A Deep Learning Approach to Brain Age Prediction in Schizophrenia, Parkinson's Disease and Post-Traumatic Stress Disorder

Author: Valeria Gomez Ramirez Supervisors: Stuart Allen and Matthias Treder

School of Computer Science and Informatics Cardiff University May 2021

Abstract

Recent publications have shown that training supervised regression methods on MRI brain imaging can be used to predict the brain age of an individual with high precision. We can use these predictions to detect diseases associated with abnormal brain ageing where the predicted age does not match the chronological age.

In this paper, we develop a convolutional neural network to predict brain age accurately. The architecture of the model is a simplified adaptation of the VGG architecture. The network is trained on healthy grey-matter segmented images and applied to clinical T1-weighted MRIs.

The model is trained on a publicly available healthy dataset and applied to a clinical dataset consisting of Schizophrenia, Parkinson's Disease, and Post-Traumatic Stress Disorder patients. We demonstrated bias in brain age prediction, and we corrected it to improve the reliability of the results. Our BrainAge model obtained a mean absolute error (MAE) of 4.03 years and 0.96 R² on the healthy dataset after correcting the bias. We used transfer learning to apply the BrainAge model to the clinical data and compared the brain age delta (predicted age – chronological age) for each condition. The results were not statistically significant p<.05, meaning that the brain age delta does not indicate abnormal brain ageing in this instance.

Acknowledgements

I would like to firstly thank both of my supervisors, Dr Matthias Treder and Prof Stuart Allen, for their guidance in this project. I am very grateful for all the help I received when I had any issues or questions about my research.

I would also like to thank Dr Stéfan Du Plessis and Dr Kamen Tsvetanov for granting me access to their datasets and helping me with issues related to the data.

I thank Dr Ann-Marie de Lange for pointing me in the right direction when I ran into problems implementing her brain bias correction equation.

Table of Contents

1. Ir	ntroduction	1
1.1.	Overview	1
1.2.	Aims and objectives	3
1.3.	Methodology Outline	3
2. B	ackground	4
2.1.	Brain Ageing	4
2.2.	Bias in Brain Age Predictions	5
2.3.	Brain Imaging Formats	6
2.4.	Machine Learning in Brain Age Prediction	7
2.5.	Deep Learning in Brain Age Prediction	8
2.6.	Regularisation	10
2.7.	Improving performance	10
2.	.7.1. Loss functions	10
2.	.7.2. Optimisation function	11
2.	.7.3. Performance metrics	11
2.8.	Analysis of Variance (ANOVA)	12
2.9.	Well-known Convolutional Neural Network Architectures	12
2.10	D. Software and Libraries	13
2.11	. Cam-CAN and Shared Roots datasets	15
2.12	2. Transfer Learning	16
3. M	Iethodology	16
3.1.	Data pre-processing	17
3.2.	Creating the model	18
3.3.	Measuring errors	20
3.4.	Application to clinical data	21
3.5.	Correction of brain age bias	21
3.6.	Analysis and statistics	22
4. R	esults and Evaluation	22
4.1.	Early Development Results	22
4	.1.1. Effect of activation function	23
4	.1.2. Effect of MRI size	23
4	.1.3. Effect of dropout	23

Z	4.1.4. Comparison with LeNet and VGG	24
Z	4.1.5. Summary	25
4.2	2. Testing the BrainAge model on clinical data	25
4.3	B. Discussion	28
5. I	Future work	29
5.1	. More data and same format	30
5.2	2. Data Augmentation	30
5.3	. Follow-up study on subjects with largest brain age delta	30
5.4	. Multivariate and Qualitative Analysis	31
5.5	. Further Optimisation	31
5.6	. Different Network Architecture	31
6. (Conclusions	32
7. I	Reflection on Learning	32
8. I	Bibliography	37
9. <i>A</i>	Appendix	43

List of Figures

Figure 1: Part of the ensemble architecture proposed by Dinsdale et al., (2021) which inspired our model 2 Figure 2: The differences between a T1-weighted image (left)(Preston, 2016) and a Grey-matter volumetric map (right)(Cam-CAN, 2011) 6 Figure 3: A depiction of the LeNet-5 architecture, as illustrated in its original paper (Lecun et al., 1998) 13 Figure 4: A depiction of the VGG-16 model architecture (Nash, Drummond and Birbilis, 2018) 13 Figure 5: Distribution of ages per gender in Cam-CAN (left) and Shared Roots (right) datasets 16 Figure 6: Brain slices showing the difference in dimensions and detail between the low-res (left) and high-res (right) images, to scale 17 Figure 7: Architecture of the proposed BrainAge model. The picture reflects the structure of the low-res model though the high-res remains the same structure with different filter sizes. The model shows five convolutional blocks, each containing a convolutional layer (orange), followed by batch-normalisation (yellow) and max pooling (green). 19 Figure 8: Graphs of MAE for each model, from left to right: BrainAge Model, LeNet and VGG 25 Figure 9: Demonstration of brain age bias; young subjects are overpredicted, old subjects under predicted 26 Figure 10: Corrected and uncorrected predicted ages against chronological ages 26

Figure 11: Both graphs plot the predicted ages against chronological ages. The dashed black line represents the identity line (x=y). The left graph shows the difference between all patients and all controls. On the right, the graph details the differences between each clinical group. 27

List of Tables

Table 1: Shared Roots data distribution __________15Table 2: Exploring the effect of using different activation functions, namely ReLU andELU. The best results for the low-res and high-res images are in bold. ________23Table 3: Results from different dropout factors (no dropout, 0.2, 0.5). Best result inbold. ________24Table 4: Comparison of the BrainAge model against LeNet and VGG. All models weretrained on the low-res images and used ReLU activation functions. Best results shownin bold. ________24Table 5: MAE and mean brain age delta results for the different clinical groups. Thehealthy control group is included for comparison. ________27

Table of Abbreviations

AD: Alzheimer's Disease

API: Application Programming Interface

BPD: Bipolar Disorder

CNN: Convolutional Neural Network

CPU: Central Processing Unit

CSF: Cerebrospinal Fluid

DNN: Deep Neural Network

DL: Deep Learning

DTI: Diffusion Tensor Imaging

ELU: Exponential Linear Unit

GM: Grey Matter

GPR: Gaussian Process Regression

GPU: Graphical Processing Unit

MAE: Mean Absolute Error

ML: Machine Learning

MNI: Montreal Neurological Institute (spatial normalisation template for MRIs)

MRI: Magnetic Resonance Imaging

MSE: Mean Squared Error

OLS: Ordinary Least Squares PCA: Principal Component Analysis PD: Parkinson's Disease PTSD: Post Traumatic Stress Disorder **ReLU: Rectified Linear Unit RMSE:** Root Mean Squared Error **RVM: Relevance Vector Machine RVR:** Relevance Vector Regression SCP: Secure Copy Protocol SD: Standard Deviation SFTP: Secure File Transfer Protocol SGL: Sparse Group Lasso SSH: Secure Shell Protocol37 SZ: Schizophrenia **TBI:** Traumatic Brain Injury VGG: Visual Geometry Group WM: White Matter

1. Introduction

We begin by presenting a general outline of the problem we aim to solve, followed by our aims and a brief summary of the methodology employed.

1.1. Overview

Ageing of the brain is a complex biological process, and unfortunately, when signs of cognitive decline become obvious, it is often too late to provide adequate treatment. Increased brain age poses a risk of neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease, as well as correlating with higher mortality rates (Levakov *et al.*, 2020). Having a method that could predict abnormal brain ageing would be a valuable tool in detecting early signs of brain deterioration that could be missed by experienced neurologists and it would allow for treatment of said abnormalities before symptoms become visible.

In recent years, machine learning techniques have enabled automatic disease prediction from imaging data. The aim is to increase the prediction accuracy beyond human performance to assist in clinical diagnosis and treatment decisions. The predicted age from these techniques can be considered to be the brain age because it is purely derived from the imaging data. However, it is not just the brain age that it is relevant, it is the difference between the predicted age and the chronological age – known as brain-age delta – that matters. This value can provide very significant insight into the ageing speed of an individual. A positive delta implies that a subject's brain looks older than their chronological age, meaning they are experiencing accelerated ageing (Peng *et al.*, 2021), on the other hand, a negative delta implies slower brain ageing. Here, brain-age delta acts as an effective biomarker able to show differences between clinical groups (Kaufmann *et al.*, 2018), and is predictive for mortality (Cole *et al.*, 2018). Thus, it is of high importance to produce accurate brain age predictions as an essential pre-requisite for considering brain-age delta a potential biomarker.

Numerous studies strive to make the most accurate brain age prediction system. Some of the methods used in the literature include machine learning approaches such as Linear Regression, Support Vector Machines and Gaussian Process Regression (Dosenbach et al., 2010; Gaser et al., 2013; Aycheh et al., 2018; Liang, Zhang and Niu, 2019; Da Costa, Dafflon and Pinaya, 2020) and more recently deep learning techniques (Cole, Poudel, *et al.*, 2017; Kawahara *et al.*, 2017; Wang *et al.*, 2019; Dinsdale *et al.*, 2021; Peng *et al.*, 2021). However, brain age prediction accuracy still needs further improvement, especially in smaller datasets where there is not enough data to train the model (Peng *et al.*, 2021). Some research suggests that deep learning performs no better than simple machine learning models in neuroimaging datasets when the sample size is too small (He *et al.*, 2020).

Traditionally, brain age prediction was performed by extracting features from brain MRIs, followed by classification or regression analysis (Jonsson *et al.*, 2019). A disadvantage of such feature extraction methods is the loss of information since the features are not explicitly selected for extracting information related to brain age. Classical

machine learning techniques rely on our current (limited) knowledge of the brain e.g., focusing on the hippocampus or other brain regions. However, there is much we do not know about the brain and nowadays, deep learning methods like convolutional neural networks (CNNs) can learn features that are important without a prior bias or hypothesis (Jonsson *et al.*, 2019). CNNs have found that smaller areas, previously deemed irrelevant in the ageing process, in fact play a bigger role than originally believed to (Jónsson, 2018).

Here, we implement a 3D Convolutional Neural Network trained on grey-matter segmentation MRIs from the Cam-CAN dataset, inspired on the structure implemented by Dinsdale *et al.*, (2021) as shown in Figure 1 which resembles a VGG-16 architecture.



Figure 1: Part of the ensemble architecture proposed by Dinsdale et al., (2021) which inspired our model

The structure from Dinsdale *et al.* is part of an ensemble architecture and is repeated three times before calculating the average. For simplicity, due to time constraints and limited computational power, our proposed architecture will be a single model with five convolutional blocks used to extract features, followed by a method to correct brain age bias. The model will be evaluated to measure how successful it is at predicting variables from the given data by measuring the mean absolute error (MAE). We use MAE for ease of comparison against literature as it is the most common metric. Cross-validation will be used to ensure validity of the model by splitting the dataset into training and testing sets differently in each iteration. The training set is used to fit the model, while the testing set is used to measure how well the model performs at making predictions on unseen data. We will do ten rounds of cross-validation using different parts of the data to reduce variance. The model will be applied to clinical data to test the hypothesis that brain age delta in subjects with neurological diseases is not the same as in healthy controls.

1.2. Aims and objectives

The ultimate goal of the field of brain age prediction is to improve longevity and quality of life through the identification of brain age delta as a biomarker to detect neurodegenerative diseases before they are too advanced. As an initial step towards this, we aim to create a deep learning model which takes MRIs as input and outputs the estimated brain age. The predictive model will detect changes in structural MRIs related to brain ageing, where changes include the loss of grey-matter, white-matter, and volume in the brain (Cole, Poudel, *et al.*, 2017).

We aim to develop a **convolutional neural network trained on a healthy dataset to predict brain age** based on neuroimaging data. We will then **apply the trained model**, **through transfer learning**, **to a clinical dataset** which includes Schizophrenia, Parkinson's disease, and PTSD data samples to **evaluate how brain age delta varies when a neurological disease is present**. Lastly, we will **evaluate our results holistically** comparing them against the literature and identifying other factors that could have played a role in the findings.

Our hypotheses are:

 $H_o - Null Hypothesis:$ brain age delta in subjects with neurological diseases is the same as in healthy controls.

 H_1 – Alternative Hypothesis: brain age delta in subjects with neurological diseases is not the same as in healthy controls.

The aims will be achieved by experimenting with different CNN architectures, activation functions, adjusting the number of filters, and evaluating which combination achieves the best results. The network will be created using Keras with a TensorFlow backend. The different combinations will be trained and tested on a healthy dataset using cross-validation with an 80/20 train/test split. Once the best model configuration is found, we will only train on the whole healthy dataset and save the model's weights so we can use transfer learning to test the model on the clinical dataset. We will correct bias in brain age prediction mathematically. Lastly, we will evaluate our results holistically by choosing metrics for quantitative evaluation and analysing the results graphically to allow us to draw conclusions from our findings.

Additionally, we aim to improve our knowledge of working with neuroimaging data and developing techniques used in data science.

1.3. Methodology Outline

To begin with, we will work on a healthy dataset to create a convolutional neural network able to predict brain age with errors in line with research or better. To do this we will firstly work on the dataset with reduced dimensions, as they allow faster computation for experimentation, before moving on to the full-size images, originally in NifTi file format which need to be converted into numpy arrays. We will experiment modifying the CNN architecture, adding different layers, adjusting the number of filters, and changing the activation functions. Next, we will compare the performance of our BrainAge model against well-known CNN architectures, such as VGG-13 and LeNet-5, which have appeared in research related to brain age prediction (Jiang *et al.*, 2020; Dinsdale *et al.*, 2021; Peng *et al.*, 2021).

Once we have a robust model, we will proceed to train it on the whole healthy dataset (as opposed to training and testing like we were doing in the previous steps) and we will save the model with its weights. Then, we will transfer the trained model to a supercomputing cluster containing the clinical data, as the data cannot be moved elsewhere due to ethical reasons. This will allow us to make use of transfer learning, meaning that the model will already be familiar with MRI data when it encounters the much smaller clinical dataset. We will freeze the feature extraction layers of the model and retrain the top layers with the healthy controls from the clinical dataset, enabling the model to familiarise itself with differences caused by having data in different formats, the different scanners used and any noise that may be present. Next, we will correct the bias in brain age prediction to obtain more accurate results. Lastly, we will test the model on the different groups in the clinical dataset, namely, Parkinson's disease, PTSD and Schizophrenia patients as well as the combination of all the aforementioned clinical groups, which we will refer to simply as patients.

At the end, we will conduct an evaluation of the results, including graphs and tables to verify the hypothesis that brain age delta in subjects with neurological diseases is not the same as in healthy controls. If the hypothesis is supported by the results, it would provide ground for brain age prediction being used in diagnosis and treatment decisions in the future.

2. Background

In this section, we provide a brief overview of the concepts relevant to this project. We discuss the usual progression of brain ageing, the bias in brain age predictions, different neuroimaging formats and current research using machine learning and deep learning methods for brain age prediction. We also outline common methods used in deep learning to improve model performance, well-known convolutional neural network architectures we will compare our model to, a summary of the software and libraries used as well as a description of our datasets.

2.1. Brain Ageing

Ageing has a direct structural impact on the brain which correlates with decreased mental and physical fitness. As a result of the ageing process, the brain experiences natural physical changes: reduced brain volume (especially in the prefrontal cortex, majorly responsible for decision making and logical reasoning), shrinking of grey and white matter, reduced volume of striatum, temporal lobe, cerebellar vermis, cerebellar hemispheres and hippocampus (Peters, 2006).

Findings suggest that brain deterioration differs in male and female brains i.e., frontal and temporal lobes are more affected in males as opposed to the hippocampus and parietal lobes in females (Peters, 2006). Also, the effects of ageing are influenced by genetic and environmental differences (Cole *et al.*, 2018) such as diet, stress levels, exercise, smoking or drinking.

