DEVELOPMENT OF A NEURAL NETWORK MODEL FOR CLASSIFYING ALZHEIMER'S DISEASE

BY

OBUNABOR CHIDERA WINNIE
(16CG021439)

A PROJECT SUBMITTED TO DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY OTA, OGUN STATE

IN PARTIAL FULFILMENT FOR THE AWARD OF THE BACHELOR OF SCIENCE (HONOURS) DEGREE IN COMPUTER SCIENCE

AUGUST 2020

CERTIFICATION

I hereby certify that this project, was carried out by Obunabor Chidera Winnie in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ogun State, Nigeria, under my supervision.

1. Name: Professor Victor Osamor
(**Supervisor)**

Signature _____ Date _____

2. Name: Prof. Ambrose A. Azeta
(**Head of Department)**

Signature _____ Date _____

DEDICATION

I would like to dedicate this research project to God by whose grace I have been able to successfully do all I needed for this project to come to a completion. And also to my family who all have believed and supported me in every way possible and to all my lecturers who were committed to teaching me and seeing to it that my work come to life.

ACKNOWLEDGEMENT

I want to appreciate my parents, Mr. and Mrs. Obunabor for giving me their support and encouragement. I also appreciate my supervisor Professor Victor Chukwudi Osamor for his assistance and supervision in achieving this project. I would also like to acknowledge my mentor and colleague, Kamorudeen Akindele and Echubiojo Udale-Ameh, who have constantly pushed me to be better in every way.

ABSTRACT

In the world as we know it today, there are so many emphasis on improving the quality of life and standard of living, so much so that we have everything being automated and programmed to function without human operation. Technology has grown so much and has impacted every aspect of our lives. Although this is true, there are still some gray areas in which we lack in this aspect. This work highlights the development of a model tailored for detecting Alzheimer's disease in brain MRI images and classifying images of Alzheimer's disease patients from normal healthy patients. The model collects image datasets of both Alzheimer's disease patients and control patients and learns some pattern from these images. The model then classifies the images based on patterns it learned on its own (unsupervised learning) and is able to detect which images belong to patients who suffer the disease and patients who are non-demented. This study also focused on exploring several features that are common to Alzheimer's disease patients to better understand indicative signs that a patient might have, or is likely to develop Alzheimer's disease. The model explores several supervised learning models as well with the use of a second dataset that is labeled. This is to help data scientist and research engineers understand what model will have a better chance at predicting the presence of Alzheimer's disease given several other related features as well as help doctors and health professionals understand what the best and most important areas to focus on when diagnosing Alzheimer's disease in the future

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND INFORMATION

There has been a lot of technological advancement in the last few decades. A lot of these technological emergences have paved the way for feats that could not have been imagined few years ago. This technology is able to detect the presence of select biomarkers, indicators of the presence of a specific disease or condition within the body.

The Alzheimer's disease which is a progressive brain disorder is the most common case of mental illness as it causes death of nerve cells and tissue loss throughout the brain, thus reducing the brain volume dramatically and affecting most of its functions. (Hosseini-Asl, Gimel'farb, & El-Baz, 2016)

This mental decline resulting from this disorder leads to dementia. For instance, the disease begins mild and gets progressively worse. (Sarraf & Tofighi, 2016)

According to NHS, this disease is common in people over the age of 65.

Alzheimer's disease may have symptoms like confusion, disorientation, difficulty making decision, problem with speech and language, problem performing self-care tasks, personality changes, aggressiveness, being paranoid, anxiety, etc., similar to other health problems like depression, chronic obstruction pulmonary disease, etc., which could disguise any suspicions of possible dementia.

Because of these effects of this disease, it is very important to begin prevention and treatment as early as possible. However, mental decline in Alzheimer's dementia is known to start showing signs at least 10 years after its onset.

Detecting this disease requires a history of patient's medical assessment and may require physical tests. Structural imaging based on magnetic resonance is an integral part of the clinical assessment of patients with suspected Alzheimer dementia. (Sarraf & Tofighi, 2016)

Nowadays, it is not reliable to depend on cognitive tests for detecting and identifying this disease progression. Thus, neuroimaging techniques such as brain Magnetic Resonance Imaging (MRI) scans can be effective in identifying patients with early signs of dementia, even before clinical symptoms begin to show.

In the serious stages of Alzheimer's dementia, severe shrinkage of the brain occurs. Although, this is not an accurate measure of Alzheimer's disease.

A machine learning algorithm such as deep learning, can classify and properly diagnose the disease in time (Sarraf, DeSouza, Anderson, Tofighi, & Initiativ, 2017)
It involves the use of a deep 3D convolutional neural network to learn generic features of this dementia. This identifies biomarkers for the disease and can be used in different domains. (Hosseini-Asl et al., 2016)

1.2 MOTIVATION

Is timely diagnosis of Alzheimer's dementia so important?

- Receiving a timely indication of Alzheimer's disease may help reduce anxiety about why you may be experiencing symptoms (Alzheimer's Association, 2019)
- Helps reduce the psychological risks associated with the more serious stages of the disease which includes depression, hallucination, and paranoia, disconnection from reality and worse, death.
- It will help victims plan ahead and express wishes about legal, financial and end-life disease (Alzheimer's Association, 2019)
- It will help reduce the number of people that end up dying from the disease

This project aims to solve this problem by implementing more advanced technology that outperforms regular cognitive assessment tests and is specifically engineered to detect the Alzheimer disease at its early stages using deep learning neural network trained prior to identify the disease and classify Alzheimer disease patients from normal patients through brain images.
The diagnosis of Alzheimer's disease, demands that i use computerized techniques to solve this problem by identifying variation changes in brain MRI scans such as severe shrinkage of brain volume which identify presence of Alzheimer's dementia.

1.3 STATEMENT OF THE PROBLEM

It is not reliable to depend on medical history and cognitive assessment of patients, since Alzheimer's disease begins to show signs 10 years after its onset, hence, in my solution, by using neural networks. I will create a model that will be able to spot the likelihood of Alzheimer's disease in brain Magnetic resonance images (MRI). This will help researchers for efficient AD diagnosis in the future.

1.4 AIM

The aim of my project is to develop a model for identifying and accurately classifying Alzheimer's dementia from patients without the disease, using Deep Artificial Neural Network.

1.5 OBJECTIVE

The objects for this project for achieving this aim are:

1. Study and analyze neuroimages of brain MRI scans
2. Build a sensitive model that can identify structural changes in the brain using deep neural networks
3. Effectively classify the diseased patients from normal healthy patients
4. Significantly prove the effectiveness and reliability of computerized diagnosis of Alzheimer's disease

1.6 METHODOLOGY

The methodology that is used in achieving the objectives for this project includes the following:

- **Literature Review:** Materials such as books. Research papers, conference proceedings, and articles were being read, studied and reviewed using academic platforms and consulting library resources.
- **Proof of Concept Implementation:**
    - Use deep learning concepts to build a model that can get information from neuroimages
    - Train my model to identify structural changes in brain MRI using neural networks
        - This method is used to create the sparse coding Algorithm that will help in feature extraction.
        - Sparse coding algorithm is an algorithm which extract features, learns from then and then reconstruct features. (Bhatkoti & Paul, 2016)
    - Test my model by classifying Alzheimer's disease patients' MRI from Normal aging patients' brain MRI
        - The research used MRI scan data for comparison with research images that were obtained from Kaggle, an open-access repository for datasets.
    - Evaluate the system
        - This will prove that the model does what it was intended for by accurately classifying Alzheimer's disease MRI from normal healthy patients.

1.7 DATA ACQUISITION

The dataset that is used was downloaded from an open access dataset repository on Kaggle. The major goal of Kaggle is to provide data scientist and researchers with accurate data used for their work. It dataset they provide is free and doesn't require legal permission to obtain.

1.8 SIGNIFICANCE OF THIS STUDY

My study will assist scientist and clinicians to effectively diagnose Alzheimer's disease at its early stages and will serve as a guide for providing means for proper administration of treatment. By means of deep neural network, it will be able to classify Alzheimer's disease patients from Normal aging patients and will give us information on the stage of dementia an Alzheimer's disease patient is experiencing. This study will help to prove the superiority of computerized medical diagnosis over traditional based cognitive assessment tests.

1.9 LIMITATIONS TO THE STUDY

Neural networks work really well on telling an image from another, or in this case detecting the presence of Alzheimer's disease from a brain MRI (magnetic resonance image). Although a challenge may arise when training your neural network. Underfitting or overfitting, may occurs (TORRES.AI, 2018). Overall, neural networks are effective, but not for everyone. Luckily, present solutions are being implemented to reduce this bottle neck of deep learning (Nicholson, 2019).

1.10 PROJECT OUTLINE

This project contains five chapters each containing different stages of the project. The first chapter contains an introduction to the project, statement of problem, aim, objectives, the significance of my study, methodology used, data acquisition and the project outline. The second chapter contains the summary of literature and consists of several journals, articles and researches related to this project. The third chapter focuses on system design and the details involved in it. The fourth chapter contains the implementation of the system, program modules, tools used and system development process. The fifth chapter contains the summary of the project, recommendation and conclusion.

CHAPTER TWO
LITERATURE REVIEW

2.1 INTRODUCTION

Alzheimer's disease is unfortunately beating the charts coming in sixth as one of the prominent causation of death in America. On the other hand, this may not be for long as it is estimated to even go up the charts as the second, after cancer (National Institute on Aging, 2019).

Alzheimer is a form of dementia and it is a disease provoked by abrasion to neurons and my result in memory loss, affect rumination and impinge regular cognitive conduct (Legg, 2018a, 2018b). It is a steady decline in thinking, reasoning, behavior and social skills resulting from the declination of brain tissues and neurons.

Reckoned with the variation of neural changes, motivation of this mental illness may vary. According to research done by Alzheimer's Association (2019), this disease makes up about 60 to 80 percent of dementia cases in America. Roughly 2/3 of the United States citizenry under the age of 65 have an immature latency of Alzheimer's.

Current medications could temporarily improve symptoms or slow decline progression. There has been several protocols delegated to the support of this disease and health care providers. In the eventual climax of the dementia, aggravation as a result of inexorable deficiency of mental activity could have a repercussion of death (Mayo Clinic, 2018).

2.2 SYMPTOMS, RISK FACTORS AND STAGES OF ALZHEIMER'S

2.2.1 SYMPTOMS FOR IDENTIFYING ALZHEIMER'S DISEASE

An individual experiencing Alzheimer's dementia could here after require complete support from another person. People who are diagnosed in Alzheimer's disease have shown to clearly exhibit these symptoms that indicate a cognitive decline in thinking, brain function, behavior and social skills. The following symptoms reveal that a person may have Alzheimer's disease.

