University of California, Merced

Classification of Mild Cognitive Impairment under Data Limitations using Deep Learning Techniques

A Thesis submitted in partial satisfaction of the requirements for the degree of Master of Science

in

Applied Mathematics

by

Ashley De Luna

Committee in charge:
Professor Roummel Marcia, Chair
Professor Arnold Kim
Professor Erica Rutter

2020

This is to certify that I have examined a copy of a technical report by

Ashley De Luna

and found it satisfactory in all respects, and that any and all revisions required by the examining committee have been made.

Applied Mathematics
Graduate Studies Chair:

_____

Professor Roummel Marcia

Thesis Committee:

_____

Professor Arnold Kim

Thesis Committee:

_____

Professor Erica Rutter

Committee Chair / Research
Advisor:

_____

Professor Roummel Marcia

_____

Date

Contents

List of Abbreviations

AD	Alzheimer's Disease
Adam	Adaptive moment estimation algorithm
ADNI	Alzheimer's Disease Neuroimaging Initiative
CN	Control Normal
CNN	Convolutional Neural Network
CSF	Cerebrospinal fluid
EMCI	Early Mild Cognitive Impairment
fMRI	functional Magnetic Reasonance Imaging
FWHM	Full width at half maximum
GLM	General linear model
GM	Gray matter
ICBM	International Consortium of Brain Mapping
LMCI	Late Mild Cognitive Impairment
MCI	Mild Cognitive Impairment
MLP	Multilayer perceptron
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
NIfTI	Neuroimaing Informative Technology Initiative
PET	Positron emission tomography
PNG	Portable Network Graphics
ReLU	Rectified linear unit
RMSProp	Root mean square propagation
SD	Standard deviation
SNR	Signal to noise ratio
SPM	Statistical Parametic Mapping
TMP	Tissue probability maps
WM	White matter
WMS	Wechsler Memory Scale

List of Figures

List of Tables

Acknowledgments

First and foremost, I want to express my sincere gratitude to my advisor Dr. Roummel Marcia. His guidance throughout my time at UC Merced helped me in my research and mathematical education without him this thesis would not be possible. I would also like to thank my committee members, Dr. Arnold Kim and Dr. Erica Rutter, for their support. I would like to thank my fellow graduate students in applied mathematics. Their conversations and support helped me grow into a better scientist. I would also like to thank my companion Eric, for persistently standing by and supporting me throughout these years. Finally, I would like to thank my parents. They have been my constant support system throughout my academic journey. I am forever grateful for all their love and encouragements.

Classification of Mild Cognitive Impairment under Data Limitations using Deep Learning Techniques

by

Ashley De Luna

Master of Science in Applied Mathematics

Dr. Roummel Marcia, Committee Chair

University of California, Merced

2020

Abstract

Mild Cognitive Impairment (MCI) is the stage between the declining of normal brain function and the more serious decline of dementia. Alzheimer's disease (AD) is one of the leading forms of dementia. Despite the fact that MCI does not always lead to AD, an early diagnosis of MCI may be helpful in finding those with early signs of AD. The Alzheimer's Disease Neuroimaging Initiative (ADNI) has utilized magnetic resonance imaging (MRI) for the diagnosis of MCI and AD. MCI can be separated into two types: Early MCI (EMCI) and Late MCI (LMCI). Furthermore, MRI results can be separated into three views of axial, coronal and sagittal planes. In this work, we perform binary classifications between healthy people and the two types of MCI based on limited MRI images using a deep learning approach. Furthermore, we implement and compare two various convolutional neural network (CNN) architectures. The MRIs of 516 patients were used in this study: 172 control normal (CN), 172 EMCI patients and 172 LMCI patients. For this data set, 50% of the images were used for training, 20% for validation, and the remaining 30% for testing. The results showed that the best classification for one model was between CN and LMCI for the coronal view with an accuracy of 79.67%. In addition, we achieved 67.85% accuracy for the second proposed model for the same classification group.

Chapter 1

Introduction

Mild cognitive impairment (MCI) is a condition where individuals experience a decline in their mental and cognitive abilities. It is the intermediate stage before the development of Alzheimer's disease (AD) and other types of dementia. Although MCI does not always transition into AD, an early diagnosis of MCI could benefit those individuals, their families and governments on a social and financial level. A study found that if all AD patients were diagnosed in the early stages, it would save a total of $7 trillion to $7.9 trillion [3]. Also, once an individual is properly diagnosed with MCI, they and their family have a better timeline for social, financial and medical decisions.

MCI is divided into early and late MCI (EMCI and LMCI). Due to the similarities between healthy and MCI brain images, the classification between EMCI and normal aging brains remains a challenging and critically important problem. Brain imaging methods, such as magnetic resonance imaging (MRI) have become a significant tool in the diagnosis of MCI and AD. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is the database used to acquire the MRI images for this study. In addition, the statistical parameter mapping (SPM) software allows us to observe the variance in brain structure and function by studying the biomarkers in brain images based on the gray matter (GM) extracted from MRI images. Then, we will use and compare two different convolutional neural networks (CNNs) to classify between individuals into healthy, EMCI or LMCI.

The work presented in this thesis aids in the prevention of AD by properly and accurately classifying between normal brains and those with EMCI and LMCI under data limitations using a deep learning approach. In Chapter 2, we will provide the necessary background to understand the project. Chapter 3 focuses on the preprocessing used on the acquired MRI images. Chapter 4 describes the proposed model architectures, the performance measurements used to compare the models and provides the results and discussion of our findings. Finally, Chapter 5 draws conclusions of the work and future work.

Chapter 2

Literature Overview

2.1 Alzheimer's Disease

Alzheimer's Disease (AD) is a type of brain disease and is the most common cause of dementia. The characteristic symptoms of dementia are difficulties with memory, language, problem-solving and other cognitive abilities that affect an individual's everyday activities. Unfortunately, AD is also a degenerative disease. It is the sixth leading cause of death in the United States and the fifth leading cause of death for Americans 65 years and older [1]. An estimated 5.8 million Americans are living with AD. By 2025, an estimated number of 7.1 million people age 65 and older with have Alzheimer's dementia. Furthermore, by 2050, the projection of those individuals age 65 and older is estimated to reach 13.8 million [4].

Unfortunately, individuals do not start noticing symptoms until years after the disease has begun. Minor symptoms include memory loss and language problems, while others develop more severe symptoms such as losing mobility in the body and ultimately succumb to the disease. This progression of AD is known as the Alzheimer's disease continuum. There are three phases to this continuum: preclinical AD, mild cognitive impairment (MCI) due to AD and dementia due to AD [5–8] (Figure 2.1). The Alzheimer's dementia phase is separated into sections that reflect the level of symptoms that interfere with an individual's daily activities. While these are the three general stages, the time that each individual spends in each part varies due to age, genetics, gender, and other factors [9].

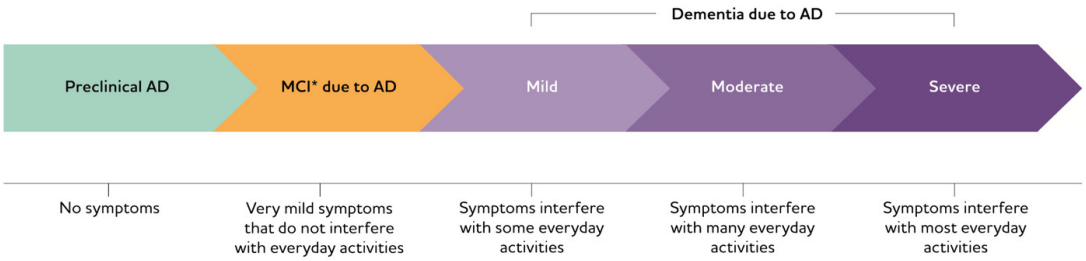


Figure 2.1: Alzheimer's disease (AD) continuum - (see [1] for details).