In the case of Traumatic Brain Injury (TBI), long-term brain alterations are common e.g., an increased risk for early cognitive decline and dementia. The behavioural and anatomical changes are similar to normal ageing however, TBI contributes to premature development of said age-associated changes (Cole, Leech and Sharp, 2015). Conditions such as Parkinson's disease (PD), Alzheimer's Disease (AD), Schizophrenia (SZ), Depression, Multiple Sclerosis (MS) and Bipolar disorders (BPD) also incite accelerated ageing (Koutsouleris *et al.*, 2014; Jonsson *et al.*, 2019). This pathological ageing is indicated by the level of deviation from the typical pattern of ageing and can be measured with a metric known as brain age delta, discussed in section 2.2.

Increased brain age, which does not always correlate with chronological age, poses a risk of neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) (Beheshti *et al.*, 2020), as well as higher mortality rates (Levakov *et al.*, 2020). Often, when cognitive decline becomes obvious, it is too late to treat it adequately. Having a method to predict brain age would be a good indicator of early signs of brain deterioration which would allow to treat abnormalities before any symptoms become visible.

Previous studies on brain age prediction discovered that brains of subjects with PD appear to be 1.5 years older than healthy brains (Beheshti *et al.*, 2020). This is due to the loss of neurons in the substantia nigra and the presence of protein deposits in the Lewy bodies (Beheshti *et al.*, 2020). In the case of SZ, the brain appears 5.5 years older (Koutsouleris *et al.*, 2014) due to reduced grey matter and chronic inflammation (Institute Progress in Mind, 2019). PTSD brains appear 1.3 years older than healthy brains as a result of reduced cortical thickness (Liang, Zhang and Niu, 2019)

2.2. Bias in Brain Age Predictions

Brain age delta, defined as the difference between chronological age and predicted brain age, has been proposed as a pathology marker linked to a range of phenotypes. It is calculated by subtracting the chronological age from the predicted age.

Research suggests the regression of age on brain features often leads to a biased model which manifests as an underestimation of brain age for older subjects, and an overestimation in younger subjects (Cole, Underwood, *et al.*, 2017; Beheshti *et al.*, 2019; Smith *et al.*, 2019; De Lange and Cole, 2020). The brain age delta correlation displays a strong negative trend.

Toolboxes such as correlation-constrained-regression (Treder *et al.*, 2021) are available to correct said bias. The bias can also be corrected mathematically as described in De Lange and Cole, (2020) by fitting a regression model using the chronological age as the x parameter and the predicted age as y. The gradient and intercept are then used to correct the bias using the equation from De Lange *et al.*, (2019)

Corrected Predicted Age = Predicted Age + $[\Omega - (\alpha * \Omega + \beta)]$

where Ω represents the chronological age, α the gradient, and β the intercept.

Correcting the brain age bias allows to reduce the variance in brain age delta values and leads to lower MAE after correction thus, whenever possible, it is recommended to do it.

2.3. Brain Imaging Formats

A Magnetic Resonance Imaging (MRI) device uses strong magnetic field and radio waves to create detailed images of the organs and tissues within the body. It is a powerful tool in diagnosis of brain injuries, cancer, multiple sclerosis, among others.

There are a few common types of MR images, e.g., spin echo MRI, diffusion tensor imaging (DTI) and functional MRI (fMRI) (Murphy and Gaillard, 2017). In this project, we focus on the spin echo images and specifically T1-weighted images and grey-matter volumetric maps.

In a T1-weighted image, the cerebrospinal fluid (CSF) appears dark, white matter (WM) appears light, and grey matter (GM) is shown in grey (Preston, 2016). There is little contrast between grey and white matter in this kind of images. The left image in Figure 2 shows an example of a T1-weighted image in MNI linear space (Preston, 2016).

On the other hand, a grey-matter volumetric map can be taken from a T1 image. This is done by a process called segmentation which splits the MR image into regions with specific properties (Ashburner and Friston, 1997). Segmentation can also be done to show only the white matter regions or cerebrospinal fluid. A grey-matter volumetric map has much more contrast between grey matter (in light grey) and white matter (in black) than a T1 image. It is sometimes preferable when looking for differences in GM tissue, such as the case of brain ageing. A grey-matter image is shown on the right in Figure 2, alongside a T1 image for comparison.



Figure 2: The differences between a T1-weighted image (left)(Preston, 2016) and a Grey-matter volumetric map (right)(Cam-CAN, 2011)

Ideally, both the training and testing sets should be in the same format i.e., both T1weighted or both Grey-matter segmentations. However, in this project, the healthy dataset is in Grey-matter format and the clinical dataset in T1 due to lack of availability of data in the same format within the timeframe established for the project. Consequently, the results will not be generalisable and the errors will be higher due to the differences in the images. Should we gain access to data in the same formats, the implementation of the model would not vary, only the inputs, and the results would have higher validity.

2.4. Machine Learning in Brain Age Prediction

Machine learning is a branch of artificial intelligence focused on building applications that are able to learn from data and can progressively improve performance independently. There are two main categories of machine learning: supervised learning and unsupervised learning. In a supervised setting, the model learns from labelled examples whereas in unsupervised learning, the model has no previous knowledge about the data and tries to find hidden patterns or intrinsic structures in the data.

Research in brain age prediction is usually of supervised learning type, which enables the model to learn from labelled examples i.e., neuroimaging data labelled with corresponding chronological age. Typically, supervised learning tasks include classification, where the goal is to learn a function that splits the inputs into two or more classes, and regression, where the goal is to infer a continuous function from labelled examples. Both approaches are studied in the literature, but regression methods are more common. For classification, ages are divided into age ranges, e.g., 18-30, 31-40, 41-50, etc. or groups e.g., young, middle aged, elderly, etc. The model should then infer which group each MRI belongs to. For regression, the ages are integers (not ranges), and the model infers the brain age based on a mathematical function. This project implements a regression method because it is more insightful to have integers as opposed to ranges for analytical purposes.

Papers using machine learning for brain age prediction extract features from MRIs followed by classification or regression analysis (Jonsson *et al.*, 2019). The features extracted are based on a hypothesis based on the current knowledge of brain structures and the areas that are relevant to ageing. However, the brain is a complex organ and pre-selecting regions leads to the loss of valuable information (Jonsson *et al.*, 2019) given that there areas of the brain that have not been thoroughly studied to determine their involvement in ageing or the atrophy of the areas themselves.

Nonetheless, research relying on machine learning has been able to provide insightful results. Gaser *et al.*, (2013) employed kernel regression methods, a class of algorithms for pattern analysis, and achieved an accuracy of up to 81% in predicting the conversion of mild cognitive impairment into Alzheimer's disease within three years follow-up. They concluded that each additional year in estimated brain age compared to chronological age resulted in 10% higher risk of developing Alzheimer's disease. They calculated the average over the absolute differences between prediction and chronological age, also known as mean absolute error (MAE), and achieved an error of 3.8 years in healthy subjects. The MAE increased up to 8.73 years in subjects with a mild cognitive impairment diagnosis highlighting accelerated brain ageing.

Franke *et al.*, (2010) published one of the first papers on brain age prediction using structural images. The T1-weighted MRIs were pre-processed to correct for bias-field inhomogeneities, followed by a registration into a common MNI space and then a grey matter segmentation. The dimension of the grey matter segmented images was reduced to 410 features using principal component analysis (PCA) (Wold, Esbensen and Geladi, 1987). They trained a Relevance Vector Machine (RVM) on the 410 features extracted from the IXI dataset which resulted in an MAE of 4.61 years on the testing set. Franke et al. attempted to replicate the results on another dataset with 108 healthy subjects aged 20-59 obtained from a different scanner and got an MAE of 5.44 years, implying that the model was sensitive to artifacts left by the imaging device. Training using both datasets produced an MAE of 4.98 on the testing set. Thus, we can assume that training on multiple datasets would make the model more robust and prevent it from relying on artifacts from a single dataset.

Wang *et al.*, (2014) explored brain age prediction based on cortical thickness and surface curvature. They used T1-weighted images for 148 distinct cortical regions to estimate the cortical thickness and surface curvature. They accomplished an MAE of 4.57 years on the testing set after training an RVM on those features on the IXI dataset.

Most models that implement a Relevance Vector Regression (RVR) model only achieve a mean error of 4.6 years (Franke *et al.*, 2010; Wang *et al.*, 2014; Kondo *et al.*, 2015). More recently, in Aycheh *et al.*, (2018), they achieved an MAE of 4.05 years. In this instance, the researchers used a Sparse Group Lasso (SGL) for feature selection using the brain's anatomical grouping (frontal, temporal, limbic, parietal, occipital and insula lobes) followed by a Gaussian Process Regression (GPR) to fit the final age prediction model. Aycheh *et al.* compared the performance of five different regression models: Ordinary Least Squares (OLS), Sparse Group Lasso (SGL), Gaussian Process Regression (GPR), Relevance Vector Regression (RVR) and Deep Neural Networks (DNN); achieving the best results with the GPR model. They also compared the performance results using hybrid approaches, that is combining more than one regression model, showing a combination of SGL and GPR was the best choice (MAE = 4.053). The errors from machine learning methods are still quite high which is why most recent approaches are using deep learning.

2.5. Deep Learning in Brain Age Prediction

Deep learning is a branch of machine learning which structures algorithms in layers to create an artificial neural network that mimics the human brain i.e., it can learn and make intelligent decisions independently. Deep learning methods learn to solve a task by forming successively complex concepts from simpler concepts. Each successive layer uses the output from the previous layer as input to learn multiple levels of representations that correspond to different levels of abstractions. For example, if using a neural network to distinguish between cats and dogs, the first layer would focus on recognising the outlines of the animal, then the next layer the fur, then ears, eyes, tail, whiskers, and so on. The method finally uses the extracted features to decide if the image looks more like a cat or a dog (provided it was trained on cat and dog images to begin with). Likewise, for the case of brain age prediction, it would firstly look at the outline of the brain, amount of grey matter and white matter, then the volume of certain regions, the sulci (grooves) and gyri (folds), and so on before being able to predict the brain age based on said features.

Convolutional neural networks (CNN) are a special type of neural network. The key difference is that they utilise weight-sharing to minimise the number of parameters in the model. This makes sense for images which have many translational invariant properties, e.g., a neuroimage remains an image of a brain if shifted a few voxels but becomes a different "brain". The CNN learns features that are invariant to shifting by making every local unit feature map perform the same operation on every part of the image. To put this simply, for a CNN trained to detect tumours in the brain, the detection should not depend on where the tumour is localised in the brain.

Convolutional neural networks are usually composed of convolutional layers, activation layers, pooling layers, and fully connected layers. They can be connected to each other and repeated in a large number of ways however, the fully connected layer has to be at the end. Convolutional layers compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. Activation layers apply an element-wise activation function such as ReLu (rectified-linear unit). Activation functions are essential as they enable the CNN to learn from complex data and spot patterns, and to map non-linear relationships between the input and output to a desired range such as between 0 and 1. The pooling layers perform a down sampling operation along the spatial dimensions. The fully connected layers transform the feature map produced by the previous convolutional layer to the desired output.

Cole, Poudel, *et al.*, (2017), one of the first papers implementing a CNN for brain age prediction, compared a CNN model (based on a VGG-16 architecture) with a GPR model and obtained significantly more accurate results on the CNN (MAE = 4.16 years as opposed to 4.66 years).

Huang *et al.*, (2017) also published a paper around the same time. They trained a 2dimensional CNN to predict brain age on T1-weighted MRIs. Their structure was once again based on a VGG net. They found out there is no need to include all the brain slices when working with a 2D network and concluded only 15 were necessary to make a prediction without hurting performance. Their model predicts subject's brain age with an MAE of 4.0 years.

More recent research using a 3D Convolutional Neural Network approach have obtained an MAE of 2.86 and 3.09 years for female and male groups respectively (Dinsdale *et al.*, 2021). However, it should be noted that, although these results are some of the lowest in the literature, it would not yield the most clinically relevant model since not all the participants were healthy in the UK Biobank dataset they used.

The lowest errors in brain age prediction have been achieved using CNNs, which is why we decided to make our own CNN model. In general, MAE results vary between 3-5 years depending on the dataset and model used. Nevertheless, a deep learning approach is not necessarily better than machine learning when a large enough dataset is unavailable (He *et al.*, 2020).

2.6. Regularisation

When training high-capacity models like CNNs, it is important to use regularisation to prevent overfitting. Overfitting occurs when the model has studied the training data too much and learned every relationship which leads to decreased performance on unseen data. This becomes obvious when there is high accuracy on the training data but low accuracy on testing data. In simple terms, regularisation is any modification done to a learning algorithm to reduce its generalisation error (Goodfellow, Bengio and Courville, 2016). There are many regularisation methods available in deep learning but, for simplicity, only dropout and early stopping are employed in this project.

Dropout is a recent regularisation method that became popular due to its simplicity and effectiveness in improving performance. It randomly drops out neurons during training which helps prevent the neuron from becoming too reliant on a small subset of important neurons (Srivastava *et al.*, 2014). It is common practice to place dropout layers exclusively in the fully connected layer as there is no significant gain in performance through adding dropout to convolutional layers (Srivastava *et al.*, 2014).

Early-stopping works well at preventing overfitting because it makes the model stop training if no improvement is seen after a certain number of epochs. Usually, training error decreases steadily over time, but the validation error starts to rise again after a certain point (Morgan and Bourland, 1990). Through early stopping, the weights at the lowest validation error are saved. Besides preventing overfitting, this method also helps prevent wasting time when the network is no longer getting smarter.

2.7. Improving performance

There are several ways to improve model performance. We provide an overview into loss functions, optimisation functions and performance metrics, all of which are used in this project to improve performance.

2.7.1. Loss functions

Neural networks learn to map a set of inputs to a set of outputs and make its predictions. The network produces a loss, defined as the distance score between the predicted values and the output values. The model backpropagates said loss to each neuron that has contributed to the output of the model, making use of this information to update the parameters weight and bias of the neural network with the aim of reducing loss and producing the best performing model. Loss functions are used to reduce the loss of a model. They work by readjusting the weight and bias parameters of the model's neurons. Generally, knowing how far off the prediction is from the true value is more significant than knowing if an incorrect prediction was higher or lower than expected.

Mean Absolute Error (MAE) measures the absolute average magnitude of the errors (true value – predicted value) in a set of predictions for a regression model. A limitation of using MAE as a loss function is that the gradient is the same throughout, which

means the gradient will be large even for small loss values. Thus, this is not always appropriate for learning.

Mean Squared Error (MSE) calculates the loss for a regression problem. MSE measures the average of the squares of the error. This value provides insight into how close the set of points of a regression model are to the regression line. Errors are more penalised than with an MAE loss function. In contrast to the gradient problem with an MAE loss, an MSE loss converges even with a fixed learning rate. The gradient of MSE loss is high for larger loss values and decreases as loss approaches zero, making it more precise at the end of training. For these reasons, we will use MSE as the loss function for our model.

2.7.2. Optimisation function

The role of optimisation functions is to find the correct values of weights and bias that minimise loss. Gradients affect how a neural network learns i.e., a higher gradient allows a model to learn faster but, if the slope is zero, the model stops learning.

Gradient descent is an optimisation strategy that involves iteratively adjusting values to minimise the loss function. The algorithm takes steps proportional to the negative of the gradient of the function at the current point and tweaks parameters iteratively with each step down the gradient. The goal is to decrease the learning steps with each iteration while adjusting the parameters to move towards the global minimum. Having a learning step that is too small could mean the algorithm misses the global minimum but, a large learning rate can result in getting trapped in a local minimum.

The Adam optimiser is a popular optimisation function. It uses an adaptive learning rate method based on the Stochastic Gradient Descent algorithm to update the network's weights and biases during training. The Adam optimiser is computationally efficient and is well suited for problems that are large in terms of data or parameters. It is also well suited for non-stationary objectives and problems with very noisy or sparse gradients. A great advantage is that it does not usually require tuning, and it scales the learning rate for individual parameters to reach the convergence of a gradient. Thus, our model employs an Adam optimiser.

