Alzheimer's disease - Multiclass prediction

Using convoluted neural network

Authors:
Ganapathy Uduvera Earappa - 21200172
Varun Nagesh - 20200599

A thesis presented for the degree of
MSc. in Data and computational Science

School of Mathematics and Statistics
University College Dublin
Ireland
Date: 31st of July 2022

Acknowledgement

A final project is a golden opportunity for learning and self-development. We consider ourselves very lucky and honoured to have so many wonderful projects and people lead us through in completion of this training.

We express our deepest thanks to University college of Dublin for allowing us to carry out the project in maths modelling and supporting us throughout with all the resources.

We are grateful and express our deepest thanks to Dr Sarp Akcay, Assistant Professor, School of Mathematics and Statistics, University College Dublin for his continuous support, committed supervision and valuable guidance while carrying out our project.

Finally we thank all other unnamed who helped us in various ways to gain knowledge and have a good training.

Abstract

Alzheimer's disease is a kind of brain disease, the neurons damaged first are those in parts of the brain that is accountable for language, memory and thinking. Therefore, detecting the disease can be made at the early stage such that we can prevent the disease using medication or treatment. The study aimed to classify Magnetic Resonance Images (MRI) of Alzheimer's disease using a Convoluted Neural Network (CNN). Convolutional neural networks (CNNs) are powerful models for learning images. They have outperformed experts in many image understanding tasks. In this project, we use the MRI data of patients who have Alzheimer's. Here we have approximately 5000 images segregated into the severity of the disease.
We aimed to deploy an advanced deep learning method to determine whether we can extract valuable biomarkers from magnetic resonance imaging (MRI). We trained Convolutional Neural Networks (CNNs) on brain MRIs from the dataset available. Finally, we classify brain images into Alzheimer's, mild demented, moderate demented non-demented, and very mild demented groups. Here we have used the MobileNet model, which is a prevalent model that is available online. We have achieved a better level of accuracy using an easily accessible model. The project has given us an in-depth, mathematical understanding of the CNN model and a greater sense of awareness of Alzheimer's disease and the problems faced in by the patients.

Contents

Chapter 1

Introduction

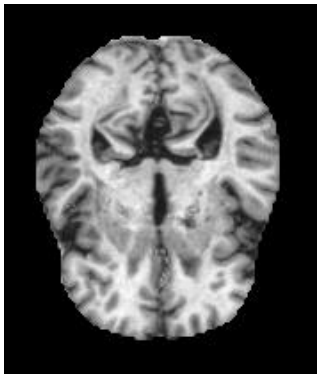
Alzheimer's disease is a chronic neurodegenerative cognitive disease. During the onset of the disease, patients will suffer from cognitive dysfunction such as memory loss and language function loss. Nowadays, with the ageing of the population, Alzheimer's Disease population is increasing rapidly, and it has become a significant disease threatening the health of the elderly. The disease will seriously affect the quality of life of the elderly, and the need for nursing and care brings heavy burdens to the patient's family and society[5].

World health organization has informed that around 55 million people worldwide have Alzheimer's, with over 60 percent living in low and middle-income countries. The percentage of older adults in the population is increasing in every country. The expected number of patients will rise to 78 million in 2030 and 139 million in 2050[15]. It is the seventh leading reason of death, and currently, no treatment is available to cure the disease. WHO also predicts that 1 in 3 seniors dies from Alzheimer's[15].

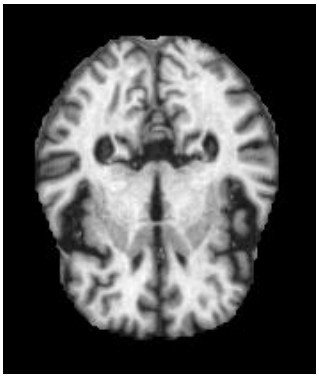
Clinically, the root cause of Alzheimer's disease is still unclear, the condition is irreversible, and a thorough treatment of the disease is not available. Only when the detection of Alzheimer's disease in the early stage can it be possible to slow down or inhibit the progression.[5] After detecting the disease by symptoms, we will be able to detect the stage of the disease. In this project, the data consists of MRI images. The data has four classes of images: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented as shown in figure 1.[4]

Magnetic Resonance Imaging (MRI) technology is the primary method of diagnosing Alzheimer's. In addition, studies have located that there are prominent biological signs in the brains of Alzheimer's patients, which is beneficial to the early prediction and diagnosis of Alzheimer's.[5]