Preclinical Alzheimer's Disease

During the preclinical AD phase, individuals do not show noticeable symptoms although, they have measurable brain changes that indicate early signs of AD biomarkers. These changes include abnormal levels of beta-amyloid as shown on positron emission tomography (PET) scans and in analysis of cerebrospinal fluid (CSF), decreased metabolism of glucose. However, the brain is able to compensate for the early changes and allow the individual to function normally [5].

2.1.1 Mild Cognitive Impairment (MCI)

Mild Cognitive Impairment (MCI) is the phase between pre-clinical AD and the more serious decline of dementia due to AD. A person with MCI will have symptoms evidence of Alzheimer's brain changes, such as subtle problems with mental and cognitive abilities. Other typical symptoms of those with MCI are memory loss or speech difficulties. For some, these symptoms may not interfere with everyday activities, while for others that do develop memory and cognitive issues, the brain can no longer compensate for the damage cause by AD [6].

Considering those with MCI, one study found that after two years' follow-up, 15% of individuals older than 65 have developed dementia [10]. Another study found that 32% of individuals with MCI developed Alzheimer's dementia within five years' follow-up [11]. Lastly, a third study found that individuals who were tracked for five years or more, 38% developed dementia [12]. However, there are cases where individuals with MCI revert back to normal cognition or remain stable. Current research goals focus on properly identifying individuals with MCI since they are more likely to develop AD.

Dementia

Individuals with dementia due to Alzheimer's disease show a noticeable decline in memory, cognitive and behavioral function that affect a person's ability to perform everyday activities. Multiple symptoms progress over the course of several years; however, the pace at which the symptoms advance varies from person to person. The symptoms reflect the degree of damage to nerve cells in different parts of the brain. Additionally, dementia can be broken down to three stages: mild, moderate and severe Alzheimer's dementia [7]. Overall, individuals with MCI are more likely to develop dementia due to Alzheimer's disease, which is why it is crucial to identify those with MCI before it reaches this final phase.

2.2 Alzheimer's Disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative (ADNI) allows researchers to work toward the prevention of Alzheimer's disease. Created in 2004 it has since continued its innovation for early detection and to identify ways to track the diseases' progression with biomarkers. It is funded by the National Institutes of Health with the purpose to support advances in AD research. Under ADNI, there are numerous research data sets available, such as diagnostic tests, MRI scans, PET scans, measurements of proteins in blood and cerebral spinal fluid, neuropsychological test data, genetics information and clinical evaluations. All ADNI data are shared through the LONI Image and Data Archive (IDA). Scientists and researchers may request access online. The data accessed for this project was acquired from the ADNI database, publicly available (http://www.loni.ucla.edu/ADNI/). ADNI utilizes various biomarkers to help the prevention and treatment of AD. A biomarker is a substance, measurement or indicator of a biological state. There are five key biomarkers that change from normal to abnormal during the progression of AD: b-amyloid (Aβ), tau protein, brain structure, memory and clinical function (see Figure 2.2) [2].

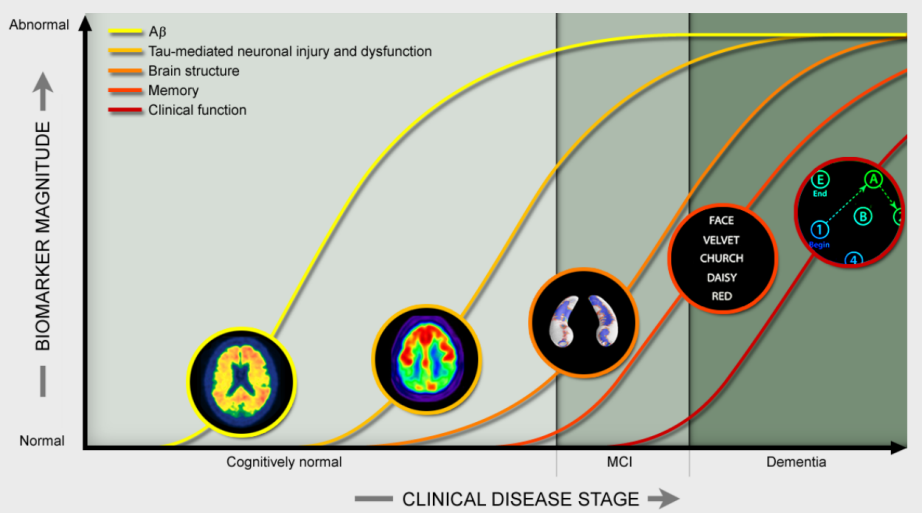


Figure 2.2: Progression of biomarkers from cognitive normal to dementia - (see [2] for details).

Using the Alzheimer's Disease Neuroimaging Initiative (ADNI) (ADNI GO/ ADNI 3), MCI patients have been classified into two groups: early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI). Levels of MCI are determined using the Wechsler Memory Scale (WMS) Logical Memory II; a subtest used to measure different memory functions in an individual. Thus, the two groups are distinguished between each other based on the degree of memory impairment. For the EMCI patients, the decline in memory is approximately between 1.0-1.5 standard deviation (SD) below the normative mean, while for LMCI patients the decline is at least approximately 1.5 SD below the normative mean [13].

2.3 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technology that produces three-dimensional detail anatomical images using magnetic and electric fields. When capturing an MRI, a patient is placed in a strong magnetic field, the protons inside the cells move out of equilibrium and align with that field. Then, when the magnetic field is removed, the MRI sensors detect the radio frequency signals released by the protons as they realign back to the equilibrium. These measurements are used to create a three-dimensional greyscale MRI. It has often been used in disease detection, diagnosis and during the prognosis of treatment [14]. One type of MRI is functional Magnetic Resonance Imaging (fMRI). It captures the brain activity by measuring the blood flow during cognitive activities. When neurons in the brain are active there is an increase in blood flow in those areas. Therefore, fMRI images are used to observe the effects of AD on the brain structure.

There are features in brain images that can be used to distinguish between AD, MCI and healthy brains. The features we consider are grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Research shows that GM measurements can be detected in brain alterations that are associated with cognitive impairment [15]. Current research on the classification of CN, LMCI and EMCI using GM extracted from fMRI images aids in the early diagnosis of Alzheimer's disease.

2.4 Statistical Parametric Mapping

The statistical parametric mapping (SPM) uses statistical techniques to examine brain activity in fMRI and PET Scans [16]. SPM are images or fields with values that are distributed according to a known probability density function. They can be considered as an "X-ray" of the significance of the regional effects. The process includes analyzing each voxel using a standard (univariate) statistical test based on the general linear model (GLM) of the data, given by

$$Y = X\beta + \epsilon \tag{2.1}$$

where Y is the acquired fMRI data, X is the design matrix, β is the calculated weight factors and ϵ is the estimated error (see [17] for details). GLM is an extension of linear regression, a method used to calculate the line of best fit for a set of data. We refer the reader to [18] for a more detailed explanation of GLM and the relationship with fMRI. SPM can also be referred to the software used for the preprocessing applied to our dataset. This software expands the field of research for measuring the spatial distribution of atrophy in brain aging research [19, 20].

2.5 Deep Learning

The field of machine learning utilizes the power of computers to learn from data [21]. Machine learning covers a wide scope of subjects, including statistical modeling, regression algorithms and random forests algorithm [22]. In this section, we will study the subsection of machine learning known as deep learning. We will define deep learning, as well as establish the building blocks needed to create deep learning algorithms. Most of the explanations in this section come from [23].

2.5.1 Feedforward Neural Networks

The basic deep learning algorithm is the feedforward neural network, also known as multilayer perceptron (MLP). The goal of the MLP is to approximate an unknown function f based on a set of data points (x_1,y_1), (x_2,y_2), ..., (x_m,y_m), where the x_i's are the inputs and the y_i's are the outputs. Mathematically, it is defined as a mapping $y = \hat{f}(x, W)$, where W is a set of learned parameters, between x and y. A feedforward neural network is composed of chaining together different functions. The most basic example contains three functions described to form the equation

$$\hat{f}(x) = \hat{f}^{(3)}(\hat{f}^{(2)}(\hat{f}^{(1)}(x))). \tag{2.2}$$

In this example, $\hat{f}^{(1)}$ is called the first layer, while the final layer $\hat{f}^{(3)}$ is called the output layer. The first layer is also known as the input layer since it starts the forward process of the network. Since the output of $\hat{f}^{(2)}$ is not shown as an output as seen in Figure 2.3, it is known as a hidden layer. These hidden layers transform a single layer perceptron into an MLP. The term "deep learning" comes from the notion of depth created in the network through multiple hidden layers. We can build upon this general framework and explain the building blocks of a network's architecture.

