Deep Convolutional GAN-based Data Augmentation for Medical Image Classification

by

Joy Datta
17301051
Bedria Durdana
17341004
Salwa Rafi
19241010

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

Department of Computer Science and Engineering
Brac University
January 2022

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

Joy Datta
17301051

Bedria Durdana
17341004

Salwa Rafi
19241010

Approval

The thesis titled "DCGAN-based Data Augmentation for Medical Image Classification" submitted by

1. Joy Datta (17301051)
2. Bedria Durdana (17341004)
3. Salwa Rafi (19241010)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 20, 2022.

Examining Committee:

Supervisor:
(Member)

Moin Mostakim
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

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

Head of Department:
(Chair)

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

Abstract

The field of medical imaging is rapidly growing with the help of machine learning, yet the problem of scarcity in labeled medical imaging still remains. Therefore training a machine learning model for medical image processing is always a difficult task. Data scarcity can be solved by using data augmentation techniques which produce and add additional data to the existing dataset. Importance of an augmented dataset also includes increasing model prediction accuracy, adding more training data to models, reducing data overfitting and creating variability in data, increasing generalization ability of models, resolving class imbalance issues in classification, and lowering data collection and labeling costs. It also helps train convolutional neural networks for increased average accuracy. This paper focuses on solving data deficiency in medical imaging through the use of an MRI dataset based on Alzheimer's affected patients. It accomplishes this by employing deep convolutional generative adversarial networks (DCGAN) for generating realistic samples from the dataset. Other approaches for making convincing new images from labeled original images differ from using a deep convolutional generative adversarial network. DCGAN learns from training samples and can generate realistic imaging data with a similar variations, distinct from the original data. We chose to further Alzheimer's research because, like most neurodegenerative disorders, the clinical diagnosis of Alzheimer's dementia had a sensitivity of 71% to 87% and a specificity of 44% to 71%, implying high rates of Alzheimer's Disease misdiagnosis among patients with cognitive impairment. Considering that alarming rate, early diagnosis of Alzheimer's disease necessitates the use of effective automated approaches.

Keywords: Data augmentation, DCGAN, Deep Learning, Classification, MRI

Dedication

When the people we love become unrecognizable due to illnesses that are not in their control, it is devastating for everyone involved. This paper is dedicated to those that suffer from neurodegenerative disorders, those that did not get the proper treatment in time and those who never got the chance to be diagnosed at all.

Acknowledgement

We wish to express our sincerest gratitude to our honorable thesis supervisor Mr. Moin Mostakim for his immense cooperation, valuable suggestions and advice whenever we needed. His dedication to guiding us in the right direction helped us when we hardly knew where to begin. We honor and respect all the faculty members of the Brac University Department of Computer Science and Engineering for giving us this opportunity to gain so much knowledge.

Table of Contents

List of Figures

List of Tables

Nomenclature

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

AD Alzheimer's Disease

$ADNI$ Alzheimer's Disease Neuroimaging Initiative

$AIBL$ The Australian Imaging, Biomarkers and Lifestyle

ANN Artificial Neural Network

CNN Convolutional Neural Network

CT Computed Tomography

DA Data Augmentation

$DCGAN$ Deep Convolutional Generative Adversarial Network

$ERGAS$ Erreur Relative Globale Adimensionnelle de Synthèse

GAN Generative Adversarial Network

$LSUN$ Large-scale Scene UNderstanding Challenge

MCI Mild Cognitive Impairment

MD Mildy Demented

MRI Magnetic resonance imaging

MS Multiple Sclerosis

MSE Mean Squared Error

$MSSSIM$ Multiscale Structural Similarity Index Measure

$NACC$ National Alzheimer's Coordinating Center

NC Normal Cognition

ND Non Demented

NDD Neurodegenerative Diseases

NN Neural Network

NSS Natural Scene Statistics

$OASIS$ Open Access Series of Imaging Studies

PET Positron Emission Tomography

$PGGAN$ Progressive Growing Generative Adversarial Network

$PSNR$ Peak Signal-to-Noise Ratio

$RMSE$ Root Mean Squared Error

SAM Spectral Angle Mapper

SCC Spatial Correlation Coefficient

$SSIM$ Structural Similarity Index Measure

UQI Universal Image Quality Index

VIF Variance inflation factor

VMD Very Mild Demented

Chapter 1

Introduction

1.1 Preamble

Alzheimer's is one of many neurodegenerative diseases that make numerous nerve cells inactive in the brain. This can result in memory loss, cognitive deterioration and eventually the inability to perform everyday tasks. Although it can affect anyone, people aged 60 and above are typically the victims of dementia [18]. Dementia can be the cause of more than 50 different types of brain malfunction, where Alzheimer's is one of the most prevalent of all cases [12]. Today, 5 million people only in the United States have Alzheimer's disease [8]. Rasmussen J. and Langerman, H. stated in their article [17] that the UK estimates the population of people affected by Alzheimer's will reach 1.2 million by 2040, while the US predicts an affected population of nearly 15 million by 2060.

Due to its irreversible properties, Alzheimer's diagnosis in pre-symptomatic or early stages can help slow down the progression of the disease [16].It has also shown the improvement of quality of life when diagnosed in early stages. Alzheimer's disease often goes undiagnosed as the symptoms are so similar to those that are seen as side effects of old age. Beach et al. used longitudinal data from the National Alzheimer's Coordination Center's (NACC) research database and found that current clinical diagnostic criteria for Alzheimer's disease had a sensitivity of 71% to 87% and a specificity of 44% to 71%, implying high rates of Alzheimer's Disease misdiagnosis among patients with cognitive impairment.

Early detection can be done through training machine learning models with datasets of brain scans such as MRIs. This would allow people to find support, take control of their symptoms and be able to live without dependency for longer. However, to train machine learning models, a large amount of data is required. The biomedical realm, on the other hand, is often setback by data scarcity. We intend on using DCGANs for data augmentation in order to reduce the data deficit. Data augmentation has lately been growing in popularity in the fields of medical science. DCGANs can produce high-quality, realistic breast ultrasound images that are virtually indistinguishable from the originals [14]. They have also been used to generate new images from scans of bladder mucosa to diagnose urinary bladder cancer [27]. Using DCGAN, researchers performed data augmentation on the Chest X-ray dataset to generate artificial chest X-rays of the under-represented class class [26].

When an abundance of MR images can be generated from real brain scans, they can be used for further research.

1.2 Motivation

Over time, machine learning has been applied in various fields for various reasons. Whether it was traffic prediction or facial recognition, the uses range from being trivial to growing towards a necessity. It is already helping advance the research for detection of high-need illnesses such as cancer. However, in the search for something more impactful to a more unique group of people, this paper holds the base of neurodegenerative diseases (NDD). Some of the more known NDDs include Parkinson's disease, multiple sclerosis and Alzheimer's disease. These disorders, although not rare, lack the amount of data needed for clinical research and detection. Patients are typically reluctant to share their data with the public making it difficult to obtain large consistent datasets which were unbiased.

Unfortunately, by the time the physical symptoms of such disorders begin to show, the treatment can become more difficult. The patient undergoes changes that are irreversible in such cases, which could have been avoided if signs were detected earlier. Along with behavioral and physical changes, oncoming neurodegenerative diseases can be detected through medical imaging. However, the massive amount of data required to train machine learning models for classification are not readily available.

When looking for appropriate datasets to train for NDD classification, it was difficult to get a hold of any for multiple sclerosis or Parkinson's disease. Even though there are databases like open access series of imaging studies (OASIS), obtaining data from similar sources was a long, tedious process that still did not allow full access. With machine learning techniques like data augmentation, this data scarcity can be overcome. Medical image data can be augmented using deep convolutional generative adversarial networks (DCGANs). The initial goal was to work with one of the more rare diseases, in order to prove that classification can be done accurately with augmented data. However, the data scarcity this paper aims to overcome became the first obstacle instead. Compared to the others, Alzheimer's was more thoroughly studied in the United States. Therefore more datasets were available for the purpose of this research. As these diseases share similar properties, this paper attempts to show that diagnosis accuracy can be increased for NDDs with limited public data.

1.3 Problem Statement

Early diagnosis of Alzheimer's disease necessitates the use of effective automated approaches [22]. Computational technologies, particularly machine learning approaches, are now valuable tools for assisting and enhancing disease diagnosis and monitoring [29]. To train machine learning models, however, a vast amount of data is required. This is especially difficult in the biomedical domain because available datasets are restricted and frequently uneven due to acquisition accessibility, costs, and pathology-related variability [23].

When it comes to research for neurodegenerative diseases, there are not enough available public datasets and the scope of data sharing is absent. Many of today's major roadblocks when it comes to data sharing are societal in nature. The ownership of data is a major problem for investigators. Individuals involved in data collection are reluctant to share with the public with the fear that their efforts will go in vain. Time, money, and experience are all needed for the collecting and integration of clinical, genetic, and imaging data. Unauthorized data use or redistribution is also a point of concern [10].

Moreover, clinical studies have to overcome many hurdles such as patients' privacy, tackling the loss of data, the inability to obtain large sample sizes and more. In practice, not many patients are willing to share their personal data for public use. Even if a number of patients happen to agree, they are still reluctant to have their data be used for teaching or merely training purposes. When gathering samples that are eligible for testing, there have been instances of ethnic biases [19] which narrow down the spectrum of study. Manual testing can also result in losing data which can be costly to recover and ultimately make researchers reach incorrect conclusions. The combination of such obstacles can lead to inaccuracy. Since large sample sizes are not a realistic approach from one single source, they can be combined with data from different clinical trials. Yet the base circumstances of each of those procedures or types of testing are difficult to gather and match for larger and efficient research in this field. It is mentioned that computer-aided-diagnosis tools can be adopted to reduce misdiagnosis as they hold the risk of worsening cognitive function. With clinical datasets, it is strenuous to secure even a thousand sample patients. Previous studies have discussed how these problems can be overcome with the help of machine learning methodologies.