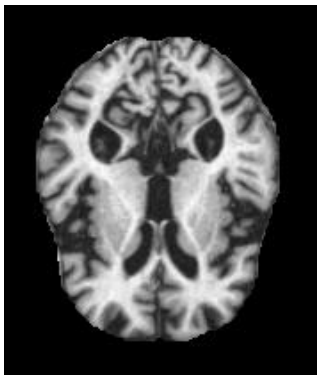
Artificial intelligence has developed rapidly in recent years, and the usage of AI technology in the medical field has increased. We incorporate Machine learning (ML) or Deep Learning (DL) algorithms into MRI imaging, but the above methods all have limitations. The machine learning algorithm analyses the MRI image after manually determining the area, but subjective factors will affect the subjectivity of the results. Deep learning algorithms can automatically extract properties of MRI images but fails to interpret these properties.



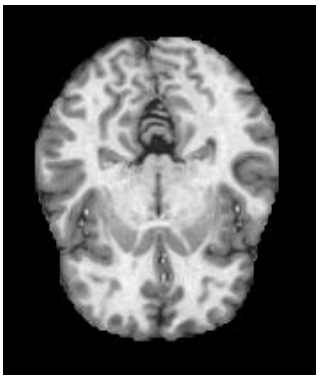
(a) Very Mild Demented



(b) Mild Demented



(c) Moderate Demented



(d) Non-Demented

Figure 1: Stages of Alzheimer's disease : MRI scans of brain

Hence, it is difficult to identify the biological characteristics of Alzheimer's disease. A Convolutional Neural network (CNN) is a Feedforward Neural network with profound algorithms and convolution computation. It is the representative algorithm of deep learning and has the ability of representation learning. It will effectively reduce the dimension of a large amount of data into a small amount while retaining image properties, in line with the principle of image processing. Here we map from the original property to the low-dimensional and dense vector space such that the distance within objects of the same class is almost close. Using the commonly used distance function in the embedding space, the distance between objects of different classes is relatively far. In this project, we use the optimized CNN models to construct a classification based on MobileNet Model using various traditional methods[7].

Chapter 2

Literature Survey

Alexander Selvikvåg Lundervoldab, Arvid Lundervold, "An overview of deep learning in medical imaging focusing on MRI." [11]
This paper describes the application of deep learning with the MRI processing chain, i.e., from acquisition to image retrieval and segmentation to disease prediction. The paper also explains CNN's detailed explanation and shows applications with MIRs of different human organs, like kidneys, prostate, spine, and brain.

November 2018.Medizinische Physik, Journal of Medical Physics.

Monika Sethi, Sachin Ahuja, "Alzheimer's Disease Classification Based on Convolutional Neural Network" [2]
This paper proposes an efficient Convolution neural network architecture that relies on brain neuroimaging (MRI) scans. This paper also describes methods adopted for Alzheimer's classification, pre-processing methods used and the sort of data inputted to the CNN model. Patch and ROI data handling practices are much more efficient than slice- and voxel-based techniques. The employment of 3D CNN to gain spatial correlations of 3D MRI images and provide better performance compared to the 2D CNN.

December 2021. BioMed Research International.

Guilherme Folego, Marina Weiler, "Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI."[6]
This paper focuses on the methodology proposed for an end-to-end deep 3D CNN for the multiclass AD biomarker identification task, using the whole image volume as input. Here the method did not use any domain-specific knowledge from Alzheimer's with an accuracy of around 53
October 2020.Front Bioeng Biotechnol, National Library of Medicine.

Chapter 3

Methodology

3.1 Data visualization

In this project, we use the MRI data of Alzheimer's disease. The classification of MRI data into four groups Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented and the data is visualized as seen in figure 1

Very Mild Demented: Very Mild demented is the early stage, where the person may function independently. In this stage, the person may feel a slight memory loss, such as forgetting words and locating objects. The primary symptoms are having difficulty performing tasks in work settings, forgetting material after reading and experiencing increased trouble with planning or organizing[3]. This is clearly observer in figure 1(a).Here the gyrus of the brain is affected slightly.

Mild Demented: Mild Demented is the next and the longest stage and can last for many years, and the person will require a more significant amount of care. Feeling moody, requiring help choosing proper clothing for the season or the occasion, having trouble controlling their bladder and bowels, experiencing changes in sleep patterns, and showing an increased tendency to wander and become lost are the symptoms in this stage[3]. As seen in the figure of the MRI images the angular gyrus and thalamus region of the brain is affected. This is clearly observer in figure 1(b).