2.5.2 Fully Connected Layers

In this section, we describe the most basic block of the MLP, namely the fully connected layer. The first layer contains the input x_i. The out of each layer k is defined by the transformation $h_k = g(W_k h_{k-1} + b_k)$, where W_k is a weight matrix of trainable parameters corresponding to the number of neurons in that layer, b_k is a bias term, g is an activation function adding non-linearity to the output of the layer and h_{k-1} is the input layer. Consider the network in Figure 2.3, which can be expressed by the following equations:

$$\begin{aligned} h_1 &= g(W_1 x_i + b_1) \\ h_2 &= g(W_2 h_1 + b_2) \\ \hat{y}_I &= f(W_o h_2 + b_o). \end{aligned} \tag{2.3}$$

Note that g is the activation function for the hidden layers and f represents the activation function in the output layer.

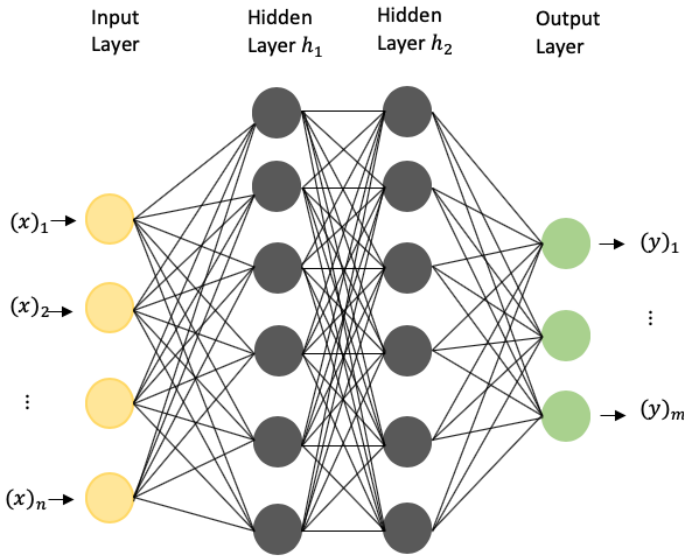


Figure 2.3: Mulit-layer fully connected network

Activation Functions

When an activation function is applied to the output of a linear transformation, it yields a nonlinear transformation to the output of that layer. It is important to note that the same activation function is used for multiple layers. Typically, the same one is used for all of the hidden layers and the output layer activation function is different than that of the hidden layers. The most commonly used activation function is the rectified linear unit (ReLU) defined by the activation function $g(z) = \max\{0, z\}$ shown in Figure 2.4. However, this function is not the only one used to build neural networks. Before using ReLU, most neural networks use the logistic sigmoid activation function (figure 2.5(a)) or the hyperbolic tangent activation function (Figure 2.5(b)) (see [23]).

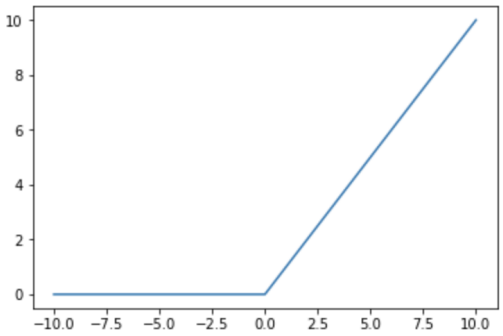
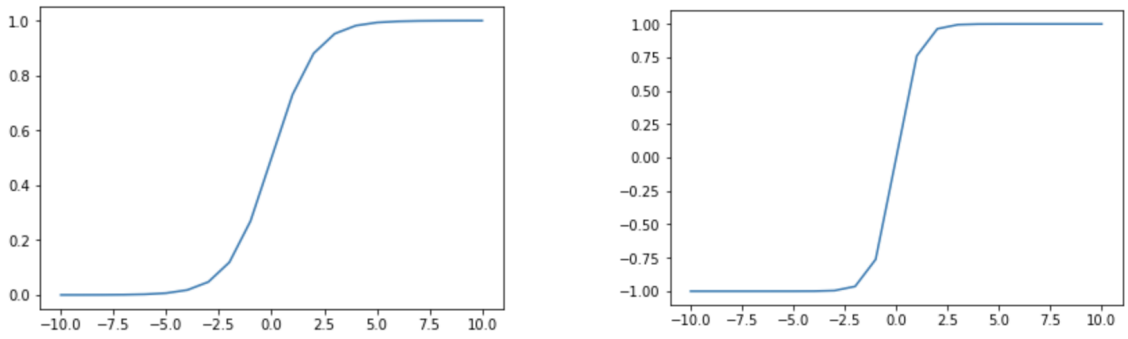


Figure 2.4: Behavior of recitifed linear unit (ReLU) in neural networks.



(a) Sigmoid

(b) Hyperbolic tangent

Figure 2.5: Behavior of activation functions used in neural networks.

2.5.3 Convolutional Layers

Convolutional layers are considered as an alternative to fully connected layers. Networks that have at least one convolutional layer as known as convolutional neural networks (CNNs). The input I of each convolutional layer is processed by a convolution. A convolution is a linear operation that involves the multiplication of a set of weights called a filter or a kernel K. The output of the convolution is a feature map S. Once we create the feature map, an element-wise activation function is applied for the outputs of the fully connected layer. Sometimes a filter does not perfectly fit the input image and must using padding, which incorporates additional pixels to the boundary of the images according to the desired kernel size. Note that stride is the number of pixels shifts over the input matrix and is used with the filter to create the feature map.

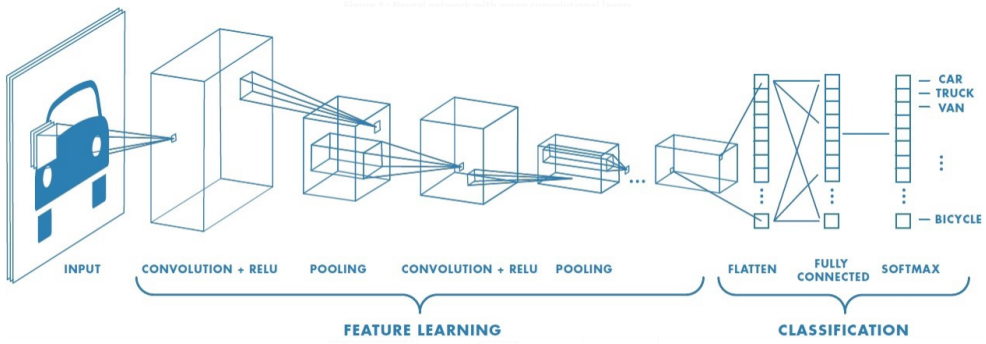


Figure 2.6: Illustration of an example of a neural network with multiple convolutional layers and fully connected layers. Each input image pass through convolutional layers with filters, pooling, fully connected layers and an activation function to classify an object to a certain class.

One major advantage of CNNs over fully connected layers are the decrease in the number of trainable parameters. It is important to note that CNNs typically still include fully connected layers into their architecture. This is a necessary component when using CNNs for classification in that the feature map of the last convolutional layer is vectorized and processed through fully connected layers and an activation function to be able to arrive at the probability that the input belongs to a certain class (see Figure 2.6 [24]).

Max Pooling

Convolutional layers and fully connected layers are two examples of mappings that provide a transformation from the input layer to the output layer. One major difference is that convolutional layer incorporates a pooling layer to the resultant feature map. The pooling layer reduces the number of parameters when the images are too large. For example, max pooling takes the largest element within a certain window from the rectified feature map (see Figure 2.7). This can also be viewed as a form of downsampling since it reduces the dimensionality of each map, but retains important information. The reason for this layer is for the output to be invariant or less sensitive to small translations in the same type of data [23].