A known early symptom of the disease manifests in strenuousness to recall information because the variation caused to the brain begins in the location that is commissioned to learn (Alzheimer's Association, 2019). According to Timothy J.Legg on Medical News Today (2018a, 2018b), the mental decline must be noticed in no less than two of the fields indexed beneath:

1. Laboriousness recalling things , resulting to:
    a. iterative questions

    b. Misplacing personal items
    c. Not remembering appointments
    d. Going astray on intimate detours
2. Degeneration affecting rationale:
    a. Not comprehending regulations
    b. Difficulty managing finance
    c. Poor decision making skills
    d. Difficulty making plans
3. Impaired vision:
    a. Finding it hard to recognize familiar faces
    b. Finding it hard to use simple tools
4. Degeneration associated with speaking, reading and writing:
    a. Difficulty in speech
5. Changes in ones behavior:
    a. Mood swings
    b. Loss of pity or care
    c. Compulsive behaviours

Other symptoms include:

- Mood Swings and crashes
- Fall in personal hygiene
- Dissocializing
- Poor judgments
- Trouble with regular tasks

(Legg, 2018a, 2018b)

Symptoms may also appear at different stages. These symptoms may help use identify what stage of Alzheimer's dementia that a patient has. The following stages may have accompanying symptoms:

- Mild
    - Going astray and mishandling finances, as well as lengthy periods on regular tasks.
- Moderate
    - Degeneration involving language, reasoning, sensory processing, and conscious thoughts and regular delusions.

- Severe
    - Here plaques are protein repose that appear around neurons and together with tangles, undoubtedly reduce neural tissues (National Institute on Aging, 2019)

2.2.2 ALZHEIMER'S DISEASE POSSIBLE RISK FACTORS

There are a couple of reasons why a person could have Alzheimer's disease. Reasons could be because of environmental factors, lifestyle, diet and genetics. According to Mayo Clinic, Alzheimer's may be provoked by unique sporogenous change able to guarantee that an individual may originate this disease. (Mayo Clinic, 2018)

Two proteins called plaques and tangles are believed to be primary killers of neurons.

1. Protein scraps, known as plaques are responsible for forming lumps around neural gaps
2. A second protein known as Tangle kills neurons by forming buildups in neurons

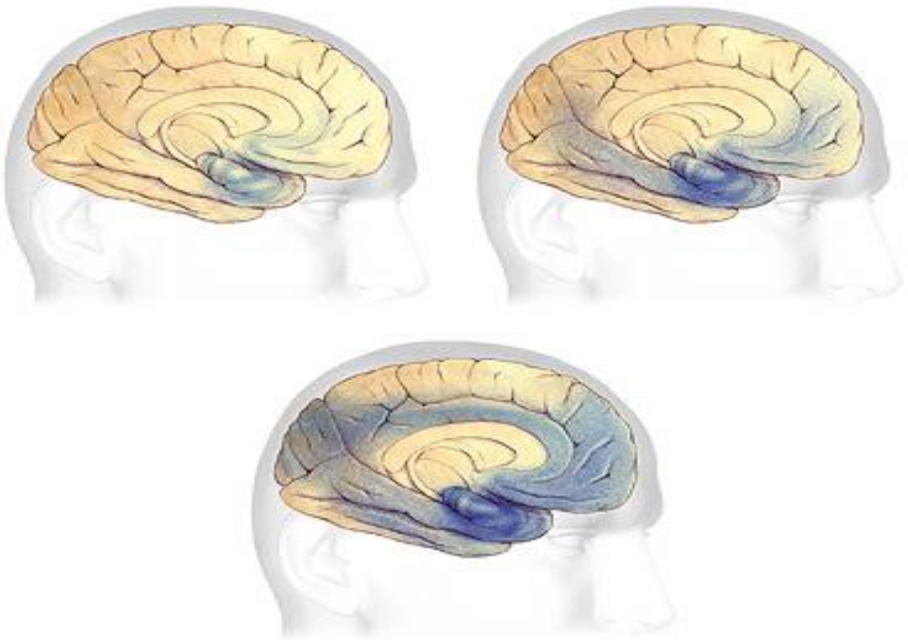

Figure 1: Showing the effect of plaques protein and tangles effect in the brain. (Source: www.alz.org)

Most scientists believe that plaques and tangles kill nerve cells. However, the deaths of nerve cells are responsible for the decline in cognitive function (Alzheimer's Association, 2019).

2.2.3 DEMENTIA IN STAGES

Dementias are neurodegenerative, what that implies is that its effects, destruction and even death of brain cells with happen gradually in stages. Alzheimer's have the following stages:

- **First:** obvious symptoms now start to show, however, early medical diagnosis could be given due to patients medical history.
- **Second**: at this stage, the most conspicuous signs start to appear forgetting information. These symptoms however could disguise any signs of possible dementia.
- **Third:** Reduced memory function and lack of concentration.
- **Fourth:** the diagnosis of the presence of Alzheimer's dementia could be made at this phase but it will be an immature stage of the disease.
- **Fifth:** slightly severe to very extreme cases will require the support of health professionals and caregivers.
- **Sixth:** now victim may need help with basic tasks which can be provided by loved ones
- **Seventh:** This stage eventually leads to death and may cause victim to lose facial expression(Legg, 2018a, 2018b)

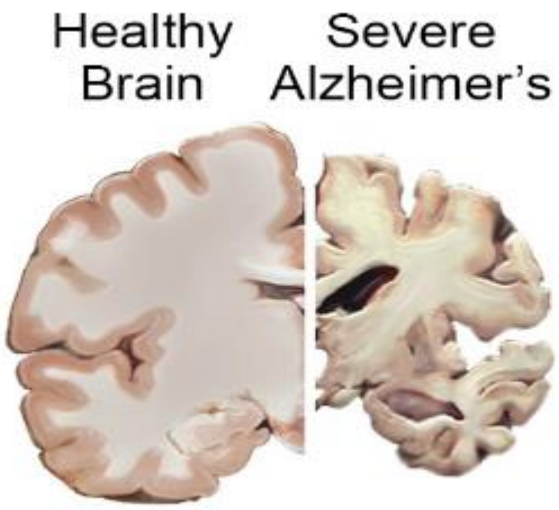

Figure 2: Showing brain image of a severe stage of Alzheimer's dementia *(*Source: www.nia.nih.gov*)*

2.3 ALZHEIMER'S: THE MOST EXPENSIVE DISEASE IN AMERICA

United States is known to be number three of the most populous counties in the world. It is estimated to have roughly about 330,149,796 occupants as at the 16th if December, 2019. The country has an alarming growth size of about 1.3% every year. It had a population size of 76 million in 1900 and has moved up to 281 million in 2000. (Wikipedia, n.d.).

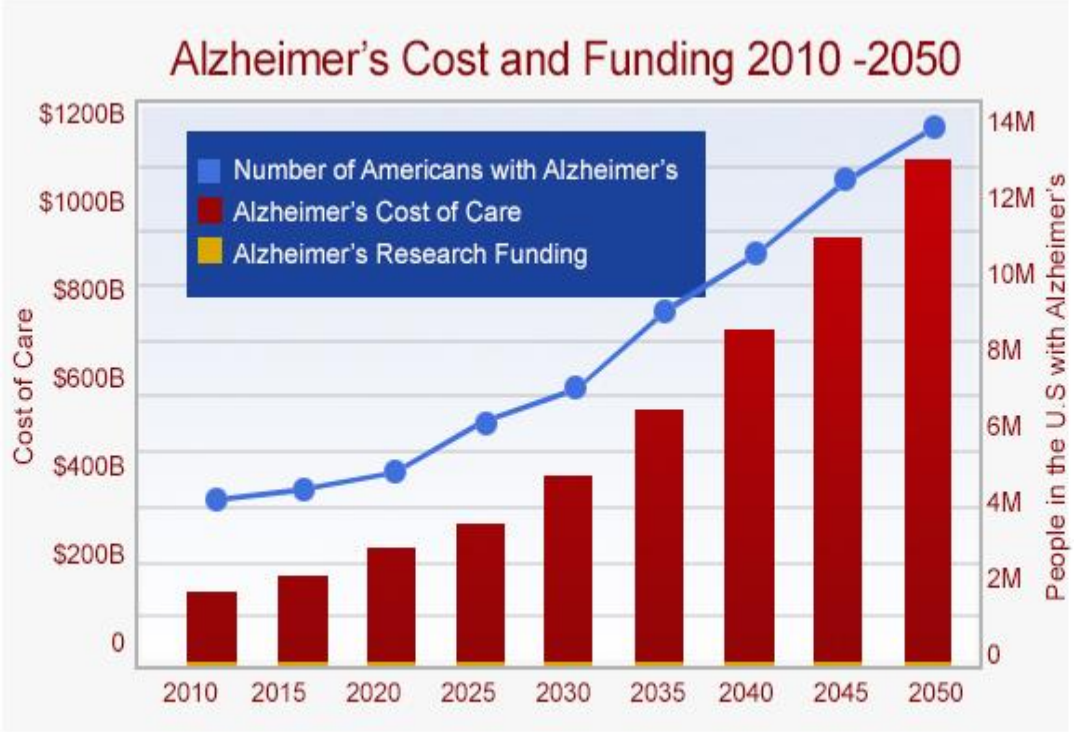

Figure 3: Showing the corresponding cost of Alzheimer's in the United States *(*Source: www.usagainstalzheimersaction.org*)*

As at Saturday, the 28th of December, based on Worldometer United Nations Data, the population of the United States was roughly at 329,064,917, which means that America's population equates to 4.27% of the the overall population of the world (Worldometers, 2019).

According to finding done by the 2019 Alzheimer's Facts and Figures Report, the summed cost of this disease in the United State in 2019 is estimated at $280 billion. 198 billion dollars is associated with Medicare while 85 billion dollars is associated with caregivers and loved ones. (Cure Alzheimer's Fund, 2020). According to the Associations, Alzheimer's disease costs about 341,000dollars for care of those experiencing this demntia and a lot of families and loves ones pay about 23,870 of that cost. According to Forbes (2018), Jody Gastfriend writes that about 18 million hours

of work time which amounts to roughly 232 billion dollars is spent taking care of a loved one experiencing this disease and caregivers have to sacrifice a lot of their time and freedom to help their loved one do daily tasks like eating and a lot more(Forbes, 2018). Many studies have shown that for these reasons, families can find themselves forced to cut back on expenditure or dip into saving in order to support members of their family suffering from this form of dementia (IOS Press, 2019)

The great news is that 15% of the total cost of care for the disease each year can be avoided by detecting it early. Overall, an early diagnosis can help families to prepare for what is to come and although this doesn't reverse the disease, current therapies can significantly improve brain activity in individuals suffering from this form of dementia. It is important to acknowledge the fact that diagnosing Alzheimer's dementia before symptoms begin to manifest will help save a lot of money and prevent other conditions that can further worsen the dementia and declining the person's quality of life.

2.4 PREVENTION AND TEST FOR ALZHEIMER'S DISEASE

2.4.1 PREVENTION OF ALZHEIMER'S DISEASE

Alzheimer's disease is not a preventable condition. There are claims that changes in diet and exercise may really make slim your chances of developing this disease in the later stages of life. They include:

- Exercising regularly
- Eating a proper balanced diet regularly

Studies have shown that this form of dementia has caused a fall in social activities and recreational activities and hubbies and reduces the used of brain power to do anything (Mayo Clinic, 2018).