2.7.3. Performance metrics

A good model has to undergo a performance evaluation. Performance metrics are different to loss functions in the sense that a metric is used to evaluate the performance of the model once training has finished, whereas the loss function is used by the optimiser during the learning process to minimise errors. Different metrics are better suited for different problems.

For a regression task, Mean Absolute Error (MAE) is usually the best metric. MAE, defined by the equation below, is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. In the formula, \hat{y}_i represents the predicted age and y_i the

chronological age, n is the total number of subjects and j represents each subject increasing up to n in steps of one.

$$MAE = \frac{1}{n} \sum_{j=1}^{n} \left| \hat{y}_j - y_j \right|$$

Since MAE is an error metric, the lower the value is, the better. In this project, the MAE will be an indication, in years, of how far off the predicted values are from the real values. We chose MAE for ease of comparison against literature since it is the most popular metric.

2.8. Analysis of Variance (ANOVA)

An Analysis of Variance (ANOVA) test is a type of inferential statistic used to determine if there is a significant difference across the means of a population. ANOVA tests are useful to test hypotheses in data such as comparing a null hypothesis with an alternative hypothesis using the difference in means.

Here, we use an independent one-way ANOVA to test the significance of the brain age delta in controls and patients. The one-way ANOVA tests the null hypothesis that two or more groups have the same population mean. The test is applied to samples from two or more groups, which can be of different sizes. It calculates the F-values by evaluating the magnitude of variance between the groups against the variance within each group of samples.

If the between-group variance is large relative to the within group variance, the F statistic will be larger than the critical value, therefore statistically significant. It means at least one of the group of means is significantly different from other group of means. The ANOVA test does not indicate which group of means is significant thus, a t-test or Tukey HSD test would need to be performed to find out. Contrary to that, if the withingroup variance is larger, and the between-group variance is smaller, the F-value would be smaller. This reflects the likelihood of no significant differences between the sample means.

The ANOVA notation is F(b, w) = x, p = y where *b* represents the degrees of freedom between the groups, *w* the degrees of freedom within the groups, *x* is the *F*-value and y the *p*-value.

2.9. Well-known Convolutional Neural Network Architectures

LeNet-5 is a CNN architecture made up of 7 layers. It consists of 2 convolutional layers followed by 2 subsampling layers and 3 fully connected layers (Lecun *et al.*, 1998). The original architecture used average pooling in the subsampling layers however, more recent implementations favour the use of max-pooling for better results as well as using ReLU instead of sigmoid as the activation function (Zhang *et al.*, 2019). Each convolutional layer uses a 5x5 kernel. The LeNet-5 architecture, as shown in Figure 3, is well-known for its small memory footprint and fast training times, making it suitable for working with large amounts of data (Zhang *et al.*, 2019) such as the case of MRIs.



Figure 3: A depiction of the LeNet-5 architecture, as illustrated in its original paper (Lecun et al., 1998)

VGG is another well-known CNN architecture. It is a very deep convolutional network used for large scale image recognition. Its structure is simple, relying only on 3x3 convolutional layers stacked on top of each other in increasing depth (Simonyan and Zisserman, 2015). One VGG block consists of sequence of convolutional layers with a padding of one (keeps height and width), followed by a 2x2 max pooling layer with stride of 2 for spatial down sampling which halves the resolution after each block (Zhang *et al.*, 2021), as shown in Figure 4 by Nash, Drummond and Birbilis, (2018).



Figure 4: A depiction of the VGG-16 model architecture (Nash, Drummond and Birbilis, 2018)

VGG remains one of the most used image-recognition architectures and it appears frequently in brain age prediction papers with some modifications (Cole, Poudel, *et al.*, 2017; Nash, Drummond and Birbilis, 2018; Jiang *et al.*, 2020; Dinsdale *et al.*, 2021). The problem with the original VGG-16 architecture is that it is too deep and therefore, requires large amounts of computational power.

This project implements a reduced version of the VGG architecture and an improved LeNet-5 to compare our BrainAge model against these well-known robust architectures. VGG was chosen for comparison because it is a popular structure for brain age prediction. On the other hand, LeNet-5 was chosen as it is noticeably light and able to handle MRIs easily. The changes made to these structures will be detailed in the methodology section.

2.10. Software and Libraries

In this section, we list all the software and library related choices. Most libraries were chosen based on being well-known tools with ample support available online.

- *Python:* The chosen programming language due to its high compatibility with most machine learning tools. Besides, it is a high-level language with strong community support, making it an ideal choice. Python is the most used language in machine learning and data science because it is easy to understand (Python, 2020).
- *Keras:* High-level API for developing neural networks. "Designed for human beings, not machines" (Keras, 2020). It makes the creation of CNNs quick and easy, with consistent APIs, and runs seamlessly on CPU and GPU. In this project, TensorFlow was used as a backend for Keras.
- *TensorFlow:* Open-source machine learning framework which enables to develop, train and test models (Abadi *et al.*, 2015).
- *Scikit-learn:* Collection of machine learning algorithms for supervised and unsupervised learning in Python (Scikit-learn, 2021). It provides efficient tools for data science such as mean absolute error (MAE) as metrics and other tools for fine tuning a model.
- *Scikit-image:* Collection of tools to manipulate images (Van Der Walt *et al.*, 2014). Useful to resize MRIs and crop unnecessary black edges.
- *Numpy:* Provides support for large, multi-dimensional arrays and matrices in Python as well as a wide range of mathematical functions to manipulate these data structures (Numpy, 2020).
- *Pandas:* Useful library to manipulate, visualise and analyse data in csv, txt or sav format (among others) such as demographics (Pandas, 2021).
- *Matplotlib.pyplot:* A collection of functions that make matplotlib work like MATLAB. Enables the creation of graphs and displaying brain slices (Matplotlib, 2021).
- *Seaborn:* A python data visualisation library based on matplotlib. It provides a high-level interface for plotting graphs with many colour themes available (Seaborn, 2020).
- *Nibabel:* Library that provides read and write access to common medical and neuroimaging file formats such as NifTi (.nii) (Brett *et al.*, 2020). We use it to convert NifTi files into numpy arrays.
- *Supercomputing Wales:* Provides access to powerful computing facilities for high performance computing tasks (Supercomputing Wales, 2021). Access is only available to authorised users, and the connection is established through SSH. Supercomputing Wales allows training and testing 3D convolutional neural networks thanks to its powerful GPUs.
- *CHPC:* Similar to the Supercomputing Wales cluster, CHPC is the supercomputing provider for Stellenbosch University (CHPC, 2016).

- *FileZilla:* Enables the transfer of scripts and files between our workstation and the supercomputing cluster using SFTP (FileZilla, 2014).
- *PuTTY:* Program that enables connection through SSH to both supercomputing clusters (PuTTY, 2020).
- *Google Colab:* Created for data science projects, allows writing and executing python code through the web browser making it easy to experiment with different configurations (Google, 2019).
- *GitHub:* We used GitHub for version control (GitHub Inc., 2020). It is a valuable tool used widely in the industry which allows to backup code and revert back to a previous working version before a bug was introduced.
- *Microsoft Excel:* Allows statistical analysis of the results with a simple interface.

2.11. Cam-CAN and Shared Roots datasets

The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) is a large scale collaboratively research project based at the University of Cambridge. The Cam-CAN project uses epidemiological, behavioural and neuroimaging data to understand how individuals best retain cognitive abilities into old age (Cam-CAN, 2011). The dataset is within public domain, upon submitting an authorisation form and it has been featured in research multiple times (Shafto *et al.*, 2014; Taylor *et al.*, 2017).

The Cam-CAN dataset, used for training the CNN models, consists of T1-weighted MRIs of 653 healthy subjects. The age of the subjects in this dataset is in the range 18-88 (mean age: 54.6) and the size of each 3D image is 181 x 217 x 181. Three types of segmentations were provided originally: grey matter, white matter, and cerebrospinal fluid. We used grey matter segmentation images because grey matter is more relevant to brain ageing (Wang *et al.*, 2019).

Shared Roots, a project run by Stellenbosch University, aims to understand the similarities of neuropsychiatric disorders and modifiable risk factors for cardiovascular disease (Shared Roots Study, 2014). The original Shared Roots dataset consists of 974 subjects divided into healthy controls and patients of Schizophrenia (SZ), Parkinson's Disease (PD) and Post Traumatic Stress Disorder (PTSD). However, we have only been provided with a reduced dataset of 290 participants out of which 124 were discarded due to bad quality, presenting motion artifacts or signs of brain atrophy in the control group. After discarding the unsuitable subjects, the distribution of individuals per group is as indicated in *Table 1*. The images are T1s in MNI linear space. The ages are in the range 19 to 81 (mean age: 45.4) and each T1-weighted image is of dimensions 91 x 109 x 91.

Group	Controls ($N = 92$)	Patients ($N = 74$)
Schizophrenia	37	21
Parkinson's Disease	15	15
PTSD	40	38

Table 1: Shared Roots data distribution

Attributes related to each participant's gender, age, smoking, alcohol intake and education are available for both datasets. Nevertheless, this project only needs the chronological age of the subjects. All MRIs are originally in NifTi (.nii) file format and have been converted into numpy arrays (.npy).

For comparative purposes, Figure 5 below shows the age distributions for male and female subjects in the Cam-CAN and Shared Roots datasets. It should be noted the male to female ratio in Cam-CAN, on the left, is fairly even whereas in Shared Roots, there are more female subjects than male. Besides, the ages in Cam-CAN are more evenly distributed than in Shared Roots, in which most subjects are middle-aged. We will discuss the implications of these distributions in more detail in the Results and Evaluation section.



Figure 5: Distribution of ages per gender in Cam-CAN (left) and Shared Roots (right) datasets

Both datasets have been approved by the ethics committees at Cardiff University as well as their origin institutions. Ethics statements can be found in the Appendix.

2.12. Transfer Learning

Transfer learning is a technique in which the best model weights acquired from training on the original dataset are selected. Then, we freeze the first few layers so that only the top layers are trainable. The reason behind this is the first layers are simple image processing filters and should not change significantly if looking at data that is similar to the training data. Next step is to train the convolutional neural network on a small randomly selected portion of the data from the new site, e.g., 10%, it should not take long because the network was already trained, and we are only fine tuning it for the new dataset. However, since our dataset is small, we will use the whole group of controls instead of 10%. Transfer learning is useful because variations in the MRI scanner used and image processing methodology may result in unwanted differences and by doing this, the network adapts to these small differences in the new dataset and the accuracy increases.

3. Methodology

The aim of this project is to develop a convolutional neural network trained on a healthy dataset to predict brain age based on neuroimaging data. We will apply the trained model, through transfer learning, to a clinical dataset which includes Schizophrenia, Parkinson's disease, and PTSD data samples to evaluate whether the brain age delta in subjects with a neurological disease is not the same as for healthy controls.

This section covers in detail the design of the proposed model and data pre-processing aspects. Due to the highly iterative nature of the project, the best methodology was an agile approach. A clear plan could not be designed from the start because information about the data is discovered as the project progresses. The agile approach enabled to tune hyperparameters and make subtle changes to the model to improve performance.

Essentially, the predictive model takes as input a multi-dimensional numpy array containing all the cropped MRIs and the labels, also as a numpy array, containing integers corresponding to the subjects' ages. The output is the estimated brain age. Mean absolute error is used to measure the accuracy of the predictions on the unseen MRIs. We will calculate the brain age delta (negative residuals) on the clinical data because an absolute value would not reflect younger-looking brains. Lastly, we will evaluate our results graphically and statistically.

3.1. Data pre-processing

For the first part, tuning the model before applying it to clinical data, the Cam-CAN dataset is used exclusively. The grey-matter segmentation MR images are stored in a shared directory on the Supercomputing Wales cluster and must be extracted from there. We iterate over all directories, which correspond to different subjects, and convert the grey matter segmented images from NifTi (.nii) into a numpy array using the Nibabel library. Then, we use scikit-image to crop 19 pixels around all the edges to remove unnecessary black edges. Next, depending on whether we use more filters with a smaller image size (low res) or fewer filters with a bigger image size (high res), we resize the MRIs down to 96 x 112 x 96 or 143 x 167 x 143, respectively, using scikit-image. Figure 6 below illustrates the difference in detail for each dimension.





Figure 6: Brain slices showing the difference in dimensions and detail between the low-res (left) and high-res (right) images, to scale

The bigger images require more computational power hence why the need to reduce the number of filters to prevent running out of memory. The demographics for the Cam-CAN dataset are stored in a file called 'CC700_mt.txt' located in the directory with the MRI files. The age for each subject is matched to their MRI by taking their unique ID, designated in the SubCCIDc column, and looking it up in the demographics text file using pandas. This value is inserted into a numpy array. All ages are added in the same order as their corresponding MRIs. The MRIs are saved in a mris.npy file and the ages, in an ages.npy file.

For the second part of the project, applying the model to the Shared Roots dataset, data is processed in a similar way as described above. The MRIs in this instance are stored in the CHPC cluster in T1-weighted format. Similarly, we iterate through the directories and convert NifTi into a numpy array using Nibabel and resize to match the data from Cam-CAN (96 x 112 x 96) with scikit-image. As a consequence of the different conditions in the Shared Roots data, subjects were split into Schizophrenia, PTSD, Parkinson's disease, and controls arrays, using the naming format [condition]_mris.npy and [condition]_ages.npy.

Details about participants in Shared Roots data are stored in a file called 'demographics.sav'. Again, the age for each subject was found matching the unique subject ID, using pandas, to the value in the text file and then added into a numpy array. It should be mentioned that ages for the Shared Roots data are in decimal numbers in contrast to integers in the Cam-CAN dataset therefore, we used python's round function to round up each value before adding it to the array.

We used 3D images because, even though they require more computational power, they are a more accurate representation of the brain than 2D slices.

3.2. Creating the model

Initially, the BrainAge model was created on Google Colab (Google, 2019) as it is better suited to experimentation. Only 50 MRIs out of 653 were used for this preliminary stage due to Google Colab's memory restrictions.

The first iteration of the model had 4 convolutional layers with ReLU as the activation function. Each convolutional layer was followed by a max-pooling layer. There was a single batch normalisation layer before the last max-pooling layer, and only one fully connected layer with one unit. The number of filters started at 32 and incremented by a scale factor of 2 per each convolutional layer. Training the CNN on 50 images using a train/test split of 80/20 with a batch size of 16 and 20 epochs, resulted in a MAE of 6.85 years.

Some changes were made to this initial model to improve performance. The structure was modified using Dinsdale *et al.*, (2021)'s model as inspiration. The convnet now had 5 convolutional layers (number of filters per successive layer: 32, 64, 128, 128, 256), each followed by a batch normalisation layer and a max-pooling layer. The activation function remained as ReLU. Two more fully connected layers were added with 256 and 128 units, respectively. The 3D convolutional layers were set to stride 1x1x1 and kernel size 3x3x3. The max pooling layers were set to stride 2x2x2 and kernel size



3x3x3. All layers had padding set to same to prevent the images from shrinking. Figure 7 displays the architecture of the BrainAge model.

Figure 7: Architecture of the proposed BrainAge model. The picture reflects the structure of the low-res model though the high-res remains the same structure with different filter sizes. The model shows five convolutional blocks, each containing a convolutional layer (orange), followed by batch-normalisation (yellow) and max pooling (green).

These changes significantly increased the number of trainable parameters and it was no longer possible to use Google Colab. Training and testing the model on the Supercomputing Wales cluster using the whole dataset (653 MRIs: 522 for training, 131 for testing) over 200 epochs with a batch size of 4 resulted in a considerably reduced MAE of around 4.4 years which is in line with values in the literature, as mentioned in section 2.

Consequently, some other changes were tested to see if they improved performance. For example, changing the loss function from the previously used mean squared error to mean absolute error, altering the batch size, changing ReLU to ELU, adding a dropout layer before the fully connected layers and changing the number of units in the fully connected layers. The best performing changes are detailed in section 4.1, omitting attempts that performed considerably worse.