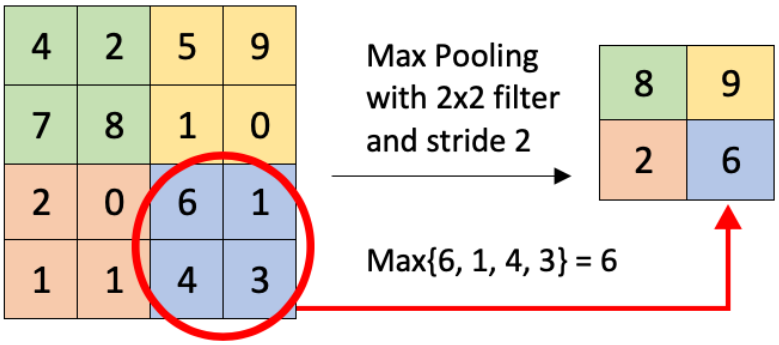


Figure 2.7: Max pooling layer applied to the output of the convolutional layer. The filter is a 2×2 filter with a stride of 2.

2.5.4 Cost Functions

When training a neural network, we use an objective function to update the parameters of the network. In a minimization problem, the objective function is also known as the cost function or loss function. The cost function is part of a feedback loop that reports how well the model is performing during the learning process in order for us to minimize the error. Since we are studying a classification problem, we will choose the binary cross entropy function as the cost function.

Cross Entropy Loss Function

Consider a binary classification problem where given information is distinguished between classes. Then, the cost function over m instances, known as the binary cross entropy function is

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y_i \log(p(x_i)) + (1 - y_i) \cdot \log(1 - p(x_i)), \tag{2.4}$$

where θ is the trainable parameters and $p(x_i)$ is the probability of y_i for the positive class ($y = 1$) given the input x_i and θ [21].

Now, let us consider problems with multiple classes. For each case, the model computes the score x for each class k. Then, we estimate the probability $\hat{p}_k$ that the instance belongs to class k by running the score through the softmax function

$$\hat{p}_k = \frac{\exp(x_k)}{\sum_{j=1}^{K} \exp(x_j)} \tag{2.5}$$

The function $J(\theta)$ in (2.4) computes the exponential of every score and then normalizes them. This allows the probabilities $\hat{p}$ to add up to one. Thus, we get the cross entropy cost function as

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} (y_i)_k \log((\hat{p}_i)_k), \tag{2.6}$$

where $(y_i)_k$ is the k-th component of the i-th instance and the target probability (either equal to 1 or 0 if it belongs to the class k) [21]. The goal is to minimize the cost function in (2.4) and (2.6) over the parameters of the network to improve classification.

Gradient-Based Learning

The purpose of gradient-based learning algorithms is to update the parameters of the neural network based on the gradient of the cost function $J(\theta)$. Gradient descent computes the gradient of the cost function with respect to the parameters $\theta = \{W_k, b_k\}_{k=1}^{m}$, where we recall that W_k is a weight matrix and b_k is the bias term, for the entire training set using the following update

$$\theta = \theta - \epsilon \nabla_\theta J(\theta) \tag{2.7}$$

where ϵ is the learning rate that defines the step size. Gradient descent terminates when every element of the gradient is sufficiently close to zero or when the loss function no longer changes.

In deep learning, the most typical optimization algorithm used is stochastic gradient descent (SGD). It is a version of gradient descent that allows for a faster gradient update by randomly selecting training set minibatches and updating the parameters based on the gradient computed in those minibatches.

2.5.5 Optimizer

We introduced the parameter known as the learning rate. Defining the learning rate appropriately is an important tool in the neural network. For instance, if the learning rate is too small, then the model is notably slow to learn, while a learning rate that is too large can lead the model to learn less information. Since the learning rate significantly affects the model performance, we will briefly review one of the stochastic gradient descent optimization algorithms.

2.5.5.1 Adam Optimization Algorithm

Adaptive moment estimation, also known as Adam, is an algorithm that computes the adaptive learning rates for each parameter [25]. The algorithm is seen as an update to the RMSProp [26]. Adam computes the decaying averages of past gradients m_t and past squared gradients v_t as

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2, \end{aligned} \tag{2.8}$$

where β_1 and β_2 are the decay rates and m_t and v_t are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients g_t, respectively. Adam then includes bias corrected first and second moment estimates to account for their initialization at the origin as

$$\begin{aligned} \hat{m}_t &= \frac{m_t}{1-\beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1-\beta_2^t} \end{aligned} \tag{2.9}$$

Just like with RMSProp, (2.7) and (2.8) are used to update the parameters and yield the Adam update rule. Some advantages of Adam, include it does not require a stationary objective, it works with sparse gradients and the magnitudes of parameter updates are invariant to rescaling of the gradient.

2.5.6 Regularization

Ideally, we want neural networks to perform well on the training data as well as unseen data. However, in some situations overfitting can occur. The strategies for reducing the test error, even at the risk of increasing the training error, are collectively known as regularization. We will discuss one regularization technique known as dropout.

Dropout

When training large neural networks, all the weights learn together and over many iterations some of the weaker connections are ignored against the stronger ones. This

concept is known as co-adaption. Dropout [27] is a computationally inexpensive technique used to address the issue of overfitting by preventing co-adaptions from occurring on the training data. While using dropout, a number of random units and their connections are dropped from the neural network during training as seen in Figure 2.8. Thus, this method reduces overfitting by training a number of "thinned" models and averaging their outputs. During backpropagation, only the units that are not dropped are updated. It is important to note one drawback of dropout is the increased amount of training time needed compared to a standard neural network of the same architecture.

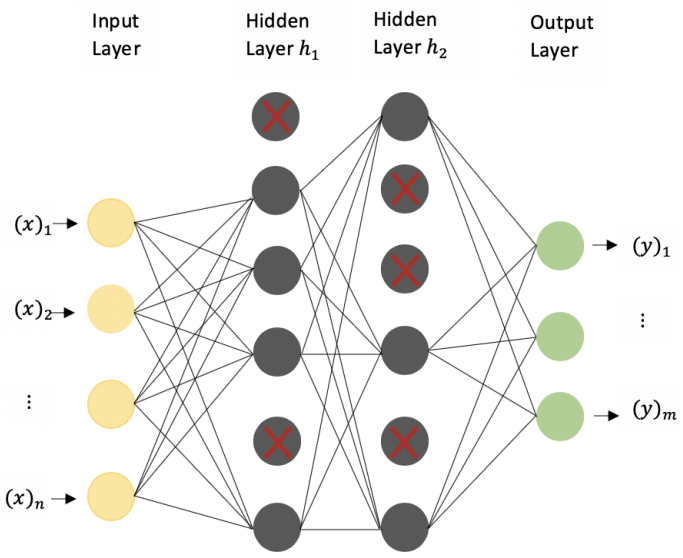


Figure 2.8: Thinned neural network with cross units dropped during training

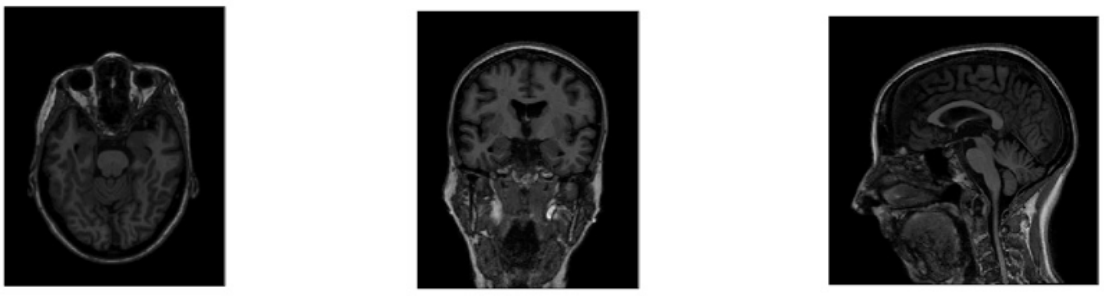
Chapter 3

MRI Data

In the following chapter, we describe the preprocessing applied to the MRI images downloaded from the ADNI database. We also discuss the additional steps and patient information to complete the training dataset.