2.4.2 DIAGNOSING ALZHEIMER'S DISEASE

There is no one size fits all approach to the diagnosis of this disease, that is why doctors look for signs here and there that can give solid information for diagnosis, (Legg, 2018a, 2018b). According to Medical News today, sometimes symptoms are related to other medical conditions. So the doctor will make sure to look out for those symptoms before proceeding with a brain cognitive examination

2.4.3 COGNITIVE ASSESSMENT TEST

The following must be present to confirm a diagnosis of Alzheimer's disease:

- Steady and constant loss of memory
- Continuous decline in brain activity

Some questions include:

- Patient's name
- Asking the person what the time is
- Asking what year it is
- Asking the person to count backward from 20 down to 1
- Asking for patient' date of birth?

(Legg, 2018a, 2018b).

2.4.4 GENETIC TESTING

Using genetic tests could help the doctor make predictions on if a patient has a chance of developing dementia in the future. Because this has very little associated with the disease, it is not very reliable to make proper diagnosis (Legg, 2018a, 2018b).

Some of these tests may not be sufficient enough to properly identify this type of dementia a person has.

Your doctor may also do the following tests:

- **Magnetic resonance imaging tests:**
    - MRIs can help identify biomarkers, which gives them information for proper diagnosis.
- **CT Scans :**
    - Brain MRI images scans are taken, allowing the doctor to identify variations in the brain
- **PET Scans:**
    - Helps your doctor detect plaques buildup and deposits (Legg, 2018a, 2018b).

2.5 MACHINE LEARNING AND NEURAL NETWORKS

2.5.1 MACHINE LEARNING AND DEEP LEARNING

Machine provides machines a way to learn and think and make decisions and predictions based on past experience and training and make future decisions based on

those learning without being programmed directly to do so (Expert System, 2017). According to Internet Society (2017), the basic process of machine learning is to give training data *to a* learning algorithm which then generates a new set of rules, based on features gotten from the data. This is in essence generating a new algorithm. The way neural networks learn is that it first identifies and learns data, it looks for features and then uses it to make predictions based on its previous training (Expert System, 2017).

A Machine Learning engine cannot provide all the capabilities on its own to deliver the results that business require (Gidney, 2018). This usually involves several technologies like:

- Natural language processing
- Latent semantic indexing
- Deep Neural networks

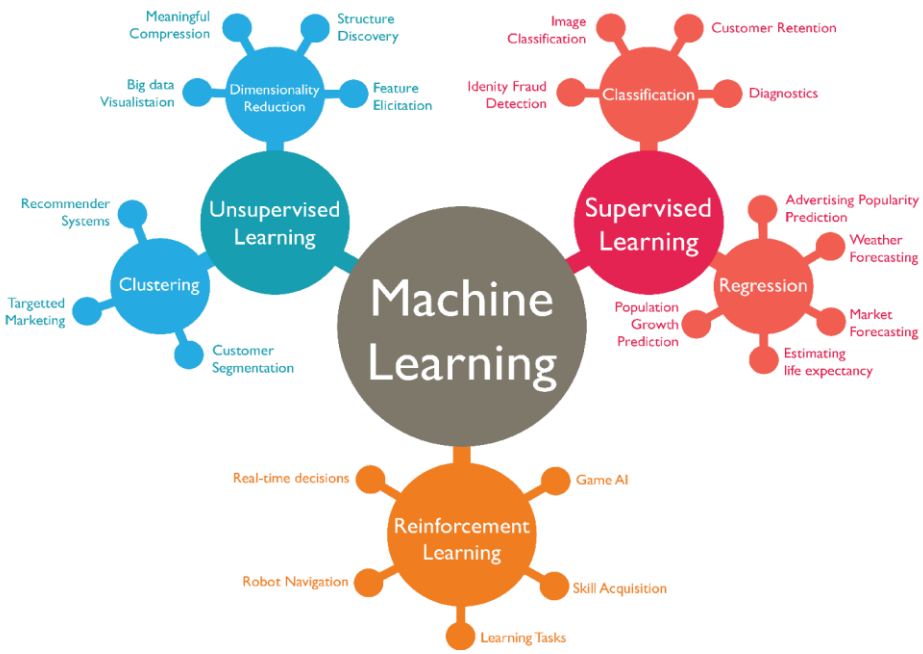

Figure 4 : Showing the divisions of Machine Learning *(*Source: www.wordstream.com*)*

Machine Learning has 3 main divisions namely Unsupervised, supervised and reinforcement learning:

- Supervised Learning:

- You feed output of your dataset into the system. This means that the machine already knows what the output is before it starts working on it or learning it.

- Unsupervised Learning:
  - The system does not have solid data sets, and the outputs of the problems are largely unknown. The model is blindly learning

- Reinforcement Learning:
  - In Reinforcement Learning involves an agent making the best action based on the current state of the results (Van Loon, 2018)

2.5.2 NEURAL NETWORKS

Neural Networks now provide the most optimal solution to many problems in Image Recognition, Speech Learning and Natural Language processing. (Salian, 2018).
Neural networks can be used to identify and recognize images and classify them based on some set standard associated with the data. A trained neural network can also make prediction on new datasets based on its training and this can be seen as using your own knowledge (Salian, 2018).

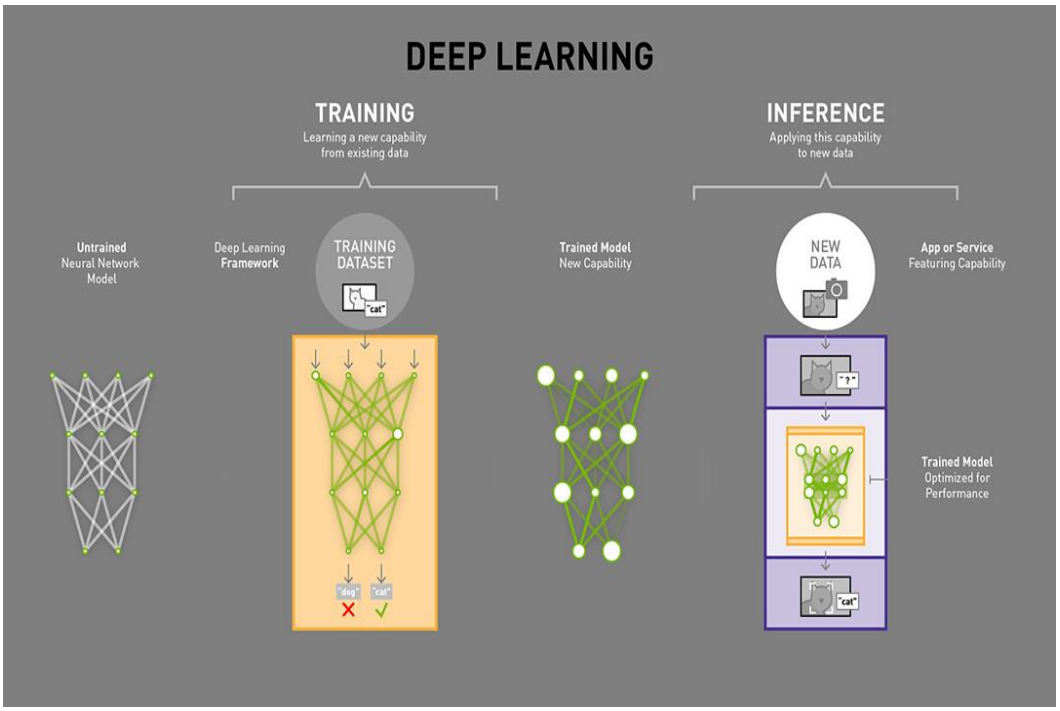

Figure 5: Showing how a Deep Neural Network works (Source: blogs.nvidia.com)

2.5.3 TRAINING YOUR NEURAL NETWORK

According to Parmar (2018) on Training Data Neural Networks, it is very important to preprocess the data before feeding it into the network so that the accuracy of the model will be good. It can be done to reduce or increase the dimensionality of the image, reduce noise or make changes to several variations associated with images data (Le, 2019). It is also a known fact that data that has not been processed properly can have a negative impact on the accuracy of the model.

Each input will be denoted as x_k where k stands for index of input. For each input x_k neuron assigns another number w_k. A vector consisting of these numbers w_k is called Weights vector. The weight of a neuron is what makes it different from every other neuron. (Bushaev, 2017). Neural networks known as a linear function that works together with a non linear function in order to operate properly. TORRES.AI (2018) mentioned in Learning Process of A neural Network, that the way neurons learn can be seen as a repetitive process involving moving through the neural network, known as forward propagation and moving backwards through the neural network known as in a repetitive way.

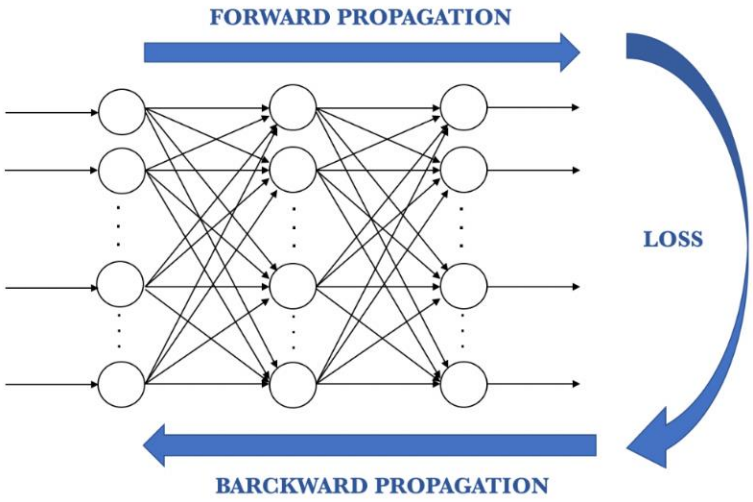

Figure 6: Showing forward and back propagation (Source: www.towardsdatascience.com)

When it comes to deep neural networks, each layer of neuron learn on a unique set of data based on the previous layer's output. This continuous process becomes more and more complex, since they advance features learned from the layer before. This whole process allows the neural network to learn from raw data and advance its knowledge of the data to be able to make complex predictions. Information learned is then fully

reconstructed. (Nicholson, 2019). Also it is very important to keep in mind that you want to reduce the losses encountered when training a neural network.

2.6 REVIEW OF EXISTING SYSTEMS

The following subsections present a review of existing systems and the different ways and methods they used to achieve their aim:

2.6.1 CONVOLUTIONAL NEURAL NETWORK FOR BINARY CLASSIFICATION FOR THE DIAGNOSIS OF ALZHEIMER'S DISEASE

In this research, Jake Billings and Gunnar Ensero used convolutional neural network to analyze open-access brain MRI scans to diagnose Alzheimer's disease. They were able to build a model that identified and classified Alzheimer's disease patients from Normal healthy patients. Their model made use of the OASIS dataset. The dataset contained both male and female and demented and non-demented. They created a model that was able to classify patients who had the disease from patients who didn't with an accuracy of 95%

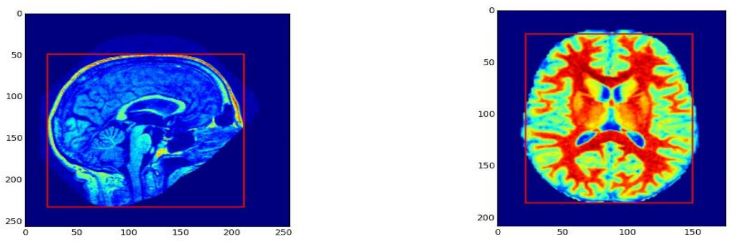

Figure 7: Showing brain MRI scans from the OASIS open-access dataset (source: www.github.com)

2.6.2 ALZHEIMER'S DISEASE PREDICTION

This was a research study carried out by Daniel Graziano, Daniel Molina-Hurtado, Esmail Fadae, Paxton Meaeder-York. In their solution, they focused on building a model that could detect Alzheimer's disease from brain Images. They made use of the ADNI open source dataset. They also focused on using the dataset on several classifying models like the logistic regression model, decision tree model, k Nearest Neighbor, support vector machine and a few other models to determine the best model suited for predicting Alzheimer's disease. While creating their model, they focused on using only the most relevant information that was used to determine if a patient had the disease. They were able to derive a visualization of all the predictive models they analyzed to give us a clear conclusion on what model what most accurate and most predictive of the disease.

