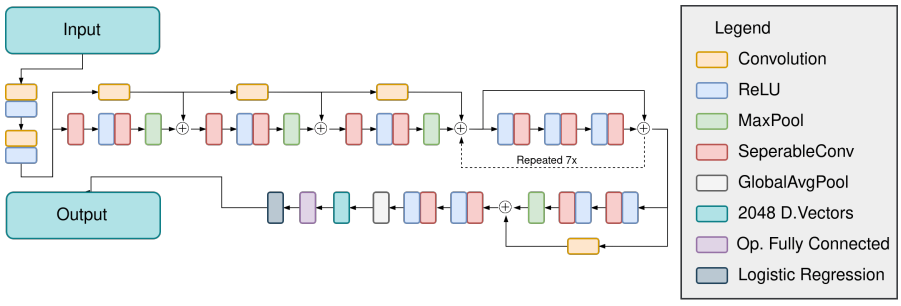


Figure 3.5: Xception Internal Architecture and Layers

3.2.5 EfficientNetV2L

The signifying characteristics of EfficientNetV2 is its fast training with smaller models due to having better parameter efficiency compared to older models. A mix of

training-aware neural network search and scale was used to achieve both its faster training speed and parameter efficiency. Thus, EfficientNetV2 models train much faster than contemporary models while being about 7x smaller, but suffers from drop in accuracy the faster it gets.

3.2.6 MobileNet

MobileNet is yet another CNN model by Google aimed at bringing efficient computer vision in devices focused on mobility. Although it consists of 4.2 million parameters, its size can be adjusted at the cost of accuracy. To achieve its characteristic portability, it made use of depthwise separable convolutions instead of standard ones. As its size are generally small, MobileNet models rarely overfit the data but its accuracy takes a dip.

3.3 Ensemble

Through ensemble models, we can reduce prediction error and bias towards a particular model by model averaging. Also, we don't have to select any particular model and therefore reduce risk of data volatility. We chose the 3 best performing models, namely VGG-19, Xception and EfficientNetV2L to be ensembled. Out of the multiple averaging methods, we employed weight based averaging, where weights were assigned to each model on their order of importance.

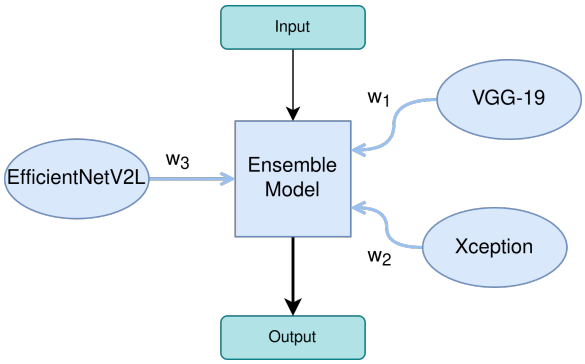


Figure 3.6: Ensemble: VGG-19, Xception and EfficientNetV2L

3.4 Explainable AI (XAI)

We employed Gradient-weighted Class Activation Mapping (Grad-CAM) for implementing our explainable model. Grad-Cam uses the weight changes in the last convolution layer of convolutional neural networks and is able to visualize which part of the image is drawing the model's attention and contributing to its prediction though a heatmap [14]. In our case, we applied the same weighted ensembling process for the explainable model to reflect what our final model actually sees while making the predictions.

3.5 Confusion Matrix

Our research employed Confusion matrix for measuring the performance of different models. This approach is appropriate for supervised machine learning algorithms and models, especially in case of uneven data-set. This set of matrices depicts a table listing correct and incorrect classifications. The following four values can appear in a confusion matrix

1. True Positive as TP
2. False Positive as FP
3. True Negative as TN
4. False Negative as FN

Few metrics based upon confusion matrix values to visualize performance of classifier are as follows:-

Accuracy is measured by the number of properly categorized sample data to the total amount of data. Equation 3.1 shows the formula for calculating accuracy.

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

Recall is the measure of true positive rate. It can be calculated by using equation 3.2.

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

The measure of positive predictive value is called Precision. Equation 3.3 shows the method of calculating precision.

$$Precision = \frac{TN}{TN + FN} \tag{3.3}$$

F1-score is the measure of harmonic average of recall and precision. It can be calculated by using equation 3.4.

$$F1 - score = \frac{2 * recall * precision}{recall + precision} \tag{3.4}$$

Chapter 4

Implementation

4.1 Data-set

Finding OCT Data for multiple sclerosis patients have been a challenging task since most of the previously done research in this topic used closed source datasets. Thankfully, Yufan et al. released a dataset in [16] which contains OCT images of 35 individuals in total, of whom 14 are healthy subjects and 21 are MS infected. The OCT images were captured using Spectralis OCT System from The Johns Hopkins Hospital, Baltimore. For each of the subjects, 49 B-Scans were captured. Hence, for each individual 49 images exist in the dataset which makes the number of total available images 1715.

4.2 Data Classification

Before building our models, we divided that dataset into 3 sets for the training, validation and testing. The training set was used to fit the dataset to the models. The validation test was for providing feedback to the models throughout the training process in order to tune the internal parameters. Finally, unbiased evaluation of the models was done using the testing set.

We used an 8:1:1 split for dividing our dataset into these three sets. As a result, approximately 80% of our dataset was used for training, 10% for validation and 10% was reserved for testing. In every set, the images were divided in two classes. The classes were 'normal' for healthy controls and 'ms' for the patients who are suffering from multiple sclerosis. The size of each set is presented in table 4.1

Table 4.1: Dataset Classification

Class	Training Set	Validation Set	Testing Set	Total
Normal	548	68	70	686
MS	823	102	104	1029
Total	1371	170	174	1715

4.3 Data Pre-processing

4.3.1 Resize

The original images in our dataset had a dimension of 1024x496 pixels. So, in order to optimize our dataset, we reduced their dimension to 224x244 pixels though resizing. We used the same dimension for all the models that we've used in our research. For the resizing purpose, the built-in functions of the TensorFlow library were used.

4.3.2 Normalization

The original images had pixel intensity values ranging from 0 to 255. In order to reduce computational complexity of the training and testing process, we have applied normalization of the images by dividing each pixel value by 255. Thus, the intensity value for all pixels fell within the range of 0 to 1, thus greatly reducing the computational complexity.

4.3.3 Augmentation

Deep Neural Networks are quite data hungry, and the size of the overall dataset plays a significant role in determining their performance. So, we used data augmentation to increase the amount of available data for our models. In our case, we used ± 5 degrees rotation and -20% to +10% of brightness variation factor for generating new data samples. All the augmentations were done using built-in functions provided in the TensorFlow library.