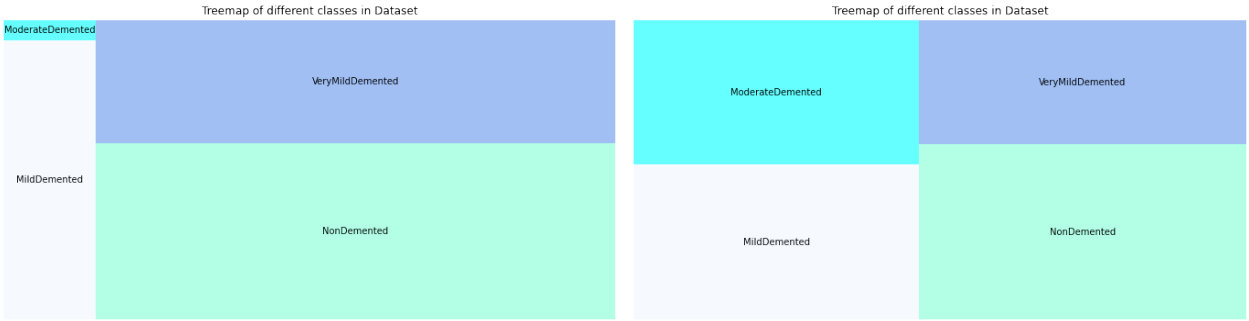
Moderate Demented: Moderate Demented leads to the final stage of the disease. In this stage, the person almost loses the ability to respond to their environment. They would require around-the-clock assistance with daily personal care. Lose awareness of experiences as well as of their surroundings, experience changes in physical abilities, and become vulnerable to pneumonia. During this stage, they may want to use support services, such as hospice care, which provides comfort and dignity at the end of life[3]. As seen in figure 1(c), the hippocampus, gyrus and insula i.e. entire part of the brain is affected.

Non-Demented: In this stage person does not experience any kind of the symptoms. There are no changes in the normal structure of the brain as seen in figure 1(d)

3.2 Data augmentation

The significance of Data augmentation is from simple transformations such as horizontal flipping, colour space augmentations, and random cropping, zooming in and out. However, these transformations encode many of the invariances to image recognition tasks. The ways of augmentation are geometric transformations, colour space transformations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, neural style transfer, and meta-learning schemes[16].

It is helpful as it improves the performance and outcomes of the models by forming new and different samples to train datasets. If the dataset in the model is rich and sufficient, the model performs better and more accurately. In our model, the distribution of the images in each class is not even as seen in figure 3(a). Hence we reshape the array and obtain different image dimensions for the particular class and ensure that all four classes have same proportion of data as seen in figure 3(b). It will help us improve the specificity and sensitivity of the model in each class.