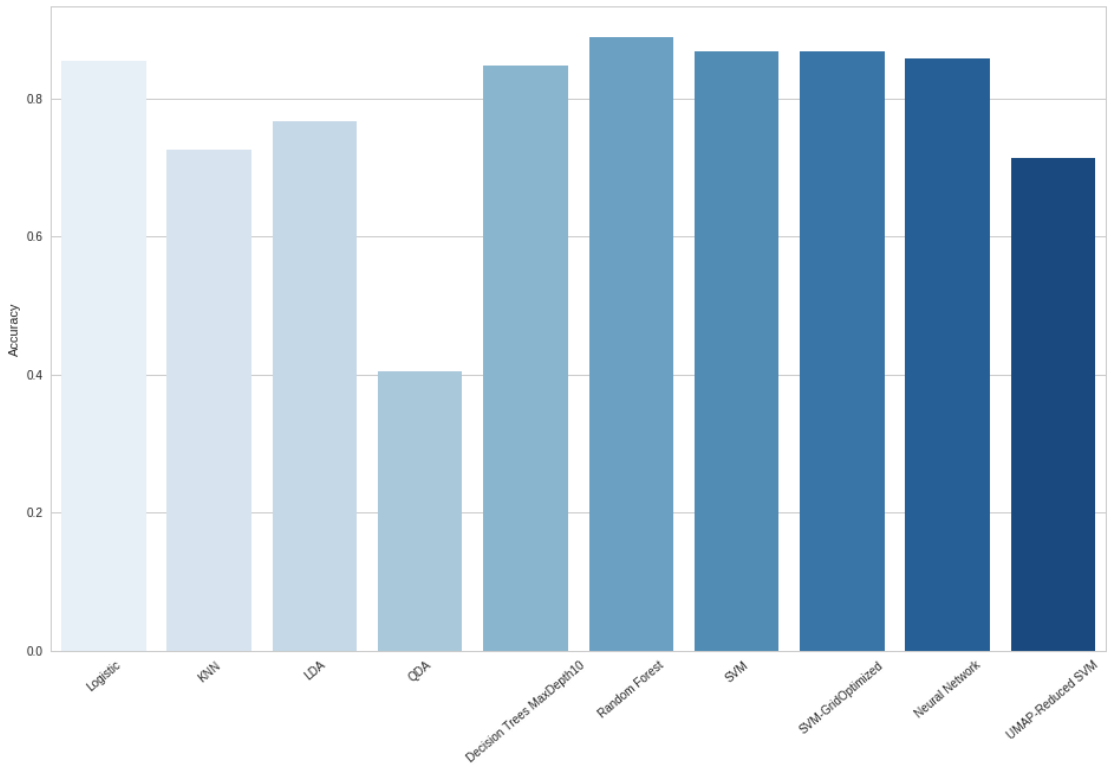

Figure 8: showing the predictive capability of several predictive models (source: www.github.com)

2.6.3 DEMENTIA CLASSIFICATION: COMPARE CLASSIFIERS

Deepak N, in his study mage use of several classifiers such as decision tree, support vector machine, etc. to classify Alzheimer's disease from the OASIS open access dataset. They aim of his study was to compare these classifiers to see which would give a better accuracy in order to recommend it for accurately diagnosing and classifying Alzheimer's disease. The dataset that was used is a labeled dataset and it was analyzed using several classifiers and the true positive and false positive values derived were visualized.

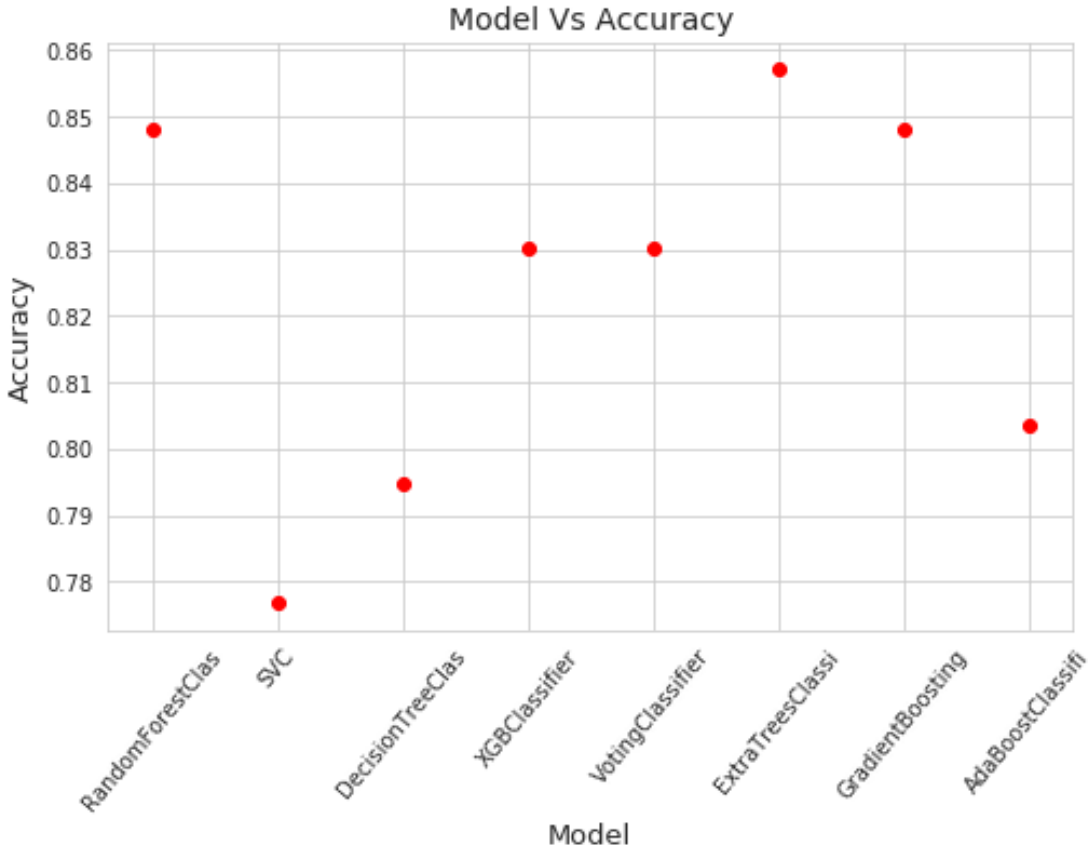

Figure 9: showing the accuracy scores of several classifiers on the OASIS dataset (source: kaggle.com)

CHAPTER THREE

DATA ANALYSIS

3.1 PREAMBLE

The method that was adopted to develop the neural network model used to identify and classify Alzheimer's disease patients from normal healthy patients will be analyzed in this chapter, stating the tools used, dataset and requirements.

3.2 DATA ANALYSIS

This is the systematic process of transforming, inspecting, evaluating, cleaning, and modeling data to get useful information required to make business decisions. It involves sorting through some large amounts of unstructured data and making sense out of it. It is done by making use of statistical and analytical tools like python and R. this is useful for exploring data in meaningful ways, exploring data, and presenting it as useful information.

3.2.1 MACHINE LEARNING

Machine learning is the study of algorithms that teaches computers to perform specific tasks without being explicitly programmed. It is a method of data analysis that provides systems the ability to automate analytical model building. The Architecture of the machine learning model is categorized into three; supervised, unsupervised, and reinforcement learning, and the processes utilized in these architectures are data ingestion, data preparation, data segregation, model training, candidate model evaluation, model deployment, model scoring, performance monitoring.

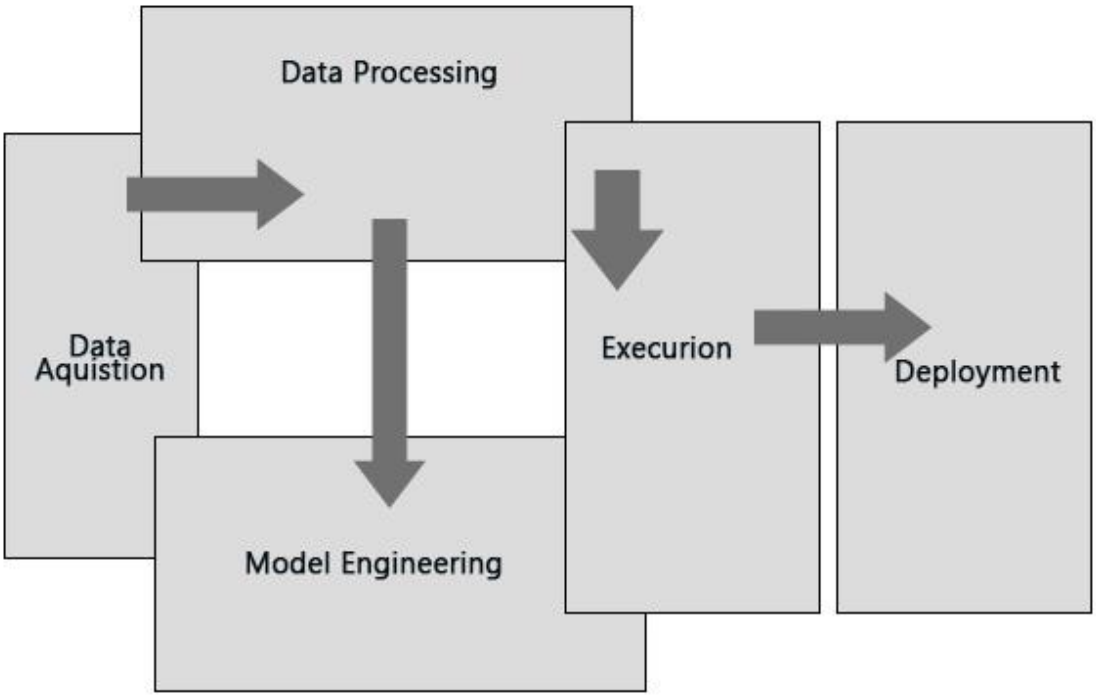

Figure 10: showing the Machine learning architecture

(Source: www.educba.com)

Some machine learning architecture processes include:

- Data collection:
    - This is the first stage in the architecture. In this process, it involves the collection, preparation, and segregation of features from data. It involves funneling incoming data into a data store.