Two main versions of the model emerged as a solution to memory issues that came about when loading the full-size MRIs as opposed to the reduced version previously in use. We will refer to them as high-res and low-res, named after the size of the MRIs. The high-res version consists of fewer filters (8, 16, 32, 32, 64, 128, 64 in order of layers) and takes as input larger MRIs (143 x 167 x 143). In contrast, the low-res version has more filters (same as Figure 7) and takes as input smaller MRIs (96 x 112 x 96).

The loss function used to measure how close the prediction is to the label is the mean squared error. The convolutional neural network was optimised with the Adam

optimiser with the default parameters. The batch size was kept at four; having a low batch size helps reduce memory consumption. Early stopping was used to monitor the loss value. It checks if the validation loss is lower, if there is no improvement for ten epochs, training stops.

The training time of the network was quite long to experiment with many design options. Both ELU and ReLU activation functions were compared, however, other ReLU variants could have better performance (e.g., leaky ReLU). The Adam optimiser was chosen because it is the most versatile. Dropout was only used in some versions and it was placed before the fully connected layers. All of the code for the convolutional neural network was written in Python using Keras.

3.3. Measuring errors

Each version of the model was run ten times in a for loop. Each time, the train and test data were split differently thanks to the train_test_split function from scikit-learn. This is done as a way to cross-validate the results since the network performs differently depending on what data falls into the train and test splits. As the model was run ten times, it was simple to calculate the mean MAE with its corresponding standard deviation to have an idea of how the model performs overall in the best and worst scenarios.

The epochs were set to 100 with an early stop watching the loss parameter over 10 epochs. If the loss does not improve for 10 consecutive epochs, the early stop is triggered, the model stops training and moves on to the next iteration in the for loop. Although the model performs better at 200 epochs, training for 100 epochs provides sufficient insight into which combination performs better and executes substantially faster. This assumes that if a model configuration is the best at 100 epochs, it will be even better at 200 epochs.

All combinations are compared in the Results and Evaluation section. Then, the best performing model will be compared against the LeNet-5 and VGG-13 implementations to demonstrate how the BrainAge model performs compared to well-known architectures.

The LeNet-5 version we implemented uses ReLU activation functions instead of sigmoid and max-pooling instead of average pooling. We used these modifications because they significantly improve the performance of the LeNet architecture (Zhang *et al.*, 2021). It has two convolutional blocks with 32 and 64 filters respectively and a kernel size of 5x5x5. The max-pooling layers have a kernel size of 2x2x2 and strides set to 2. It has 3 fully connected layers with 120, 84 and 1 units, respectively.

The VGG-13 adaptation we used is based on Jiang *et al.*, (2020). It consists of five convolutional blocks, each with two convolutional layers, a batch normalisation layer, activation layer and max pooling. The number of filters started at 8 and doubled after each convolutional block (8, 8, 16, 16, 32, 32, 64, 64, 128,128). The activation was set to ReLU, kernel size to 3x3x3 and the padding to same. The max-pooling layers have

kernel size set to 2x2x2 and strides to 2. The VGG adaptation has 3 fully connected layers with 128, 64 and 1 units respectively and activation function ReLU.

3.4. Application to clinical data

The best performing version of the BrainAge model was selected. The model was trained on the Supercomputing Wales cluster for 200 epochs using the whole Cam-CAN dataset (N = 653, mean age = 54.6 years, SD = 18.6, age range 18-88 years) and all the weights were saved into a h5 file. Saving the weights allows to perform transfer learning, as outlined in section 2.12. This method enables the BrainAge model to retain the knowledge it gained from the Cam-CAN dataset when predicting brain age and apply it to the Shared Roots dataset.

The h5 file containing the trained model with its weights was transferred to the CHPC cluster via SCP using FileZilla. We loaded the BrainAge model on the CHPC cluster and froze the first 17 layers so that only the top layers are trainable. We trained the CNN on all the healthy controls from the Shared Roots dataset (N = 92, mean age = 45.6 years, SD = 16.9, age range 19-81 years), to allow the network to familiarise itself with the differences from the new dataset.

The model was tested for four different groups to highlight differences between neuropsychiatric conditions: PTSD (N = 38, mean age = 45.0, SD = 11.8, age range 22-72 years), SZ (N = 21, mean age = 32.3, SD = 8.0, age range 19-50 years), PD (N = 15, mean age = 61, SD = 6.8, age range 49-73 years) and all of them combined (N = 74, mean age = 44.6, SD = 14.0, age range 19-73 years), referred to as patients. We also tested on the controls with a 80/20 train/test split to compare patients against them.

All participants included in the training set were healthy according to local study data. All data was visually quality controlled to ensure quality and accuracy of image processing. Demographics were error-checked, and exclusions made if age values were unavailable. We discarded 124 in total.

We ran the model individually for each condition a for loop for 10 iterations, epochs were set to 200 and batch-size to 4. Like before, we included an early stop monitoring the loss for 10 epochs to prevent overfitting.

3.5. Correction of brain age bias

As mentioned in the background section, the predicted ages displayed a bias; the age for younger subjects was overestimated, and it was underestimated for older subjects. Therefore, we corrected the predicted ages using the equation in section 2.2 (De Lange *et al.*, 2019) after making predictions with the CNN.

We performed a Lasso regression on x = chronological age, y = uncorrected predicted age. We chose Lasso because it improves prediction accuracy and reduces overfitting. Then, we used the gradient and intercept from the regression on the equation below where Ω is the chronological age, α the gradient and β the intercept.

Corrected Predicted Age = Predicted Age + $[\Omega - (\alpha * \Omega + \beta)]$

We calculated the correlation coefficients between chronological and predicted ages both before and after correction, as well as the correlation between the chronological ages and brain age delta. This was done both for training and testing to allow for comparison.

It should be noted that the chronological ages array needs to be reshaped into a column vector. This is because the Lasso regression expects a 2D array as input. The chronological and uncorrected ages need to be reshaped back into a 1D array before putting them in the equation to avoid any dimensions errors when calculating the MAE or correlation coefficients.

3.6. Analysis and statistics

All results were saved in csv files to facilitate the creation of graphs and tables. We read the csv values using pandas and created graphs with seaborn.

We computed the mean brain age delta values for each condition by subtracting the chronological age from the predicted age and averaging it over all the subjects. Also, we calculated the correlation coefficients between chronological ages and predicted ages both before and after correction as well as the correlation coefficients between chronological ages and brain age delta. Lastly, we used Microsoft Excel to perform a one-way ANOVA test.

4. Results and Evaluation

The aim of the project has been achieved. We created a convolutional neural network able to predict brain age with a MAE of 4.03 years in healthy subjects in the Cam-CAN dataset and we applied it to clinical data where it obtained a MAE of 17.76 years in healthy subjects from Shared Roots dataset. It is important to bear in mind the images used for each dataset are in different formats (grey-matter and T1-weighted) and this is reflected in the significantly higher MAE for the healthy controls in Shared Roots. The mean brain age delta (predicted age – chronological age on clinical patients suggests abnormal brain ageing, with the largest difference seen on Schizophrenia patients which have a numerically older brain (+5.19 years), followed by PTSD (-1.72 years) and Parkinson's disease (-0.65 years) which appear to have numerically younger brains. Yet, none of the results are significant at p<.05 meaning the null hypothesis is accepted.

In this section, we will outline the results obtained from the preliminary stages which helped to refine the model, and lastly, we will evaluate the final results.

4.1. Early Development Results

The results in this section reflect some of the assumptions made early on and justify the design choices made. At an early stage, the Cam-CAN dataset was used for evaluation, using a train/test split of 80/20. Each variation of the model was run for 100 epochs in a for loop for 10 times as a cross-validation method. The train_test_split function from scikit-learn selected a different division of data into training and testing groups in each iteration of the for loop, ensuring different combinations of training and testing sets were run using different parts of the data, as mentioned in section 3.3.

Each variation of the model was run for high-res and low-res MRIs, see section 3.2. However, we will discuss only the most significant results for clarity.

It should be noted that running each model for 200 epochs instead of 100 significantly improves the performance of each one, reducing the MAE value by around one year. However, it would be very time consuming to execute 200 epochs 10 times for each variation so instead, 100 epochs provide very substantial insight into the performance of each model. This was based on the assumption that the best model at 100 epochs will remain the best model at 200 epochs.

4.1.1. Effect of activation function

We discussed how the activation function are essential to obtain results in section 2.5. Jonsson *et al.*, (2019) used ELU activation functions in their ResNet architecture and achieved a good MAE of 4.006 years. Motivated by this result, we decided to compare the ELU function against the most versatile and famous ReLU function.

The results, as shown below on Table 2, suggest that for this particular task, ReLU is a better choice (MAE = 5.6019) when using the low-res version whereas ELU performs better (MAE = 6.8334) in the high-res version of our model.

Version	Activation	MAE	± SD
Low-res	ReLU	5.6019	0.811
Low-res	ELU	6.3696	1.531
High-res	ReLU	8.4138	6.110
High-res	ELU	6.8334	1.969

 Table 2: Exploring the effect of using different activation functions, namely ReLU and ELU. The best results for

 the low-res and high-res images are in bold.

4.1.2. Effect of MRI size

Following on from the results on Table 2, we can observe that the MAE is lower for the low-res images regardless of the activation function used (MAE = 5.6019 with ReLU; MAE = 6.3696 with ELU). This is due to the low-res version having more filters in the CNN which can therefore extract more features relevant to brain age. Contrary to that, the high-res version uses slightly more detailed images but has to compromise on the number of filters on each convolutional layer due to the high computational power that would be required for such operation. The errors remain the lowest for the ReLU function. Also, the standard deviation is lower for ReLU which implies there is less variance in the results.

4.1.3. Effect of dropout

Findings from the two previous sections indicated the best choice of activation function for this task is ReLU and that the low-res version performs better than the highres due to including more filters. Thus, we focused on the low-res with ReLU functions model to examine the effects of dropout.

A single dropout layer was employed in the model, located before the fully connected layers to reduce the number of active neurons. We compared not having a dropout layer with using a factor $\rho = 0.2$ and 0.5, where ρ is the fraction of neurons randomly dropped in a layer. The best results were achieved when setting ρ to 0.2 (MAE = 5.4660) (see Table 3).

It is possible that better results could have been achieved with a lower ρ value, such as 0.1. This is because lower dropout values perform best due to the small dataset size available, in relation to the network architecture.

Dropout	MAE	± SD	
none	5.6019	0.811	
0.2	5.4660	0.857	
0.5	5.9104	2.364	

Table 3: Results from different dropout factors (no dropout, 0.2, 0.5). Best result in bold.

4.1.4. Comparison with LeNet and VGG

We compared our model against LeNet and VGG architectures to demonstrate the BrainAge model performs better. For the comparison, we used the low-res images, ReLU activation functions and no dropouts for the LeNet and VGG models, but a dropout of 0.2 for the BrainAge model since it achieved the best results in the previous section.

Cross-validation was implemented in the same way as the previous section, running 100 epochs of each model for 10 times in a for loop.

The BrainAge model performed slightly better than VGG (MAE = 5.4660 vs MAE = 5.4929) as shown in Table 4, which makes sense given that the BrainAge model is an adaptation of VGG with the main advantage being it does not take as long to run because there is only one convolutional layer per convolutional block as opposed to two in VGG. The LeNet model showed the worst performance in this instance, probably due to the fact that the architecture is not deep enough to extract all the features relevant to brain age.

Model	MAE	± SD	
BrainAge	5.4660	0.857	
LeNet	5.8748	0.545	
VGG	5.4929	1.031	

 Table 4: Comparison of the BrainAge model against LeNet and VGG. All models were trained on the low-res

 images and used ReLU activation functions. Best results shown in bold.

Although the number of epochs was set to 100 for each of the models, the early stopping call-back was set to monitor the loss, and if no improvements were seen after 10 epochs, stop. As mentioned before, the purpose of early stopping is to prevent overfitting and reduce the training time when no improvement is seen. Both LeNet and VGG models stopped significantly before 100 epochs, at 34 and 52 epochs, respectively as shown in Figure 8. On the other hand, the BrainAge model continued to make progressive improvement and the early stopping did not get triggered until the 98th epoch.



Figure 8: Graphs of MAE for each model, from left to right: BrainAge Model, LeNet and VGG

These graphs only illustrate the behaviour of each model on the last iteration of the for loop and are therefore not representative of all the other iterations which could have had better or worse performance.

4.1.5. Summary

Based on the results from this section, it was assumed that the best possible variation of the BrainAge model uses the low-res version, employs ReLU functions and has a dropout layer set to 0.2.

At the beginning of section 4, we established the assumption that the best performing model at 100 epochs will still be the best one at 200. We can prove this is the case as the BrainAge model achieves an MAE as low as 4.4406 (SD: 0.513) when trained for 200 epochs and using cross-validation. The correlation between chronological and predicted age is 0.95, indicating they are very closely related.

4.2. Testing the BrainAge model on clinical data

We plotted a graph of the training set of healthy subjects' predicted ages against their chronological age (see Figure 9) and experienced the brain age bias phenomenon discussed in section 2.2. The black line represents the identity line (x=y), the blue regression line should be close to it if the predictions are accurate, but they are not. Predicted ages for subjects aged between 20-40 are overpredicted and ages for individuals over 50 are severely underpredicted. This phenomenon occurred in all the patients groups as well.



Figure 9: Demonstration of brain age bias; young subjects are overpredicted, old subjects under predicted

Once we implemented the brain age bias equation from De Lange and Cole (2020) to correct the ages, we plotted the corrected and uncorrected predicted ages against the chronological ages. We were successful in our implementation as the orange regression line representing the corrected ages lies perfectly on the x=y line (see Figure 10). This means the predictions are much more accurate now. The correlation between predicted and chronological ages increased, on average, from around 0.2 to around 0.5 meaning that they are moderately related. The correlation between brain age delta and chronological ages also changed from around -0.4 to close to 0. This is ideal as we want their correlation to be as close as possible to zero so that the predicted values are accurate.



Figure 10: Corrected and uncorrected predicted ages against chronological ages

We tested the brain age bias correction equation on the Cam-CAN data as well and it reduced the MAE from 4.44 to 4.03 years, it also increased the correlation from 0.95 to 0.96. This demonstrates it is highly beneficial to use the equation to correct the bias.

Applying the BrainAge model to the Shared Roots dataset yielded interesting results. The training set was the same for every group (healthy controls) and the testing set was the corresponding clinical group. Predictions on the healthy controls group were made using an 80/20 train/test split.

A graphical representation of the results can be found in Figure 11 and a table with the full results in Table 5.



Figure 11: Both graphs plot the predicted ages against chronological ages. The dashed black line represents the identity line (x=y). The left graph shows the difference between all patients and all controls. On the right, the graph details the differences between each clinical group.

Group	N	MAE	± SD	Δ Brain Age	± SD
Schizophrenia	21	13.1273	5.056	5.1873	15.572
Parkinson's	15	11.0247	4.866	-0.6493	11.476
PTSD	38	13.5461	2.132	-1.7188	17.823
Patients	74	10.3269	7.598	0.0822	13.342
Controls	18	16.5921	6.423	0.9364	18.168

Table 5: MAE and mean brain age delta results for the different clinical groups. The healthy control group isincluded for comparison.

From the graph on the right, we can see that out of the 21 Schizophrenia subjects, 8 appear significantly older than chronological age but the rest are younger. The mean brain age delta for Schizophrenia subjects is 5.1873 (Table 5) which is numerically larger than for the controls. It suggests that SZ brains are, on average, 5 years older. However, the dataset is small (N = 21), and the oldest Schizophrenia subject is 50 consequently, it is difficult to determine whether older subjects would present accelerated brain ageing too. Looking at the trend, it gives the impression that for subjects over 50, the predicted age would be slightly higher than for healthy subjects, but fairly close to the x=y line. These results are in line with previous findings from literature (Koutsouleris *et al.*, 2014) implying Schizophrenia patients have older brains by 5.5 years.