1.4 Research Objective

Considering the many benefits of data augmentation in addressing data scarcity and enriching biomedical research, this research aims to generate synthetic MR images. The augmented dataset can be used to improve the accuracy of classification tasks. Instead of using affine transformation approaches, this study uses a Deep Convolutional GAN for data augmentation since DCGAN produces better outcomes. The primary objectives of this research are-

1. Generating realistic samples from our obtained dataset
2. Generating datasets that can be used to train machine learning models
3. Addressing data deficiency in machine learning
4. Contributing to the growth of medical imaging data
5. Contributing to the advancement of Alzheimer's research

1.5 Thesis Structure

As mentioned above, in this paper, the goal is to overcome the problem of data deficiency while simultaneously showing that doing so will increase diagnostic accuracy of Alzheimer's. Firstly, neurodegenerative diseases were discussed with a brief mention of how it affects patients' lives. Followed by this, will be short summaries of relevant papers that did similar studies in recent years. The summaries will dive into the different models and datasets that were used as well as the outcome reached by each. After this, the third portion gives a background of the models used for this paper. This includes the base topics of generative adversarial networks and neural networks as well as their branches of DCGAN and CNN upon which the models are built. The dataset used for these models is mentioned in methodology which goes on to elaborate the python codes and libraries used. Nearing the end of the paper, how the results were obtained from the deep learning codes will be analysed in detail with diagrams for visual representation. To conclude this paper, the limitations and the scope of this research will be brought to light.

Chapter 2

Related Work

Our research has three main objectives. Firstly, we have thoroughly investigated the importance of the correct diagnosis of neurodegenerative disorders in their early stages. It is seen on multiple occasions where misdiagnosis increases the risks of fatality. The lack of data for research in these fields can be made up for through data augmentation. The proof of success by data augmentation using GANs in different fields of study is being strengthened with each novel experimentation. Based on the usage of DCGANs in medical fields such as ultrasounds [15], cancer detection [27], x-ray images [26] and others, we want to prove that it can also help increase the accuracy of machine learning models to diagnose Alzheimer's. Once the generated data is verified, it will need the proper neural network that can detect abnormalities in MRIs with utmost precision.

With a thorough understanding of the setbacks that may arise while doing further research, it is necessary to bring to light the studies that have been done in similar conditions. While some work will show a direct relation to our topic, others will be relevant in only certain areas. We have attempted to deeply analyze a small portion of the broad variety of studies in similar fields below.

The authors in [30] attempted to use multiple datasets of T1-weighted MRI scans from the ADNI, AIBL and NACC cohorts. Among the others only the ADNI dataset of 151 participants' scans were taken further into the study as protocols for obtaining MRIs were inconsistent among the other images. MRI scanners come with different magnetic strengths which are measured in Teslas. The strengths range from 0.5T (Tesla) to 3T, with the higher value providing higher quality images with more cost. Both 1.5T and 3T scans were taken from the same subjects under the same circumstances. A GAN and fully convolutional model were both trained simultaneously with different data. MRIs from participants with mild cognitive impairment were used for training the GAN. The same scans were however not used for the FCN as it only performed binary classification between normal cognition and Alzheimer's disease. The portion used for the GAN models was separated into training, testing and validation subgroups in the ratio of 3:1:1. Using this mode of image-to-image translation, this paper shows that 1.5T MRIs can be reconstructed for enhanced clarity of the early stages of Alzheimer's. This would allow predictions of cognitive status to be made with higher accuracy.

In [31] several deep learning methods were overviewed each of which used one of the three most used scans in the diagnosis of neurodegenerative diseases. The techniques include computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET). While each has their own advantages and drawbacks, MRIs are the most popular scanning technique as it has good resolution to soft tissue and does not cause ionizing radiation damage to the human body. The authors mention that deep learning models can automatically learn image features that are useful for Alzheimer's classification in comparison to other traditional models. Between unsupervised and supervised learning processes, supervised models are preferred due to their accuracy with which it is stated that convolutional neural networks (CNNs) are the most successful deep learning model. As this paper reviews classification methods, there is an emphasis on the generalization problems that may occur due to the lack of ample data that can be overcome with data augmentation.

Before diving into the deep convolutional neural network model for Alzheimer's classification [21] discusses the affected portion of the brain. AD causes the hippocampus area to decrease which shrinks the brain cortex and enlarges the ventricles. These changes are what bring about a patient's difficulty in the ability to think, remember or even perform tasks in daily life. As AD cannot be detected until at least the stage of mild cognitive impairment, the irreversible properties of the disease make it difficult to treat the patient before further damage as no cure has been found till date. The authors have applied a deep convolutional neural network which transforms 3D MRI scans into three different 2D scans. Each scan then becomes a 300x300 size image of the axial (top), coronal (side) and sagittal (back) views. The CNN layers include the ReLU activation functions, max pooling layers and the fully connected layers. An accuracy of 99.89% was reached using the proposed model to classify images into the three classes of normal cognition, mild impairment and Alzheimer's disease. They conclude that deep CNN models provide accurate predictions and help learn important features from medical imaging data.

In a study to augment chest x-rays through DCGAN, Kora Venu, S. and Ravula, S. (2020) concluded that traditional data augmentation methods result in an accuracy above 92%, while there is a clear increase of accuracy for GAN based data augmentation at 95.5%. While other research solely focused on image quality [26] was able to show a precision of 96.2% and a sensitivity of 97.7%. This recent study was able to demonstrate that even a small dataset of only 1,341 images was able to generate realistic x-ray images.

Variations of GANs such as PGGAN have also proved that MR images themselves can be generated with the dimension of 256x256 [20]. When combining PGGAN with classic data augmentation, there is increased efficiency and accuracy to the point where even experts in MRIs become unable to differentiate between real and fake images.

In a paper about image recognition to detect weather conditions [20], DCGAN was used with a combination of CNN. The augmented images turned out to be valuable for increased training data. Alzheimer's Disease Neuroimaging Initiative (ADNI), a public dataset used for Alzheimer's research, reached an average accuracy rate of almost 98% in Alzheimer's detection [25]. It is important to note that the sample size used for AD consisted of only 138 patients. This study also includes how the

CNN model used was precise enough to not misclassify any MRIs diagnosed with Parkinson's disease. This itself is one of the most crucial steps in NDD detection which further proves the importance of having a larger dataset for DA.

A study used a methodology to make synthetic structural brain networks in Multiple Sclerosis [23]. The study included 29 patients with relapsing-remitting MS and 19 patients with secondary-progressive MS. The paper's findings prove that advanced generative models can be applied directly to structural brain networks. The research shows, both quantitatively and qualitatively, that newly created data are very similar to real data. This paper supports the expected outcome of our research.

Chapter 3

Model Backgrounds

This section delves into the architecture and mechanism of the models we have utilized. Two models were employed in this paper- DCGAN and CNN.

3.1 GAN

GANs (Generative Adversarial Networks), first proposed by Ian Goodfellow in 2014 [9], have ushered in a revolution in Deep Learning, and they are now one of the most studied areas in Artificial Intelligence. GANs have been used to solve real-world problems [28]. Image generation, super resolution, and data augmentation are just a few of the applications where GANs have shown to be effective. A standard GAN system is made up of two neural networks- the generator (G) and the discriminator (D). The generator generates quality images that should match ground truth. The discriminator determines whether the image generated by the generator is real or fake [2]. Both networks are in direct competition with one another allowing them to optimize the training until a balanced state is achieved. This mathematically specifies a min max game for G and D's value function V.

$$\min_G \max_D E_{x \sim p_{\text{data}}(x)}[\log D(x)] + E_{z \sim p_{\text{generated}}(z)}[1 - \log D(G(z))] \tag{3.1}$$

Here, x is the data, z denotes the latent space from which G takes samples, and p represents the respective probability distributions. Eventually, a Nash equilibrium (Nash, 1951) is reached, in which neither the generator network G nor the discriminator network D can better their ability to distinguish between fake and real samples.

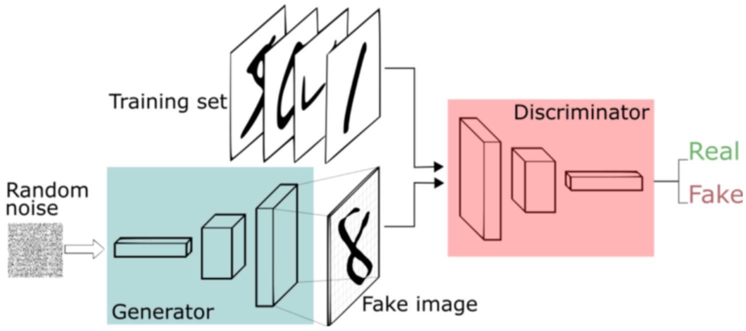

Figure 3.2: GAN Architechture

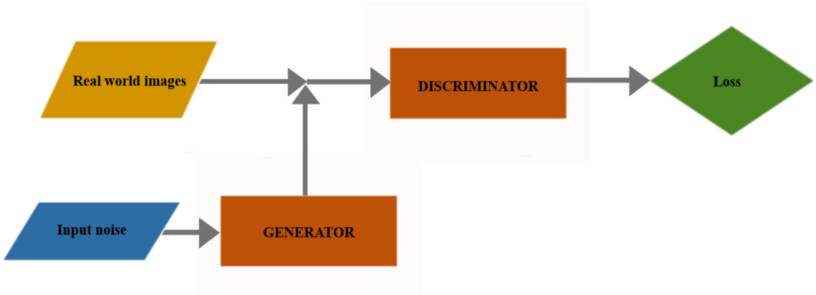

Figure 3.1: Scheme of a Generative Adversarial Network

This smart design allows GAN to learn autonomously. GAN is capable of producing high-quality music, speech, and images. The generated outputs come very close to that produced by humans. This popular network model has recently evolved and has various versions [28]. This paper uses the GAN-based system DCGAN (Deep Convolutional GAN) where the generator and the discriminator are more complex neural networks, composed of several convolutional layers.

3.2 Deep Convolutional GAN

The DCGAN is an extension of GANs which includes convolutional as well as transpose convolutional layers. It was introduced by Alec Radford in 2015 [11] as an effort to bridge the gap between the success of CNNs for supervised and unsupervised learning. Its convolutional design helps it balance GAN training. DCGAN omits the pooling layer in CNNs which allows the framework to spatially up and down sample on its own. This model compresses a picture into a vector for the discriminator part of the model. DCGAN emerged as a cutting-edge approach for creating images, sounds, and videos. Due to the layers it also takes the benefits of CNN feature extraction. DCGAN has shown promising results in real large-scale datasets such as CelebA, LSUN, and Google Image Net [13]. This paper designs a model for MRI generation using the DCGAN network structure.