- Data processing:
    - This stage involves the normalization of data, data exploration, encoding, transformation, cleansing, and feature engineering. This step is dependent on the learning been used and on the type of processing, it may require. This stage in the architecture also involves splitting data into training and testing subsets to train the model and further test it with new incoming data.

- Data modeling:

- In this layer of the architecture, it will involve the use of various machine learning algorithms to recognize patterns in the dataset. This will help to properly model the data and adapt the system to easily recognize features. This layer prepares the system for the execution stage.

- Execution:
  - This stage is where the experiment is done and the system is accessed for its performance and checked against the test dataset to see how accurate the prediction is. This stage also involves optimizing the algorithm and maximizing the system's performance

- Deployment:
  - After the chosen model is produced, it is then embedded and deployed into the decision-making system. It may further involve applying the machine learning model to new datasets to extract features that could be beneficial to the decision making process. It is still being observed at this point to understand how it operates in the real world.

3.2.2 CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are a class of deep neural networks that are used in the area of visual imagery analysis, recognition, and classification. This is a very unique character as it can identify features in images without being supervised and learn distinct features by itself. Convolutional neural networks made outstanding improvements in the field of deep learning when it was invented by Yann LeCun in the year 1994.

3.3 DATASET

A dataset is a collection of a related set of information that can be manipulated by a computer program. It may be structured or unstructured. It is usually represented in a tabular form and may be organized into some type of data structure usually for a unique body of work. A dataset may correspond to a database where every column of

the table represents a particular variable and each row represents a particular record of the dataset. In statistics and data analysis however, we have different types of datasets that are available:

- Numerical dataset
    - This type of data deals only with numbers. The entire dataset is composed of only numerical values. Some examples include:
        - The number of pages in a book
        - The heights and weights of the employees at a company
        - The number of citizens in a country
- Bivariate dataset
    - This type of dataset is called bivariate because it deals with two variables. This dataset deals with the relationship between the two variables. For example:
        - An employee's height and age
- Correlation dataset
    - The correlation dataset consists of values that are dependent on each other. The values demonstrate some type of relationship they have with each other. For example:
        - When the weight of a person is dependent on his height
- Multivariate dataset
    - This is a dataset that consists of many variables. For example:
        - The height of a ball
        - The volume of a ball
        - The weight of a ball
        - The color of a ball
- Categorical dataset
    - This type of data represents the characteristics of a person or an object. Some examples are:
        - A person's race
        - A person's nationality

To build a model in Machine learning you must have a good dataset. A good dataset is a crucial aspect of Machine learning; otherwise, it will not be possible for your model to learn. During AI development, we use three different datasets: training dataset, validation dataset (used to tune the final model), and testing dataset. Gathering data is not enough; cleaning and labeling the data takes up the bulk of the work. Usually, your testing set should take up to 20 – 30% of your entire data, and your training set should be about 60 – 70% of your entire data. This is to ensure that the machine learning model is provided with as large enough data sufficient for learning. The size of your data will depend on how complicated the task is but you must have enough of it.

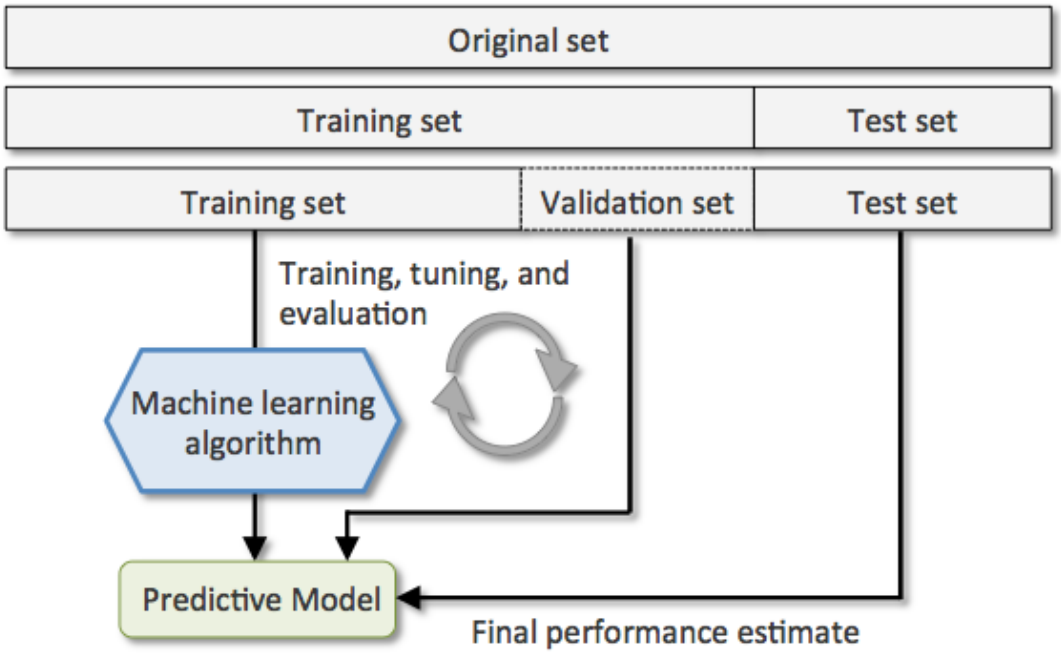

Figure 11: showing how training and testing of dataset flows
(Source: towardsdatascience.com)

3.3.1 DATASET ACQUISITION

Data acquisition involves the collection and gathering of sufficient data relevant to your work. For this project, it was important to acquire data that was relevant to the classification and diagnosis of Alzheimer's disease. This involved getting brain MRI scans from an open-source repository that contained the brain images of patients with

Alzheimer's disease. Kaggle provides an open access repository for data scientists, researchers and machine learning engineer.

Some sample images from the dataset used are listed below:

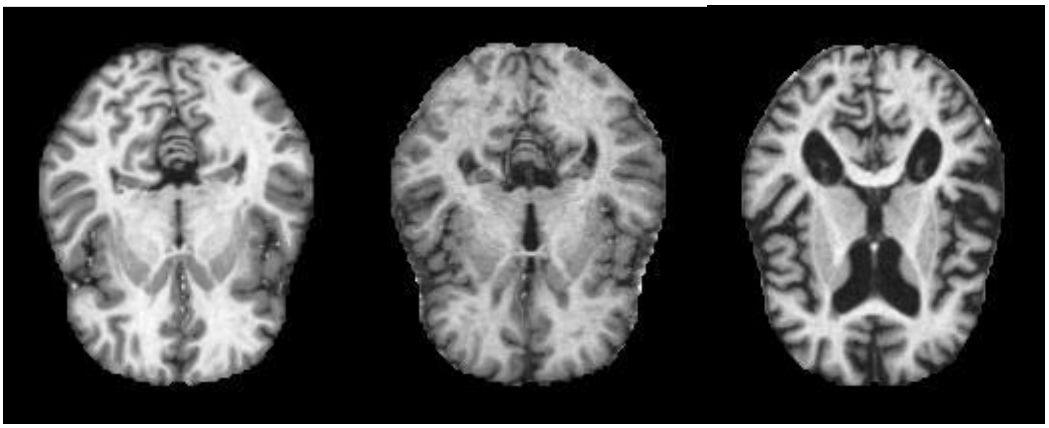

Figure 12: showing some brain images used for training the model (source: www.Kaggle.com)

3.3.2 DATASET DESCRIPTION

The dataset that was used for this project can be publicly downloaded from (https://www.Kaggle.com/tourist55/alzheimers-dataset-4-class-of-images). Kaggle is an open access repository for datasets. The dataset used contained 6400 images. 3200 of the images are non-demented and the rest are demented, mildly demented or moderately demented. The data contained images of both male and female and was compiled from several websites containing open access brain MRI images.

3.4 REQUIREMENTS

The language used for this project is python and the following are the frameworks used:

- Keras and Tensor flow which was used in building the autoencoder as well as the classifier.
- Pydicom was also used in loading and manipulating the 3d images

- Plotly and Scikit-image and also Matplotlib was used for visualizing the images
- Numpy was also used in the project for data manipulation and conversion to arrays
- The Glob library was used for also reading in the data.
- Jupyter notebook, an open source web application for data visualization, machine learning and so on.
- Anaconda, an open source software distribution of R and python used for scientific computing
- Tensor flow, an open source python library used for machine learning applications such as neural network. It is used as a backend to Keras

3.4.1 MINIMUM SOFTWARE REQUIREMENT

To create and build the image classifier for this project, you would need a good programming language that allows you to analyze the images. Programming languages like Python and R are mainly used for data analysis and could both be used to create the image classifier for the brain images. I used Python programming language and several third party applications that for this project and these are my recommendations for the minimum software requirements:

- Python 3.6 (minimum)
- VGG19
- OS
- Keras version 2.24
- Numpy version 1.6.24
- Scikit learn version 0.16.2
- Matplotlib version 3.03
- Tensor flow version 1.15.2

3.5 DATA PREPROCESSING

The images collected were first read in and preprocessed. The images were counted and then the images were also reshaped using Numpy and then fed into the autoencoder. The data was also split into training and testing datasets so we can train and evaluate the performance of the model on unseen images.

The pre-trained autoencoder was used to find patterns and abstractions in the data and then a convolutional neural network was then trained to classify which brain scan is diagnosed with Alzheimer's disease. The CNN architecture that was used to perform this was the VGG19.

CHAPTER 4
DATA MODEL IMPLEMENTATION

4.1 INTRODUCTION

Data model implementation is a collection of components needed to implement, create and build the data model. This chapter explores the dataset used for the model implementation. It includes all the software components, hardware components and the programming language that was utilized to bring the proposed neural network model to completion. Images will be used to display the process of the model creation.

4.2 MODEL REQUIREMENT

The model requirements are required to build out the proposed model, the minimum software and hardware that must be available in the system. it is important that before we begin to implement the model, that few measures are in place to ensure that the entire process runs smoothly. After doing extensive research and taking a deep look at the data, I have decided to limit my scope to the following:

- Medical History:
    - For this study the patient's age, gender, race and status was collected. Some family related information was collected also to find out if there was a history of Alzheimer's disease or any other form of dementia in the family line.
- Neuropsychological Tests:
    - One of the most essential information that was required for this study was the neuropsychological tests. It consists of a set of tests that the patients have to go through in other to monitor their cognitive function. Usually it is to evaluate a patient's ability to remember certain things and a patient's ability to perform simple tasks that required very little cognitive capacity. At the end of these tests, the patient's performance is given a score. Some of the most widely used tests include CDR (Clinical Dementia Rating), FAQ (Functional Activities Questionnaire) or MMSE (Mini–Mental State Examination).
- Biomarkers:
    - A biomarker is described as a biological indicator or marker that is usually linked to a disease in the body. In the case of this study, some

of the biomarkers used for Alzheimer's disease detection are the Amyloid and Tau protein.

4.2.1 RECOMMENDED HARDWARE REQUIREMENTS

- NVIDIA GPU
- Intel core i7 4600U processor
- 8 GB RAM
- 14 inch full HD screen
- 256GB SSD

4.3 IMPLEMENTATION TOOLS USED

The central idea to building an auto encoder is that it has two sections: the encoder and the decoder, but before we come to all that, the data in this case images have to be preprocessed and transformed into something that is workable.