3.1 Preprocessing

The format of the downloaded MRI images from the ADNI database are the Neuroimaging Informatics Technology Initiative (NIfTI) files. It is important to note that the data image types downloaded are semi-processed, meaning that the images are aligned and centered and not the original raw MRI data. This is important because the additional preprocessing in this study will not be possible on the original data since they are not properly aligned. The MRI can be viewed as two dimensional orthogonal projections of the brain (i.e. coronal, sagittal and axial (Figure 3.1)). A software known as Statistical Parametric Mapping (SPM) was used for the preprocessing of the neuroimages. For this part of the preprocessing, the most current version of SPM12 in unison with MATLAB(2019b) was used. The three main steps of the preprocessing are segmentation, normalization and finally smoothing. The explanations described in this section can be found in [28].



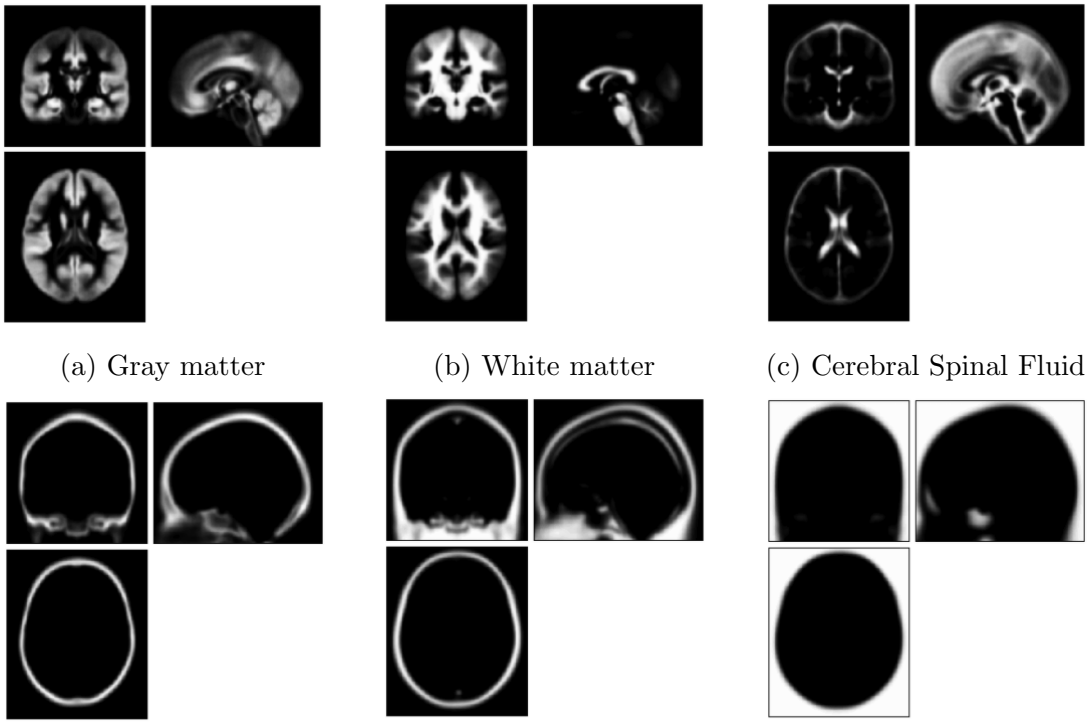
(a) Axial plane (b) Coronal plane (c) Sagittal plane

Figure 3.1: Two dimensional orthogonal projection of MRI images before preprocesed.

3.1.1 Segmentation

The first part of the preprocessing is known as segmentation. The 3D MRI images can be classified into different tissues types. The tissue types are defined based on the tissue probability maps (TPM) provided by SPM12. This can be found in SPM - tpm/TPM.nii, which is a multivolume NIfTI file (one volume for each of the 6 tissues classes).The TPM reflect the probability of a voxel, which represents a value on a regular grid in three dimensional space, belonging to each tissue class based on the segmentation of a large number of young adult brains that have been normalized to standard space. The order of the tissue is the following: grey matter (GM), white matter (WM), cerebral spinal fluid (CSF), bone, soft tissues and air/background (Figure 3.2). For this study we will focus on gray matter (GM). After segmentation, a native-space image is produced that reflect the voxel's probability of belonging to the gray matter tissue class (Figure 3.3(a)).

In order to complete segmentation through SPM12, the bias regularization is set on the light regularization (0.001), the bias full width at half maximum (FWHM) is set on the 60 mm cutoff and the affine regularization on the International Consortium of Brain Mapping (ICBM) space template. Lastly, for the spatial normalization of the data to the Montreal Neurological Institute (MNI) spaces, the deformation field was set in the forwarding mode.



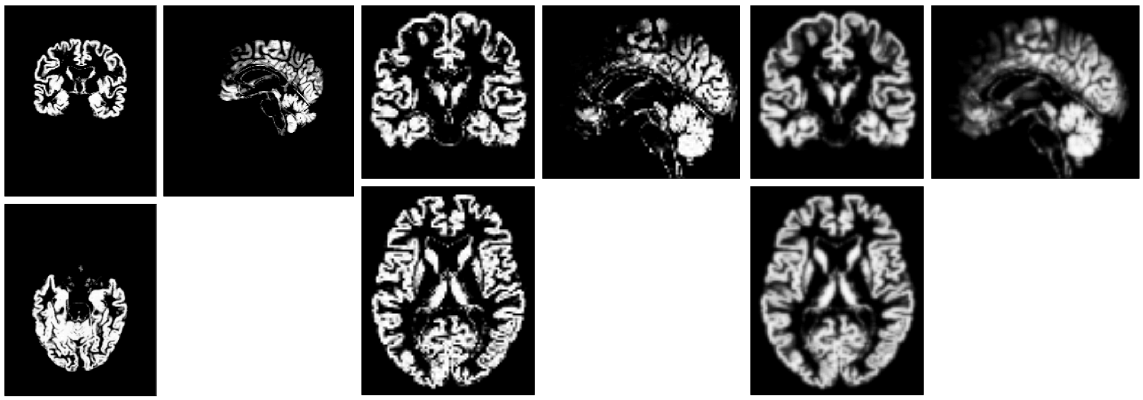
(a) Gray matter (b) White matter (c) Cerebral Spinal Fluid

(d) Bone (e) Soft Tissue (f) Air/Background

Figure 3.2: Tissue probability maps (TPM) - Source: SPM12.

3.1.2 Normalization

After segmentation, the GM images were further analyzed with the next step, normalization as seen in Figure 3.3(b). Before SPM12, spatial normalization was based on minimizing the mean squared difference between a template and a warped version of the image. Now, spatial normalization involves warping all the segmented images to the same space, which is achieved by matching to a common template. The TPM defines the space that the segmented images will be warped to. This process produces a deformation field image file (y_*.nii), which records the non-linear transformation between spaces. During the non-linear transformation, it is calculated how the image should move and shrink or expand to fit the template. Thus, the deformation field contains three image volumes encoding the x, y, and z coordinates (in mm) of where each voxel maps to in the standard space. In order to normalize the GM images to MNI space, we set the written normalized images voxel size on (2 2 2) mm and the interpolation to the 4th Degree B-Spline.



(a) Segmentation (b) Normalization (c) Smoothing

Figure 3.3: Pre-processing phases

3.1.3 Smoothing

The final preprocessing steps is smoothing (Figure 3.3(c)), which is completed in order to increase signal to noise ratio (SNR) and the ability of statistical techniques to detect true and task related changes in the signal. All normalized GM images were set with the Gaussian smoothing kernel set to (2 2 2) mm.

The original size of the data was 176 x 240 x 256. After segmentation the dimensions of the image remain the same. Then, after the normalization process, all the GM images were reduced to the size of 79 x 95 x 79.