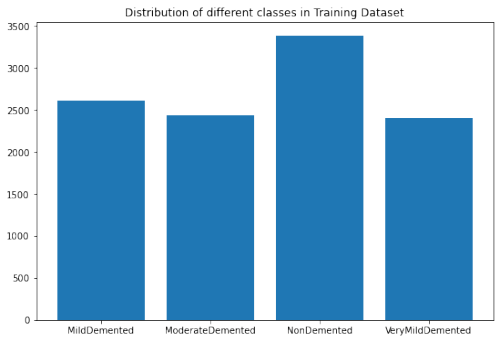


(a) Before data augumentation (b) After data augumentation

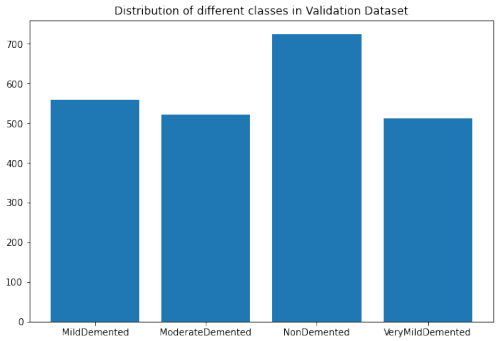
Figure 2: Graphs of data augmentation

3.3 Data splitting

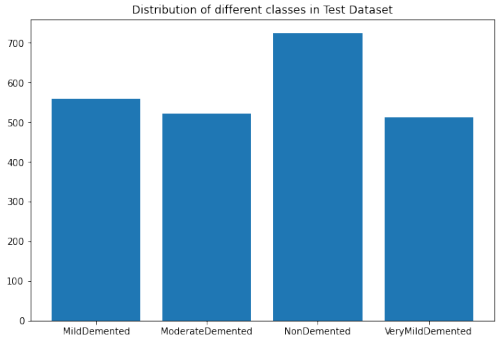
In this step, we divided the dataset into three subsets for validation purposes. We use the subset for model prediction (i.e., test data), and the other two sets (i.e., training and validation) to assess model performance by training against new data. After data pre-processing, we randomly split the whole dataset into an 85:15 ratio, where we used 85 percent for training and 15 percent for testing. It will enable the machine to create new combinations every time to run the model and make it possible to predict it with the highest accuracy.
Before model training, we split the training dataset into two subsets for training(70 percent) and validation(15 percent). The validation dataset helps to choose hyper tuning parameters, such as regularization and learning rate. These hyper tuning parameters can limit the model over-fitting and improve accuracy. From the bar graph in figure 3.2 we can observe the variation of data in training, validation and test data for all four classes of Alzheimer's disease.
Though we observer all the images are the same,from image 3.2(a) we can observer that the value in Y axis is greater than the graphs 3.2(b) and 3.2(c).
Training Data: 70 %
Validation data: 15 %
Test data: 15 %



(a) Training data.



(b) Validation data.



(c) Test data.

Figure 3: Data splitting.

3.4 Convoluted Neural Network

A convoluted neural network is a multilayered perceptron that consists of an input layer, several hidden layers, and an output layer. The hidden layers are further composed of numerous other layers, including convolutional (C-layers), subsampling (S-layers), and fully connected layers (FC-Layers)[10]. C-layers are the foundation of the CNN model, which is used to extract low- to complex-level features from images. The S-layer reduces the amount of data in each feature map obtained after the C-Layer while retaining the most important features. Finally, an FC layer takes the C- and S-layer results and converts them into a single long vector, which is then used to classify the various images into corresponding labels. [10] Here for our project we use MobileNet CNN model as it is a easily available model online and is a faster and smarter model that has reduced number of parameters compared to others.

3.4.1 MobileNet architecture

MobileNet, as shown in figure 3.3, has a smaller structure, less computation, and higher precision, and used for mobile terminals and embedded devices. Based on depthwise separable convolutions, MobileNets use two global hyperparameters to keep a balance between efficiency and accuracy. We obtain Depthwise convolution from two operations
1. Depthwise convolution
2. Pointwise convolution

MobileNet is a class of CNN that Google open-sourced, giving us an excellent starting point for training our classifiers that are very small and fast[19].
The core idea of MobileNet is the decomposition of convolution kernels. By using depthwise separable convolution, we decompose the standard convolution into a depthwise convolution and a pointwise convolution with a 1 x 1 convolution kernel, as shown in the FIGURE. The depthwise convolution filters perform convolution to each channel, and we use the 1 x 1 convolution to merge the outputs of the depthwise convolution layers. In this way, we replace N standard convolution kernels with M depthwise convolution kernels and N pointwise convolution kernels . A standard convolutional filter combines the inputs into a new set of outputs. At the same time, depthwise separable convolution divides the inputs into two layers, one for filtering and the other for merging[19].

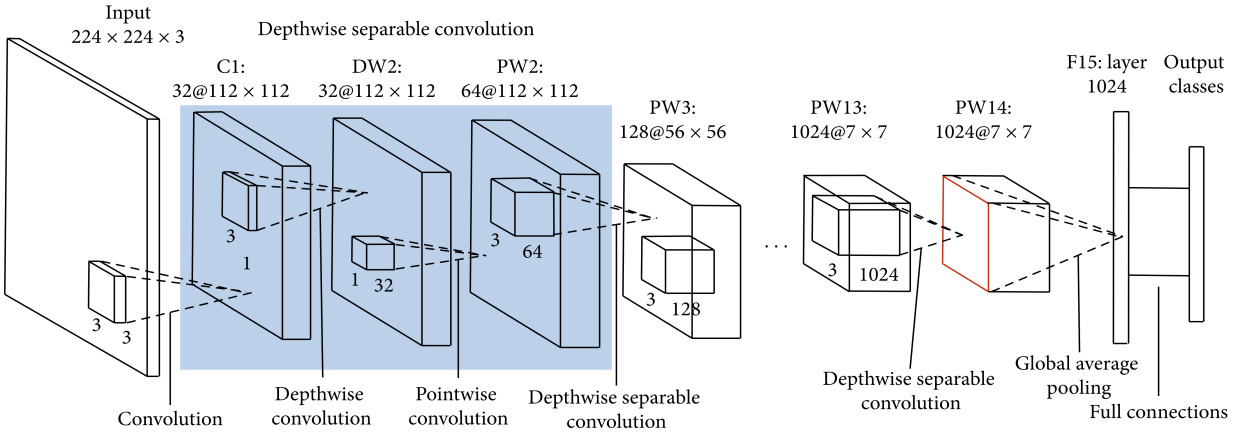


Figure 4: MobileNet Architecture

3.4.2 Computational analysis

In the standard convolution layer, assuming the height, width, and input channel number of the input feature maps I are h, w, and the convolution filter K is s x s, the output channel number is n, and the output feature map O = K x I can be obtained by the convolution of I and K with no padding zeros and stride I, as shown in the following formula[19]

$$O(y,x,j) = \sum_{i=1}^{m} \sum_{u,v=1}^{s} k(u,v,i,j) I(y+u-1, x+v-1, i) \tag{1}$$