Chapter 5

Result and Analysis

5.1 Result

For our purpose of classifying the OCT images of healthy and MS patients, we used six models in our analysis. They were trained using VGG-16, VGG-19, InceptionV3, Xception, EfficientNetV2L and MobileNet architectures. After training through transfer learning and fine-tuning our models for 30 epochs, we evaluated them through our testing dataset. All the six different architectures that we trained yielded different results in our testing process when it comes to classifying OCT images of healthy and multiple sclerosis subjects. Although their performance varied from one architecture to another, all of them were able to gain accuracy above 90%.

5.1.1 Learning Performance Curve Analysis

VGG-16

When it comes to the training process of VGG-16, we observed a gradual increase in accuracy and disease in loss. As figure 5.2 indicates, after approximately 25 epochs, our mode's accuracy went above 95% and remained similar for the following epochs. The loss became nearly 0% at epoch 27 and remained such for the next epochs. At the end of 30 epochs, the model's training and validation accuracy and loss curves became consistent. It seems that the various convolutional layers of the model is indeed helping it to detect features quickly from the OCT images, and the small filter size of 3x3 is also helping for the overall optimization.

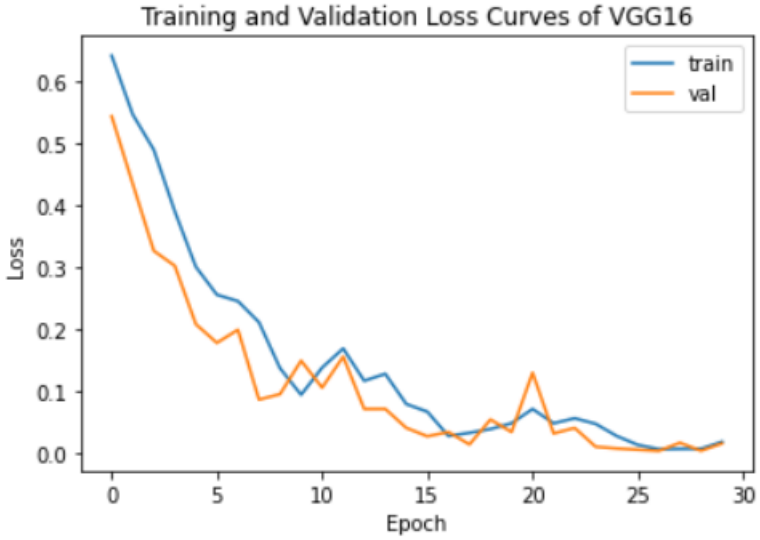


Figure 5.1: VGG-16 Loss Curve

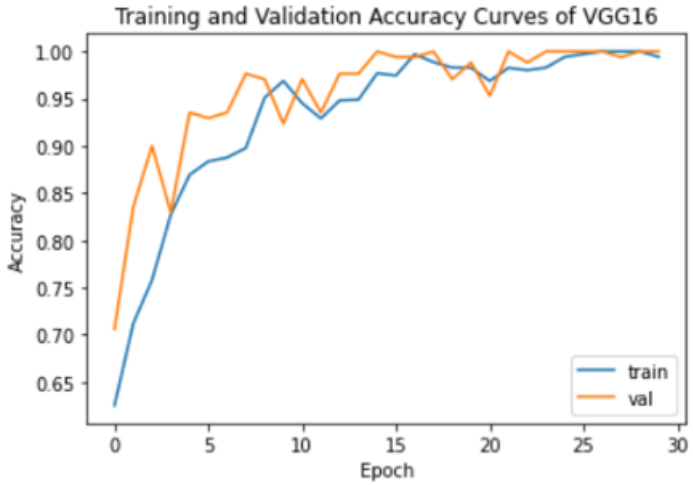


Figure 5.2: VGG-16 Accuracy Curve

VGG-19

As figure 5.4 shows, the accuracy rate of our model reached closed to 100% near epoch 25 and remained steady after that for both training and validation processes. The loss function became almost zero and after a slight peak near epoch 25 as well and stayed so with small fluctuations.

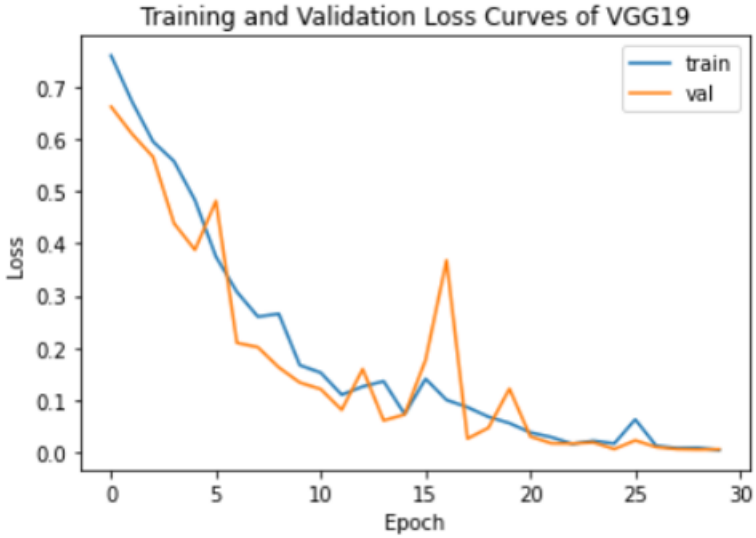


Figure 5.3: VGG-19 Loss Curve

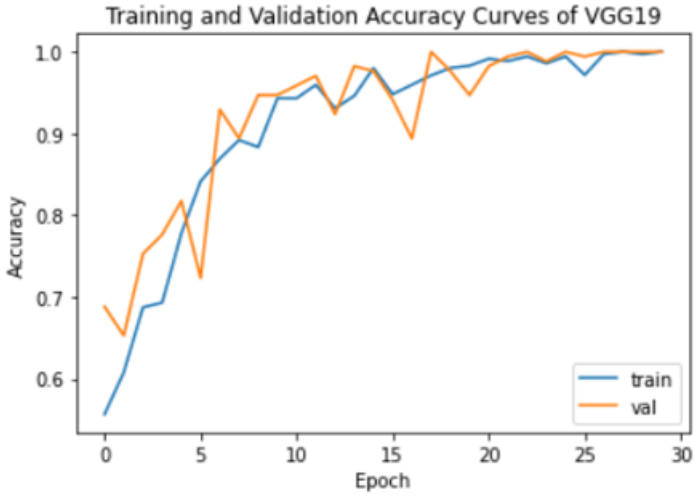


Figure 5.4: VGG-19 Accuracy Curve

MobileNet

When it comes to MobileNet, the training accuracy went above 95% right after 20 epochs as evident by figure 5.6. The validation accuracy however didn't catch up with training but still managed to exceed the 90% mark. The training loss curve of MobileNet reached closed to 0 at the end of the training process, whereas the validation loss went as down as 20%. It looks like MobileNet's focus on optimization and reduction in computational complexity is causing such loss of accuracy.