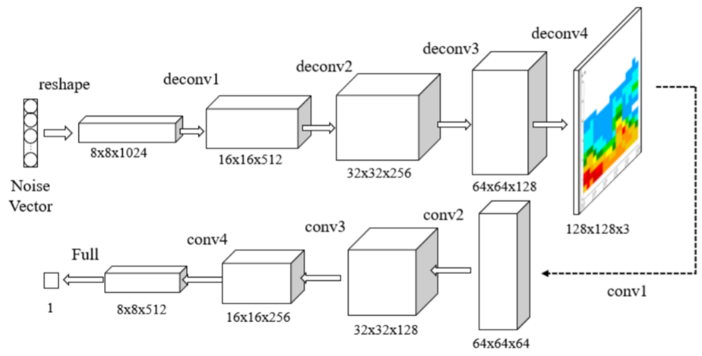

Figure 3.3: The structural association of generative model and discriminative model in DCGAN

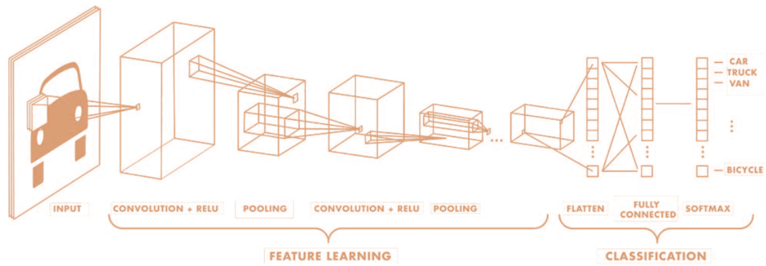

Figure 3.4: CNN Architecture

3.3 CNN

In medical applications, neural networks are becoming increasingly popular. Among existing neural networks, the one used in this paper for the purpose of MRI classification is Convolutional Neural Networks (CNN)

CNN is a sort of feed-forward neural network that enables more detailed feature extraction from acquired images. It is considered the most successful deep model among conventional networks in supervised models [31]. Unlike other methods, CNN can learn and concatenate low-level edge and shape features from a huge amount of labeled input to generate high-level semantic representations. There are three sorts of layers in CNN's architecture- convolution, pooling, and fully connected. Because the upper layers of CNN are more sensitive to semantics and the middle layers are more sensitive to underlying patterns like colors and gradients, using the upper or middle layers is a popular and effective CNN method. CNN has several different forms as a result of many practices and research. Image recognition, facial recognition, and video analysis are all applications that use CNNs.

Chapter 4

Methodologies

At the beginning of our research, our first task was to collect a dataset to feed the DCGAN to generate realistic synthetic MR images. Initially, a dataset was collected from an open source platform. The training dataset was passed through a DCGAN to generate MR images for the data augmentation task. Secondly, the generated images using 10 different image quality assessment metrics was evaluated. After that, he most realistic images were chosen and compared to real images from the original dataset using imagehash to assess the performance of a CNN that classifies the input MR images from the original dataset. It is known that deep learning models usually outperform most of the learning algorithms if enough data is given. Therefore, the generated realistic MR images can improve the performance of CNN using the augmented dataset.

4.1 Input Dataset

The dataset used in this paper was acquired from an online Kaggle challenge of MRI brain images which was uploaded in 2019. It contained images of JPG format with a typical resolution of 176 x 208 pixels. There are 5121 training images and 1279 test images in each having three color channels. The dataset consisted of two folders of images classified by 'train' and 'test' images. Each training and test dataset has four subcategories namely 'Mild demented', 'Moderate demeneted', 'Non demented' and 'Very mild demented'. The images are 2D MRI cross sections of the brain. The brain cross-section was placed in the middle with the backdrop cropped out. The images are planes of the brain taken at various heights.

We deemed this dataset ideal for this research because of the vast number of images it contains and the uniformity of the images.

4.2 Data Pre-processing

The training data was used to augment the dataset using a DCGAN. MRI data were difficult to collect since medical images are often kept private and the few datasets that were available were not easily accessible and also keeping in mind that if our proposed method works on this dataset then the method would also perform somewhat similar to the new datasets. We normalized the dataset since normalization often helps to reduce the training time. In our first experiment, we converted the (208, 176, 3) images to (64, 64, 3) images. Although satisfactory results could not be obtained, after generating the images using the original (208, 176, 3) proportions realistic MRIs began to generate. The data would look like the data shown in the figure below. The rows depicting MR images for mild demented, moderate demented, non demented and very mild demented respectively.

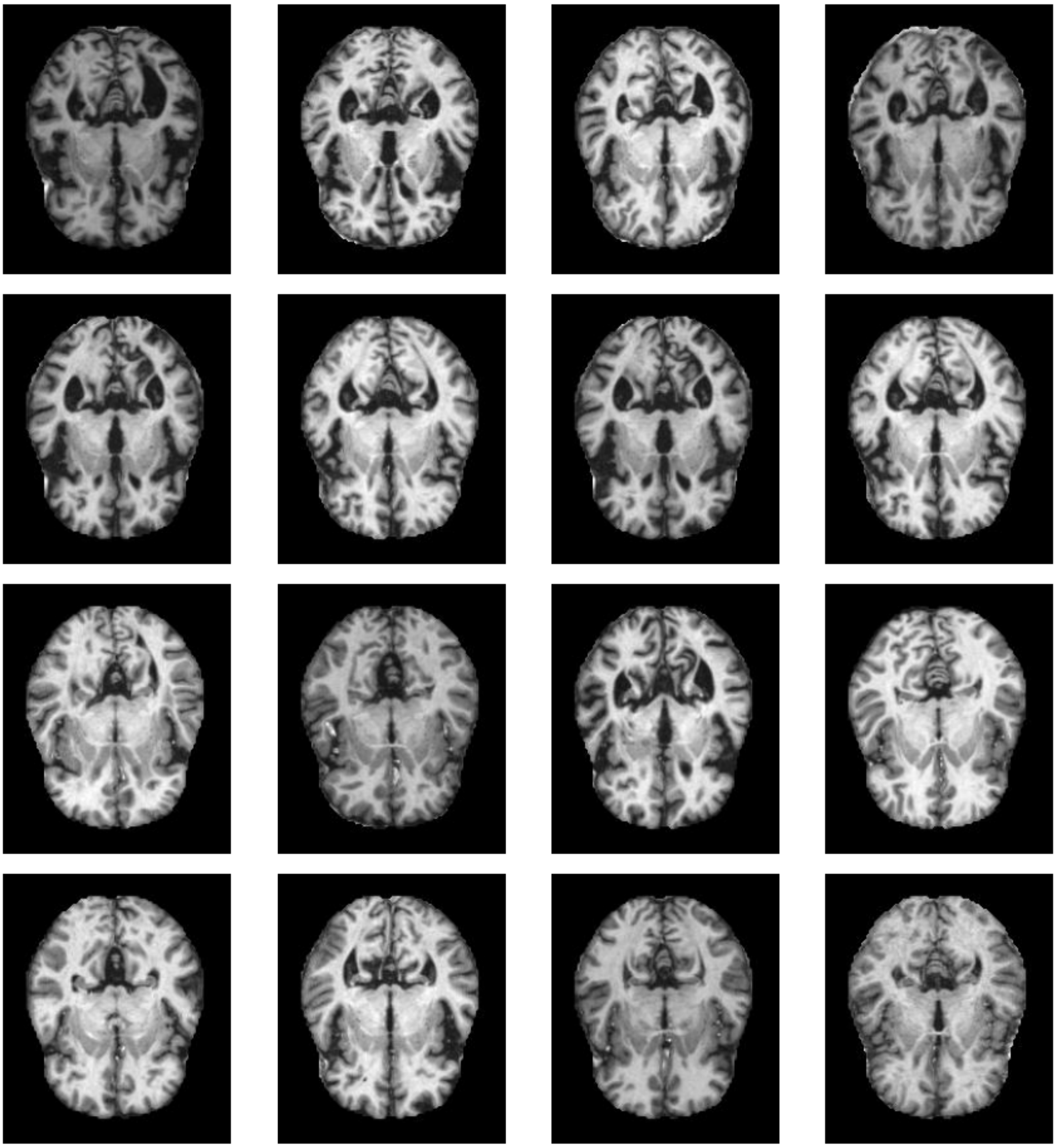

Figure 4.1: Samples from the dataset

4.3 DCGAN Implementation

A Deep Convolutional Generative Adversarial Network (DCGAN) was implemented to generate realistic MR images for data augmentation. It was found that in many image generation tasks DCGAN is widely used [25]. The DCGAN consisted of two neural networks; a discriminator and a generator, both of which compete with each other as a zero-sum or minimax game. The discriminator had three convolutional layers, three LeakyReLu, one flatten, one dropout and one dense layer. The input of the discriminator was the images from both the generated images from the generator and real images from the original dataset. Hence the input shape of the first Conv2D layer was (208, 176, 3). The output shape of the first Conv2D layer was (None, 104, 88, 64) and the total number of parameters in this layer was 3136. Number of parameters for the second and third convolution layers were 131200 and 262272. The dense layer had 73217 parameters which had an output shape of (None, 1) and the activation function for the dense layer was a sigmoid function. The total number of parameters in the discriminator was 469,825. The task of the discriminator was to tell whether an image came from the original dataset or from the generator and it gave feedback to the generator to fine tune the parameters. The discriminator also fine tuned itself at each iteration. The generator network consisted of three transpose convolutional layers, three LeakyReLu, one dense, one reshape and one Conv2D layer. The output shape of the reshape layer was (None, 26, 22, 128). The latent dimension of the input layer of the generator was 128. The output shape of the dense layer was (None, 73216) with 7394816 number of parameters. The first Conv2D transpose layers converted the shape into (None, 52, 44, 128) and the number of parameters was 524544. The second and third Conv2d transpose layers converted the shape into (None, 104, 88, 256) and (None, 208, 176, 512) respectively. The optimizers of both the generator and discriminator were Adam optimizers and learning rate for both of those were 0.0001. Binary cross entropy was used as the loss function of the model. The total number of parameters in the generator was 10,317,699. A free Nvidia K80 GPU on Kaggle was used to run the code. Since convolutional neural networks are suitable for working with images we tried to integrate that approach for the data augmentation task for our research. The images were generated 3 times for experimentation. First, (64, 64, 3) images were produced with 80 epochs. Second, (208, 176, 3) 2064 images with 172 epochs which took around 9 hours to train the DCGAN. Finally, 320 images were generated with 160 epochs, 2 images per epoch which was used to assess the quality of the generated images. It also took 9 hours to train. Below are the graphs for the discriminator and the generator architecture and the first graph is about how a GAN works.