Where O(y, x, j) represents the value of the point (y, x) in the jth output feature map, K(u, v, i, j) represents the value of the point (u, v) on the channel I in the jth convolution filter and I(y, x, i) represents the value of the point (y, x) on ith input feature map. From formula (1), it is known that an output value needs s x s x m times multiplication, so the total amount of calculations is s x s x m x (h – s + 1) x (w – s + 1) x n and the number of parameters is s x s x m x n[19].
When introducing dilated convolution filters to the depthwise convolution, the above feature maps I is firstly convoluted with the depthwise convolution filter K, and the output feature graph Odc is obtained through the following formula[19]

$$O_{dc}(y,x,j) = \sum_{u,v=1}^{s} K(u,v,j) I(y+u+(u-1)(r-1)-1, x+v+(v-1)(r-1)-1, j) \tag{2}$$

The total computation of the depthwise separable convolution is (s x s x n) x (h - s - (s - 1)(r - 1) + 1) x (w - s -(s -1)(r - 1) + 1) x m, amd the total number of parameters is s x s x m + m x n. It can be seen that the parameter of the depthwise separable convolution are reduced with the standard convolution[19]

$$\frac{s \times s \times m + m \times n}{s \times s \times m \times n} = \frac{1}{n} + \frac{1}{s^2} \tag{3}$$

In the same way when carrying out the depthwise convolution with padding zeros, the reduction ratio of parameters is

$$\frac{(s \times s + n) \times m \times h \times h \times w}{s \times s \times m \times n \times h \times w} = \frac{1}{n} + \frac{1}{s^2} \tag{4}$$

From this analysis it can the observed that the receptive field of the deep convolution with expansion rate r and convolution kernal size s x s

3.4.3 Layers and parameters in the architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
$5\times$ Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
$5\times$ Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Figure 5: MobileNet Body Architecture

From figure 5 we can understand the number of layers in the model and the filter shape of each layer with the input shape. Since the model is very long with many numbers of functions and layers, we try to understand some of the essential layers and parameters in this section.

Batch normalization layer:

Batch normalization is a technique for training deep neural networks that standardizes the inputs to a layer for each mini-batch. It stabilizes the learning process and dramatically reduces the number of training epochs required to train deep networks[17]. We can define the normalization formula of Batch Norm as

$$z^N = (\frac{z - m_z}{s_z}) \tag{5}$$

Batch normalization is a general technique to normalize the inputs to a layer. It works with most network types, such as Multilayer Perceptrons, Convolutional Neural Networks and Recurrent Neural Networks. Batch normalization is used in inputs to the layer before or after the activation function in the previous layer. It is more appropriate after the activation function for s-shaped functions like the hyperbolic tangent and logistic function[17].

Activation Function:

The activation function compares the input value to a threshold value. The neuron is activated if the input value exceeds the threshold value. It's disabled if the input value is less than the threshold value, which prevents the output to the next or hidden layer.
Relu is a half rectifies activation function, i.e. f(z) is zero when z is less than zero and f(z) is equal to z when z is above or equal to zero[20].

$$R(z) = \begin{cases} z & z > 0 \\ 0 & z <= 0 \end{cases} \tag{6}$$

Softmax function turns a vector of K real values into a vector of K real values that sum to 1. The input values can be positive, negative, zero, or greater than one, but the softmax

transforms them into values between 0 and 1 so that the interpretation is in probabilities[20].

$$S(z) = \frac{1}{1 + e^{-z}} \tag{7}$$

Average pooling and Zero padding:

Average pooling involves calculating the average for each patch of the feature map. Each 2×2 square of the feature map is down-sampled to the average value in the square[9].
Zero padding refers to the process of symmetrically adding zeroes to the input matrix. It's a commonly used modification that adjusts the input size to our requirements in designing the CNN layers when the dimensions of the input volume need to be preserved in the output volume[9].