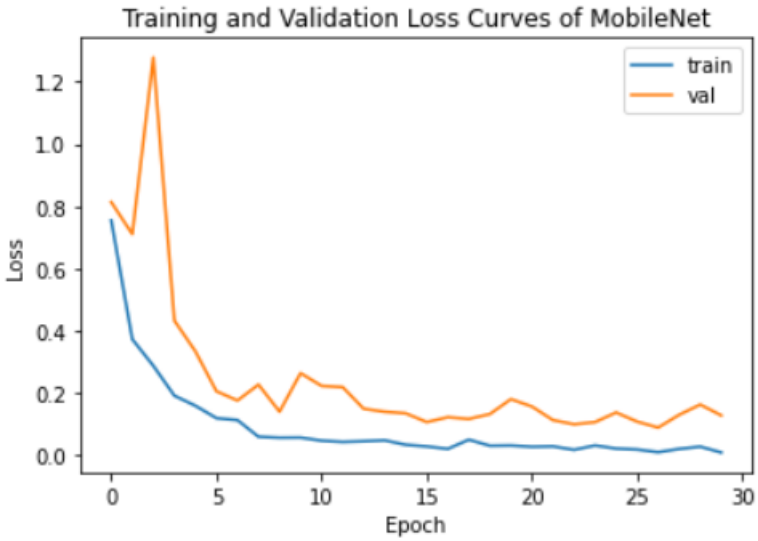


Figure 5.5: MobileNet Loss Curve

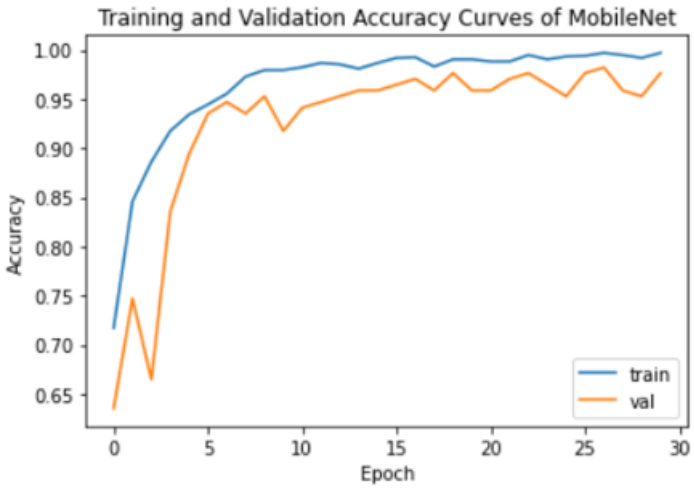


Figure 5.6: MobileNet Accuracy Curve

InceptionV3

Our InceptionV3 based model also showed decent performance in both training and validation process. The model gained above 90% accuracy for both training and validation. As shown in figure 5.7, the model's loss for both training and validation went as low as 10%.

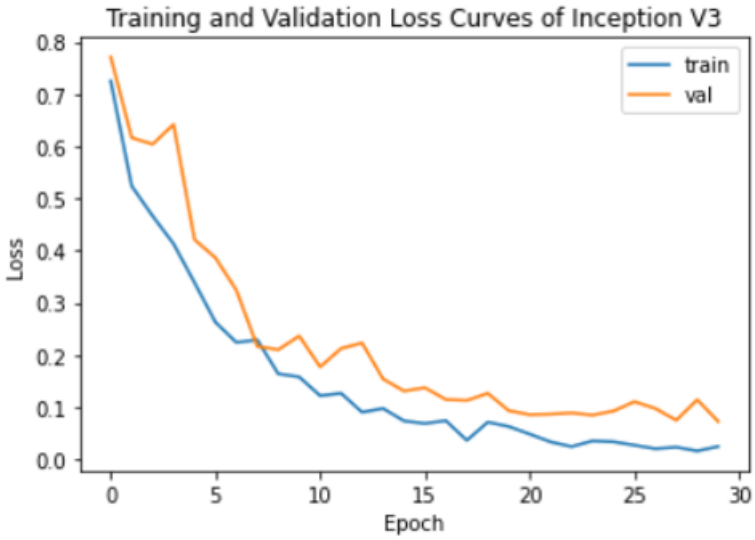


Figure 5.7: InceptionV3 Loss Curve

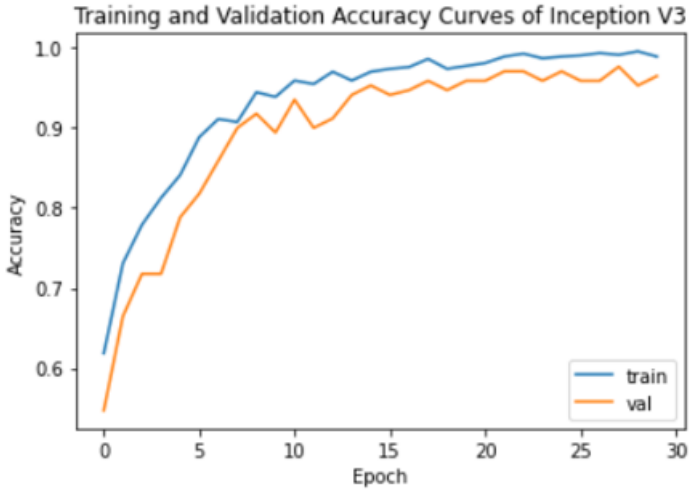


Figure 5.8: InceptionV3 Accuracy Curve

Xception

As figure 5.10 demonstrates, our Xception based model reached closed to 100% accuracy just after 10 epochs for both training and validation. It's losses also went as low as 10% around the same time.