For Parkinson's disease, the graph on the right shows 5 subjects out of 15 display slightly accelerated brain ageing but the rest are younger than healthy subjects. The mean brain age delta is -0.65 years which is numerically smaller than for controls, indicating subjects with PD present slightly decelerated brain ageing. The green line

of best fit suggests that younger patients would have, in theory, severely accelerated brain ageing. Contrary to the Schizophrenia dataset where there were no patients older than 50, in the Parkinson's dataset the youngest participant is 50. It would be note-worthy to see what the predicted ages would be in younger people with Parkinson's, but it is usually undiagnosed until the age of 60.

In the case of PTSD, the graph on the right suggests decelerated brain ageing on almost half of the subjects. In this instance, the graph correlates with the brain age delta value quite well as it indicates PTSD patients have, on average, a 1.72-year younger brain. The PTSD data set is better distributed in terms of age than the previous two groups. This means the results are more generalisable. Nonetheless, taking into account PTSD is a result of an external stressor, it is difficult to determine if the changes in brain age delta are caused by the stressor itself as it may have appeared earlier or later on in life for each individual and thus, impacted the brain more if the trauma occurred at a younger age.

The graph on the left shows the controls are further away from the identity line than the group of all patients together. This is also reflected by the brain age delta on the table which implies the controls have the second oldest brains after Schizophrenia. We believe this abnormal result is fruit of having a smaller training set for the controls due to splitting all the controls (N = 92) into training and testing sets whereas, for the patients' groups, the whole controls group was used for training. It would be a good idea to retest on a larger dataset to determine if the abnormal results were due to the data itself or whether the subjects appeared younger due to other factors such as lifestyle choices. We should also keep in mind the 'healthy' controls group does not necessarily mean the subjects were overall healthy, it simply means they have not been diagnosed with SZ, PD or PTSD.

In order to test the hypothesis that brain age delta in subjects with neurological diseases is not the same as in healthy controls, a one-way between-groups ANOVA test was performed. The control and patient distributions were sufficiently normal for conducting the ANOVA (i.e., skew <|2.0| and kurtosis <|9.0|; Schmider *et al.*, 2010).

The one-way between-groups ANOVA suggested that there was not a statistically significant difference in brain age delta across the groups, F(4, 162) = 0.71, p = .583. Thus, we have to accept the null hypothesis that brain age delta is the same in subjects with neurological conditions as in healthy subjects.

4.3. Discussion

Despite obtaining results suggesting abnormal brain ageing, the results are not statistically significant at p<.05. We accept the null hypothesis. We believe these unexpected results were caused by having training and testing data in different formats. The Brain-Age model was trained on grey-matter maps and despite trying to minimise the differences by training the model again on the healthy controls from Shared Roots in T1-weighted format, both datasets should have been in the same format to have more valid results. Another issue was that the Shared Roots dataset is fairly small. We were provided with 290 images but after discarding 124 due to low-quality, we were left with only 166 images. 92 were healthy controls used for further training which meant the clinical groups only had 21 (SZ), 15 (PD) and 38 (PTSD) scans each. Having such a small dataset does not allow to generalise the findings unless they can be replicated on a bigger dataset. The subjects' ages were not distributed evenly either. For example, in SZ the oldest individual was 50 (range 19-50) and in PD the youngest was 49 (range 49-73) meaning that in both conditions an extensive range of ages was unaccounted for. On the other hand, ages in PTSD were distributed better (range 22-72). The global average life expectancy is around 73 years (Max Roser, Esteban Ortiz-Ospina and Hannah Ritchie, 2019). Schizophrenia reduces life expectancy by 15 to 25 years (Wildgust, Hodgson and Beary, 2010) which explains why the oldest individual in the SZ group was 50. Mortality in PD patients occurs around 12 years after initial diagnosis which is usually in the early 60s (Morgan *et al.*, 2014), making it difficult to gather better age distributed sample.

The solution does not account for differences between male and female brains. As mentioned in the Background section, male and female brains age differently. Therefore, it would have been interesting to have two separate models, one for male and one for female subjects, or an ensemble architecture like Dinsdale *et al.*, (2021) to allow to see differences. We could have also accounted for these differences by passing a sex label alongside the chronological age during training and have the model predict both.

There could be ethical implications related to the results as the data is not distributed the same way in both datasets. In Cam-CAN the ratio of male to female subjects is even but, in Shared Roots, most individuals are female. This makes it once again difficult to generalise the results as they are not an accurate representation of the general population and also females' brain predicted ages are younger than chronological age whereas males' are older (Cole *et al.*, 2018). Moreover, ages in Cam-CAN are better distributed whereas in Shared Roots, it is mostly middle-aged subjects as evident in Figure 5. Besides, both datasets come from distinct locations so individuals will have been exposed to different lifestyle factors such as diet and schooling. Most subjects in the Shared Roots dataset are bilingual in Afrikaans and English which is shown to affect brain development and is linked with more grey matter and higher white matter integrity (Pliatsikas *et al.*, 2020). On the other hand, most subjects in Cam-CAN are monolingual.

Lastly, the solution cannot predict age for one person alone, we have to use mean errors because the model is off by 4.03 years so if a person's chronological age is 70 but their predicted age is 74, is that difference because of the model error or person experiencing accelerated ageing?

5. Future work

The discussion made evident areas that need improvement to use brain age delta as a biomarker in the detection of neurological conditions. In this section, we explore the
changes that could be made to this project in the future to obtain more accurate and generalisable results.

5.1. More data and same format

We mentioned in the discussion that the datasets we used were small, especially the testing one. We believe training the BrainAge model on a larger dataset (or combination of datasets) would help reduce the MAE and generate more generalisable results on the clinical data. There are several healthy datasets publicly available, e.g., IXI, UK Biobank, OASIS, that have been previously featured in several papers (Wang *et al.*, 2014; Jonsson *et al.*, 2019; Smith *et al.*, 2019; Dinsdale *et al.*, 2021; Peng *et al.*, 2021). We were unable to use said datasets due to time constraints as they require the submission of an application form which takes two to three weeks for approval.

It is more difficult to find publicly available datasets with neuropsychiatric conditions as it involves more ethical concerns. Nonetheless, there are some available such as ADNI which contains subjects with mild cognitive impairment and dementia and has also appeared in the literature (Kaufmann *et al.*, 2018).

It is worthy emphasising again that the healthy and clinical datasets should be in the same format e.g., both T1-weighted. This would help reduce anomalous results and reduce the testing MAE.

It would be interesting to not only use larger datasets but also to explore different conditions i.e., autism, anxiety, or depression which have not been researched thoroughly yet to be able to say if they affect brain ageing.

5.2. Data Augmentation

Following on from the previous point, if it were impossible to access a larger dataset, data augmentation would be an appropriate solution. Data augmentation consists of modifying the available data by reflecting images horizontally or vertically, rotating them or adding noise. Through this technique, we could triple the number of images to be used in training and testing. Adding noise would be the most suitable way to augment data in the training set because, if used in a clinical scenario as a biomarker, the images would at times have worse quality than others. Ideally, the model should be robust all around and be able to manage worse quality images almost as accurately as better-quality MRIs.

5.3. Follow-up study on subjects with largest brain age delta

A large brain age delta means the predicted brain age is significantly bigger or smaller than the subject's chronological age. It would be a promising idea to have a follow-up study after three or five years on subjects which exhibited the largest brain age delta to see whether they got a diagnosis for a neurodegenerative disease. With a follow-up study we could determine if we predicted the disease before it happened which could make predicted brain ages through deep learning a valid biomarker to be used in diagnosis. However, we should keep in mind that we did not collect the data ourselves and the subjects' identities are anonymous for confidentiality. The follow-up study would therefore need to be conducted by the researchers who originally collected the data, providing the participants agreed to be contacted again.

5.4. Multivariate and Qualitative Analysis

Despite having access to demographic information about the subjects we decided not to undertake any multivariate or qualitative analysis as it would require substantial amounts of time we did not have. We believe we could gather valuable knowledge from including variables such as gender and smoking/drinking habits in our neural network. Instead of passing these parameters alongside the chronological age, we could have different networks for each of these binary choices i.e., separate female and male networks, smokers and non-smokers.

Further qualitative analysis could also be done after getting the results from the network. For example, looking at the subjects who had a remarkably high or exceptionally low predicted ages and exploring their demographics. Perhaps their predicted age was very low because they have a very healthy diet, meditate and exercise regularly and do not drink or smoke, all of these being factors often linked with longer lifespans. On the other hand, for those with higher predicted ages maybe it is not a sign of a neurodegenerative disease but an indicator of mortality due to poor lifestyle choices. It is important to take a holistic approach and look at all the possible variables that could have affected the predicted age.

5.5. Further Optimisation

As a result of time constraints, we could not optimise the model as much as we would have liked to. It is possible that adjusting more hyperparameters such as the learning rate and the batch size could result in lower MAEs.

Other parameters such as kernel size, number of filters, padding and strides could have been optimised further by running different variations and comparing them. However, this would be a very time-consuming task because CNN require large amounts of computational power.

Key decisions such as the model architecture, loss functions and activation functions were choices made on experimentation to find the best results. But values for other parameters were only chosen because they were the widely accepted default value across different papers. Training the model with different kernel sizes for instances could have made smaller steps to a better performing model.

5.6. Different Network Architecture

The network architecture we implemented for our BrainAge model was inspired by classic architectures like VGG that stacks layers on top of each other sequentially. Most recent papers in brain age prediction have implemented ResNet or Ensemble architectures (Couvy-Duchesne *et al.*, 2020; Da Costa, Dafflon and Pinaya, 2020; Levakov *et al.*, 2020; Dinsdale *et al.*, 2021; Peng *et al.*, 2021). Although most of these papers

only perform brain age prediction using a healthy dataset instead of applying it to clinical data, they have been able to achieve MAE values as low as 2.86 years by implementing an ensemble architecture (Dinsdale *et al.*, 2021). We believe a more modern network architecture would allow for more efficient learning and better performance overall.

6. Conclusions

Recent publications have shown that by training supervised regression methods on MRI brain imaging, age-related brain changes can be used to predict the age of an individual with high precision. These predictions can be used to estimate the biological age of the brain and to detect diseases and genetic components associated with abnormal brain ageing.

In this paper, we developed a method to predict brain age on healthy and clinical subjects, corrected brain age bias, and evaluated the model's performance. We developed and trained a convolutional neural network on grey-matter segmented MRIs to predict brain age. The CNN architecture implemented a single dropout layer for regularisation. We compared our model against VGG-13 and LeNet-5 and discovered our model performed better. After correcting for brain age bias (the underprediction for old subjects and overprediction for young subjects), our model had an MAE of 4.03 years and 0.96 R² on the test set of Cam-CAN (healthy), which is in line with values from the literature even though our dataset was smaller. We used transfer learning to transfer the knowledge gained from the Cam-CAN data to Shared Roots. We obtained an MAE of 17.71 years on healthy controls and 13.13 years on Schizophrenia, 11.02 years on Parkinson's disease, 13.55 years on PTSD and 10.33 years on the combination of all patients. However, for clinical data, the negative residual or brain age delta was a better measurement. Subjects with Schizophrenia had a larger positive brain age delta compared to healthy controls. Schizophrenia patients were predicted, on average, a brain age 5.19 years older. On the other hand, subjects with Parkinson's and PTSD were predicted a brain age 0.65 and 1.72 years younger, respectively. However, after conducting an ANOVA test, we found that our results were not statistically significant. Thus, we accepted the null hypothesis that brain age delta is the same in subjects with neurological conditions as in healthy subjects. We concluded that our datasets were quite limited and biased and thus, the results cannot be generalised. Besides, other variables such as biological sex and lifestyle factors should be considered in the future.

7. Reflection on Learning

This project was a challenge from start to end. It has enabled me to develop technical and soft skills along the way.

Before starting the project, I had no tangible experience in deep learning. I had taken a couple of in person and online courses as I am very interested in the field but had never produced anything due to lack of data. Finding appropriate datasets is one of the biggest challenges in data science because data is sensitive and often needs ethical approval which is more difficult to obtain without an affiliation to an academic institution. Also, DNNs require high computational power which I normally do not have access to. While doing this project, I was granted access to two different supercomputing clusters. It was quite difficult at first to understand how to load modules, submit jobs and get the GPUs to work, but reading the documentation and asking the IT helpdesks was extremely useful. I am grateful for the experience I gained from doing my final year project. Now, I feel confident in programming a CNN and using transfer learning and believe this will allow me to apply said knowledge to future deep learning tasks as I hope to pursue a career in data science. I also feel more confident programming in python and finding libraries that help me make tasks easier.

Choosing the topic of research was a fairly easy decision. I knew I wanted to do a data science project, but also something that could have an impact on people, such as allowing early diagnosis of complex diseases. Having studied some psychology and neuroscience in the past boosted my fascination for the brain and encouraged me to take on a project studying its ageing process.

Throughout the whole developmental process, I was faced with multiple issues. This gave me a valuable insight into data science research: it is never a linear process, and many unexpected problems arise, the key is to stay motivated. I was incredibly lucky to have the guidance of my supervisors and their colleagues' who provided support at the times I found myself stuck. Knowing when to ask for help was especially important too. For example, I contacted the researcher who published the brain age correction formula I used, and she was very kind and pointed me in the right direction; I was regressing on the wrong values.

This task enabled me to build upon my soft skills significantly. Project management was a key skill while undertaking this project. I stuck fairly well to my initial plan although I saw it more as a general guideline than a strict plan due to the iterative nature of the project. It was very important to set small, achievable, weekly goals along the way to ensure some progress was made regularly but also, it is even more important to know when to move on when something is not working. For example, half-way through my project I tried to implement a toolbox to correct brain age prediction bias. I tried to implement it as a single layer at the end of the CNN, taking the inputs of the last convolutional layer and passing them through, but no matter what I did, I kept seeing dimensional errors. Despite believing this tool would make my model more accurate, I was unable to implement it. Most likely due to lack of knowledge but perhaps incompatibility as it had only been tested on machine learning tasks and not convnets. Thus, I had to make the decision of not giving it more than five days and move on to the next part in order to meet deadlines. I also got dimensional errors when I corrected the bias mathematically. It turned out that fitting a regression model requires a 2D array, but I was passing a 1D array. However, I had to reshape it back into a 1D array to pass the values through the equation. Looking back, I could have probably used the toolbox to correct brain age bias as the dimensional errors were the same.

It was a reoccurring problem to keep wanting to tinker hyperparameters and different layers combinations. This was particularly bad because each small tweak would result in 3 hours or more of runtime on the Supercomputing cluster due to the large amount of data. Therefore, it was crucial to move on once results in line with literature values were achieved.

I used GitHub (GitHub Inc., 2020) as a version control for my project to minimise the risks of losing my work. It is a valuable tool that saved me in multiple occasions when I needed to revert back to a previously working stage of the code after making a mistake. I pushed my code at least once a day to ensure it was all up to date. I believe this is a very important practice in the workplace and I am glad I feel confident pushing and pulling code from the command line.

In order to get the most out of the meetings with my supervisor, I wrote down nonurgent questions the days prior to each meeting to make sure I did not forget to ask anything. I took notes during the meetings as well so that I could refer back to them as I was working. Moreover, I kept a daily project journal with bullet points of the struggles, achievements and to-do lists which proved to be useful while writing the whole report and especially this reflection.

Attention to detail is essential when working on data science projects. For example, when I worked out the residuals for the clinical data, I calculated them the usual mathematical way (true value – predicted value) but in the case of brain age delta, we want the negative residuals. That means the calculation is reversed (predicted value – true value) in order to make the results more intuitive. Otherwise, the results would seem to indicate 5 years younger when they are actually older.