This is the entire process to which the model was built.

4.3.1 READING IN THE PACKAGES

The first thing to do is to import or load all the necessary packages and libraries into the development environment or IDE. Several packages were used for several different things and accomplished what they set out to do, some of the packages include, Keras, Scikit-Learn, glob, Numpy etc.

Keras has several modules that do different things; we loaded the Keras layers module which had functions like Input, Dense, merge, Reshape, Conv2D, MaxPooling2D etc. The Input function was used to load in the input shape of the images which will be touched in a minute, Dense was used to build a fully connected layer in the model, rescale the tensor from values between 0 and 225 to values between 0 and 1. Reshape was used to reshape the input images, Conv2D was used in building a 2d convolutional layer, Max pooling was used to grab the essential representations found in the images etc. To train the model, I used the fit_generator method of the ImageDataGenerator class to train the network. From sklearn we made use of the train test split feature to split the data to train and test, so we could train the data and also test it on unseen images. The VGG19 is the CNN architecture that was used to train the model. It is a variant of VGG which consists of 19 layers. 16 convolution layers, 3

fully connected layers, 5 MaxPool layers and one soft max layer. VGG19 is a deep CNN used to classify images.

Turnitin

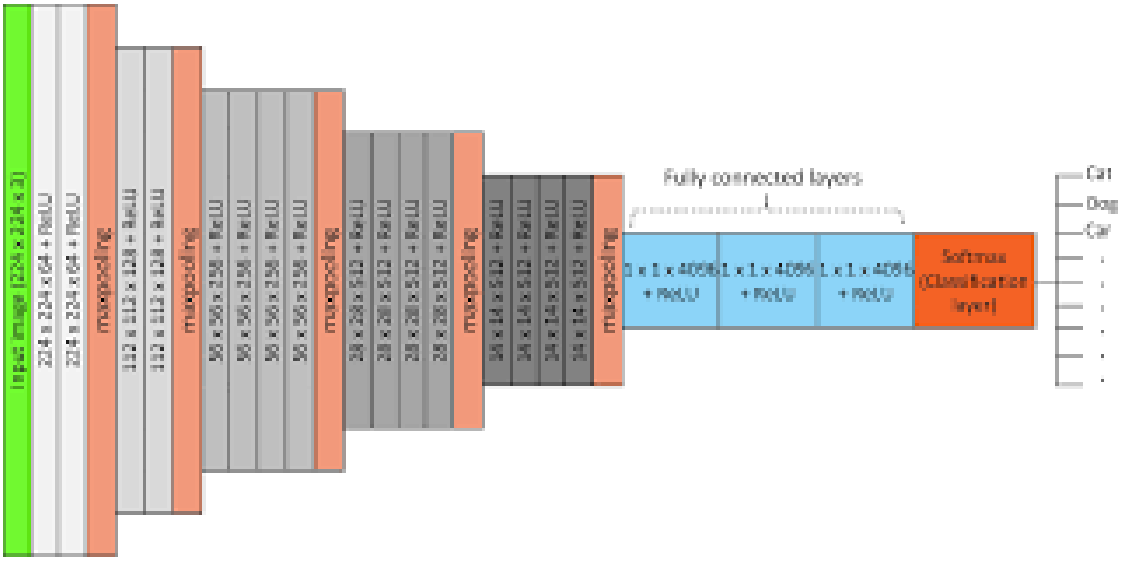

Figure 13: Showing the VGG19 Architecture (source: link.springer.com)

The model was fine-tuned and then I fitted the dataset to it. I used the softmax activation function to get a result of either demented or non-demented. Softmax is mostly used in neural network architectures to map the normalized output to a probability distribution over predicted output classes, in this case, demented or non-demented. After the dataset was fitted to the model, the model was let alone to learn patterns from this never seen data. At the end of the learning phase, the model was able to classify the images with an accuracy of 73 %

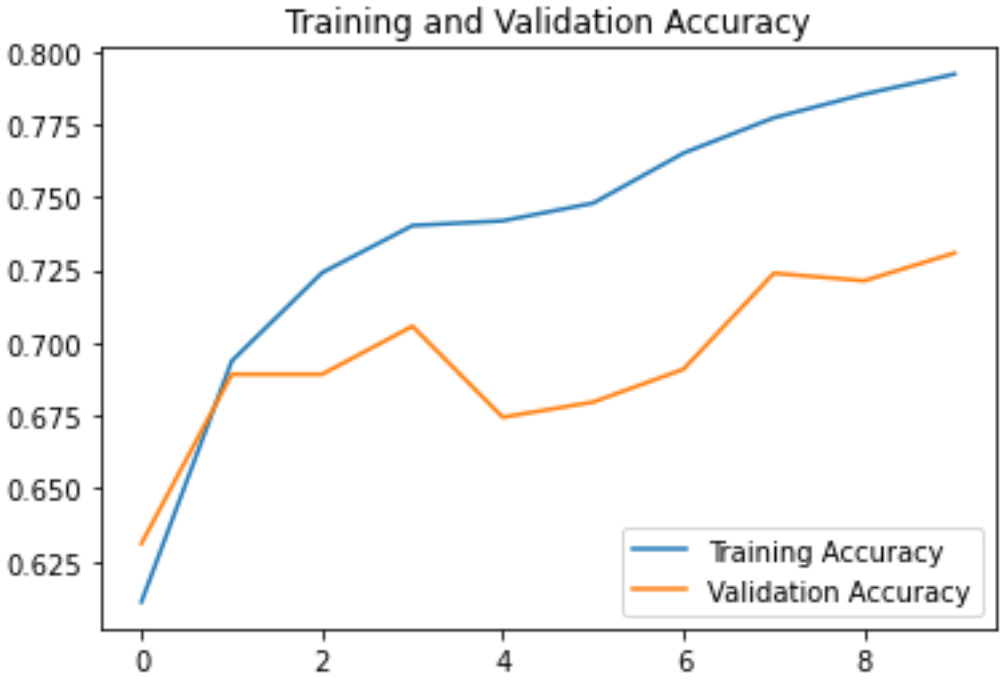

Figure 14: showing the accuracy plot of the neural network model

Figure 15: showing the training and validation loss of the neural network model

4.4 EXPLORATORY DATA ANALYSIS (EDA)

In this portion of the project, I am going to explore Alzheimer's disease in greater depths. Alzheimer's disease is known to be one of the leading causes of death in the

United States. The disease is known to affect individuals usually around the age of 65 to 90 years. There have been a couple of controversies around the main cause of this form of dementia and the most identifying area. In this part of the chapter, we will explore the Open Access Series of Imaging Studies (OASIS) dataset to help determine the relationship between individuals who suffer from this disease and the features that are most significant to identify the symptoms early on.

4.4.1 OPEN ACCESS SERIES OF IMAGING STUDIES (OASIS)

The OASIS-2 dataset is an open source dataset that contains the longitudinal MRI data in non-demented and demented older adults, with 150 subjects and 373 MR Sessions. This contains individuals with ages ranging from 60 to 95each subject was scanned on two or more visits separated by a year at least for a total of 373 imaging sessions. All the individuals are right handed and include both men and women. This dataset was used to explore the disease.

Columns	**Full description**
EDUC	Years of Education
SES	Socioeconomic Status
MMSE	Mini Mental State Examination
CDR	Clinical Dementia Rating
eTIV	Estimated Total Intracranial Volume
nWBV	Normalize Whole Brain Volume
ASF	Atlas Scaling Factor

Table 1: showing OASIS dataset description

Columns	**Min**	**Max**	**Mean**
EDUC	6	23	14.6
SES	1	5	2.34
MMSE	17	30	27.2
CDR	0	1	0.29
eTIV	1123	1989	1490
nWBV	0.66	0.837	0.73
ASF	0.883	1.563	1.2

Table 2: showing the min, max and mean values of the OASIS dataset column

	Subject ID	Group	M/F	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
0	OAS2_0001	0	1	87	14	2.0	27.0	0.0	1987	0.696	0.883
1	OAS2_0002	1	1	75	12	3.0	23.0	0.5	1678	0.736	1.046
2	OAS2_0004	0	0	88	18	3.0	28.0	0.0	1215	0.710	1.444
3	OAS2_0005	0	1	80	12	4.0	28.0	0.0	1689	0.712	1.039
4	OAS2_0007	1	1	71	16	2.0	28.0	0.5	1357	0.748	1.293

Table 3: showing the first 4 rows of the OASIS dataset

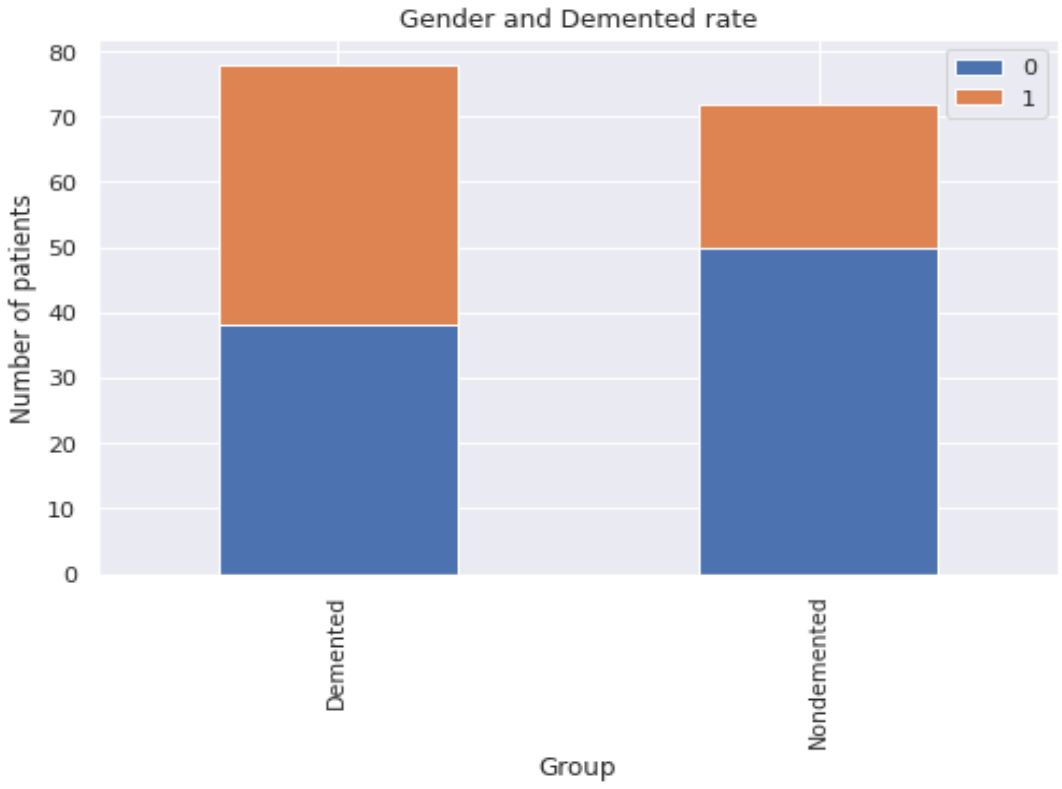

Figure 16: Showing the ratio of demented (1) and non-demented (0) group in the dataset