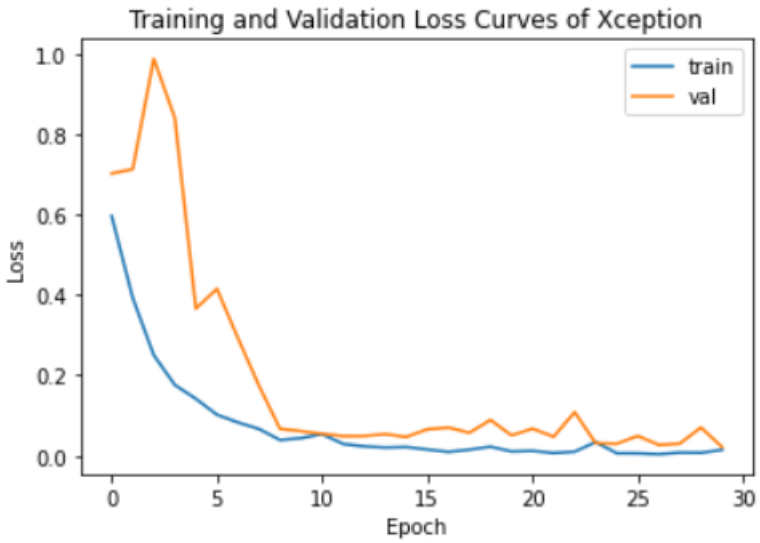


Figure 5.9: Xception Loss Curve

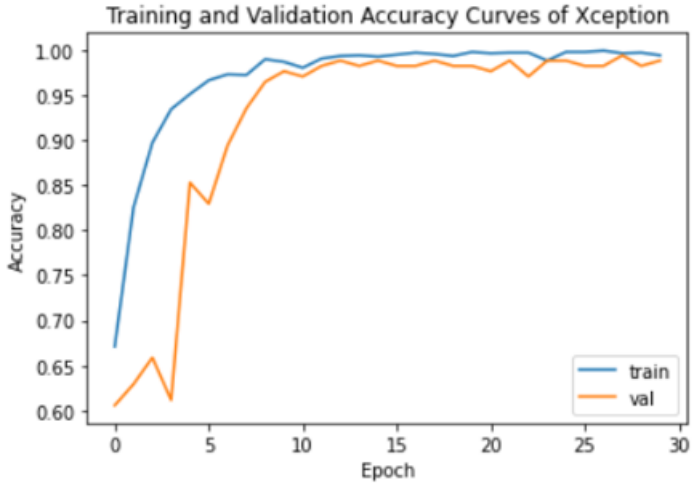


Figure 5.10: Xception Accuracy Curve

EfficientNetV2-L

The EfficientNetV2L based model showed decent performance and was outperformed only by VGG-16 and VGG-19. As shown in figure 5.12, the model reached near 100% accuracy just after 25 epochs for both training and validation. The losses also went down close to 0% by that time.

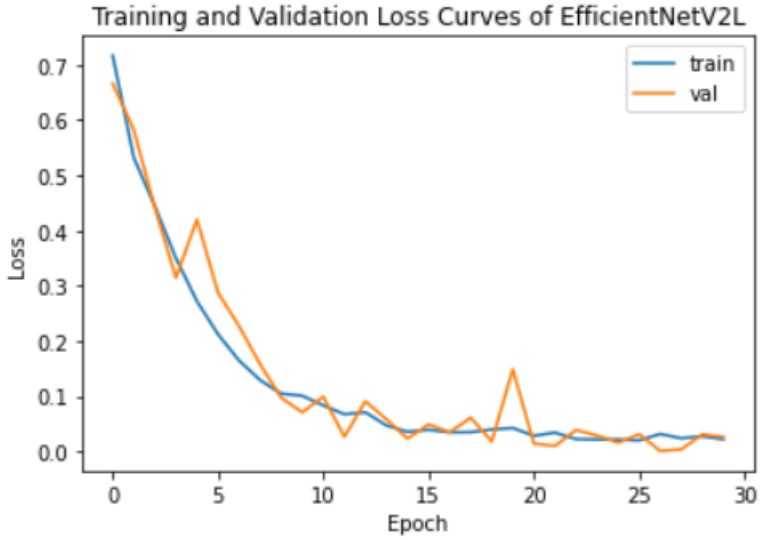


Figure 5.11: EfficientNetV2-L Loss Curve

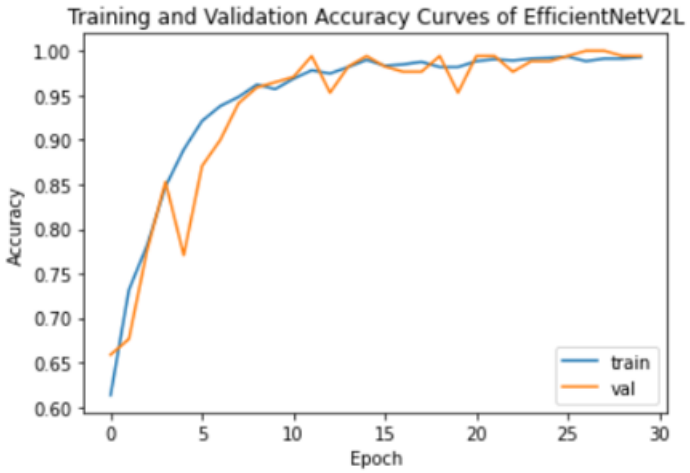


Figure 5.12: EfficientNetV2-L Accuracy Curve

5.1.2 Analysis of Confusion Matrix

Although all the six architectures that we applied were able to gain above 90% accuracy for our training dataset, we needed to evaluate them using our testing dataset that none of the models had seen before. In order to evaluate them, we have used a confusion matrix to plot both their correct and incorrect predictions. We judged the models though this evaluation process in order to select the top three best performing ones that we finally used for the ensembling process.

VGG-16

The confusion matrix of the model trained with VGG-16 is shown in figure 5.13. Here, we can see that the among 104 OCT images of multiple sclerosis patients, the model was able to correctly classify all of them. When it comes to images of healthy subjects, the model incorrectly classified one of them as an MS patient while correctly classifying the rest of the 69 images.

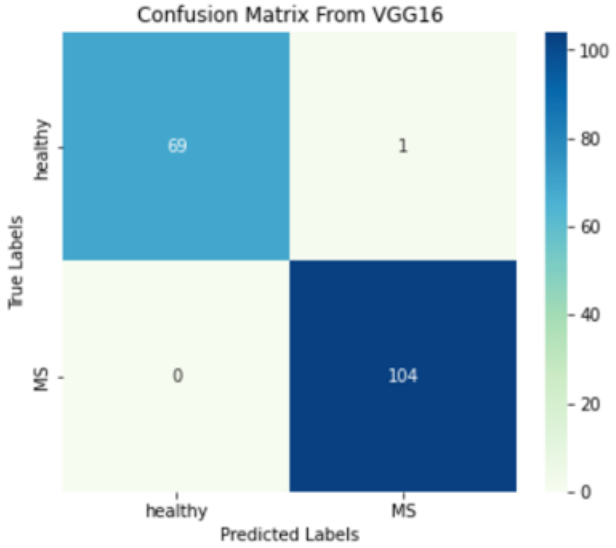


Figure 5.13: VGG-16 Confusion Matrix

VGG-19

Figure 5.14 shows the confusion matrix of the model trained with VGG-19. We can see that the model was able to correctly classify all the images and made no errors.