Problems related to the datasets also hindered progress. For instance, I was not given access to the Shared Roots dataset until significantly further on in my project and when I did, the dataset was in a very different format (T1) than the Cam-CAN dataset (grey-matter segmentations) I had used for training. This led to many emails to see if the person responsible for each dataset could provide the other format i.e., have both datasets in T1 or both in grey-matter segmentations. I was not sure how long it would take to receive the datasets in matching formats, so I had to set up a contingency plan: keep the training set the same (grey-matter segmentation from Cam-CAN), freeze the first layers of the model and train it again on the controls from the Shared Roots dataset so that the model was somewhat familiar with this new format, lastly, test on the patients from Shared Roots. The results would not be as accurate as using the same format in both training and testing, but it was the only way to overcome the obstacle in a short amount of time. In the future, I would ensure the datasets are available in the same format before starting the project.

Thanks to this project, I have improved my communication skills too. Doing a project in collaboration with Cam-CAN and Stellenbosch University meant I had to give presentations on what my project was about, the goals and methodology so that they could provide me with the data I needed and guide me on the best approach. This was a key part of the project, a deliverable early on which gained me access to both datasets. I include the slides of these presentations in the Appendix.

Writing the report was also part of the communication skills I developed. Before this project I had not written anything of similar length or detail. In school we were taught

to start reports or essays with the introduction, then the main body with the results, and lastly the evaluation and conclusions. I learnt that I do not like writing reports this way and I prefer starting with the main body, results and evaluation, conclusions and lastly the introduction and abstract. I feel that writing reports this way is easier as otherwise I would have to keep changing the introduction as my methodology and results evolve. However, I was not very strict with the order of writing these sections and I found that jumping back and forth between methodology and results while writing bullet points of what needed to be included in the introduction and background worked well for me. Also, formatting the report at the beginning and including references as I went along saved me a lot of time and prevented forgetting where I got each reference from. In the future, I would stick to these methods.

To ensure good organisation, I kept all my files in subfolders within the same folder. All the literature was organised using Mendeley (Mendeley, 2018) which also helped me format references. I also had a table on Notion (Notion Labs Inc., 2021) with the papers I read which divided the paper into the network structure, methodology and findings as well as my thoughts about each paper. This was particularly useful as it allowed me to compare all papers at a glance and made writing the background section significantly easier.

The national lockdown due to Covid-19 made it very difficult to separate working on the project from personal life. I like to keep my workspace separate from my personal space by working in the library or labs. It was difficult to be disciplined because being at home meant I had more distractions but also no clear division from working area to personal area. This led me to work over ten hours a day on many occasions which impacted my sleeping and eating habits. I learnt that I need to work on my discipline and have a strict working schedule by specifying my working hours on my calendar or other methods.

Halfway through my project I had a change of supervisors. It was scary at first not knowing whether my new supervisor was going to be familiar with my research topic. However, it is important to be flexible and adapt to challenges as they arise. Therefore, we had a handover meeting where I briefed Stuart on my progress to date and the overall aims of the project. In the end, the change of supervisors ended up being quite beneficial. I was very lucky that Matthias remained available for me to ask any questions more centred around neuroimaging data and Stuart helped me with more general questions regarding python, analysis of results and the report itself.

Once I obtained my results, I realised I needed a statistical test to find out if my results were significant and supported the alternative hypothesis. I do not know much about statistics, so it was very tough choosing an appropriate test. At first, I chose a t-test, but it is not very sophisticated and does not test if all the samples are in the same population. After discussing with my supervisor, I chose an ANOVA test which evaluates the variance between and within groups. I believe getting further training on statistics would be very beneficial in order to pursue a data science career. Overall, I believe I handled the difficulties successfully and am proud of my final product as well as the skills and knowledge I gained throughout the project.

8. Bibliography

- Abadi, M. et al. (2015) TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.
- Ashburner, J. and Friston, K. J. (1997) 'Image Segmentation', in Human Brain Function. 2nd edn.
- Aycheh, H. M. et al. (2018) 'Biological brain age prediction using cortical thickness data: A large scale cohort study', Frontiers in Aging Neuroscience, 10(AUG), pp. 1–14. doi: 10.3389/fnagi.2018.00252.
- Beheshti, I. et al. (2019) 'Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme', NeuroImage: Clinical, 24(102063). doi: 10.1016/j.nicl.2019.102063.
- Beheshti, I. et al. (2020) 'T1-weighted MRI-driven Brain Age Estimation in Alzheimer's Disease and Parkinson's Disease', Aging and disease, 11(3), p. 618. doi: 10.14336/AD.2019.0617.
- Brett, M. et al. (2020) 'nipy/nibabel: 3.2.1'. doi: 10.5281/ZENODO.4295521.
- Cam-CAN (2011) Cambridge Centre for Ageing and Neuroscience Research strategy. Available at: https://www.cam-can.org/index.php?content= (Accessed: 4 February 2021).
- CHPC (2016) Centre for high performance computing. Available at: https://www.chpc.ac.za/ (Accessed: 3 May 2021).
- Cole, J. H., Underwood, J., et al. (2017) 'Increased brain-predicted aging in treated HIV disease', Neurology, 88(14), pp. 1349–1357. doi: 10.1212/WNL.0000000003790.
- Cole, J. H., Poudel, R. P. K., et al. (2017) 'Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker', NeuroImage, 163(November), pp. 115–124. doi: 10.1016/j.neuroimage.2017.07.059.
- Cole, J. H. et al. (2018) 'Brain age predicts mortality', Molecular Psychiatry, 23(5), pp. 1385–1392. doi: 10.1038/mp.2017.62.
- Cole, J. H., Leech, R. and Sharp, D. J. (2015) 'Prediction of brain age suggests accelerated atrophy after traumatic brain injury', Annals of Neurology, 77(4), pp. 571–581. doi: 10.1002/ana.24367.
- Da Costa, P. F., Dafflon, J. and Pinaya, W. H. L. (2020) 'Brain-Age Prediction Using Shallow Machine Learning: Predictive Analytics Competition 2019', Frontiers in Psychiatry, 11. doi: 10.3389/fpsyt.2020.604478.
- Couvy-Duchesne, B. et al. (2020) 'Ensemble Learning of Convolutional Neural Network, Support Vector Machine, and Best Linear Unbiased Predictor for Brain Age Prediction: ARAMIS Contribution to the Predictive Analytics

Competition 2019 Challenge', Frontiers in Psychiatry, 11, p. 593336. doi: 10.3389/fpsyt.2020.593336.

- Dinsdale, N. K. et al. (2021) 'Learning patterns of the ageing brain in MRI using deep convolutional networks', NeuroImage, 224, p. 117401. doi: 10.1016/j.neuroimage.2020.117401.
- Dosenbach, N. U. F. et al. (2010) 'Prediction of individual brain maturity using fMRI', Science, 329(5997), pp. 1358–1361. doi: 10.1126/science.1194144.
- FileZilla (2014) Download FileZilla Client for Windows (64bit x86). Available at: https://filezilla-project.org/download.php (Accessed: 3 May 2021).
- Franke, K. et al. (2010) 'Estimating the age of healthy subjects from T 1-weighted MRI scans using kernel methods: Exploring the influence of various parameters', NeuroImage, 50, pp. 883–892. doi: 10.1016/j.neuroimage.2010.01.005.
- Gaser, C. et al. (2013) 'BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease', PLoS ONE, 8(6). doi: 10.1371/journal.pone.0067346.
- GitHub Inc. (2020) GitHub: Where the world builds software, GitHub Inc. Available at: https://github.com/ (Accessed: 18 May 2021).
- Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning: Regularization. MIT Press.
- Google (2019) Welcome to Colaboratory Colaboratory. Available at: https://colab.research.google.com/notebooks/intro.ipynb#recent=true (Accessed: 12 April 2021).
- He, T. et al. (2020) 'Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics', NeuroImage, 206(July 2019), p. 116276. doi: 10.1016/j.neuroimage.2019.116276.
- Huang, T.-W. et al. (2017) AGE ESTIMATION FROM BRAIN MRI IMAGES USING DEEP LEARNING.
- Institute Progress in Mind (2019) Inflammation, accelerated aging and structural brain changes in schizophrenia | Lundbeck Institute Campus. Available at: https://institute.progress.im/en/content/inflammation-accelerated-agingand-structural-brain-changes-schizophrenia (Accessed: 6 May 2021).
- Jiang, H. et al. (2020) 'Predicting Brain Age of Healthy Adults Based on Structural MRI Parcellation Using Convolutional Neural Networks', Frontiers in Neurology, 10(January). doi: 10.3389/fneur.2019.01346.
- Jonsson, B. A. et al. (2019) 'Brain age prediction using deep learning uncovers associated sequence variants', Nature Communications, 10(1), pp. 1–10. doi: 10.1038/s41467-019-13163-9.

- Jónsson, B. A. (2018) 'Brain Age Prediction using Magnetic Resonance Imaging and Deep Learning', Faculty of Electrical and Computer Engineering, University of Iceland.
- Kaufmann, T. et al. (2018) 'Genetics of brain age suggest an overlap with common brain disorders', bioRxiv. bioRxiv. doi: 10.1101/303164.
- Kawahara, J. et al. (2017) 'BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment', NeuroImage, 146, pp. 1038–1049. doi: 10.1016/j.neuroimage.2016.09.046.
- Keras (2020) Why choose Keras? Available at: https://keras.io/why_keras/ (Accessed: 15 April 2021).
- Kondo, C. et al. (2015) 'An age estimation method using brain local features for T1weighted images', in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. Institute of Electrical and Electronics Engineers Inc., pp. 666–669. doi: 10.1109/EMBC.2015.7318450.
- Koutsouleris, N. et al. (2014) 'Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders', Schizophrenia Bulletin, 40(5), pp. 1140–1153. doi: 10.1093/schbul/sbt142.
- De Lange, A.-M. G. and Cole, J. H. (2020) 'Commentary: Correction procedures in brain-age prediction', NeuroImage: Clinical, 26(102229). doi: 10.1016/j.nicl.2020.102229.
- De Lange, A. M. G. et al. (2019) 'Population-based neuroimaging reveals traces of childbirth in the maternal brain', Proceedings of the National Academy of Sciences of the United States of America, 116(44), pp. 22341–22346. doi: 10.1073/pnas.1910666116.
- Lecun, Y. et al. (1998) Gradient-Based Learning Applied to Document Recognition.
- Levakov, G. et al. (2020) 'From a deep learning model back to the brain—Identifying regional predictors and their relation to aging', Human Brain Mapping, 41(12), pp. 3235–3252. doi: 10.1002/hbm.25011.
- Liang, H., Zhang, F. and Niu, X. (2019) 'Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders', Human Brain Mapping, 40(11), pp. 3143–3152. doi: 10.1002/hbm.24588.
- Matplotlib (2021) matplotlib.pyplot Matplotlib 3.4.1 documentation. Available at: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.html (Accessed: 3 May 2021).
- Max Roser, Esteban Ortiz-Ospina and Hannah Ritchie (2019) Life Expectancy Our World in Data, Our World in Data. Available at: https://ourworldindata.org/life-expectancy (Accessed: 14 May 2021).

- Mendeley (2018) Free Reference Manager & Citation Generator Mendeley, Elsevier. Available at: https://www.mendeley.com/reference-management/mendeleydesktop (Accessed: 18 May 2021).
- Morgan, J. C. et al. (2014) 'Mortality in levodopa-treated Parkinson's disease', Parkinson's Disease, 2014. doi: 10.1155/2014/426976.
- Morgan, N. and Bourland, H. (1990) 'Generalization and parameter estimation in feedforward nets | Advances in neural information processing systems 2', in Advances in neural information processing systems, pp. 630–637.
- Murphy, A. and Gaillard, F. (2017) MRI sequences (overview), Radiopaedia.org. Available at: https://radiopaedia.org/articles/mri-sequences-overview (Accessed: 15 April 2021).
- Nash, W., Drummond, T. and Birbilis, N. (2018) 'A review of deep learning in the study of materials degradation Scientific Figure on ResearchGate'.
- Notion Labs Inc. (2021) Notion The all-in-one workspace for your notes, tasks, wikis, and databases., Notion Labs Inc. Available at: https://www.notion.so/ (Accessed: 18 May 2021).
- Numpy (2020) NumPy. Available at: https://numpy.org/install/ (Accessed: 3 May 2021).
- Pandas (2021) pandas documentation pandas 1.2.4 documentation. Available at: https://pandas.pydata.org/docs/ (Accessed: 3 May 2021).
- Peng, H. et al. (2021) 'Accurate brain age prediction with lightweight deep neural networks', Medical Image Analysis, 68, p. 101871. doi: 10.1016/j.media.2020.101871.
- Peters, R. (2006) 'Ageing and the brain', Postgraduate Medical Journal. BMJ Publishing Group, pp. 84–88. doi: 10.1136/pgmj.2005.036665.
- Pliatsikas, C. et al. (2020) 'The effect of bilingualism on brain development from early childhood to young adulthood', Brain Structure and Function, 225(7), pp. 2131–2152. doi: 10.1007/s00429-020-02115-5.
- Preston, D. (2016) MRI Basics. Available at: https://case.edu/med/neurology/NR/MRI Basics.htm (Accessed: 15 April 2021).
- PuTTY (2020) Download PuTTY: latest release (0.74). Available at: https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html (Accessed: 3 May 2021).
- Python (2020) Download Python | Python.org. Available at: https://www.python.org/downloads/ (Accessed: 3 May 2021).
- Schmider, E. et al. (2010) 'Is It Really Robust?: Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption', Methodology, 6(4), pp. 147–151. doi: 10.1027/1614-2241/a000016.

- Scikit-learn (2021) scikit-learn: machine learning in Python scikit-learn 0.24.2 documentation. Available at: https://scikit-learn.org/stable/ (Accessed: 3 May 2021).
- Seaborn (2020) Installing and getting started seaborn 0.11.1 documentation. Available at: https://seaborn.pydata.org/installing.html (Accessed: 3 May 2021).
- Shafto, M. A. et al. (2014) 'The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing', BMC Neurology, 14(1), p. 204. doi: 10.1186/s12883-014-0204-1.
- Shared Roots Study (2014) About | Shared Roots study. Available at: https://sharedrootsstudy.wordpress.com/about/ (Accessed: 21 April 2021).
- Simonyan, K. and Zisserman, A. (2015) Very deep convolutional networks for largescale image recognition.
- Smith, S. M. et al. (2019) 'Estimation of brain age delta from brain imaging', NeuroImage, 200(February), pp. 528–539. doi: 10.1016/j.neuroimage.2019.06.017.
- Srivastava, N. et al. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research.
- Supercomputing Wales (2021) Home Supercomputing Wales. Available at: https://www.supercomputing.wales/ (Accessed: 3 May 2021).
- Taylor, J. R. et al. (2017) 'The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample', NeuroImage, 144, pp. 262– 269. doi: 10.1016/j.neuroimage.2015.09.018.
- Treder, M. S. et al. (2021) 'Correlation constraints for regression models : controlling bias in brain age prediction', Frontiers in Psychiatry, 12, pp. 1–23. doi: 10.3389/fpsyt.2021.615754.
- Van Der Walt, S. et al. (2014) 'Scikit-image: Image processing in python', PeerJ, 2014(1). doi: 10.7717/peerj.453.
- Wang, J. et al. (2014) 'Age estimation using cortical surface pattern combining thickness with curvatures', Medical and Biological Engineering and Computing, 52(4), pp. 331–341. doi: 10.1007/s11517-013-1131-9.
- Wang, J. et al. (2019) 'Gray matter age prediction as a biomarker for risk of dementia', Proceedings of the National Academy of Sciences of the United States of America, 116(42), pp. 21213–21218. doi: 10.1073/pnas.1902376116.
- Wildgust, H. J., Hodgson, R. and Beary, M. (2010) 'The paradox of premature mortality in schizophrenia: new research questions.', Journal of

psychopharmacology (Oxford, England), 24(4 Suppl), pp. 9–15. doi: 10.1177/1359786810382149.