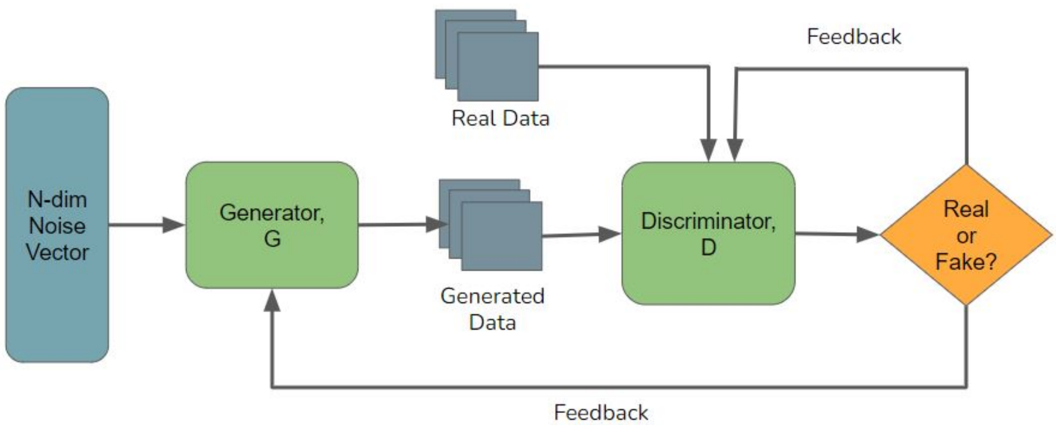

Figure 4.2: A simple GAN

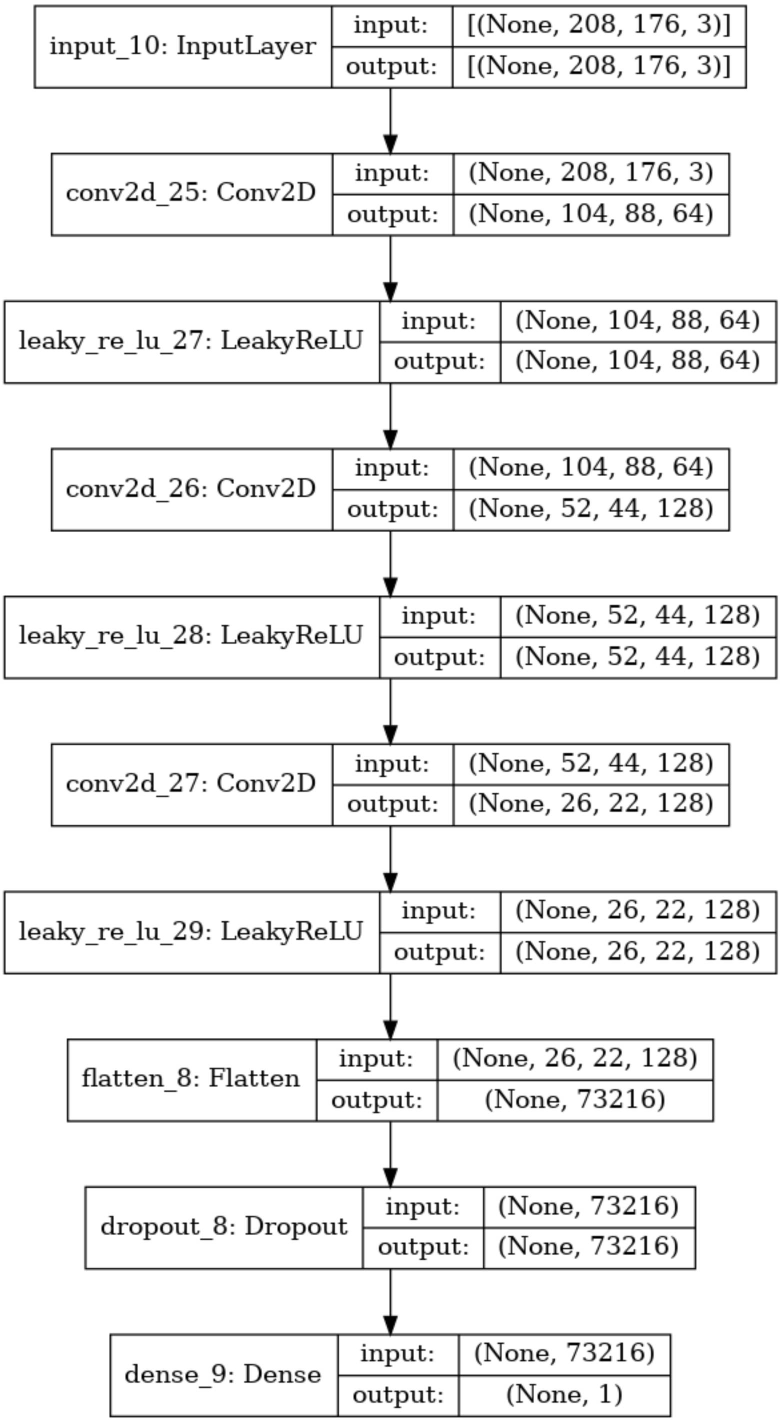

Figure 4.3: Architecture of the Discriminator of the DCGAN

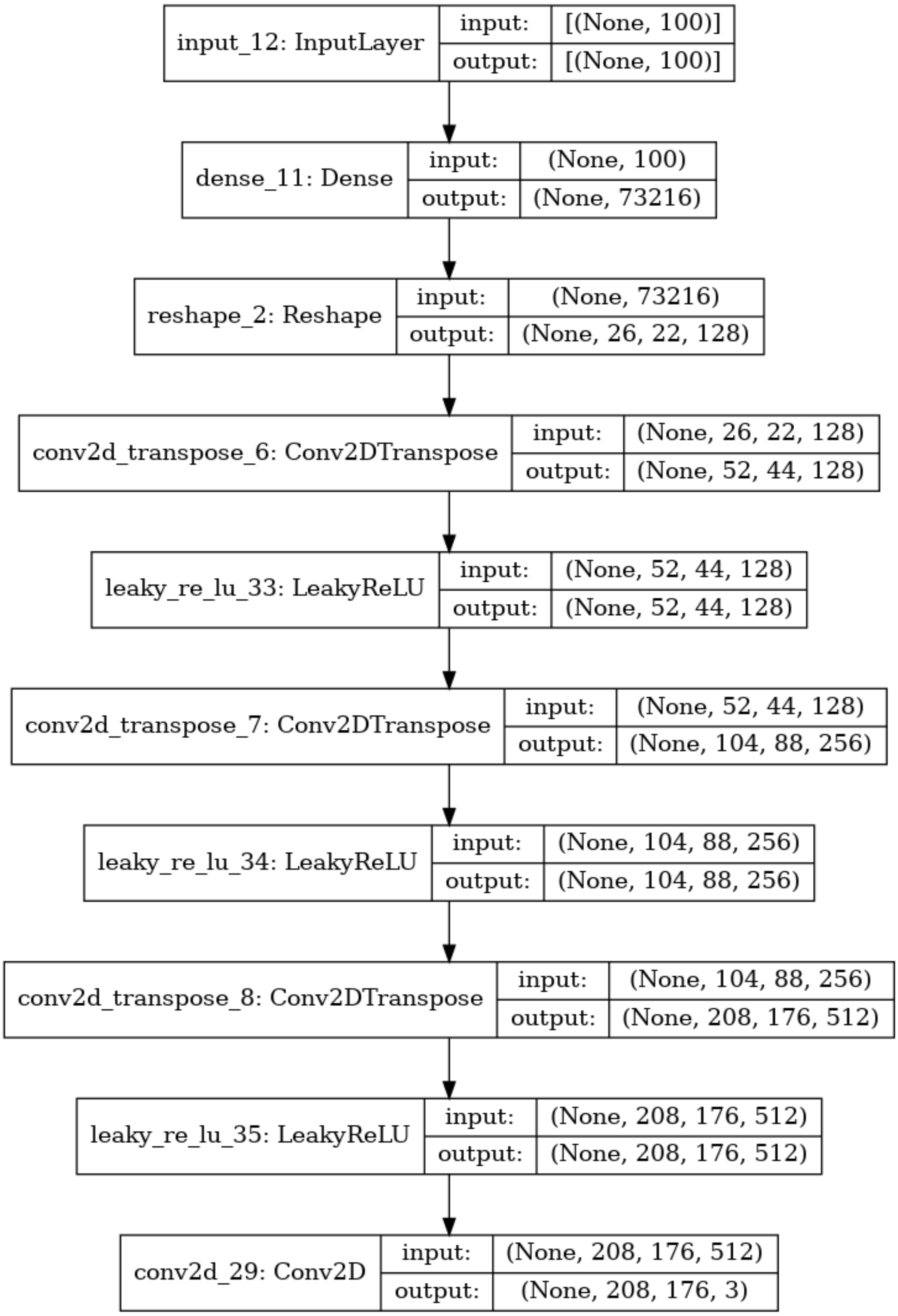

Figure 4.4: Architecture of the Generator of the DCGAN

4.4 CNN Implementation

The architecture of the CNN model transformed all the images into numpy arrays by appending it into a Python list and resizing it. The dataset was already separated into "test" and "train" sets. As classification models with several layers like this can take a very long time to train on datasets, transfer learning was used to train this model instead [24]. This sped up the process by re-using weights from two ImageNet datasets, EffecientNetb0 and ResNet50. These datasets were specifically developed for standard computer vision datasets for tasks such as image recognition. Unlike traditional CNN layers, this model uses Global Average Pooling instead of Max Pooling so that the average values are used instead of the maximum values when pooling. This helped to significantly reduce the computational load on GoogleCollab GPU which was used to run the Python code.To classify the images into one of the four classes mentioned above the softmax function is used as a generalization of the sigmoid function. The table below shows the output shape and parameter for the first ten layers of the CNN model used.

Type of Layer	Output Shape	Parameter
Input	(None, 150, 150, 3)	0
Rescaling	(None, 150, 150, 3)	0
Normalization	(None, 150, 150, 3)	7
ZeroPadding2D	(None, 151, 151, 3)	0
Conv2D	(None, 75, 75, 32)	864
BatchNormalization	(None, 75, 75, 32)	128
Activation	(None, 75, 75, 32)	0
DepthwiseConv2D	(None, 75, 75, 32)	288
Activation	(None, 75, 75, 32)	0
GlobalAveragePooling2D	(None, 32)	0
Reshape	(None, 1, 1, 32)	0
. . .	. . .	. . .

Table 4.1: CNN Model Summary: First 10 layers

One fourth of the full existing dataset was fed to the CNN for 10 epochs. When analyzing the epoch versus training validation accuracy and loss, it was clear that the lack of data created the problem of overfitting causing the large gap between the losses. The same was done using both EffecientNetb0 and ResNet50 datasets.