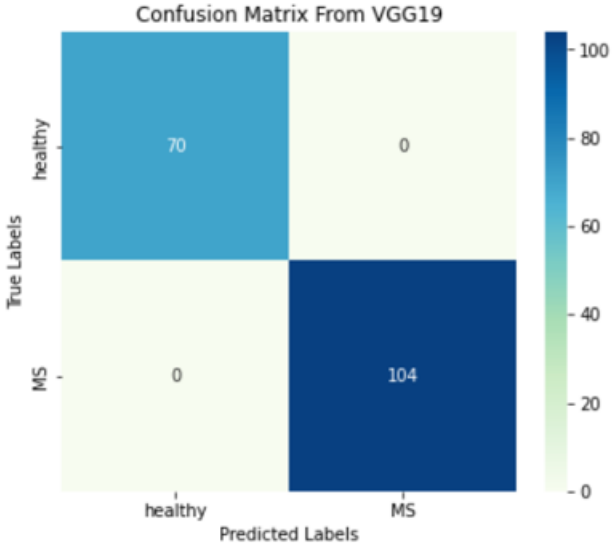


Figure 5.14: VGG-19 Confusion Matrix

MobileNet

The MobileNet based model was able to show decent performance on our testing dataset, as evident by figure 5.15. It was able to correctly classify all 104 MS patients. Out of 70 OCT images of healthy individuals, it misclassified 4 of them as MS patient.

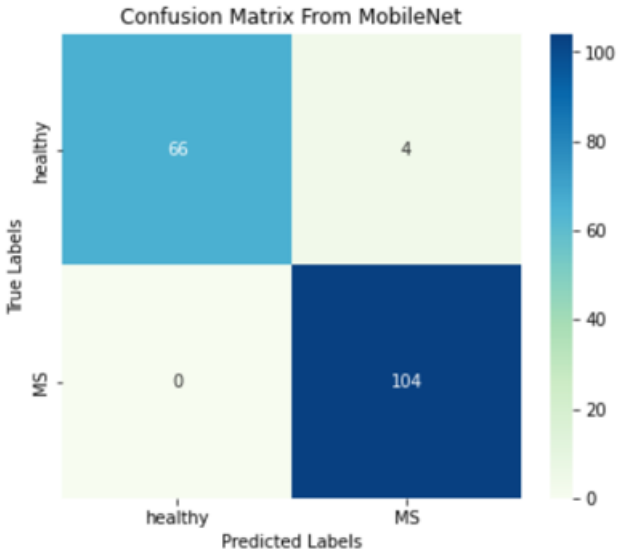


Figure 5.15: MobileNet Confusion Matrix

InceptionV3

The InceptionV3 based model struggled a bit compared to other models, as it wrongly classified 8 healthy subjects as MS patients and 1 MS patient as healthy. Thus, out of total 174 testing samples, it had 9 wrong classifications in total as shown in figure 5.16.

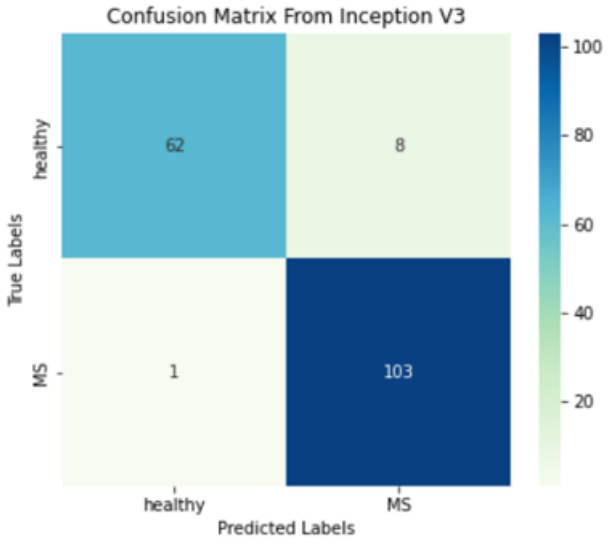


Figure 5.16: InceptionV3 Confusion Matrix

Xception

The testing performance of Xception reflect that of its training and validation performance, as it was outperformed only by the VGG based models. It correctly classified all the healthy subjects as shown in figure 5.17. Among 104 images of MS images, however, it misclassified 2 of them as healthy.

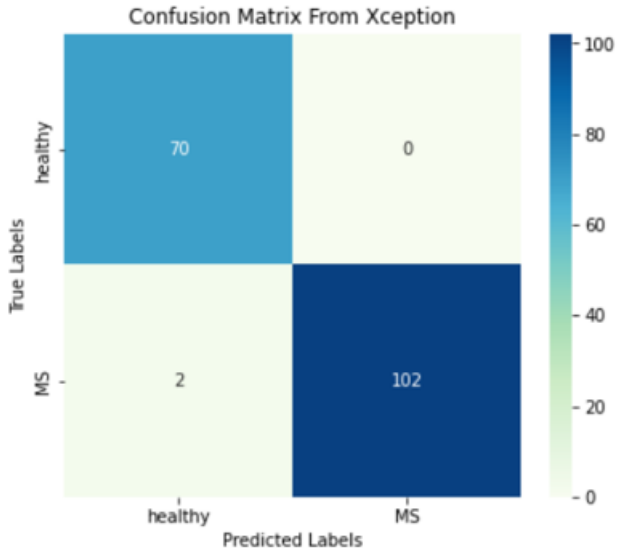


Figure 5.17: Xception Confusion Matrix

EfficientNetV2-L

The EfficientNetV2-L model's performance was also decent as evident by figure 5.18, the model only had 4 misclassifications in total among the 174 test samples.

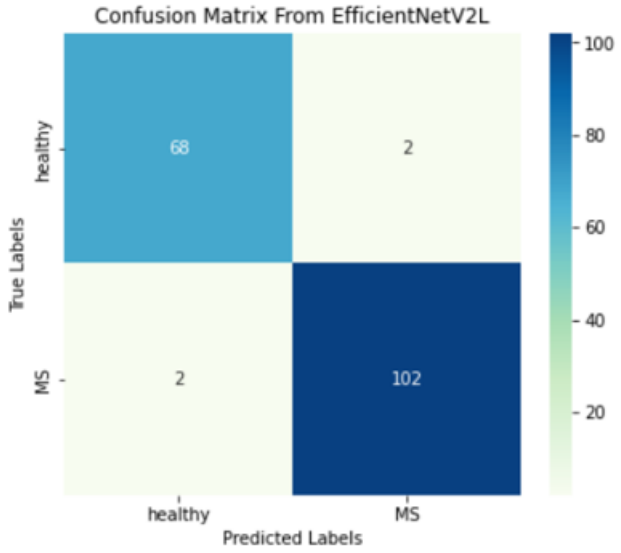


Figure 5.18: EfficientNetV2-L Confusion Matrix

5.2 Result Comparison

All the six different architectures that we trained yielded different results in our testing process when it comes to classifying OCT images of healthy and multiple sclerosis subjects. We considered categorical accuracy (3.1), recall (3.2), precision (3.3) and f1-score (3.4) as our evaluation parameters. The results of our evaluation are provided in table 5.1