Regularization:

Regularization is a set of techniques that can prevent over-fitting in neural networks and thus improve the accuracy of a Deep Learning model when facing completely new data. Dropout regularization is a technique to prevent neural network from overfitting. Dropout works by randomly disabling neurons and their corresponding connections. It prevents the network from relying too much on single neurons and forces all neurons to learn to generalize better[18].

Flatten and dense layer:

The Flatten layer converts the 28x28x32 output of the convolutional layer into a single one-dimensional vector, used as input for a dense layer. The last dense layer has the most parameters. This layer connects every output 'pixel' from the convolutional layer to the output classes. That results in a large number of connections, so a large number of parameters. It undermines the expressiveness of the convolutional layers, which have much fewer parameters[8].

In this project we use MobileNet CNN model with keras application. Some of the important arguments we use here are input shape and include top. In input shape we define the 3 inputs channel, width and height and also the value should be no smaller than 32. In include top we declare true or false and it would define whether to include the fully-connected layer at the top of the network[8].

3.4.4 Model Validation

Model validation is the practice of identifying an optimal model through skipping the train and testing on the same data and helps to reduce complex overfitting issues. To overcome such an issue, we performed the early stopping method to train the model and, after that, calculate the accuracy. It is always a challenge to validate the model with a trained dataset, and to ensure the model is noise-free, we use early stopping techniques. Early stopping is a method that allows you to specify an arbitrarily large number of training epochs and stop

training once the model performance stops improving on the validation dataset. In this work, we applied the Early stopping technique because it is a popular ML technique and produces low-bias models[13].

3.4.5 Cross Entropy Loss function

The cross-entropy loss function commonly used in multiclassification problems is used as the loss function of MobileNet convolutional neural training. The model is iteratively trained with Adam as the model optimizer. It is also known as softmax loss and has been proven quite effective in eliminating outliers in image classification tasks[14]. In this project we use categorical cross entropy and it is calculated as:

$$L_{CE} = -\sum_{i=1}^{n}(t_i \times log(p_i)) \tag{8}$$

where ti is the truth label and pi is the softmax probability for the ith class.

Chapter 4

Results

4.1 Accuracy and loss

The accuracy metric assesses the algorithm's performance in an informative way. The accuracy of a model is usually determined after the evaluation of model parameters. It assesses how accurate your model's prediction is compared to the actual data.
An algorithm is optimized using a loss function. The loss is computed using training and validation data. Furthermore, its interpretation is determined by how well the model performs in these two sets. It is the total of all errors made in training or validation sets for each example. The loss value indicates how poorly or well a model performs after each optimization iteration.
After performing the CNN model training and validations we plot the check the accuracy and loss for the models. From the graph in figure 6 for we can find that the accuracy is approximately 80%.

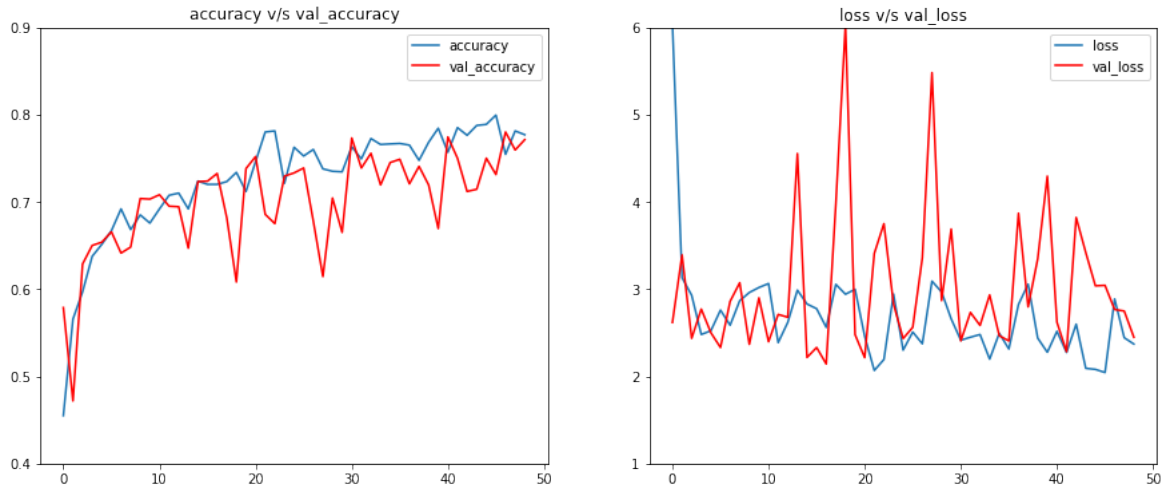


Figure 6: Accuracy and loss graph of training and validation data

4.2 Confusion matrix

A Confusion matrix is a N x N matrix used to assess the performance of a classification model, where N represents the number of target classes. For each classification, the matrix compares the actual target values to those predicted by the machine learning model. Confusion matrices are useful because they directly compare values like True Positives, False Positives, True Negatives, and False Negatives.