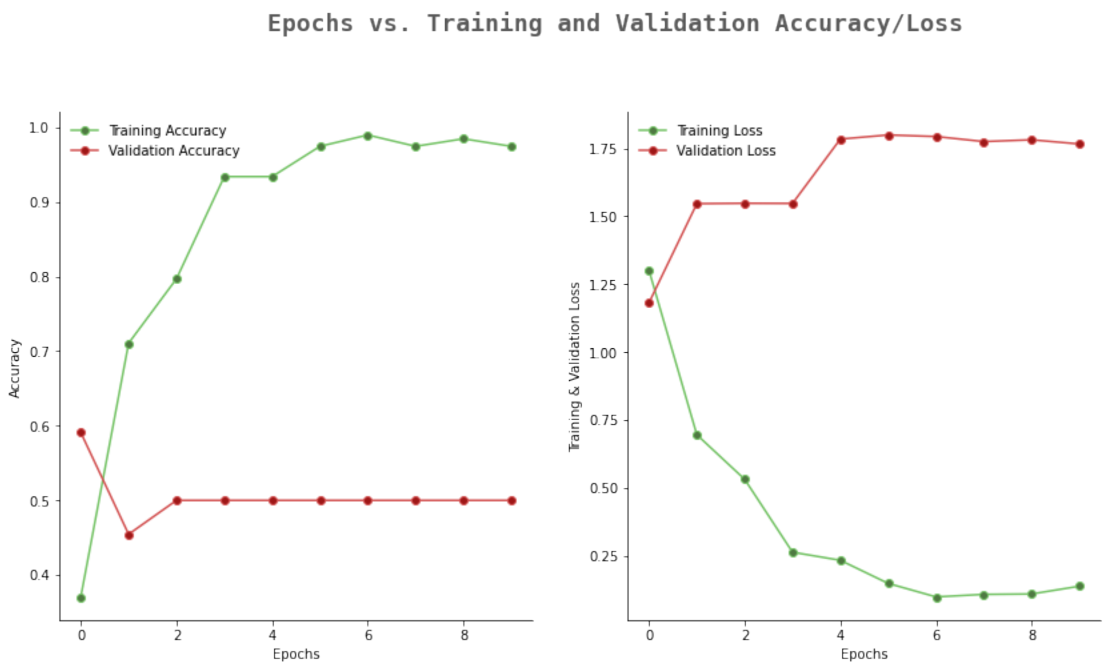

Figure 4.5: EfficientNetb0 CNN: Less Data for 10 epochs

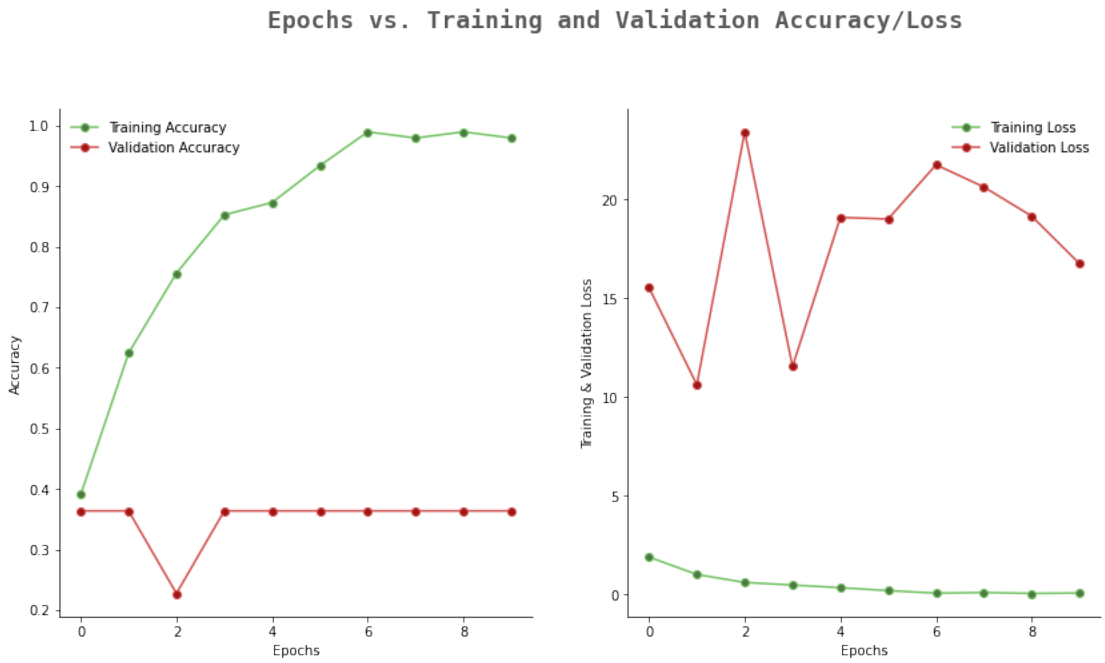

Figure 4.6: ResNet50 CNN: Less Data for 10 epochs

The full data of more than 5000 images was taken and run for the same number of epochs for both ImageNet pre-trained weights. The difference in the results can be seen clearly as the curves of training and validation loss start to meet.

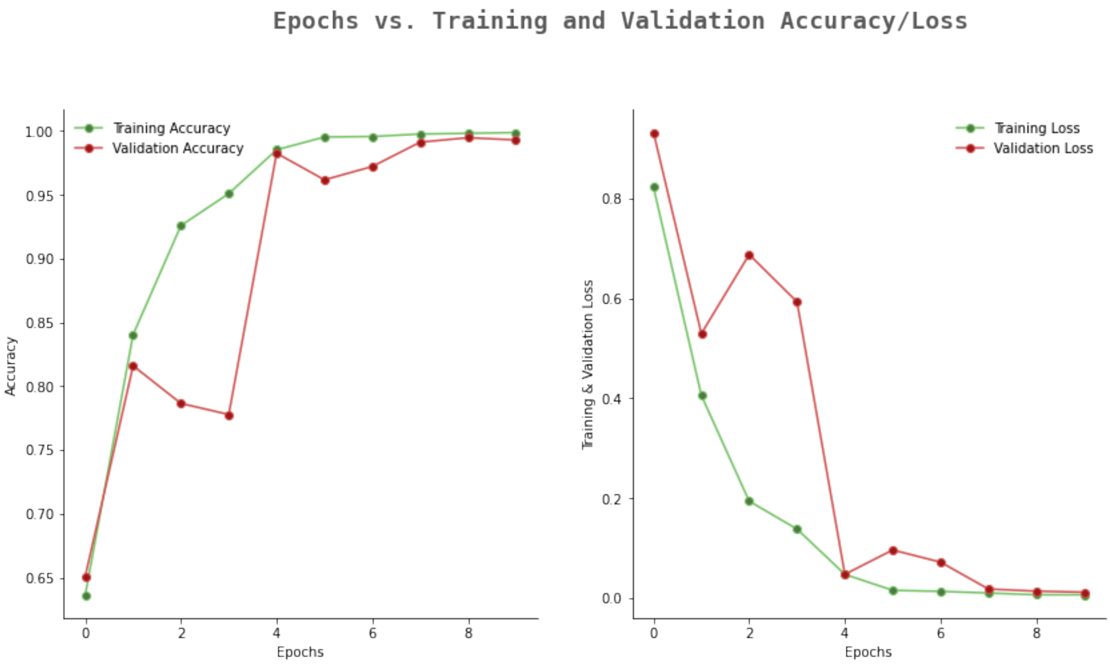

Figure 4.7: EfficientNetb0 CNN: FullData for 10 epochs

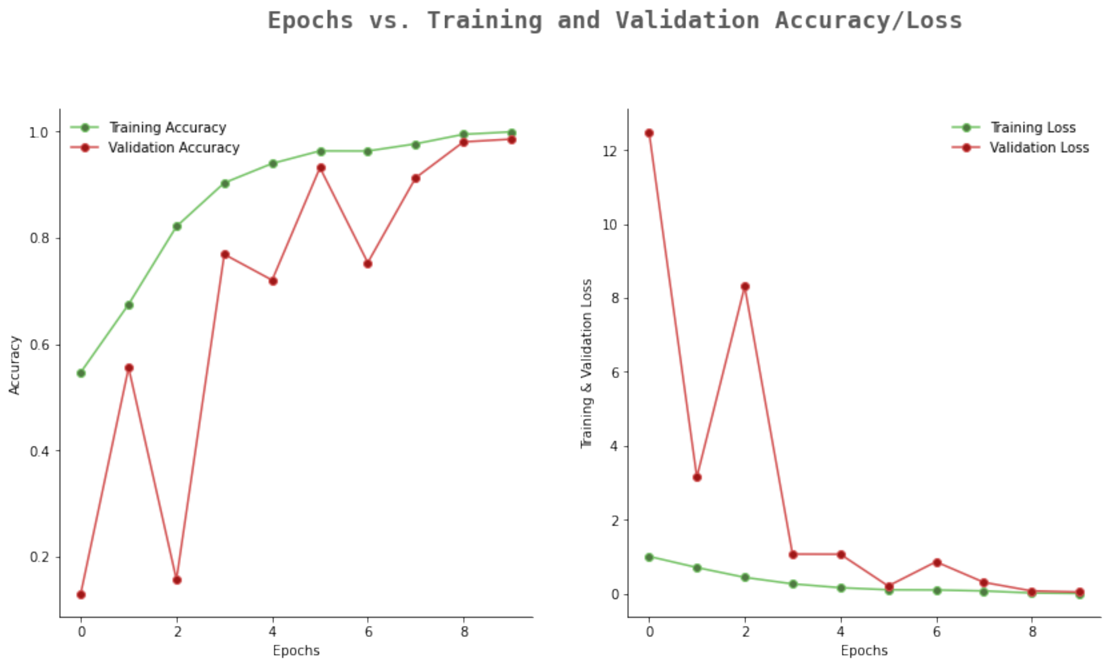

Figure 4.8: ResNet50 CNN: Full Data for 10 epochs

When the results already improved, just by using a larger dataset, the classifier was tested a little further in order to see the results for double the epochs.