Table 5.1: Comparison Between The Performance of Different Models

Arch.	Precision	Recall	F1-Score	Accuracy
VGG-19	100.00%	100.00%	100.00%	100.00%
VGG-16	99.52%	99.29%	99.40%	99.43%
Xception	98.61%	99.04%	98.81%	98.85%
Efficient-NetV2-L	97.61%	97.61%	97.61%	97.70%
MobileNet	98.15%	97.14%	97.59%	97.70%
InceptionV3	95.60%	93.80%	94.52%	94.83%

5.3 Ensembling

It is evident by table 5.1 that VGG-19 is the best performing model for the dataset, followed by VGG-16. Xception and EfficientNetV2-L also showed impressive results and hence they occupied the third and fourth place respectively. Though they individually yield impressive results, different architectures have different biases. Moreover, just because one model is performing well on our dataset doesn't mean it will attain that performance on OCT images gathered from different machines and of different qualities. So in order to create a generalized and bias free model, we ensembled the top 3 best performing architectures. Since VGG-19 and VGG-16 are structurally very similar and VGG-19 outperformed VGG-16, we decided to skip VGG-16. So, we ensembled the outputs produced by VGG-19, Xception and EfficientNetV2-L based models. After considering different ensembling mechanisms, we decided to go with weighted ensembling that allowed us to adjust the preference of each of the base models by assigning weights to their outputs. In our case, we assigned 40% preference to VGG-19's output. The rest 60% of preference was divided among Xception and EfficientNetV2-L equally. The ensembled model performed well as shown in figure 5.19. The model correctly classified all the images of our testing dataset.

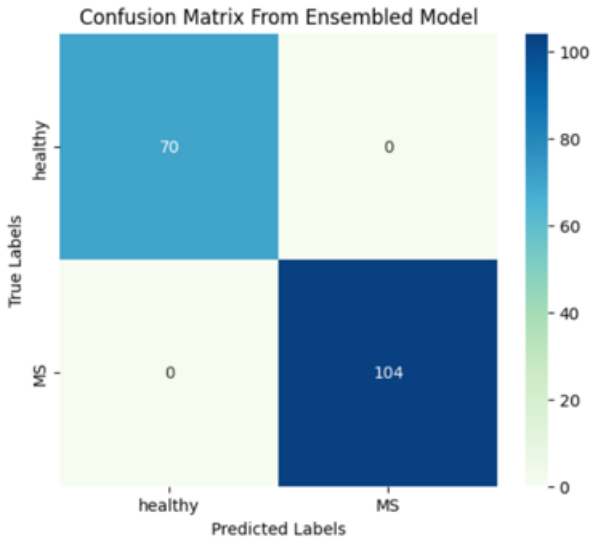


Figure 5.19: Ensembled Model Confusion Matrix

5.4 Explainable AI (XAI)

Although our ensembled model gained 100% accuracy in terms of classifying all the OCT images, the reasoning behind its predictions are not apparent to us. As mentioned before, we want the models to look anomalies in the retinal layer portion of the OCT images for making their prediction. In order to verify that the model is paying attention to where we want it to, we have implemented an explainable model using the Grad-CAM approach.

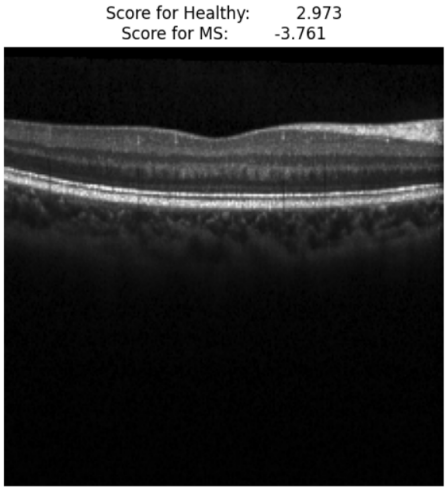


Figure 5.20: Input (Healthy).

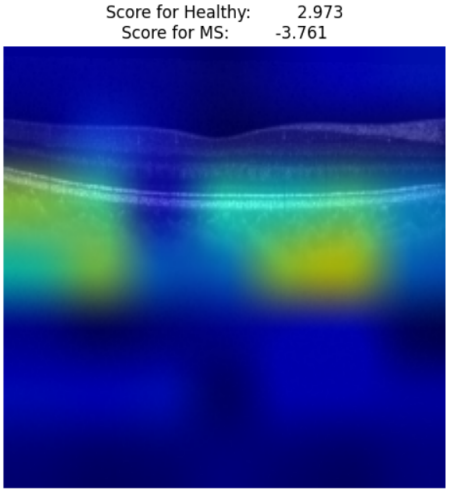


Figure 5.21: Grad-CAM Output.

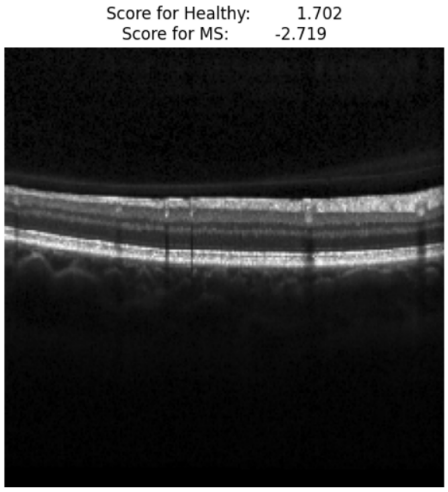


Figure 5.22: Input (Healthy).

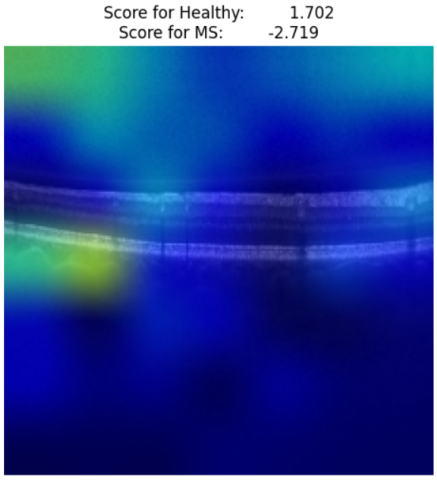


Figure 5.23: Grad-CAM Output.