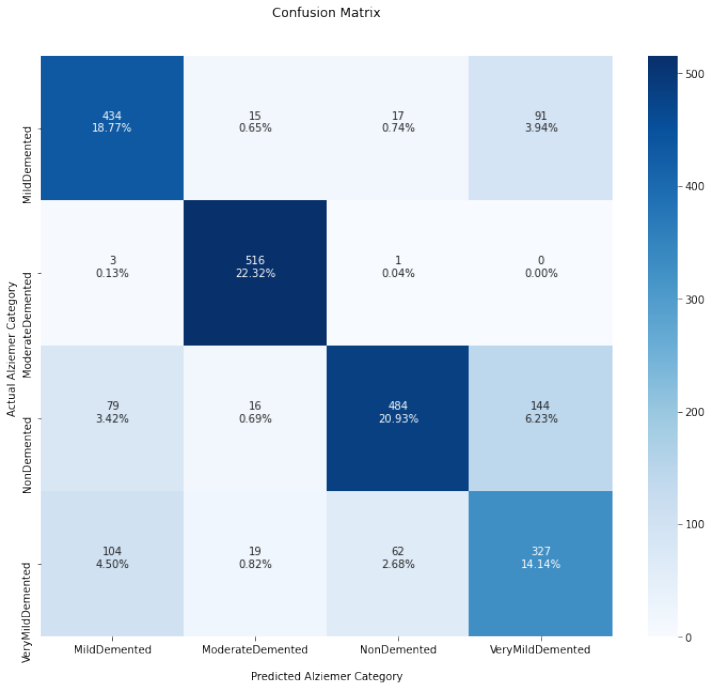


Figure 7: Accuracy and loss graph of training and validation data

From the above Figure 7 we will be able to draw the following conclusions as we have the matrix plotted for the different Alzheimer's classes for actual and predicted values.

- For Mild Demented class of Alzheimer's disease the prediction is correct for 434 images among 557 MRI scans in test data

- for very mild demented stage we have 327 of correctly predicted data among 512 MRI images of the training data, however we have 104 and 62 images incorrecly predicted data as mild demented and Non demented.

- For Moderate demented we have 516 correctly predicted values among 520 MRI scans in the test data and here we have have 3, 1 and 0 images predicted for mild demented, non demented and very mild demented. From this we can conclude that there are prominent difference observed in Moderate Demented.

- For non-demented we have correctly predicted value equal to 484 of 723 MRI scans in the test data. Here incorrect predictions for very mild demented and mild demented is 144 and 79 images and this is again because of poor changes in the MRI images.

4.3 Sensitivity and specificity

Sensitivity: The sensitivity of a test is its ability to determine the patient cases correctly. We should estimate the proportion of true positives inpatient cases to estimate it. Mathematically:

$$Sensitivity = \frac{TruePositive}{TruePositive + falseNegative} \tag{9}$$

Specificity: The specificity of a test is its ability to determine the healthy cases correctly. We should estimate the proportion of true negatives in healthy cases to estimate it. Mathematically:

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \tag{10}$$

	precision	recall	f1-score	support
MildDemented	0.70	0.78	0.74	557
ModerateDemented	0.91	0.99	0.95	520
NonDemented	0.86	0.67	0.75	723
VeryMildDemented	0.58	0.64	0.61	512
accuracy			0.76	2312
macro avg	0.76	0.77	0.76	2312
weighted avg	0.77	0.76	0.76	2312

Figure 8: Classification report

On an average the overall precession and sensitivity of all the classes is around 0.77.

Chapter 5

Conclusion

Machine Learning research associated with neurological studies will offer a more precise analysis of Alzheimer's. In this project, based on brain MRI features, we used a framework based on supervised learning models to classify different stages of Alzheimer's patients into four categories, i.e., very mild, mild, moderate, and Non-Demented.The MRI demographic information and the patient's pre-existing conditions can help improve classification accuracy when predicting the AD subject status. The highest average accuracy of 0.80 (approx). We proposed a simple and efficient dementia identification technique. Future research should look into more sophisticated prediction models and alternate ways.