Individuals who suffer from Alzheimer's disease or any similar dementia have been tested to have lower MMSE score compared to individuals who don't suffer from the disease. While exploring the dataset, we discovered that the individuals who had higher MMSE scores were patients who did not suffer from this disease. This is to confirm that an individual having Alzheimer's disease will continue to experience steady decline in cognitive functions

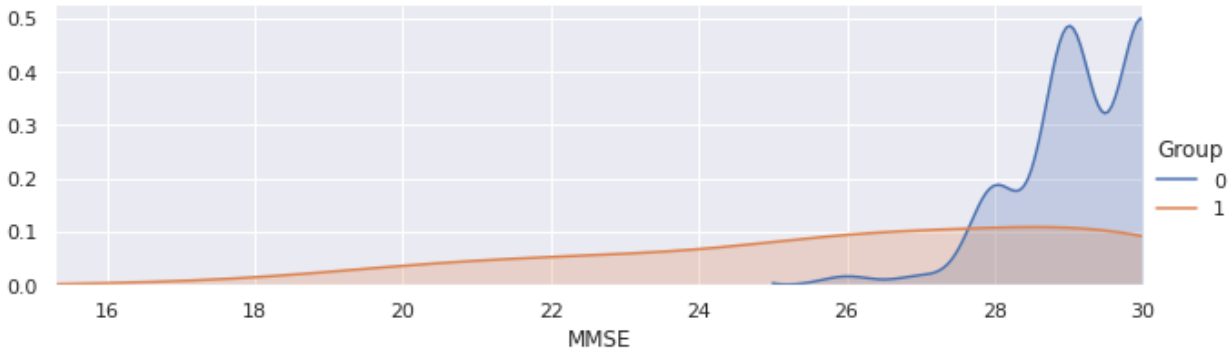

Figure 17: Showing the non-demented patients (0) having higher MME score

The neurodegenerative ability of Alzheimer's disease is discovered to be significant with individuals around the ages of 65 – 90years. While exploring the OASIS dataset and visualizing the relationship between the individuals and their ages, it was shown that individuals that suffered this disease were around the ages of late 60's up to their 90's. Something else that was discovered was that the individuals who were tested with this disease didn't live past their early 90's. Alzheimer's disease have proven to be a leading cause of death by progressively diminishing cognitive abilities in its victims until they can't perform basic everyday tasks and eventually causing the death of such individual. The image below is an illustration of this.

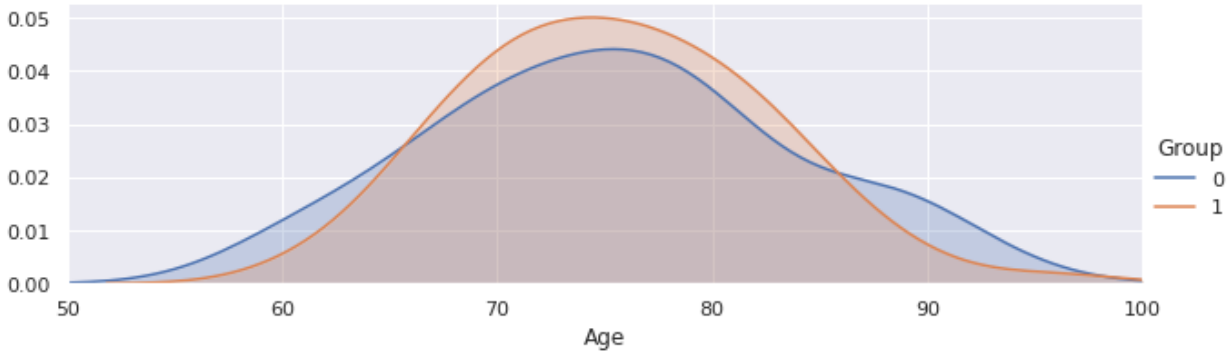

Figure 18: Showing the effect of Alzheimer's disease in individuals of ages 70 to 90 years

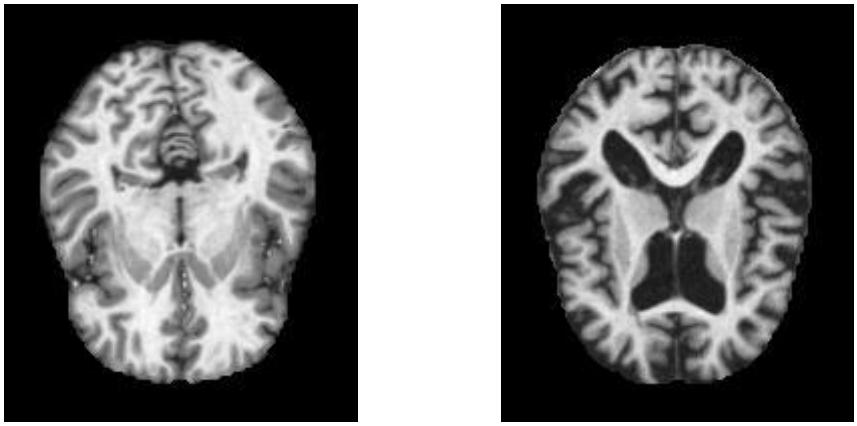

Figure 19: Showing the effect of Alzheimer's disease on the brain

Alzheimer's dementia is a kind of dementia that causes brain cells to waste away. This degenerative property of the disease causes the brain volume to shrink significantly just like in the image displayed above. The effect of this can lead to memory loss, decline in thinking and reasoning, difficulty making decisions or planning and behavioural changes. The dataset that was explored shows that the individuals who had Alzheimer's disese and were 65 years and older, had relatively low brain volume.

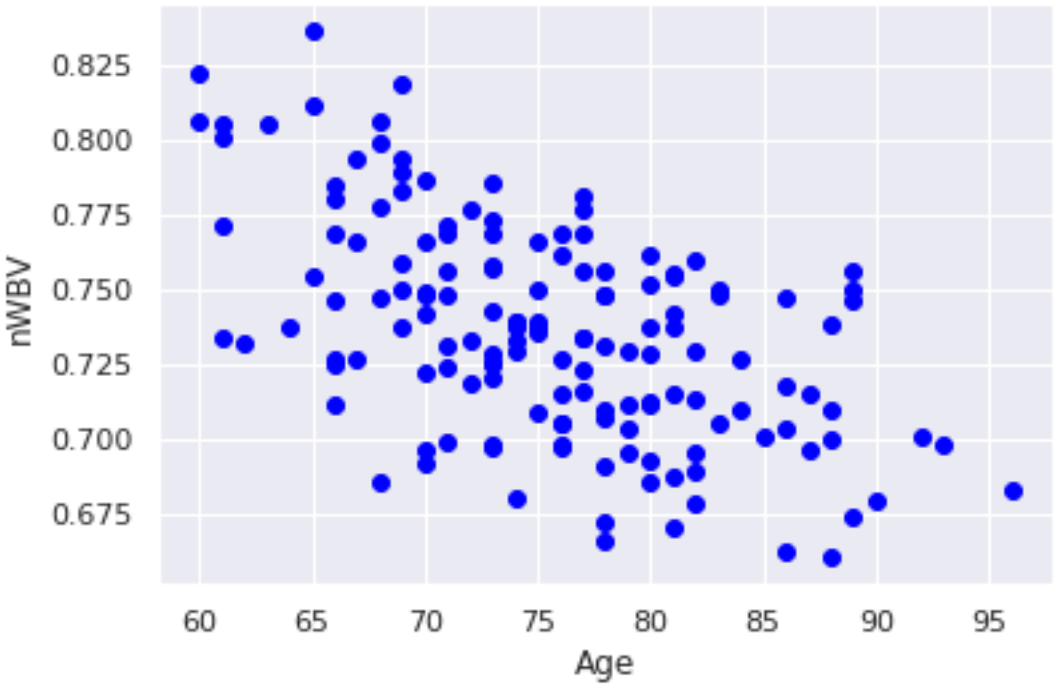

Figure 20: Showing the relationship between the ages of individuals with AD and their brain volume

.

4.5 PERFORMANCE MEASURES

To test the performance of the models explored, I use the receiver operating characteristic curve (AUC) as my performance measure. For neurodegenerative diseases like Alzheimer's disease and most forms of dementia, it is very important that I we obtain a high true positive value so that the patients who suffer from Alzheimer's disease can be easily diagnosed and treated early. With this said, it is just as important to ensure that we obtain a very low false positive value since we don't want to diagnose a healthy patient as demented. Therefore, this performance measure, the AUC is a very ideal option.

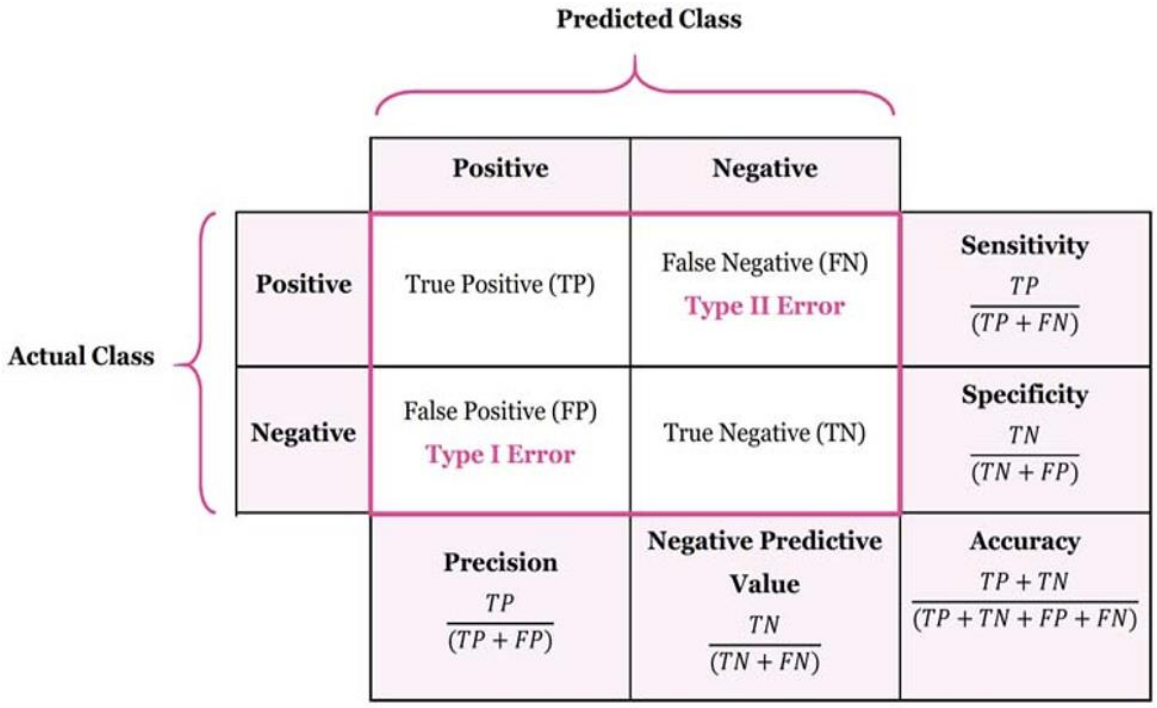

Figure 21: showing confusion matrix (Source: Medium.com)

4.5.1 DECISION TREE

This is a model that arranges its information in the form of a tree like structure. It classifies the information along various branches. It is a non-parametric supervised learning method used for classification and regression.