Fig. 5.20 and 5.22 depicts OCT images of healthy individuals that were fed into our final ensembled model. Fig. 5.21 and 5.23 shows the Grad-CAM output produced by those inputs. From the above images, can see that the model is indeed paying considerable amount of attention near the retinal layers, but no significant activity is happening overall. However, the situation changes when it encounters the OCT images of MS patients, as evident by the figures shown below. Here, fig. 5.24 and 5.26 shows OCT images of MS patients and fig. 5.25 and 5.27 shows their Grad-CAM outputs. We can see that, a lot of activities are happening in the retinal layer

region, which indicates that some anomalies exist in that part of the OCT image. As a result, the model is able to distinguish OCT image of healthy individuals and multiple sclerosis patients.

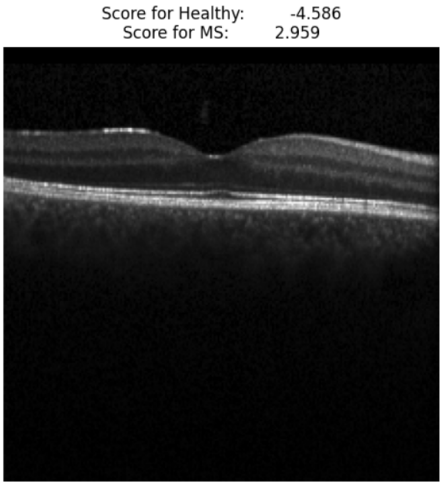


Figure 5.24: Input (MS).

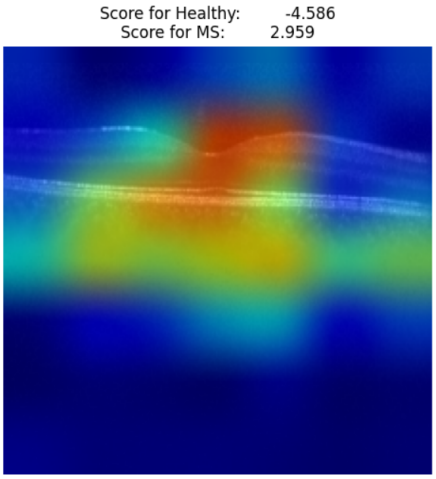


Figure 5.25: Grad-CAM Output.

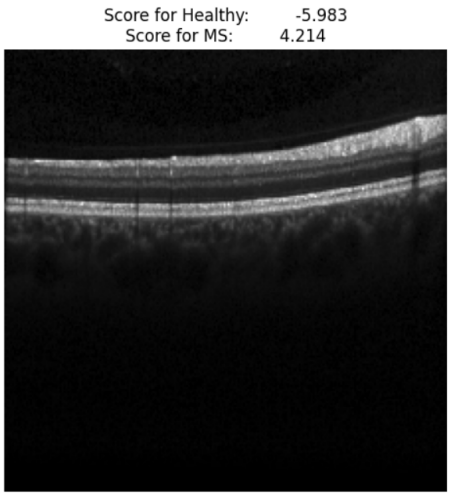


Figure 5.26: Input (MS).

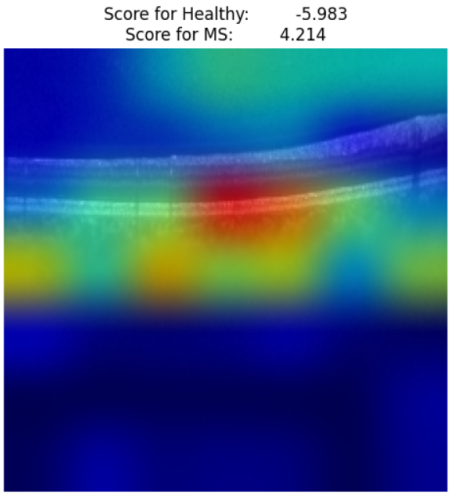


Figure 5.27: Grad-CAM Output.

Therefore, it can be concluded that our model is indeed paying attention to different retinal layers featured in the OCT images, extracting their features for predicting whether an OCT image is from a healthy individual or Multiple Sclerosis patient. So, the retinal layers are indeed working as a biomarker for detecting MS according to our proposed model.

Chapter 6

Conclusion

The significance of Optical Coherence Tomography (OCT) in detecting neurodegenerative disorder is huge and requires due attention. Unlike traditional imaging technologies like MRI and CT scans, OCT is economical, faster, and intelligible. To detect neurodegenerative diseases, we suggested an efficient deep learning approach. The system uses retinal images instead of brain images and utilizes the combined effectiveness of multiple models in accordance with explainable AI. Due to the ease with which symptoms can be detected early on, our approach allows us to actively look for neurodegenerative diseases instead of responding to them.

Through the proposed system, we can detect life-altering brain diseases in their infancy using OCT and improve their quality of life through early treatment. Since the accuracy of our model is demonstrated both by evaluation parameters and explainable AI, it has the potential of paving the way to the diagnosis of neurodegenerative diseases like Multiple Sclerosis that impacts different retinal layers. Its effectiveness can be further improved by training it with larger dataset composed of OCT images from different machines and different demographics. Due to lower costs associated with OCT, it will be within reach of almost everyone, regardless of their financial condition.

Bibliography

[1] J. FISHER, D. JACOBS, C. MARKOWITZ, S. GALETTA, N. VOLPE, M. NANOSCHIAVI, M. BAIER, E. FROHMAN, H. WINSLOW, and T. FROHMAN, "Relation of visual function to retinal nerve fiber layer thickness in multiple sclerosis," *Ophthalmology*, vol. 113, no. 2, pp. 324–332, Feb. 2006. DOI: 10.1016/j.ophtha.2005.10.040. [Online]. Available: https://doi.org/10.1016/j.ophtha.2005.10.040.

[2] J. Sepulcre, M. Murie-Fernandez, A. Salinas-Alaman, A. Garcia-Layana, B. Bejarano, and P. Villoslada, "Diagnostic accuracy of retinal abnormalities in predicting disease activity in MS," *Neurology*, vol. 68, no. 18, pp. 1488–1494, Apr. 2007. DOI: 10.1212/01.wnl.0000260612.51849.ed. [Online]. Available: https://doi.org/10.1212/01.wnl.0000260612.51849.ed.

[3] E. Grazioli, R. Zivadinov, B. Weinstock-Guttman, N. Lincoff, M. Baier, J. R. Wong, S. Hussein, J. L. Cox, D. Hojnacki, and M. Ramanathan, "Retinal nerve fiber layer thickness is associated with brain MRI outcomes in multiple sclerosis," *Journal of the Neurological Sciences*, vol. 268, no. 1-2, pp. 12–17, May 2008. DOI: 10.1016/j.jns.2007.10.020. [Online]. Available: https://doi.org/10.1016/j.jns.2007.10.020.