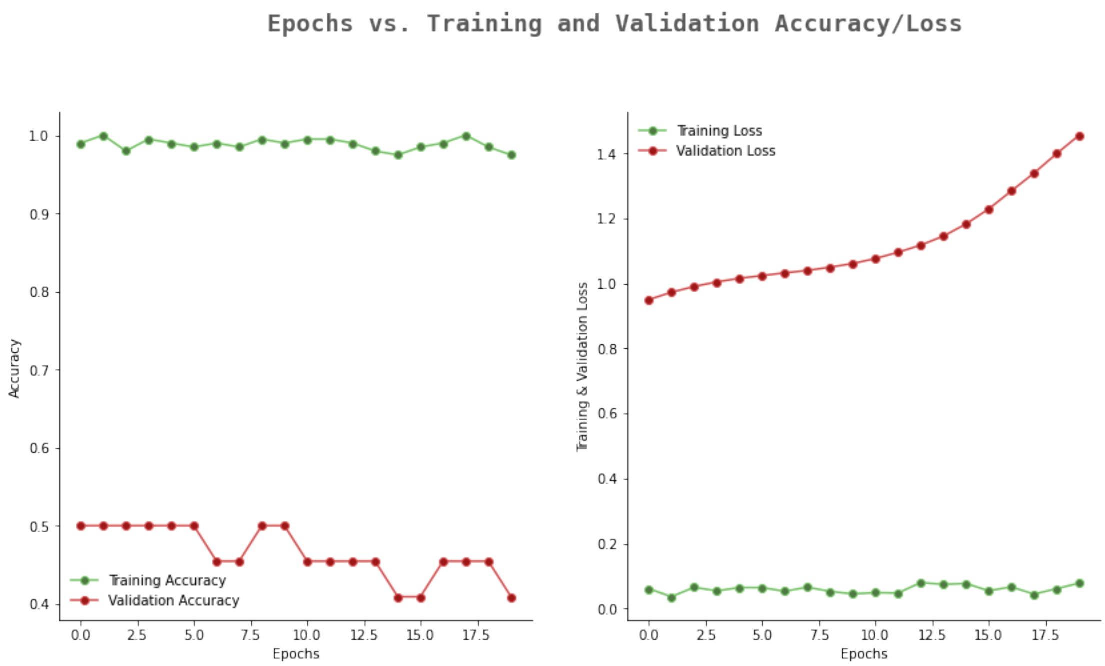

Figure 4.9: EffecientNetb0 CNN: Less Data for 20 epochs

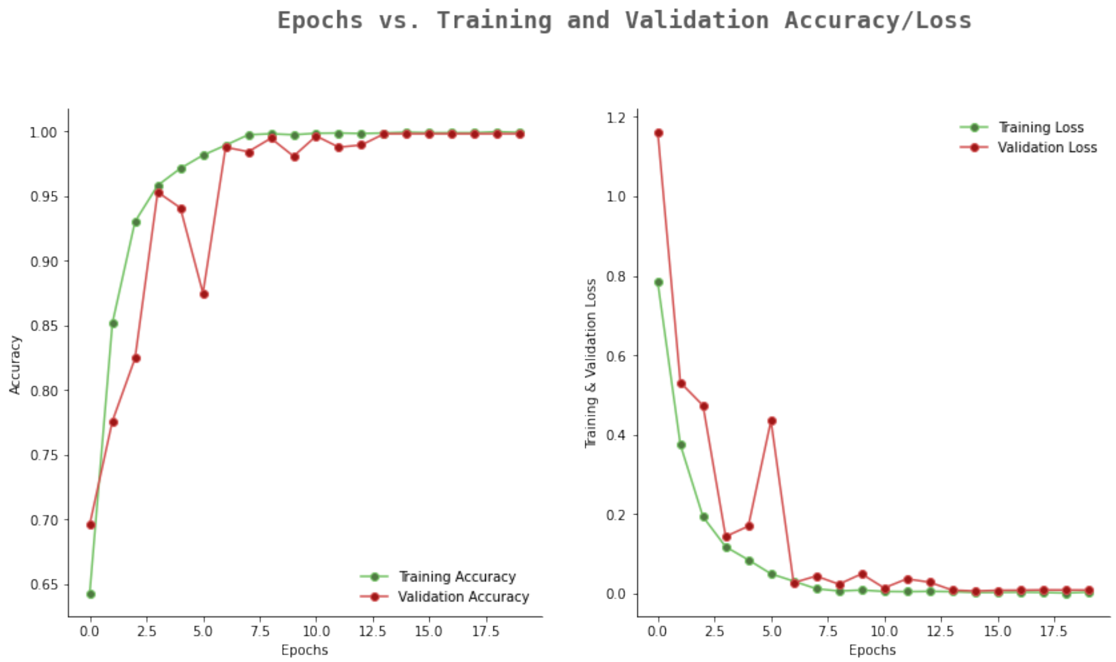

Figure 4.10: EffecientNetb0 CNN: Full Data for 20 epochs

With these test runs, the concluded results were simple. It showed that larger datasets training machine learning models for classifications yielded results with higher accuracy. With this in mind, the goal to augment data for better diagnosis rates was further solidified.

Chapter 5

Result Analysis

5.1 Performance Evaluation of the DCGAN

Training a GAN is often difficult since the discriminator and the generator play a zero-sum game, determining the number of epochs the GAN requires in order to generate high quality results. Therefore, it was experimented with different image size, batch size and number of epochs. Initially, conversion of the original images to (64, 64, 3) images was trained with the DCGAN with batch size 64 and 80 epochs. Then the DCGAN was trained with a batch size of 32 and 160 epochs. Although this study experimented with the converted MRIs of shape (64, 64, 3), that data was not used for augmenting the dataset. The converted images underwent a loss of information as they had far less pixels than the original MRIs.The input images in that case were (208, 176, 3). Below are the graphs for the loss of the discriminator and the generator for each of the above mentioned scenarios.

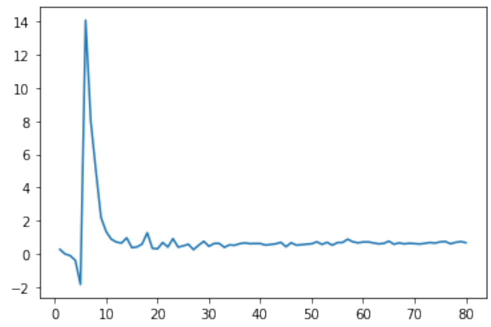

Figure 5.1: Discriminator loss (epoch 1 to 80) with batch size 64

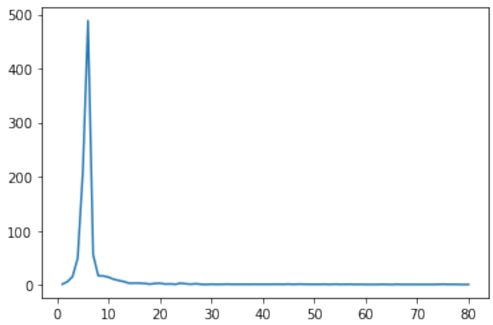

Figure 5.2: Generator loss (epoch 1 to 80) with batch size 64

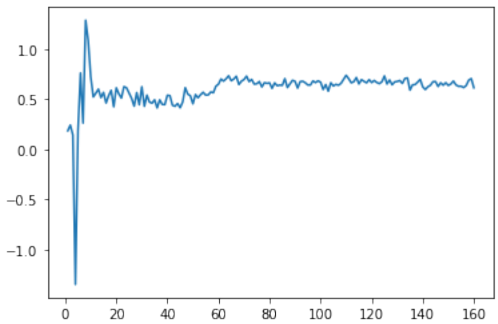

Figure 5.3: Discriminator loss (epoch 1 to 160) with batch size 32

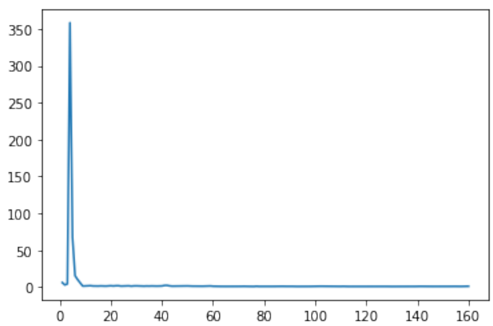

Figure 5.4: Generator loss (epoch 1 to 160) with batch size 32

As the expected results could not be obtained during the first experiment, the DC-GAN was trained with a batch size of 32 and 172 epochs. This is when realistic MRIs began to generate which reulted in a total of 2064 usable images. Since the loss of the discriminator was lower than the loss in 80 epochs training, the generator also tried to generate more realistic images. Therefore, high quality images began to generate. It took approximately 9 hours to train the DCGAN with a free Nvidia K80 GPU. Finally, 320 MRIs were generated with batch size 32 and 160 epochs. A total of 2384 (208, 176, 3) MR images using the DCGAN were finally generated. The graph below shows the generated images at each epoch for both training from epoch 1 to 80 and 1 to 160 with (208, 176, 3) pixel images.

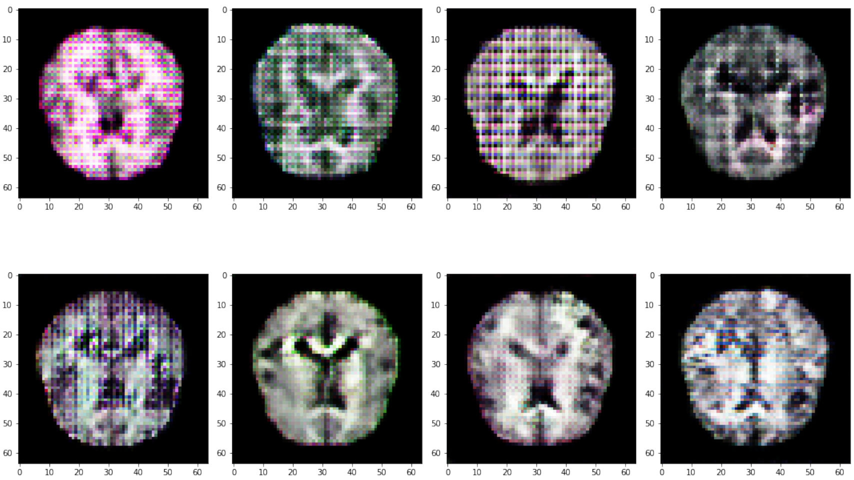

Figure 5.5: Development of output images by number of epochs (10 to 80)