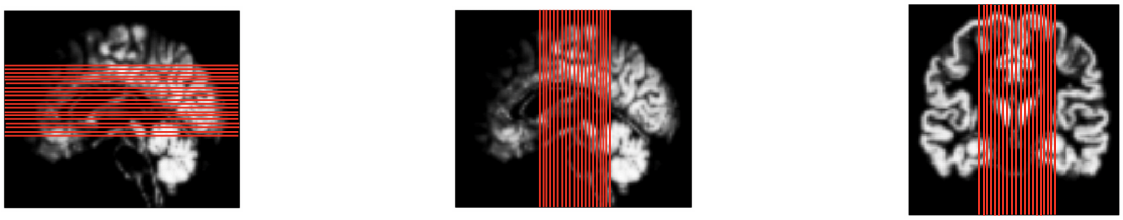
3.2 2D MRI Data

Now, the preprocessed 3D MRI NIfTI images are sliced into 2D portable network graphics (PNG) images for the model using MATLAB (2019b). We sliced the 3D MRI images into 2D MRI images along the three planes: axial, coronal and sagittal planes (see Figure 4.1)). Then, we resize them to 64×64 pixels to be used for the convolutional neural network (CNN).

(a) Axial plane (b) Coronal plane (c) Sagittal plane

Figure 3.4: Two dimensional orthogonal projection of MRI images after preprocesed.

Depending on the anatomical plane, we consider what we believe is the appropriate range of slices of the brain for the model. For example, in the axial plane, the top and bottom portion of the brain will not show a significant difference between healthy brains and those with MCI. Also, in the coronal plane and sagittal plane, the same consideration is applied for the front/back and left/right sides of the brain, respectively. Since we want to study the classification of MCI under data limitations, we choose a small range of images per plane. Therefore, we selected 20 images of each plane; a total of 60 images were considered per subject. The respective slices for each plane are shown in Figure 4.2 through the various anatomical planes. We make the assumption that this range of images will allow us to appropriately differentiate between CN, EMCI and LMCI.



(a) Axial slices - Sagittal plane (b) Coronal slices - Sagittal plane (c) Sagittal slices - Coronal plane

Figure 3.5: MRI images with slices.

3.3 Patient Data

For this study, we obtained a total of 516 subjects: 172 EMCI, 172 LMCI and 172 control normal (CN) subjects. The demographic information of all the subjects of the three groups are shown in Table 4.1, where N represents the total number of subjects and F and M are the number of females and males, respectively.

Table 3.1: Demographic charateristic of the subjects.

	CN (N = 172) 90 F/ 82 M		EMCI (N=172) 76 F/ 96 M		LMCI (N=172) 77 F/ 95 M	
	Mean	SD	Mean	SD	Mean	SD
Age	76.1	6.9	71.3	7.7	72.3	7.5

Then, 10,320 images were used from each of the three planes: axial, coronal and sagittal from each group (Control normal, Early MCI and Late MCI) for a total of 36,000 images used in the study. Additionally, the images were separated as 50% images used for training, 30% images used for testing and 20% images used for the validation set.

Chapter 4

Numerical Experiments

For this project, we performed a binary classification between Control Normal (CN), Early MCI (EMCI) and Late MCI (LMCI) within each of the axial, coronal and sagittal planes. We used two different models described in the following sections. We will also discuss the performance measurements used to analyze the results and compare the models.

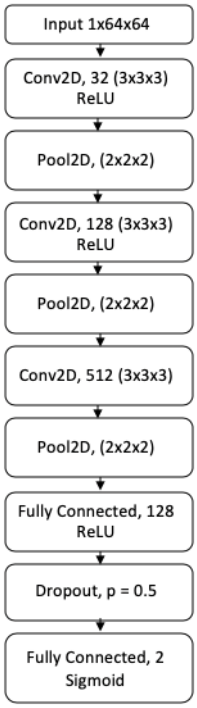
4.1 Model I: Simple CNN

The first model we will be considering is a variation of the model found in [29]. It is composed of three convolutional layers with max pooling between each layer (see Figure 5.1a). For the first convolutional network, the 32 filters with a kernel size of 3 x 3 and a max-pooling kernel size set on 2×2 were considered. The second and third layer consists of 128 and 512 filters respectively, followed by the same max-pooling kernel size from the first layer. Throughout all the convolutional layers, ReLU was used as the activation function. Then, a fully connected network with 128 input neurons and a ReLU activation function was used. Finally, a sigmoid activation function was used to conduct the binary classification. The model is compiled with the Adam optimizer with a learning rate of 1×10^{-4}. A binary cross entropy function was used to measure the performance of the model. The batch size was set to 128 images and used 100 epochs for the CNN.

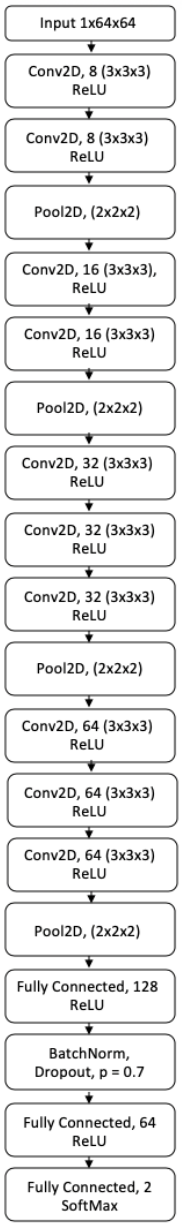
4.2 Model II: VoxCNN 2D

For the second model, we consider the VoxCNN in 2D, which is similar to the model found in [30]. This model is composed of four volumetric convolutional blocks of 8, 16, 32 and 64 filters respectively, followed by the same max pooling in between, along with ReLU used as an activation function (see Figure 5.1b). After, a fully connected network with 128 input neurons and a ReLU activation function were used. Then, a batch normalization followed by a drop out with a probability of

0.7 were used. Finally, a fully connected layer with 64 input neurons followed by a softmax activation function. In addition, we define a kernel initializer as the glorot uniform and a bias initializer of zeros for initializing the layer's weights throughout the network. This model was also compiled with the Adam optimizer using a learning rate of 2.7×10^{-6} and a binary cross entropy function to measure the performance. The batch size for this model is set to 5 for 50 epochs.



(a) Simple CNN



(b) VoxCNN 2D

Figure 4.1: Simple CNN and VoxCNN architectures

4.3 Experiments

In this section, we define the experiments that were conducted using the models described in the previous sections. Experiment I: the classification of Control Normal (CN) and Early MCI (EMCI) along the three different MRI planes: axial, coronal and sagittal. This first experiment classifies between the first two stages leading to dementia. A high performance for this experiment would be crucial for early detection of Alzheimer's disease (AD). Experiment II: the classification of Control Normal (CN) and Late MCI (LMCI) along the three different MRI planes. For this second experiment, there should be a bigger difference between the first stage of the healthy brains and the third stage of LMCI. Therefore, we should expect a noticeable different in the classification for this second experiment. Finally, Experiment III: the classification of Early MCI (EMCI) and Late MCI (LMCI) along the three different MRI planes. In this third experiment, we study the distinction between the characteristics of EMCI and LMCI.

4.4 Performance Measurements

The performance of each model will be evaluated using the following five metrics: accuracy, specificity, recall, precision and F-score. The parameters definitions are as followed:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times \ 100$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100$$

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100$$

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100$$

$$\text{F-score} = 2 \times \frac{\text{P} \times \text{R}}{\text{P} + \text{R}} \times 100$$

Experiment I is the binary classification of CN and EMCI, while Experiment II is the binary classification of CN and LMCI. Thus, for Experiment I and II, we define true positive (TP) as the number of EMCI or LMCI subjects who were correctly classified. Then, we define true negative (TN) as the number of healthy brains correctly classified as CN. The false positive (FP) is defined as the number of EMCI or LMCI classified as CN. Finally, the false negative (FN) is defined as the number of healthy brains classified as either EMCI or LMCI.