- Wold, S., Esbensen, K. and Geladi, P. (1987) 'Principal component analysis', Chemometrics and Intelligent Laboratory Systems, 2(1), pp. 37–52. doi: https://doi.org/10.1016/0169-7439(87)80084-9.
- Zhang, A. et al. (2019) 6. Convolutional Neural Networks Dive into Deep Learning, Dive Into Deep Learning. Available at: https://d2l.ai/chapter_convolutionalneural-networks/lenet.html (Accessed: 16 April 2021).
- Zhang, A. et al. (2021) Dive Into Deep Learning. 0.16.2. Available at: https://d2l.ai/d2l-en.pdf (Accessed: 16 April 2021).

9. Appendix

Version	Activation	Dropout	MAE	± SD	RMSE	± SD
Low-res	ReLU	0.2	5.4660	0.857	6.7949	0.925
High-res	ReLU	0.5	5.5147	0.572	7.0488	0.707
Low-res	ELU	0.2	5.5317	0.843	6.9057	1.172
Low-res	ReLU	none	5.6019	0.811	6.9103	0.842
Low-res	ELU	0.5	5.7307	0.910	7.0729	0.969
Low-res	ReLU	0.5	5.9104	2.364	7.3959	2.508
High-res	ELU	0.5	6.2304	1.632	7.8383	1.831
Low-res	ELU	none	6.3696	1.531	7.9870	1.871
High-res	ELU	none	6.8334	1.969	8.1601	1.981
High-res	ELU	0.2	7.4948	3.185	9.1071	3.274
High-res	ReLU	none	8.4138	6.110	9.9449	6.125
High-res	ReLU	0.2	10.3688	9.714	11.9076	9.897

Raw results early-development

Raw results on clinical data combined

	Healthy			Parkinson'	s		PTSD		S	chizophrer	nia		Patients	
chrono	corrected	delta	chrono	corrected	delta	chrono	corrected	delta	chrono	corrected	delta	chrono	corrected	delta
69	78.590	9.590	54	47.718	-6.282	30	30.974	0.974	31	27.305	-3.695	54	48.962	-5.038
22	16.321	-5.679	57	59.152	2.152	26	19.133	-6.867	30	82.907	52.907	57	56.129	-0.871
41	53.671	12.671	51	47.703	-3.297	46	42.557	-3.443	44	62.055	18.055	51	50.297	-0.703
59	38.847	-20.153	61	59.133	-1.867	60	47.945	-12.055	36	57.348	21.348	61	58.953	-2.047
56	72 668	16 668	69	48 326	-20 674	31	22 486	-8 514	30	27 744	-2 256	69	47 005	-21 99
26	29.027	3 027	65	52 771	-12 229	59	54 101	_/ 800	14	41 563	-2 /37	65	59 31/	-5 686
20	29.027	-10 872	65	55 890	-9 110	51	27 254	-23 7/6	10	28 /16	9 /16	65	54 720	-10 280
55	45.242	12 (57	65	02.052	20.052	22	27.234	-23.740	50	45.040	4.052	65	02.450	25 450
58	45.343	-12.057	67	93.952	20.952	32	25.024	-0.970	50	45.048	-4.952	67	92.459	25.455
55	43.289	-11./11	/3	66.937	-6.063	34	46.201	12.201	37	26.401	-10.599	/3	/4.352	1.352
27	20.841	-6.159	64	68.072	4.072	62	53.131	-8.869	27	20.577	-6.423	64	63.524	-0.476
31	33.879	2.879	49	63.641	14.641	51	53.291	2.291	46	43.766	-2.234	49	59.661	10.661
78	62.624	-15.376	54	64.904	10.904	36	39.066	3.066	33	19.276	-13.724	54	60.164	6.164
45	61.728	16.728	59	61.424	2.424	32	-8.429	-40.429	23	21.496	-1.504	59	59.559	0.559
37	28.306	-8.694	60	47.991	-12.009	38	20.219	-17.781	27	44.828	17.828	60	46.931	-13.069
33	93.586	60.586	67	67.646	0.646	41	26.434	-14.566	35	66.182	31.182	67	66.090	-0.910
52	52 729	0 729				48	119 892	71 892	32	37 720	5 720	30	30 324	0 324
80	58 236	-21 764		Kurtosis:	0.954	44	46 171	2 171	28	26 514	-1 486	26	21 124	-4 876
55	52 691	2 2 2 10		Skow	0.334	61	62 501	2.171	20	25.954	1 1 1 6	16	41 400	4.600
20	26.200	10 200		Skew.	0.703	72	03.301	2.301	27	20.402	-1.140	40	41.400	-4.000
20	30.298	10.298		iviean:	-0.649	72	91.226	19.226	20	20.492	-5.508	60	48.255	-11.74
			-	SD:	11.476	51	61.312	10.312	30	24.300	-5.700	31	25.738	-5.262
	Kurtosis:	5.022				22	21.581	-0.419	23	37.140	14.140	59	53.041	-5.959
	Skew:	1.814		-		46	29.316	-16.684				51	34.238	-16.762
	Mean:	0.936				27	24.395	-2.605		Kurtosis:	2.782	32	20.860	-11.140
	SD:	18.168				44	52.250	8.250		Skew:	1.588	34	41.640	7.640
						51	52.531	1.531		Mean:	5.187	62	53.571	-8.429
						61	72.935	11.935		SD:	15.572	51	54.809	3.809
						46	55,960	9.960				36	35.516	-0.484
						22	49 581	17 591				30	2 820	-29 171
						48	49 701	1 701				32	25 /11	-12 590
						-+0 F0	45.701	21.701				41	20.002	11.000
						50	28.722	-21.278				41	29.002	-11.998
						46	39.483	-6.517				48	105.713	57.713
						56	47.897	-8.103				44	42.027	-1.973
						54	50.901	-3.099				61	67.891	6.891
						61	31.947	-29.053				72	88.578	16.578
						33	37.064	4.064				51	55.413	4.413
						44	25.774	-18.226				22	20.366	-1.634
						50	70.301	20.301				46	33.066	-12.934
						35	23.860	-11.140				27	25,112	-1.888
												44	51 070	7 070
							Kurtocic	6 740	1			51	50 714	0.296
							Kurtosis.	0.740				51	50.714	-0.200
							Skew:	1.559				61	65.719	4.719
							Mean:	-1./19				46	57.099	11.099
							SD:	17.823				32	50.436	18.436
												48	46.892	-1.108
												50	33.968	-16.032
												46	41.142	-4.858
												56	47.160	-8.840
												54	50.387	-3.613
												61	39,272	-21.728
												22	33 824	0.824
												ΔΛ	30 512	-13 /19
												50	67 760	17 760
												21	27.700	2 5 2 5 2 5
												31	27.475	-3.525
												30	/1.580	41.580
												44	59.555	15.555
												36	53.154	17.154
												30	27.810	-2.190
												44	41.821	-2.179
												19	27.178	8.178
												50	39.940	-10.060
												37	28.045	-8.955
												27	22,665	-4.335
												46	45 200	-0.800
												22	22 120	10.000
												35	22.133	1 017
												23	21.983	-1.01/
												27	42.821	15.821
												35	60.501	25.501
												32	36.841	4.841
												28	27.129	-0.871
												27	25.578	-1.422
												35	25.814	-9.186
												26	23.111	-2.889
												30	24.674	-5.376
												22	25 121	12 121
												23	55.121	12.121
													Kuntaalo	4.024
													Kurtosis:	4.624
													Skew:	1.476
													Mean:	0.082
													SD:	13.342

ANOVA results

healthy	parkinsons	ptsd	schizo	patients							
9.590295	-6.282111	0.973942	-3.69477	-5.03803							
-5.67886	2.1523445	-6.86711	52.90684	-0.87075							
12.67093	-3.296674	-3.44348	18.05532	-0.703							
-20.1533	-1.867406	-12.0547	21.34838	-2.04694							
3 027158	-20.6/3//	-8.51302	-2.25592	-21.9948	Anova: Single Factor						
-10.8718	-9.110175	-23.7464	9.415572	-10.2799	Anova. Single ractor						
-12.6571	26.952387	-6.97624	-4.95186	25.45923	SUMMARY						
-11.711	-6.063355	12.20059	-10.5986	1.351618	Groups	Count	Sum	Average	Variance		
-6.15947	4.0719573	-8.8694	-6.42272	-0.47641	healthy	19	17.79131259	0.936385	348.3960411		
2.878679	14.640757	2.291121	-2.23422	10.66083	parkinsons	15	-9.739422095	-0.64929	141.1155218		
-15.3762	10.904058	3.065614	-13.7239	6.163596	ptsd	38	-65.31484365	-1.71881	326.2500829		
16.72755	2.4243182	-40.4286	-1.50365	0.558788	schizo	21	108.932991	5.187285	254.598768		
-8.69391	-12.00931	-17.7805	17.82783	-13.0686	patients	74	6.08625137	0.082247	180.4446498		
60.586	0.6463415	-14.5664	31.18178	-0.90981							
0.728907		71.89213	1 49602	1 97624							
-2 31877		2.170304	-1 14641	-4.59976	Source of Variation	55	df	MS	F	P-value	E crit
10.29769		19.22595	-5.50782	-11.7447	Between Groups	680.8394789	4	170.2099	0.714677538	0.583029	2.427460599
		10.31155	-5.69975	-5.26247	Within Groups	38582.43391	162	238.1632			
		-0.41897	14.13966	-5.95945	•						
		-16.6842		-16.7615	Total	39263.27339	166				
		-2.60513		-11.1398							
		8.249657		7.639591							
		1.530672		-8.42892							
		11.93467		3.809253							
		9.95958		-0.48357							
		17.58139		-29.1714							
		-21 2779		-12.5667							
		-6 51651		57 71314							
		-8.10279		-1.97305							
		-3.09932		6.890803							
		-29.0529		16.57773							
		4.06431		4.4129							
		-18.2258		-1.63386							
		20.30058		-12.9337							
		-11.14		-1.88764							
				7.069792							
				4 719305							
				11.0994							
				18.43568							
				-1.10774							
				-16.0322							
				-4.85754							
				-8.84037							
				-3.61257							
				-21./28							
				-13 4965							
				17,76027							
				-3.52526							
				41.58042							
				15.55514							
				17.15364							
				-2.19026							
				-2.17943							
				8.178227							
				-10.0599							
				-0.95545							
				-0.80008							
				-10.8614							
				-1.01695							
				15.82132							
				25.50121							
				4.840876							
				-0.87121							
				-1.42198							
				-9.18566							
				-2.88938							
				12 12102							



School of Computer Science & Informatics

Ethical Approval Request Form Form valid until 1st April 2021

Instructions

Do not use this form if your research is with the NHS or NHS-linked: please refer instead to the NHS Local Research Ethics Committee.

Do not use this form if your research involves adults who do not have the capacity to consent. Such projects have to be submitted to the National Research Ethics Service (NRES) system: http://nres.nhs.uk/

Please carefully review:

- School Research Ethics documentation
- Data management, collecting personal data, data protection act requirements
- Information Security Framework
- Research Integrity and Governance
- Research Ethics

Please complete the Research Integrity Online Training Programme (<u>Staff link</u>, <u>Student</u> link) prior to submitting this form.

Please complete this form at least **2 weeks** before starting your data collection/human involvement activities and send to <u>comsc-ethics@cardiff.ac.uk</u> along with **all** the relevant attachments:

- Full Project plan/proposal
- Participant Information Form, either:
 - hard copy, e.g <u>briefing</u> and <u>debriefing</u> (if appropriate)
 online equivalent
- Consent Form or online equivalent (or justification as to why this is not possible)
- Certificate(s) of completion of the Research Integrity Online Training Programme (RIOTP) for all staff associated with a project (and students if applicable).
- (If applicable) Details concerning external funding
- (If an extension is requested) Provide a list of motivations and list of amendments to any previous approvals

Submissions will be reviewed at the next COMSC Research Ethics Group meeting held approximately fortnightly.

Page 1 of 11

REFERENCE ID: COMSC/Ethics/2021/009



School of Computer Science & Informatics

Ethical Approval Request Form Form valid until 1st April 2021

1 General Information

Title of Project:

Predicting brain age from mri data using convolutional neural networks

If this submission relates to a previous approval request (e.g. a revision or extension):

Previous ID:

If this approval refers to an Undergraduate or Masters Student Project:

Student(s) Names and IDs:

Valeria Gomez Ramirez 1732102

Supervisor Name(s):

Matthias Treder

If this approval refers to a research project (e.g. Staff, Postgraduate Research Student):

Principle Researcher:

Other Researchers:

Project Start Date: 01/02/2021

- End Date: 14/05/2021

Attachments:	Yes	NA	Document Version ID
Full project plan/proposal			
Participant Information Form			
Consent Form			
RIOTP Completion Certificates			
Details concerning external funding			
Motivations for and list of amendment	nts		

Page 2 of 11



School of Computer Science & Informatics

Ethical Approval Request Form Form valid until 1st April 2021

2 Recruitment Procedure

		Yes	No	NA
1	Does your project include children under 18 years of age?			
	If "Yes," have you read and understood Cardiff University's Code of Practice for researchers Working With Children and Young People which forms part of the Safeguarding Children and Vulnerable Adults Policy? The Interim Guidance is at Appendix 1, Page 9 of this <u>Policy</u>			
2	Does your project include people with learning or communication difficulties?			
3	Does your project include people in custody?			
4	Is your project likely to include people involved in illegal activities?			
5	Does your project involve people belonging to a vulnerable group, other than those listed above?			
6	Does your project include people who are, or are likely to become your clients or clients of the department in which you work?			
7	Does your project provide for people for whom English / Welsh is not their first language?			
If ar will	ny of the blue boxes has been ticked, please explain how the potentia be handled:	al ethic	cal iss	ue(s)
-				
Piea	ase describe now do you plan to recruit participants:	osch U	niversi	tv
I WII				ty.

Page 3 of 11



School of Computer Science & Informatics

Ethical Approval Request Form Form valid until 1st April 2021

3 Consent Procedures



REFERENCE ID: COMSC/Ethics/2021/009

CARDIFF School of Computer Science & Informatics PRIFYSGOL Ethical Approval Request Form Form valid until 1st April 2021
4 Possible Harm to Participants
Yes No NA
13 Is there any realistic risk of any participants experiencing either physical or psychological distress or discomfort?
14 Is there any realistic risk of any participants experiencing a detriment to their interests as a result of participation?
If any of the blue boxes has been ticked, please explain how the potential ethical issue(s) will be handled:
If there are any risks to the participants, please explain how you intend to minimise these risks:
Page 5 of 11











School of Computer Science & Informatics

Ethical Approval Request Form Form valid until 1st April 2021

9 Other Ethical Considerations

If there are other potential ethical issues that you think the Committee should consider please explain them in the following space. It is your obligation to bring to the attention of the Committee any ethical issues not covered on this form.

Page 10 of 11



School of Computer Science & Informatics

Ethical Approval Request Form Form valid until 1st April 2021

10 Any other comments

If there is additional information that you think the Committee should consider please explain in the space below:

Page 11 of 11



Increased brain age frequently poses a risk of neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease, as well as higher mortality rates (Levakov *et al.*, 2020). Often, when cognitive decline becomes obvious, it is too late to treat it adequately. Having a method to predict brain age would be a good indicator of early signs of brain deterioration which would allow to treat abnormalities before any symptoms become visible.

Over the recent years, machine learning techniques have enabled automatic disease prediction from imaging data. The aim is to increase the prediction accuracy beyond human performance to assist in clinical diagnosis and treatment decisions. The predicted age from these techniques can be considered to be the "brain age" because it is purely derived from the imaging data. However, it is not just the brain age that it is relevant, it is the difference between the predicted age and the actual age – known as brain-age delta – that matters. This value can provide very valuable insight into the ageing speed of an individual. A positive delta implies that a subject's brain looks older than their real age, meaning they are experiencing accelerated ageing (Peng *et al.*, 2021). Here, brain-age delta acts as an effective biomarker and is able to show differences between clinical groups (Kaufmann *et al.*, 2018), and is predictive for mortality (Cole *et al.*, 2018). Thus, it is of high importance to generate accurate brain age predictions as an essential pre-requisite for considering brain-age delta as a biomarker.