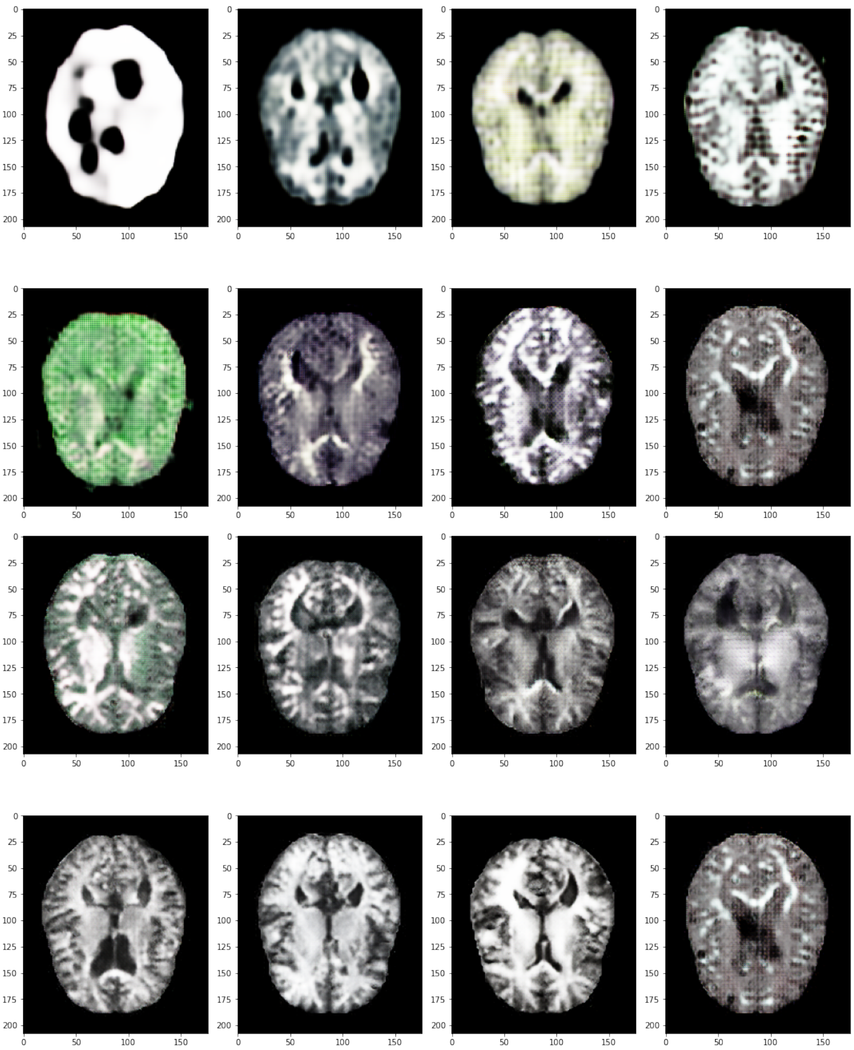

Figure 5.6: Development of generated images by number of epochs passing (10 to 160)

5.2 Generated Image Quality Assessment

Using the Deep Convolutional GAN a total of 2384 MRIs was generated and then the quality was assessed by 10 different image quality assessment metrics. The images were analyzed by using an open source package for image quality assessment named Sewar. These metrics were used for assessing the quality of generated images.

5.2.1 UQI

The quality of MRIs was assessed using UQI(Universal Image Quality Index) metrics [4]. UQI evaluates the quality of images by looking at the loss of correlation, luminance and contrast distortion etc. After evaluating the generated MRIs they displayed an average UQI score of 0.7713. The minimum and maximum UQI scores from the generated images were 0.1332 and 0.84 respectively. Below is the graph of UQI scores for each of the 320 generated images from the third training phase of the DCGAN.

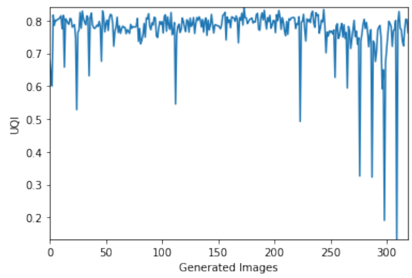

Figure 5.7: UQI scores of generated images.

5.2.2 MSE

Mean Squared Error or MSE was also used to assess the image quality. It's a commonly used technique to find errors that calculate the mean squared deviation of an estimator or how much deviated the output is from the expectation. The average MSE score for the generated images was 2857.33. The minimum and maximum MSE scores were 1812.12 and 11629.22 respectively. Here is the MSE score graph for each of 320 generated MRIs.

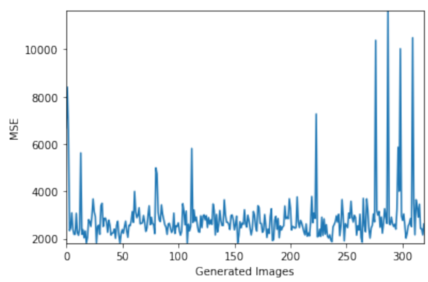

Figure 5.8: MSE scores of generated images.

5.2.3 RMSE

RMSE or Root Mean Squared Error is the square root of the MSE. It estimates the difference between two values. The average RMSE value of our generated MRIs was 52.5. The minimum RMSE score was 41.1 and the maximum was 104.91. Here is the graph for RMSE scores of the synthetic MRIs.

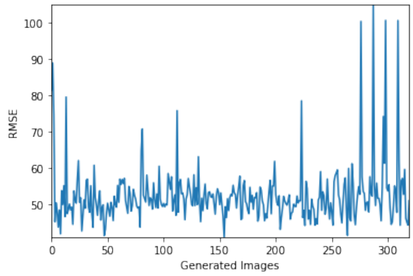

Figure 5.9: RMSE scores of generated images.

5.2.4 PSNR

Peak Signal-to-Noise Ratio or PSNR is a ratio between the maximum possible power of an image and the power of noise that affects the quality. [6] PSNR is expressed as a logarithmic quantity and is easily defined via MSE. The average PSNR value was 13.83. Whereas minimum and maximum values of PSNR for the generated MRIs were 7.13 and 15.87 respectively.

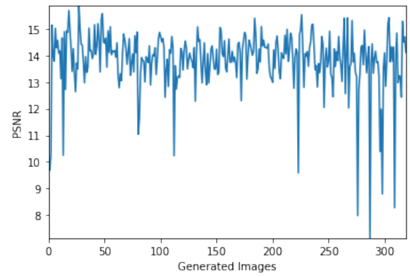

Figure 5.10: PSNR scores of generated images.

5.2.5 SSIM

Structural Similarity Index Measure (SSIM) is an image quality assessment method that takes an image as reference and measures the quality of a given image. [6] The SSIM values range from 0 to 1 where 1 means perfect and if the values approaches 0 that means the image is not good enough. Usually SSIM values for good quality image reconstruction are around 0.9. Though we got 0.523 as the average SSIM value of our generated dataset. It is a fairly good SSIM value for generated MRIs and there is hope that the images can be improved by training the DCGAN with more epochs to get the best parameters for the image generation task. The minimum SSIM value among all the images was 0.111. Whereas the maximum was 0.6. Here is the graph of the SSIM values for the generated MRIs.

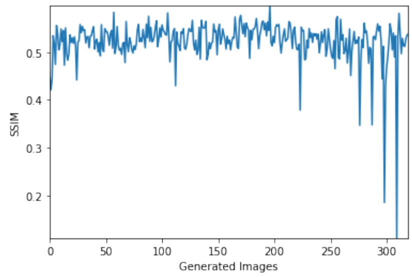

Figure 5.11: SSIM scores of generated images.

5.2.6 MSSSIM

Multiscale Structural Similarity Index Measure (MSSSIM) is a more advanced image quality assessment method that uses a process of multiple stages of subsampling that is conducted over multiple scales. [5] It's a variant of SSIM but performs better than that. The average MSSSIM value was 0.53 with a maximum of 0.66.

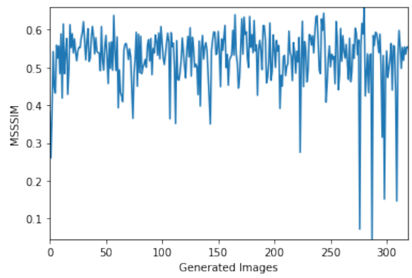

Figure 5.12: MSSSIM scores of generated images.

5.2.7 ERGAS

Erreur Relative Globale Adimensionnelle de Synthèse or ERGAS [3] was used to assess the generated MRIs and got an average of 56381.5 was obtained. Whereas the minimum and maximum values are respectively 36092.96 and 327175.42.

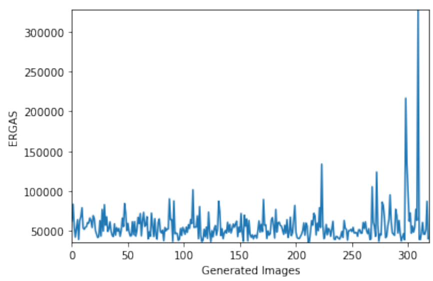

Figure 5.13: ERGAS scores of generated images.

5.2.8 SCC

Spatial Correlation Coefficient (SCC) is a measure where images are expressed in terms of the correlation coefficient. [2] The correlation coefficient tells whether the images are correlated or not. Generated images should have a positive SCC value and we got 0.05 as our average SCC value for the generated MRIs.

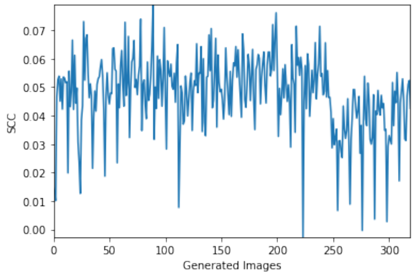

Figure 5.14: SCC scores of generated images.

5.2.9 SAM

Spectral Angle Mapper or SAM is a spectral classification that matches pixels using n-D angle with reference spectra. [1] The average SAM value was 0.47 with a minimum and maximum value 0.37 and 1.46 respectively.

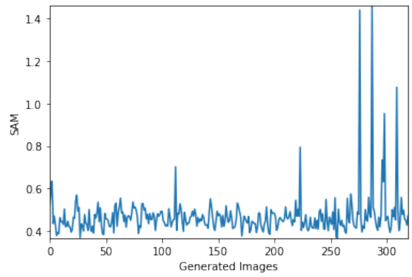

Figure 5.15: SAM scores of generated images.

5.2.10 VIF

VIF is based on Natural Scene Statistics (NSS). [7] The average VIF score was 0.09. The minimum and maximum VIF scores were 0.002 and 0.113 respectively.

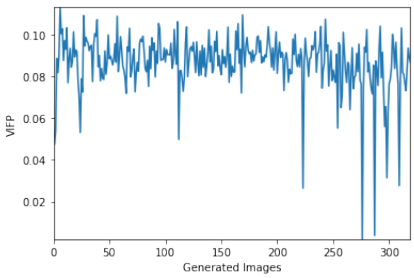

Figure 5.16: VIF scores of generated images.

The table shows the average values for each of the metrics used with the minimum and maximum values.