For Experiment III, we define true positive (TP) as the number of EMCI correctly classified as itself. Furthermore, true negative (TN) is defined as the number of LMCI

correctly classified. Then, false positive is the number of EMCI classified as LMCI and false negative is defined as the number of LMCI classified as EMCI.

Based on the described definitions above, accuracy represents the model's ability to correctly classify between all groups: CN, EMCI or LMCI for each experiment, respectively. The specificity measures the proportion of all negatives are correctly identified. Thus, for Experiment I and II, it reflects how many CN cases were classified correctly. Then, for Experiment III, it reflects the number of LMCI cases correctly identified. Furthermore, recall, also known as sensitivity, measures the proportion of all positives correctly identified. In Experiment I and II, it reflects how many EMCI or LMCI subjects were identified correctly, respectively. In Experiment III, it reflects how many EMCI subjects were correctly classified. The precision is the proportion of true positive results that were actually correct. Lastly, F-score is a measure of the test's accuracy and uses the average of precision and recall to accomplish this.

4.5 Results

For this project, we studied the classification of healthy brains and brains diagnosed with mild cognitive impairment (MCI) under a limited range of data. We performed binary classifications between the control normal (CN), Early MCI (EMCI) and Late MCI (LMCI). In this section, we evaluate the results of the performed experiments for the two proposed models.

Experiment I: Control Normal/Early MCI Classification

Table 4.1: Performance measure results for Control Normal/Early MCI classification using Simple CNN (Model I) and VoxCNN 2D (Model II) along three different MRI planes.

	MRI Views	**Accuracy** (%)	**Specificity** (%)	**Recall** (%)	**Precision** (%)	**F-Score** (%)
Model I	Axial	76.57	75.27	77.87	75.96	76.86
	Coronal	**78.90**	**79.30**	**78.49**	**79.39**	**78.77**
	Sagittal	77.07	79.05	75.10	78.30	76.58
Model II	Axial	65.85	66.72	64.98	66.47	65.45
	Coronal	**66.69**	66.30	**67.07**	66.61	**66.74**
	Sagittal	66.21	**75.15**	57.27	**70.36**	62.58

Experiment I classifies between CN and EMCI for all three planes: axial, coronal and sagittal. From Table 4.1, we can see that Model I outperforms Model II in all metrics along the three anatomical planes. Furthermore, in Model I, the coronal plane provides higher range for all five performance measures, while in Model II,

the coronal plane presents a higher range in only three out of the five metrics. The highest accuracy of 78.90% for Model I and 66.69% for Model II was achieved along the coronal plane.

We can further evaluate the proposed models considering the performance measures described in the previous section. The metrics amongst Model I are all fairly similar. Along the coronal plane, we have a highest specificity of 79.30%. The specificity reflects how many Control Normal cases were classified correctly. Therefore, the higher the specificity, the fewer normal subjects were classified as EMCI. We also have a high recall of 78.49% and a high precision of 79.39% for the coronal plane, which means our first model is good at detecting and discriminating the differences between healthy brains and EMCI.

For Model II, the performance measures are not as consistent as Model I. As stated above, only three out of the five metrics rank high for the coronal plane. Conversely, we have a higher precision of 70.36% for sagittal plane. Therefore, we obtained a high number of the false negatives, which means a greater number of healthy brains were incorrectly classified as EMCI. Likewise, we have a high specificity of 75.15% for the sagittal plane. Thus, a higher number of Control Normal brains were classified correctly. Although, the results indicate a high recall and F-Score for the coronal plane. Since we are classifying between the first two stages leading up to dementia, there is a noticeable distinction that the models learned in order to properly classify.

Experiment II: Control Normal/Late MCI Classification

Table 4.2: Performance measure results for Control Normal/Late MCI classification using Simple CNN (Model I) and VoxCNN 2D (Model II) along three different MRI planes.

	MRI Views	**Accuracy** (%)	**Specificity** (%)	**Recall** (%)	**Precision** (%)	**F-Score** (%)
Model I	Axial	77.17	80.04	74.30	78.84	76.48
	Coronal	**79.67**	**82.44**	**76.90**	**81.50**	**79.06**
	Sagittal	77.45	79.09	75.81	78.48	77.04
Model II	Axial	67.02	65.58	68.45	66.96	67.16
	Coronal	**67.85**	66.24	**69.47**	67.81	**68.12**
	Sagittal	67.41	**72.71**	62.11	**69.76**	65.38

Experiment II classifies between Control Norman (CN) and Late MCI (LMCI) along the three anatomical planes: axial, coronal and sagittal. Model I exceeds Model II for the second experiment as seen in Table 4.2. For Model I, the coronal plane shows a higher range in all five metrics. On the other hand, Model II has a higher range in the coronal plane for three out of five metrics, which is the same result as Experiment I. Additionally, the coronal plane also has higher accuracy compared to

the other anatomical planes. The classification of CN and LMCI for the coronal plane results in a high accuracy of 79.67% for Model I compared to the highest accuracy of 67.85% obtained for Model II.

Overall, the performance measures from Model I are similar. However, there are some metrics that are more distinct than others. The results indicate a higher specificity of 82.44% for the coronal plane. In this experiment, a higher specificity indicates that fewer healthy brains were misclassified as LMCI. On the other hand, a higher precision, for instance, in the coronal plane with 81.50%, shows that we were able to classify a greater amount of those labeled as positive. The increase in specificity and precision is attributed to a greater anatomical difference between the first stage of healthy brains to the third stage of Late MCI (LMCI).

We see an inconsistency in the performance measures for Model II compared to Model I. The classification of CN and LMCI for the sagittal plane shows a high specificity of 72.71%. In addition, we have a high precision of 69.76% for the sagittal plane. Compared to Experiment I, Model II shows similar results regarding a high specificity and precision for the sagittal plane.

Experiment III: Early MCI/Late MCI Classification

Table 4.3: Performance measure results for Early MCI/Late MCI classification using Simple CNN (Model I) and VoxCNN 2D (Model II) along three different MRI planes.

	MRI Views	**Accuracy** (%)	**Specificity** (%)	**Recall** (%)	**Precision** (%)	**F-Score** (%)
Model I	Axial	74.36	74.07	74.65	74.27	74.43
	Coronal	**77.02**	**76.47**	77.56	**76.86**	**77.10**
	Sagittal	76.82	75.87	**77.77**	76.49	76.99
Model II	Axial	62.41	61.69	63.14	62.30	62.53
	Coronal	**64.76**	**63.41**	**66.10**	**64.64**	**64.99**
	Sagittal	63.35	62.31	64.40	63.31	63.51

Lastly, Experiment III classifies between EMCI and LMCI along the three anatomical planes: axial, coronal and sagittal. Just like Experiment I and II, Model I surpasses Model II for all metrics for all three planes as seen in Table 4.3. However, unlike the previous experiments, Model I has a higher range in the coronal plane for four out of five metrics, while Model II has a higher range in the coronal plane for all five metrics. Note that the coronal planes has a higher accuracy compared to the axial and sagittal planes for both models. The classification of EMCI and LMCI for the coronal plane shows a high accuracy of 77.02% and 64.76% for Model I and Model II, respectively.

For Model I, the performance measures are alike to each other. Yet, there are some metrics that more apparent than others. For instance, the classification of EMCI and

LMCI for the sagittal plane has a recall of 77.77%. A higher recall means fewer cases of LMCI were incorrectly classified compared to the EMCI cases. Contrarily, we have a high specificity, precision and F-score for the coronal plane. On the other hand, the results are more consistent for Model II. The classification of EMCI and LMCI for the coronal plane has a high recall of 66.1% for the coronal plane, where we were able to classify more LMCI cases versus the EMCI cases. We also have a high specificity, precision and F-score for the coronal plane. The EMCI and LMCI patients share similar characteristics, which is why the results for this experiment are more consistent between both models.

4.6 Discussion

In this chapter, we presented and analyzed the results of the experiments for two different models. Our work concentrates on the classification of Control Normal (CN), Early MCI and Late MCI with limited data range. Using the performance metrics above, we evaluated the proposed models. For our project, the results show that Simple CNN (Model I) model outperforms VoxCNN 2D (Model II) for each experiment.