[4] V. Pueyo, J. Martin, J. Fernandez, C. Almarcegui, J. Ara, C. Egea, L. Pablo, and F. Honrubia, "Axonal loss in the retinal nerve fiber layer in patients with multiple sclerosis," *Multiple Sclerosis Journal*, vol. 14, no. 5, pp. 609–614, Jun. 2008. DOI: 10.1177/1352458507087326. [Online]. Available: https://doi.org/10.1177/1352458507087326.

[5] J. Toledo, J. Sepulcre, A. Salinas-Alaman, A. García-Layana, M. Murie-Fernandez, B. Bejarano, and P. Villoslada, "Retinal nerve fiber layer atrophy is associated with physical and cognitive disability in multiple sclerosis," *Multiple Sclerosis Journal*, vol. 14, no. 7, pp. 906–912, Jun. 2008. DOI: 10.1177/1352458508090221. [Online]. Available: https://doi.org/10.1177/1352458508090221.

[6] R. I. Spain, M. Maltenfort, R. C. Sergott, and T. P. Leist, "Thickness of retinal nerve fiber layer correlates with disease duration," *The Journal of Rehabilitation Research and Development*, vol. 46, no. 5, p. 633, 2009. DOI: 10.1682/jrrd.2008.11.0156. [Online]. Available: https://doi.org/10.1682/jrrd.2008.11.0156.

[7] L. S. Talman, E. R. Bisker, D. J. Sackel, D. A. Long, K. M. Galetta, J. N. Ratchford, D. J. Lile, S. K. Farrell, M. J. Loguidice, G. Remington, A. Conger, T. C. Frohman, D. A. Jacobs, C. E. Markowitz, G. R. Cutter, G.-S. Ying, Y. Dai, M. G. Maguire, S. L. Galetta, E. M. Frohman, P. A. Calabresi, and L. J. Balcer, "Longitudinal study of vision and retinal nerve fiber layer thickness in

MS," *Annals of Neurology*, NA–NA, 2010. DOI: 10.1002/ana.22005. [Online]. Available: https://doi.org/10.1002/ana.22005.

[8] C. Fjeldstad, M. Bemben, and G. Pardo, "Reduced retinal nerve fiber layer and macular thickness in patients with multiple sclerosis with no history of optic neuritis identified by the use of spectral domain high-definition optical coherence tomography," *Journal of Clinical Neuroscience*, vol. 18, no. 11, pp. 1469–1472, Nov. 2011. DOI: 10.1016/j.jocn.2011.04.008. [Online]. Available: https://doi.org/10.1016/j.jocn.2011.04.008.

[9] P. Albrecht, M. Ringelstein, A. Müller, N. Keser, T. Dietlein, A. Lappas, A. Foerster, H. Hartung, O. Aktas, and A. Methner, "Degeneration of retinal layers in multiple sclerosis subtypes quantified by optical coherence tomography," *Multiple Sclerosis Journal*, vol. 18, no. 10, pp. 1422–1429, Mar. 2012. DOI: 10.1177/1352458512439237. [Online]. Available: https://doi.org/10.1177/1352458512439237.

[10] E. Garcia-Martin, V. Polo, J. M. Larrosa, M. L. Marques, R. Herrero, J. Martin, J. R. Ara, J. Fernandez, and L. E. Pablo, "Retinal layer segmentation in patients with multiple sclerosis using spectral domain optical coherence tomography," *Ophthalmology*, vol. 121, no. 2, pp. 573–579, Feb. 2014. DOI: 10.1016/j.ophtha.2013.09.035. [Online]. Available: https://doi.org/10.1016/j.ophtha.2013.09.035.

[11] G. Soufi, E. AitBenhaddou, Z. Hajji, S. Tazrout, A. Benomar, M. Soufi, A. Boulanouar, R. Abouqal, M. Yahyaoui, and A. Berraho, "Evaluation of retinal nerve fiber layer thickness measured by optical coherence tomography in moroccan patients with multiple sclerosis," *Journal Français d'Ophtalmologie*, vol. 38, no. 6, pp. 497–503, Jun. 2015. DOI: 10.1016/j.jfo.2014.11.008. [Online]. Available: https://doi.org/10.1016/j.jfo.2014.11.008.

[12] E. Garcia-Martin, M. P. Bambo, M. L. Marques, M. Satue, S. Otin, J. M. Larrosa, V. Polo, and L. E. Pablo, "Ganglion cell layer measurements correlate with disease severity in patients with alzheimer's disease," *Acta Ophthalmologica*, vol. 94, no. 6, e454–e459, Feb. 2016. DOI: 10.1111/aos.12977. [Online]. Available: https://doi.org/10.1111/aos.12977.

[13] D. Coric, L. J. Balk, M. Verrijp, A. Eijlers, M. M. Schoonheim, J. Killestein, B. M. Uitdehaag, and A. Petzold, "Cognitive impairment in patients with multiple sclerosis is associated with atrophy of the inner retinal layers," *Multiple Sclerosis Journal*, vol. 24, no. 2, pp. 158–166, Feb. 2017. DOI: 10.1177/1352458517694090. [Online]. Available: https://doi.org/10.1177/1352458517694090.

[14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. DOI: 10.1109/ICCV.2017.74.

[15] M. Cellerino, C. Cordano, G. Boffa, G. Bommarito, M. Petracca, E. Sbragia, G. Novi, C. Lapucci, E. Capello, A. Uccelli, and M. Inglese, "Relationship between retinal inner nuclear layer, age, and disease activity in progressive MS," *Neurology - Neuroimmunology Neuroinflammation*, vol. 6, no. 5, e596, Aug. 2019. DOI: 10.1212/nxi.0000000000000596. [Online]. Available: https://doi.org/10.1212/nxi.0000000000000596.

[16] Y. He, A. Carass, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Retinal layer parcellation of optical coherence tomography images: Data resource for multiple sclerosis and healthy controls," *Data in Brief*, vol. 22, pp. 601–604, Feb. 2019. DOI: 10.1016/j.dib.2018.12.073. [Online]. Available: https://doi.org/10.1016/j.dib.2018.12.073.

[17] A. López-Dorado, M. Ortiz, M. Satue, M. J. Rodrigo, R. Barea, E. M. Sánchez-Morla, C. Cavaliere, J. M. Rodríguez-Ascariz, E. Orduna-Hospital, L. Boquete, and E. Garcia-Martin, "Early diagnosis of multiple sclerosis using swept-source optical coherence tomography and convolutional neural networks trained with data augmentation," *Sensors*, vol. 22, no. 1, p. 167, Dec. 2021. DOI: 10.3390/s22010167. [Online]. Available: https://doi.org/10.3390/s22010167.