Numerous studies strive to reach the goal of making the most accurate brain age prediction system. Some of the methods used in the literature include machine learning methods such as linear regression, support vector machines and Gaussian process regression (Dosenbach *et al.*, 2010; Gaser *et al.*, 2013; Aycheh *et al.*, 2018; Liang, Zhang and Niu, 2019; Da Costa, Dafflon and Pinaya, 2020) and more recently, deep learning techniques (Cole *et al.*, 2017; Kawahara *et al.*, 2017). However, brain age prediction accuracy still needs further improvement, especially in smaller datasets where there is not enough data to train the model (Peng *et al.*, 2021). Also, some research suggests that deep learning performs no better than simple machine learning models in neuroimaging datasets (He *et al.*, 2020).

Traditionally, brain age prediction was performed by extracting features from brain MRIs, followed by classification or regression analysis (Jonsson *et al.*, 2019). A disadvantage of such feature extraction methods is the loss of information since the features are not explicitly for extracting information related to brain age. Nowadays, deep learning methods like convolutional neural networks (CNNs) can learn features that are important without a bias or a hypothesis (Jonsson *et al.*, 2019).

This project will implement a 3D Convolutional Neural Network trained on T1-weighted MRIs from the Cam-CAN dataset, inspired on the structure implemented by Cole *et al.* (2017) shown in *Figure 1*.



Figure 1 CNN architecture implemented by Cole *et al.* (2017)

The proposed structure for this project will have five layers, four of them to be used for feature extraction and the last layer with a constrained regression model (Treder, 2020; Treder *et al.*, 2021) to reduce bias. The model will be evaluated to measure how successful it is at predicting variables from the given data by measuring the mean absolute error (MAE) and the mean squared error (MSE). Cross-validation will be used to split the dataset into subsets and perform an analysis on one subset at a time. Data is split into two sets: training and test. The training set is used to train the model, while the test set is used to measure how well the model performs at making predictions on that test set. For the purpose of reducing variability of results, ten rounds of cross-validation will be carried out using different portions of the dataset.

Ethics

The data for this project has been provided by Cam-CAN, The Cambridge Centre for Ageing and Neuroscience, which is a large-scale collaborative research project based at the University of Cambridge. The Cam-CAN project uses epidemiological, behavioural, and neuroimaging data to understand how individuals can best retain cognitive abilities into old age (Cam-CAN, 2011). Access to the dataset is within public domain, and an authorisation form has been submitted to gain access to the dataset for the purposes of this project. The data has appeared several times in the literature (Shafto *et al.*, 2014; Taylor *et al.*, 2017) which means that ethical approval was granted in order to publish those papers.

Clinical MRI data provided by the SharedRoots project at Stellenbosch University will also be used to test the CNN. The dataset is anonymised but still contains biometric data pertaining to the participants such as age and sex. The Shared Roots project has been approved by the Health Research Ethics Committee at Stellenbosch University. Besides, the SharedRoots data will be analysed on the Stellenbosch Cluster, which is where it is stored, to ensure its safety and confidentiality.

Given the data contains sensitive information, precautions to ensure the safety of the data are required. The datasets are to be analysed through a password and security protected supercomputing cluster (SCW, 2019) and Stellenbosch's Cluster. Also, both datasets have received ethical approval documents from their corresponding institutions which will be forwarded to COMSC Ethics before starting the project. Training and Ethics forms are to be completed and sent to COMSC Ethics prior to starting analysis on the data.

Aims & Objectives

The ultimate aim of the project is to develop a Convolutional Neural Network to predict brain age, given raw T1-weighted MRIs as input. This predictive model will detect changes in structural MRIs related to ageing, where changes include the loss of grey-matter, white-matter, and volume in the brain, all common signs of ageing of the brain (Cole *et al.*, 2017). The aims are outlined as follows:

Aims

- Develop a CNN trained on the Cam-CAN dataset to predict brain age based on neuroimaging data
- Apply the model to the Shared Roots dataset which includes Schizophrenia, Parkinson's disease, and HIV data samples to evaluate the brain age delta on neurodegenerative data
- Desirable aim: Identify which brain regions play a greater role when predicting brain age

The following objectives suggest how the above aims will be achieved:

Objectives

- Research activation functions and evaluate which model is best suited for the problem e.g., Tanh, ReLu, Leaky ReLu, SoftMax
- Use TensorFlow to develop a Convolutional Neural Network (CNN) to extract features from a raw MRI in 3D
- Train the CNN using the Cam-CAN dataset using cross-validation
- Experiment with a range of CNN architectures to determine which gives the best results
- Use the MVPA toolbox (Treder, 2020) in the last layer of the network to reduce bias in brain age prediction
- Select a method of quantitative evaluation to determine how interpretable and useful the data produced by the model is, and how it increases the transparency of the CNN
- Create data visualisations from the results of the data analysis using matplotlib
- Draw conclusions about the limits of CNN interpretability and whether deep learning models can truly be transparent in the context of clinical MRI predictions

Work Plan

Supervisor Meetings

I have scheduled individual meetings at 11:00 every Thursday with my supervisor, Matthias Treder, up until mid-March. A different supervisor will be taking over from then with presumably weekly meetings as well. These meetings will be held over Microsoft Teams and will be used to discuss progress and any problems encountered.

Weekly Plan

Here is an outline of the key tasks to be completed by the end of each week. This is not a strict plan since the project is an iterative process and I will have to revisit previous tasks in later weeks. Also, there will be unaccounted challenges and issues which could take longer than a week.

- Week 1: 01/02 07/02
 - Background research: Convolutional Neural Networks and literature on brain age prediction
 - Initial meeting with supervisor to discuss initial plan and ask technical questions
 - Write initial report
- Week 2: 08/02 14/02
 - Continued background research
 - Finish Convolutional Neural Networks course
 - Decide how many layers, how many filters per layer, activation function and general CNN structure
- Week 3: 15/02 21/02
 - Continued background research, create a summary to be used in final report
 - Build core of 3D CNN using TensorFlow
- Week 4: 22/02 28/02
 - Build core of 3D CNN using TensorFlow to predict age and get training and test error
 - Train CNN using Cam-CAN images
- Week 5: 01/03 07/03
 - Further training and improvements to the CNN
 - Experiment with different CNN architectures (LeNet-5, VGG, ResNet)
- Week 6: 08/03 14/03
 - Experiment with different CNN architectures and evaluate which one performs best
- Week 7: 15/03 21/03
 - Test using an unseen portion of the Cam-CAN dataset and/or the Shared Roots data. The CNN should take in unseen MRIs and correctly predict age
 Evaluate how well it performs using appropriate testing techniques
- Week 8: 22/03 28/03
 - Evaluate the strengths and weaknesses of the solution
 - Demonstrate with data what it shows
 - Think about what could be improved
- Easter Recess: 27/03 18/04
 - Catching up if needed

- Week 9: 19/04 25/04
 - Start writing final report: Background, Approach, Implementation, Evaluation
- Week 10: 26/04 02/05
 - Continue writing: Results, Introduction, Abstract
 - Week 11: 03/05 09/05
 - Continue writing: Future work, Conclusion, Appendices
- Week 12: 10/05 14/05
 - Last minute tweaks, testing, debugging, refactoring code
 Polish report and hand it in

Bibliography

Aycheh, H. M. *et al.* (2018) 'Biological brain age prediction using cortical thickness data: A large scale cohort study', *Frontiers in Aging Neuroscience*, 10(AUG), pp. 1–14. doi: 10.3389/fnagi.2018.00252.

Cam-CAN (2011) *Cambridge Centre for Ageing and Neuroscience Research strategy*. Available at: https://www.cam-can.org/index.php?content= (Accessed: 4 February 2021).

Cole, J. H. *et al.* (2017) 'Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker', *NeuroImage*, 163(November), pp. 115–124. doi: 10.1016/j.neuroimage.2017.07.059.

Cole, J. H. *et al.* (2018) 'Brain age predicts mortality', *Molecular Psychiatry*, 23(5), pp. 1385–1392. doi: 10.1038/mp.2017.62.

Da Costa, P. F., Dafflon, J. and Pinaya, W. H. L. (2020) 'Brain-Age Prediction Using Shallow Machine Learning: Predictive Analytics Competition 2019', *Frontiers in Psychiatry*, 11. doi: 10.3389/fpsyt.2020.604478.

Dosenbach, N. U. F. *et al.* (2010) 'Prediction of Individual Brain Maturity Using fMRI', *Science*, 329(5997), pp. 1358–1361. doi: 10.1126/science.1194144.

Gaser, C. *et al.* (2013) 'BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease', *PLoS ONE*, 8(6). doi: 10.1371/journal.pone.0067346.

He, T. *et al.* (2020) 'Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics', *NeuroImage*, 206(July 2019), p. 116276. doi: 10.1016/j.neuroimage.2019.116276.

Jonsson, B. A. *et al.* (2019) 'Brain age prediction using deep learning uncovers associated sequence variants', *Nature Communications*, 10(1), pp. 1–10. doi: 10.1038/s41467-019-13163-9.

Kaufmann, T. *et al.* (2018) 'Genetics of brain age suggest an overlap with common brain disorders', *bioRxiv*. bioRxiv. doi: 10.1101/303164.

Kawahara, J. *et al.* (2017) 'BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment', *NeuroImage*, 146, pp. 1038–1049. doi: 10.1016/j.neuroimage.2016.09.046.

Levakov, G. *et al.* (2020) 'From a deep learning model back to the brain—Identifying regional predictors and their relation to aging', *Human Brain Mapping*, 41(12), pp. 3235–3252. doi: 10.1002/hbm.25011.

Liang, H., Zhang, F. and Niu, X. (2019) 'Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders', *Human Brain Mapping*, 40(11), pp. 3143–3152. doi: 10.1002/hbm.24588.

Peng, H. *et al.* (2021) 'Accurate brain age prediction with lightweight deep neural networks', *Medical Image Analysis*, 68, p. 101871. doi: 10.1016/j.media.2020.101871.

Peters, R. (2006) 'Ageing and the brain', *Postgraduate Medical Journal*. BMJ Publishing Group, pp. 84–88. doi: 10.1136/pgmj.2005.036665.

SCW (2019) *Supercomputing Wales Portal*. Available at:

https://portal.supercomputing.wales/index.php/ (Accessed: 4 February 2021).

Shafto, M. A. *et al.* (2014) 'The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing', *BMC Neurology*, 14(1), p. 204. doi: 10.1186/s12883-014-0204-1.

Taylor, J. R. *et al.* (2017) 'The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample', *NeuroImage*, 144, pp. 262–269. doi: 10.1016/j.neuroimage.2015.09.018.

Treder, M. S. (2020) 'MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data', *Frontiers in Neuroscience*, 14(June), pp. 1–19. doi: 10.3389/fnins.2020.00289.

Treder, M. S. *et al.* (2021) 'Correlation constraints for regression models : controlling bias in brain age prediction', *Frontiers in Psychiatry*, 12, pp. 1–23. doi: 10.3389/fpsyt.2021.615754.

REFERENCE ID: COMSC/Ethics/2021/009



Page 1 of 1


Project Goal

- Develop a CNN able to predict brain age accurately using MRIs as input
- Apply the CNN model to the SharedRoots dataset to suggest if a brain shows early signs of a neurodegenerative disease when the predicted age varies largely from the chronological age

DATASETS

Cam-CAN: 653 MRIs from healthy individuals

SharedRoots: MRIs from individuals with neurodegenerative diseases (Parkinson's, Schizophrenia, PTSD)

1

2

APPROACH

Create a CNN with 5 layers and see how it performs on Cam-CAN data

Tweak kernel sizes, strides, optimizer, activation function to see what gives better results using recommended values from literature

Use cross-validation to evaluate the model

Use Matthias' correlation constrained regression module after the NN to reduce bias in brain age prediction

Create a function that can take MRIs in NIfTI format and pre-process them (normalize and convert into .npy)

Compare my NN with LeNet and VGG architectures and a model trained already for MRIs, best performing used to test on SharedRoots data

Apply the model to the SharedRoots dataset to evaluate changes in brain age delta in brains with neurodegenerative diseases

CNN MODEL ARCHITECTURE

(Conv3D -> Batch Normalization -> Max Pooling) * 5 -> Flatten -> (Dense) * 3





Implemented cross-validation with 5 folds (MAE: 4.59)

Working on adding correlation-constrained-regression to reduce bias on predictions • At the moment, the MAE seems to only be calculated for the NN and not for the final Linear Regression

Model seems to overfit at 200 epochs, will reduce to around 40

6

Loss evaluation Mean Absolute Error (MAE) Training MAE 5000 4000 3W 40 S 3000 2000 1000 50 75 100 125 150 175 200 Epochs 75 100 125 150 175 200 Epochs ó 25 25 50 Root Mean Squared Error (RMSE) Training RMSE
Val RMSE 60 3SMM 20 0 25 50 75 100 125 150 175 200 Epochs

PROJECT GOAL

To develop a CNN able to predict brain age using MRIs as input

Ideally suggest if brain shows early signs of a certain neurodegenerative disease if the predicted age varies largely from chronological age

APPROACH

Firstly, create a basic CNN with 5 layers and see how it performs on Cam-CAN data – data used is in .npy format

Run on google colab first on limited dataset (50 MRIs) and then run it on the supercomputing cluster with whole dataset

Tweak kernel sizes, strides, optimizer, activation function to see what gives better results using recommended values from literature

Create a function that can take MRIs in NIfTI format and pre-process them (normalize and convert into .npy)

Use the correlation-constrained-regression toolbox on the last layer to reduce bias in brain age prediction

Apply the model to the SharedRoots dataset to evaluate changes in brain age delta on neurodegenerative brains

1

2

WORK DONE SO FAR

Downloaded 50 MRIs from the Cam-CAN dataset in .npy format and familiarised myself visualising cross-sections of the MRIs

Combined all the MRIs in an array and another array for the corresponding chronological $\ensuremath{\mathsf{ages}}$

Created the core of a basic CNN structure (conv3D -> max pool)*4 -> conv3D -> BatchNormalization ->max pool -> flatten -> dense

Using 80/20 train/test split at the moment

 $\ensuremath{\mathsf{Evaluating}}$ using mean_squared_error as loss function, Adam optimizer and MSE and MAE as metrics

With batch size 16 and epochs 20:

Loss: 57.3935661315918 Mean Squared Error: 57.3935661315918 Mean Absolute Error: 6.8465118408203125



Layer (type)	Output Shape	Param #
conv3d_40 (Conv3D)	(None, 96, 112, 96, 32)	896
max_pooling3d_40 (MaxPooli	ng (None, 48, 56, 48, 32)	
conv3d_41 (Conv3D)	(None, 48, 56, 48, 64)	
max_pooling3d_41 (MaxPooli	ng (None, 24, 28, 24, 64)	
conv3d_42 (Conv3D)	(None, 24, 28, 24, 128)	
max_pooling3d_42 (MaxPooli	ng (None, 12, 14, 12, 128)	0
conv3d_43 (Conv3D)	(None, 12, 14, 12, 128)	442496
<pre>max_pooling3d_43 (MaxPooli</pre>	ng (None, 6, 7, 6, 128)	0
conv3d_44 (Conv3D)	(None, 6, 7, 6, 256)	884992
<pre>batch_normalization_8 (Bat</pre>	ch (None, 6, 7, 6, 256)	1024
<pre>max_pooling3d_44 (MaxPooli</pre>	ng (None, 3, 4, 3, 256)	0
flatten_8 (Flatten)	(None, 9216)	0
dense_12 (Dense)	(None, 1)	9217
Total params: 1,615,297 Trainable params: 1,614,78 Non-trainable params: 512		