Tables 4.1 and 4.2 show a noticeable difference within the accuracies between individuals with healthy brains and those diagnosed with MCI. Additionally, the accuracies for the classification of EMCI and LMCI for all three anatomical planes show considerably lower performance measures.

The coronal plane has the highest accuracy between the axial and sagittal planes in all three experiments for both Model I and Model II, while the axial plane had the lowest percentages for the performance measurements out of the three anatomical planes. This tells us that there was less information for the classification of each brain present in the axial slices compared to the coronal and sagittal planes in these experiments.

Recall, that Model I has three convolutional layer and two fully connected layers, while Model II has four convolutional blocks and three fully connected layers. A CNN is used to increase the performance of a model by extracting discriminative features through images. Given the promising results obtained from the Simple CNN, we had hoped that the VoxCNN would have a better performance through consecutive convolutional layers. However, the results showed us that was not the outcome possibly due to the limited data range. Overall, we obtain lower percentages for the performance measurements of Model II compared to the results of Model I.

Chapter 5

Conclusion

The study of mild cognitive impairment (MCI) is critical for the early diagnosis of Alzheimer's disease (AD). Therefore, an accurate and reliable diagnosis of MCI will aid in identifying those individuals at an increased risk of the progression to dementia. In recent years, deep learning has contributed to solving such complex problems. Thus, a convolutional neural network (CNN) can provide important information to classify between CN, EMCI and LMCI patients.

In this thesis, we applied two different classification methods to investigate the performance of 2D MRI images under data quantity limitations. Additionally, we studied the binary classifications of CN, EMCI and LMCI with regard to the three anatomical planes: axial, coronal and sagittal. We performed a thorough analysis to gain insight into what the proposed models had learned. The best results were achieved for the classification of CN and LMCI for the coronal plane with a 79.67% accuracy. Overall the results indicate that the simpler CNN architecture outperforms a more sophisticated CNN under a limited dataset.

We performed the experiments with the assumptions that the dataset acquired characterized a noticeable difference between the baseline of healthy brains and brains diagnosed with mild cognitive impairment (MCI). We would have like to perform additional experiments, such as a classification of each group (CN, EMCI and LMCI) containing all three anatomical planes. Further experimentation could include an increase of the slices taken within each image. Unfortunately, due to computational limitations, we were unable to complete these experiments.

This work contributes to the study of two deep learning frameworks that explores the possibility of gray matter extracted from MRI images of the axial, coronal and sagittal planes for MCI diagnosis. Then, we evaluated our work with a relatively small dataset and achieve promising results. Furthermore, we exhibit the effective increase in classification performance with the simple CNN architecture consisting of three convolutional layers. In future work, it would useful to study different ranges of images selected for each anatomical plane. Finally, we aim to continue this study with a classification of CN, EMCI and LMCI with the combination of all three anatomical planes.

References

[1] A. Association, "2020 Alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 16, no. 3, pp. 391–460, 2020. vii, 2

[2] "ADNI : Study Design," 2017. vii, 4

[3] A. Association et al., "2018 alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 14, no. 3, pp. 367–429, 2018. 1

[4] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans, "Alzheimer disease in the United States (2010–2050) estimated using the 2010 census," Neurology, vol. 80, no. 19, pp. 1778–1783, 2013. 2

[5] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr, J. Kaye, T. J. Montine, et al., "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," Alzheimer's & dementia, vol. 7, no. 3, pp. 280–292, 2011. 2, 3

[6] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, et al., "The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," Focus, vol. 11, no. 1, pp. 96–106, 2013. 2, 3

[7] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, et al., "The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," Alzheimer's & dementia, vol. 7, no. 3, pp. 263–269, 2011. 2, 3

[8] C. R. Jack Jr, M. S. Albert, D. S. Knopman, G. M. McKhann, R. A. Sperling, M. C. Carrillo, B. Thies, and C. H. Phelps, "Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups

on diagnostic guidelines for Alzheimer's disease," Alzheimer's & Dementia, vol. 7, no. 3, pp. 257–262, 2011. 2

[9] L. Vermunt, S. A. Sikkes, A. Van Den Hout, R. Handels, I. Bos, W. M. Van Der Flier, S. Kern, P.-J. Ousset, P. Maruff, I. Skoog, et al., "Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype," Alzheimer's & Dementia, vol. 15, no. 7, pp. 888–898, 2019. 2

[10] R. C. Petersen, O. Lopez, M. J. Armstrong, T. S. Getchius, M. Ganguli, D. Gloss, G. S. Gronseth, D. Marson, T. Pringsheim, G. S. Day, et al., "Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology," Neurology, vol. 90, no. 3, pp. 126–135, 2018. 3

[11] A. Ward, S. Tardiff, C. Dye, and H. M. Arrighi, "Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature," Dementia and geriatric cognitive disorders extra, vol. 3, no. 1, pp. 320–332, 2013. 3

[12] A. J. Mitchell and M. Shiri-Feshki, "Rate of progression of mild cognitive impairment to dementia–meta-analysis of 41 robust inception cohort studies," Acta Psychiatrica Scandinavica, vol. 119, no. 4, pp. 252–265, 2009. 3

[13] P. S. Aisen, R. C. Petersen, M. C. Donohue, A. Gamst, R. Raman, R. G. Thomas, S. Walter, J. Q. Trojanowski, L. M. Shaw, L. A. Beckett, et al., "Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans," Alzheimer's & Dementia, vol. 6, no. 3, pp. 239–246, 2010. 4

[14] "Magnetic Resonance Imaging (MRI)," Dec 2019. 5

[15] Y. Zhang, N. Schuff, M. Camacho, L. L. Chao, T. P. Fletcher, K. Yaffe, S. C. Woolley, C. Madison, H. J. Rosen, B. L. Miller, et al., "MRI markers for mild cognitive impairment: comparisons between white matter integrity and gray matter volume measurements," PloS one, vol. 8, no. 6, 2013. 5

[16] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," Human brain mapping, vol. 2, no. 4, pp. 189–210, 1994. 5

[17] G. Flandin and K. J. Friston, "Statistical parametric mapping (SPM)," Scholarpedia, vol. 3, no. 4, 2008. 5

[18] J.-B. Poline and M. Brett, "The general linear model and fMRI: does love last forever?," Neuroimage, vol. 62, no. 2, pp. 871–880, 2012. 5

[19] G. Chételat, B. Desgranges, V. De La Sayette, F. Viader, F. Eustache, and J.-C. Baron, "Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment," Neuroreport, vol. 13, no. 15, pp. 1939–1943, 2002. 5

[20] S. Xie, J. Xiao, G. Gong, Y. Zang, Y. Wang, H. Wu, and X. Jiang, "Voxel-based detection of white matter abnormalities in mild alzheimer disease," Neurology, vol. 66, no. 12, pp. 1845–1849, 2006. 5

[21] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019. 6, 10

[22] J. Friedman, T. Hastie, and R. Tibshirani, The Elements of Statistical Learning, vol. 1. Springer series in statistics New York, 2001. 6

[23] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016. 6, 7, 9

[24] MathWorks, "Introduction to Deep Learning: What Are Convolutional Neural Networks?," 2017. 9

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014. 11

[26] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," COURSERA: Neural networks for machine learning, vol. 4, no. 2, pp. 26–31, 2012. 11

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The journal of machine learning research, vol. 15, no. 1, pp. 1929–1958, 2014. 12

[28] J. Ashburner, G. Barnes, C. Chen, J. Daunizeau, G. Flandin, K. Friston, S. Kiebel, J. Kilner, V. Litvak, R. Moran, et al., "SPM12 manual," Wellcome Trust Centre for Neuroimaging, London, UK, p. 2464, 2014. 13

[29] H. Taheri Gorji and N. Kaabouch, "A deep learning approach for diagnosis of mild cognitive impairment based on mri images," Brain sciences, vol. 9, no. 9, p. 217, 2019. 18

[30] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 835–838, IEEE, 2017. 18