4.5.2 RANDOM FOREST

This is a supervised learning algorithm used for classification and regression. It runs effectively on large database. It consist of a large number of individual decision trees that operate as an ensemble

4.5.3 SUPPORT VECTOR MACHINE (SVM)

Support vector machine is supervised learning algorithm that is a discriminative classifier formally defined by a separating hyper plane. It is used for classification, regression and outliers detection.

4.5.4 OTHER MODELS

Some other models that were explored were the extra tree classifier, which aggregates the results of multiple de-correlated decision trees collected in a forest to output its classification result; gradient boosting classifier, which is used for classification and regression; and the AdaBoost or Adaptive boosting classifier, which can be used with other algorithms to improve performance.

Figure 22: showing the ROC curve of all the models used in the dataset analysis

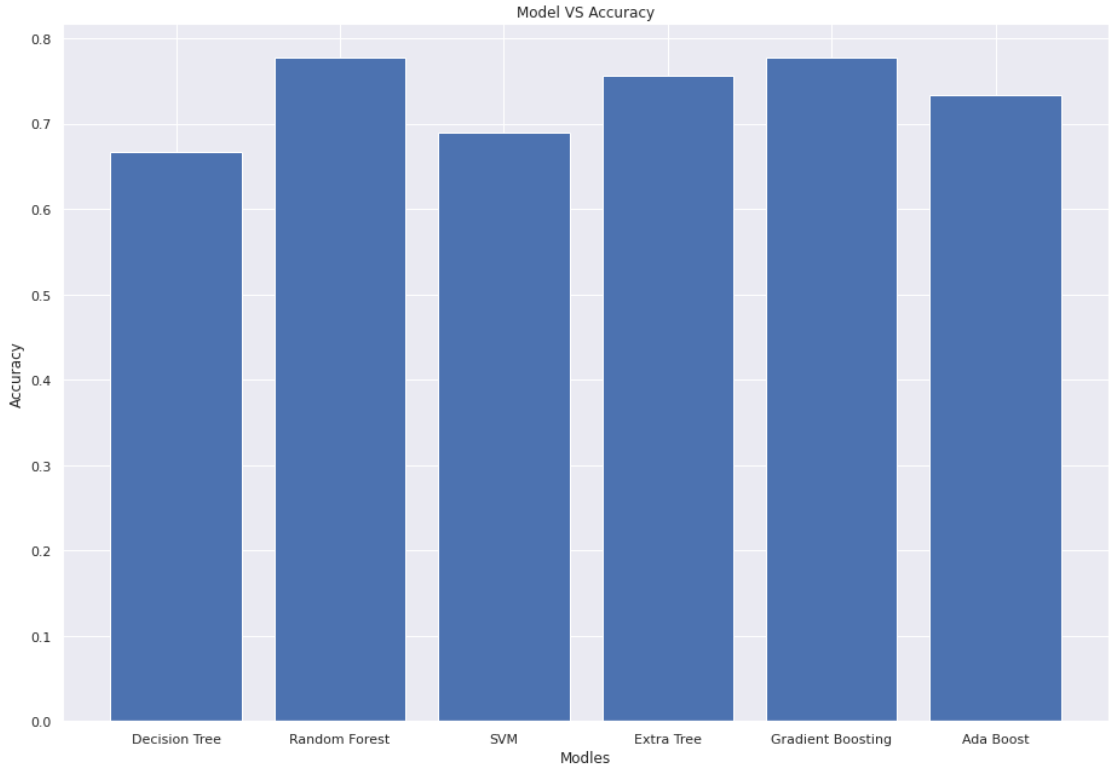

Figure 23: showing the accuracy scores of each model evaluated

CHAPTER 5
CONCLUSION

5.1 SUMMARY

There have been studies that have showed the impact of Alzheimer's disease in the lives of middle aged to elderly individuals. These impacts help us to understand the negative effect of dementia on mental functionality, memory and behavior. With the continuous emergence of the disease there has been an increase in computerized methodologies for the early detection of Alzheimer's disease to help diagnose and treat the disease in early stages of its onset, even before health professionals can begin to notice clinical symptoms. In this study, I was able to build a sensitive Convolutional Neural Network, the VGG19, which could detect the presence of Alzheimer's disease in brain MRI images and classify Alzheimer's disease patients from normal healthy patients with an accuracy of 73%. I was able to use a second dataset form the Open Access Series for Imaging Studies (OASIS) to explore the disease. I proved the relationship between Alzheimer's disease and certain features that were common to the patients who suffered from this dementia. I was able to provide only the most relevant information for the diagnosis of Alzheimer's disease for future use by doctors and health professionals. Finally I was able to prove the advantage and effectiveness of computerized diagnosis of Alzheimer's disease dementia over traditional medical diagnosis. The model can be downloaded from **https://github.com/winniebuna**

5.2 RECOMMENDATIONS AND FUTURE STUDIES

There are so many things I would love to explore in future studies. One of them would be that I would like to take my detecting model a few steps forward and create a predictive model. I would focus on using so many popular algorithms and test the data set on these algorithms to find out which one gives the best accuracy. This will help me understand the predictive power of these algorithms. I would also like to focus on a particular category like a particular test such as neuropsychological tests vs. imaging vs. biomarkers. The goal of this is to provide doctors and health practitioners with the most effective test for identifying the disease and for accurately predicting the development of Alzheimer's disease in patients.

Once the predictors are covered extensively, another area that would be wise to explore is the time. I would hopefully achieve this by researching and applying some time series approaches since the data would be inherently of that nature.

Lastly I would like to improve the data analysis process. The data cleaning process and even the training process where I would better tune my model. The results of your model is only as good as the data; garbage in equals garbage out. I would tune my model and try to obtain a better accuracy score in detecting the presence of Alzheimer's disease in the brain images and better predictions in the further studies. I strongly believe that there is a lot more areas that I barely touched, more so because I know that Neural Networks have so much more potential that I have yet to explore, for this dataset than what I have seen.

5.3 CONCLUSION

This study focused on improving health and longevity of elderly patients most especially by collecting the MRI brain images of patients and detecting the presence of Alzheimer's disease in these images before physical signs begin to appear. Studies have proven over time that individuals with Alzheimer's disease or any other form of Dementia only begin to show clinical signs 10 years after the onset of the disease, which might be very late to do much. This study would help doctors and health practitioners detect the disease early on, and slow down further progression in later stages of the individual's life.

REFERENCES

Alzheimer's Association. (n.d.). What Is Alzheimer's Disease?. Retrieved from https://www.alz.org/alzheimers-dementia/what-is-alzheimers

Alzheimer's Disease Neuroimaging Initiative (2017). Retrieved from http://adni.loni.usc.edu/

Being Patient. (2018, March 22). Alzheimer's: The Most Expensive Disease in America. Retrieved from https://www.beingpatient.com/alzheimers-cost/

Bushaev, V. (2017, November 27). How do we 'train' neural networks?. Retrieved from https://towardsdatascience.com/how-do-we-train-neural-networks-edd985562b73

Cure Alzheimer's Fund. (n.d.). The Cost Of Alzheimer's Disease. Retrieved from https://curealz.org

Expert System. (2017, March 7). What is Machine Learning? A Definition. Retrieved from https://expertsystem.com/machine-learning-definition/

Gastfriend, J. (2018, October 22). The Hidden Costs Of Alzheimer's Disease. Retrieved from https://www.forbes.com/sites/jodygastfriend/#1593a78c621d

Gidney, K. (2018, April 11). Demystifying machine learning. Retrieved from https://www.infoworld.com/article/3268524/demystifying-machine-learning.html

Hosseini-Asl, E., & Gimel'farb, G., & El-Baz, A. (2016). Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. Retrieved from http://scholar.google.com

Internet Society.(2017, April 18). Artificial Intelligence and Machine Learning: Policy Paper. Retrieved from https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/?gclid=CjwKCAiAgqDxBRBTEiwA59eENxQLVaKGhmbCq46PIyamzFffTdF2wexm4xm-SYhVR9p-E11Wekk0WRoCjKcQAvD_BwE

IOS Press. (n.d.). Analysis Reveals Economic Cost of Alzheimer's Disease Dementia are "Tip of the Iceberg". Retrieved from https://www.iospress.nl/ios_news/47479/

Le, J. (2019, October 29). The 4 Research Techniques to Train Deep Neural Network Models More Efficiently. Retrieved from https://heartbeat.fritz.ai/the-4-research-techniques-to-train-deep-neural-network-models-more-efficiently-810ea2886205

Legg, T, J. (2018, February 13). What to know about Alzheimer's disease. Retrieved from https://www.medicalnewstoday.com/articles/159442.php

Legg, T, J. (2018. December 4). Everything You Need to Know About Alzheimer's disease. Retrieved from https://www.healthline.com/health/alzheimers-disease

Loon, R, V. (2018, February 5). Machine learning explained: Understanding supervised, unsupervised, and reinforcement learning. Retrieved from https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/

Mayo Clinic.(n.d.). Alzheimer's disease. Retrieved from https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447

National Institute on Aging. (2019, May 2019). Alzheimer's Disease Fact Sheet. Retrieved from https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet

NHS. (2018). Alzheimer's disease. Retrieved from http://nhs.uk/conditions/alzheimers-disease/

Nicholson, C. (n.d.). A Beginner's Guide to Neural Networks and Deep Learning. Retrieved from https://pathmind.com/wiki/neural-network

Parmar, R. (2018, September 11). Training Deep Neural Networks. Retrieved from https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964

Petersen, R C., & Donohue, M C., & Gamst, A C., & Harvey, D J., & Jack, C R., & Jagust, W J., & Toga, A W. (2010). Alzheimer ' s Disease Neuroimaging Initiative (ADNI) Clinical characterization. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2809036/pdf/7302.pdf

Pushkar, B., & Paul, M. (2016). Early diagnosis of Alzheimer's disease: A multi-class deep learning framework with modified k-sparse autoencoder classification. Retrieved from https://ieeexplore.ieee.org/document/7804459

Salian, I. (2018, August 2). SuperVize Me: What's the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning?. Retrieved from https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/

Sarraf, S., & D.DeSouza, D., & Anderson, J., & Tofighi, G. (2016). DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. Retrieved from http://dx.doi.org/10.1101/070441

Sarraf, S., & Tofighi, G. (2016). Classification of Alzheimer's disease structural MRI data by deep learning convolutional Neural Networks. Retrieved from http://scholar.google.com

TORRES.AI, J.(2018, September 21). Learning process of a neural network. Retrieved from https://towardsdatascience.com/how-do-artificial-neural-networks-learn-773e46399fc7

Wikipedia (2020). Alzheimer's Disease Neuroimaging Initiative. Retrieved from https://en.wikipedia.org/wiki/Alzheimer%27s_Disease_Neuroimaging_Initiative

Wikipedia.(n.d.).Demographics of the United States. Retrieved from https://en.wikipedia.org/wiki/Demographics_of_the_United_States

Worldometer (2019, July 1). United States Population. Retrieved from https://www.worldometers.info/world-population/us-population/#