Metrics	Average	Minimum	Maximum
UQI	0.7713	0.1332	0.8407
MSE	2857.3290	1812.1200	11629.2233
RMSE	52.4996	41.0997	104.9130
PSNR	13.8253	7.1343	15.867
SSIM	0.5227	0.111	0.597
MSSSIM	0.5260	0.0451	0.6595
ERCAS	56381.4915	36092.9596	327175.417
SCC	0.0475	-0.0027	0.0789
SAM	0.4696	0.3672	1.4580
VIF	0.0868	0.0019	0.1133

Table 5.1: Metric Comparison

The original and generated images are given below side by side to compare the generated images with the original images.

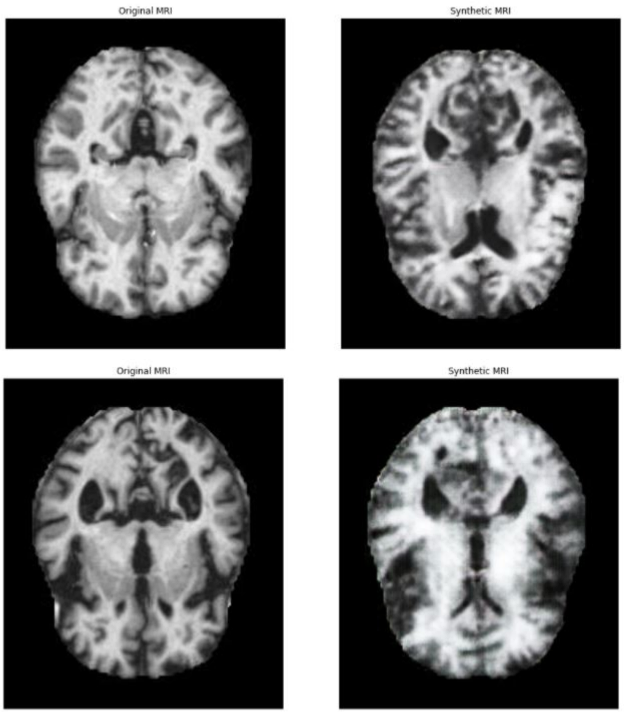

Figure 5.17: Original and generated MRIs side by side

It is important to note that the generated images were analyzed by comparing them with the original images. The original images were chosen randomly from each of the 4 classes at each iteration to avoid a bias. Therefore, the average of each metric value should be adjusted to an approximated value. At the end of the image generation task the images were further assessed using Imagehash with different cutoff values and removed some images from our generated dataset to keep only the best images. This table shows cutoff values and percentage of images removed.

Cutoff	Removed MRIs	Perfect MRIs
1	15%	85%
2	34.375%	65.625%
3	74.6875%	25.3125%

Table 5.2: Cutoff Values vs. The Percentage of Selected Images

Chapter 6

Conclusion

Usually affine transformation techniques are used for improving the performance of neural networks but the transformed data does not have much difference with the original samples. Therefore, using those techniques cannot ensure better results as expected in the medical imaging tasks. A Deep Convolutional GAN was used to generate the synthetic MRIs that were heavily correlated with the original MR images. Therefore, the proposed DCGAN based data augmentation can ensure a better generalization of deep neural networks. This research aims to solve the data deficiency problem by augmenting the original dataset. Since the generated MRIs are adequately realistic and have sufficient scores in different image quality assessment metrics there is hope that these generated images can improve the medical imaging classification accuracy. This research, like most, was no stranger to obstacles that limit its possibilities. But this paper deems acknowledgement of these limitations indispensable for advancing future work in this domain. The initial stumbling block in this research was obtaining datasets. Despite the severity of Alzheimer's disease and the need for its early diagnosis, MRI datasets are not readily available to the public for a multitude of reasons. The problem of data scarcity is even worse for other Neurodegenerative Disorders. Fortunately, the dataset finally acquired for the purpose of this research yielded good quality training image datasets allowing the CNNs to perform better with the training data. However, it would have been preferable if the data came with the details of patient demography along with the MRI scanning strength. Given that there were some limitations, the generated datset could not be validated. However, the quality of the generated MRIs was sufficient, as expected. Therefore, the generated high fidelity MRIs could improve the performance of any relevant medical image classification task.

Bibliography

[1] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," 1992.

[2] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge landsat tm and spot panchromatic data," *International Journal of Remote Sensing*, vol. 19, no. 4, pp. 743–757, 1998. DOI: 10.1080/014311698215973. eprint: https://doi.org/10.1080/014311698215973. [Online]. Available: https://doi.org/10.1080/014311698215973.

[3] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion ?" In *Third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, T. Ranchin and L. Wald, Eds., Sophia Antipolis, France: SEE/URISCA, Jan. 2000, pp. 99–103. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00395027.

[4] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002. DOI: 10.1109/97.995823.

[5] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, 2003, 1398–1402 Vol.2. DOI: 10.1109/ACSSC.2003.1292216.

[6] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.

[7] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006. DOI: 10.1109/TIP.2005.859378.

[8] H. N. Center, *The challenge of neurodegenerative diseases*, 2012.

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, *Generative adversarial networks*, 2014. arXiv: 1406.2661 `[stat.ML]`.

[10] N. Ashish, P. Bhatt, and A. W. Toga, "Global Data Sharing in Alzheimer Disease Research," *Alzheimer Dis Assoc Disord*, vol. 30, no. 2, pp. 160–168, 2016.

[11] A. Radford, L. Metz, and S. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks*, 2016. arXiv: 1511.06434 `[cs.LG]`.

[12] M. W. Bondi, E. C. Edmonds, and D. P. Salmon, "Alzheimer's disease: Past, present, and future," *Journal of the International Neuropsychological Society*, vol. 23, no. 9-10, pp. 818–831, 2017.

[13] W. Fang, F. Zhang, V. S. Sheng, and Y. Ding, "A method for improving cnn-based image recognition using dcgan," *Computers, Materials and Continua*, vol. 57, no. 1, pp. 167–178, 2018.
[14] T. Fujioka, M. Mori, K. Kubota, *et al.*, "Breast Ultrasound Image Synthesis using Deep Convolutional Generative Adversarial Networks," *Diagnostics (Basel)*, vol. 9, no. 4, Nov. 2019.
[15] T. Fujioka, M. Mori, K. Kubota, *et al.*, "Breast ultrasound image synthesis using deep convolutional generative adversarial networks," *Diagnostics*, vol. 9, no. 4, 2019, ISSN: 2075-4418. [Online]. Available: https://www.mdpi.com/2075-4418/9/4/176.
[16] T. Jo, K. Nho, and A. J. Saykin, "Deep learning in alzheimer's disease: Diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers in aging neuroscience*, vol. 11, p. 220, 2019.
[17] J. Rasmussen and H. Langerman, "Alzheimer's disease–why we need early diagnosis," *Degenerative neurological and neuromuscular disease*, vol. 9, p. 123, 2019.
[18] "Alzheimer's disease." (2020), [Online]. Available: https://www.nih.gov/about-nih/what-we-do/nih-turning-discovery-into-health/alzheimers-disease.
[19] S. Golriz Khatami, S. Mubeen, and M. Hofmann-Apitius, "Data science in neurodegenerative disease: its capabilities, limitations, and perspectives," *Curr Opin Neurol*, vol. 33, no. 2, pp. 249–254, Apr. 2020.
[20] C. Han, L. Rundo, R. Araki, *et al.*, "Infinite brain mr images: Pggan-based data augmentation for tumor detection," in *Neural Approaches to Dynamics of Signal Exchanges*, A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero, Eds. Singapore: Springer Singapore, 2020, pp. 291–303, ISBN: 978-981-13-8950-4. DOI: 10.1007/978-981-13-8950-4_27. [Online]. Available: https://doi.org/10.1007/978-981-13-8950-4_27.
[21] A. Nawaz, S. M. Anwar, R. Liaqat, J. Iqbal, U. Bagci, and M. Majid, "Deep convolutional neural network based classification of alzheimer's disease using mri data," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 2020, pp. 1–6. DOI: 10.1109/INMIC50486.2020.9318172.
[22] M. Tanveer, B. Richhariya, R. U. Khan, *et al.*, "Machine learning techniques for the diagnosis of alzheimer's disease: A review," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 1s, Apr. 2020, ISSN: 1551-6857. DOI: 10.1145/3344998. [Online]. Available: https://doi.org/10.1145/3344998.
[23] B. Barile, A. Marzullo, C. Stamile, F. Durand-Dubief, and D. Sappey-Marinier, "Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis," *Computer Methods and Programs in Biomedicine*, vol. 206, p. 106 113, 2021, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2021.106113. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260721001887.
[24] "Brain tumor mri classification." (2021), [Online]. Available: https://www.kaggle.com/jaykumar1607/brain-tumor-mri-classification-tensorflow-cnn/notebook.
[25] "Generating fake faces with dcgan." (2021), [Online]. Available: https://github.com/AarohiSingla/DCGAN-Generating-Fake-Faces-with-DCGAN.
[26] S. Kora Venu and S. Ravula, "Evaluation of deep convolutional generative adversarial networks for data augmentation of chest x-ray images," *Future In-*

ternet, vol. 13, no. 1, 2021, ISSN: 1999-5903. DOI: 10.3390/fi13010008. [Online]. Available: https://www.mdpi.com/1999-5903/13/1/8.

[27] I. Lorencin, S. Baressi Šegota, N. Anđelić, *et al.*, "On urinary bladder cancer diagnosis: Utilization of deep convolutional generative adversarial networks for data augmentation," *Biology*, vol. 10, no. 3, 2021, ISSN: 2079-7737. DOI: 10.3390/biology10030175. [Online]. Available: https://www.mdpi.com/2079-7737/10/3/175.

[28] A. Öcal and L. Özbakır, "Supervised deep convolutional generative adversarial networks," *Neurocomputing*, vol. 449, pp. 389–398, 2021, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2021.03.125. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231221005178.

[29] A.-M. Tăuţan, B. Ionescu, and E. Santarnecchi, "Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques," *Artificial Intelligence in Medicine*, vol. 117, p. 102 081, 2021, ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2021.102081. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365721000749.

[30] X. Zhou, S. Qiu, P. S. Joshi, *et al.*, "Enhancing magnetic resonance imaging-driven alzheimer's disease classification performance using generative adversarial learning," *Alzheimer's research & therapy*, vol. 13, no. 1, pp. 1–11, 2021.

[31] S. Gao and D. Lima, "A review of the application of deep learning in the detection of alzheimer's disease," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 1–8, 2022, ISSN: 2666-3074. DOI: https://doi.org/10.1016/j.ijcce.2021.12.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666307421000280.