Chapter 6

Discussion and Future scope

Recent advances in clinical screening and technologies such as brain imaging have made it possible to detect Alzheimer's early. Although brain imaging is effective in most cases, the results are not always accurate. As a result, the focus of the research has shifted to the identification of molecular biomarkers, which aids in distinguishing between genotype and phenotype characteristics. Molecular data-based of blood plasma research proves to be effective. Still, it generates huge volumes of data consisting of transcripts, transcriptomes, etc. It creates a "curse of dimensionality" problem[12]. There is also work being done to determine the risk of developing Alzheimer's dementia later in life using a simple blood test and getting the gene data and develop diagnosis using convoluted neural network[1]. Thus, machine learning-based feature selection techniques are implemented to select only the relevant genes affecting the outcome. We are headed further in a direction without MRI images and instead gene technology and blood plasma.

Bibliography

[1] Ivan Arisi, Mara D'Onofrio, Rossella Brandi, Armando Felsani, Simona Capsoni, Guido Drovandi, Giovanni Felici, Emanuel Weitschek, Paola Bertolazzi, and Antonino Cattaneo. Gene expression biomarkers in the brain of a mouse model for alzheimer's disease: mining of microarray data by logic classification and feature selection. *Journal of Alzheimer's Disease*, 24(4):721–738, 2011.

[2] Xiaowang Bi, Wei Liu, Huaiqin Liu, and Qun Shang. Artificial intelligence-based mri images for brain in prediction of alzheimer's disease. *Journal of Healthcare Engineering*, 2021, 2021.

[3] Richard J Caselli and Eric M Reiman. Characterizing the preclinical stages of alzheimer's disease and the prospect of presymptomatic intervention. *Journal of Alzheimer's Disease*, 33(s1):S405–S416, 2013.

[4] Ivanoe De Falco, Giuseppe De Pietro, and Giovanna Sannino. A two-step approach for classification in alzheimer's disease. *Sensors*, 22(11):3966, 2022.

[5] Michael A DeTure and Dennis W Dickson. The neuropathological diagnosis of alzheimer's disease. *Molecular neurodegeneration*, 14(1):1–18, 2019.

[6] Guilherme Folego, Marina Weiler, Raphael F Casseb, Ramon Pires, and Anderson Rocha. Alzheimer's disease detection through whole-brain 3d-cnn mri. *Frontiers in bioengineering and biotechnology*, 8:534592, 2020.

[7] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[8] Emtiaz Hussain, Mahmudul Hasan, Syed Zafrul Hassan, Tanzina Hassan Azmi, Md Anisur Rahman, and Mohammad Zavid Parvez. Deep learning based binary classification for alzheimer's disease detection using brain mri images. In *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 1115–1120. IEEE, 2020.

[9] Jagadeesh Kakarla, Bala Venkateswarlu Isunuri, Krishna Sai Doppalapudi, and Karthik Satya Raghuram Bylapudi. Three-class classification of brain magnetic resonance images using average-pooling convolutional neural network. *International Journal of Imaging Systems and Technology*, 31(3):1731–1740, 2021.

[10] Jihun Kim, Jonghong Kim, Gil-Jin Jang, and Minho Lee. Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection. *Neural Networks*, 87:109–121, 2017.

[11] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.

[12] Jue Mo, Sana Siddiqui, Stuart Maudsley, Huey Cheung, Bronwen Martin, and Calvin A Johnson. Classification of alzheimer diagnosis from adni plasma biomarker data. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, pages 569–574, 2013.

[13] Modupe Odusami, Rytis Maskeliunas, Robertas Damaševičius, and Sanjay Misra. Comparable study of pre-trained model on alzheimer disease classification. In *International Conference on Computational Science and Its Applications*, pages 63–74. Springer, 2021.

[14] Muhammad Wildan Oktavian, Novanto Yudistira, and Achmad Ridok. Classification of alzheimer's disease using the convolutional neural network (cnn) with transfer learning and weighted loss. *arXiv preprint arXiv:2207.01584*, 2022.

[15] World Health Organization et al. Global action plan on the public health response to dementia 2017–2025. 2017.

[16] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

[17] Russell Reed and Robert J MarksII. *Neural smithing: supervised learning in feedforward artificial neural networks*. Mit Press, 1999.

[18] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[19] Wei Wang, Yiyang Hu, Ting Zou, Hongmei Liu, Jin Wang, and Xin Wang. A new image classification approach via improved mobilenet models with local receptive field expansion in shallow layers. *Computational Intelligence and Neuroscience*, 2020, 2020.

[20] Chao Zhang and Philip C Woodland. Parameterised sigmoid and relu hidden activation functions for dnn acoustic modelling. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.