University of Cape Town

Masters Thesis

Investigation of Brain Ageing in HIV-Positive Individuals Using Convolutional Neural Networks

---

Supervisor:
Assoc. Prof. Jonathan Shock

Co-Supervisor:
Assoc. Prof. Deshen Moodley

Author:
Rachel Catzel

A thesis submitted in fulfilment of the requirements for the degree of Master of Science in the

Department of Mathematics and Applied Mathematics

June 2, 2024

Declaration of Authorship

I, RACHEL CATZEL, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate Date: 7-02-2024.

Abstract

Developments in the field of Deep Learning (DL) have provided new means of tracking healthy ageing, and have established DL-predicted brain age as an accurate and reliable biomarker for brain health. Deviations from a healthy brain ageing trajectory, indicated by an increased predicted brain age relative to chronological age, and thus positive brain age delta, have been associated with cognitive impairments. This thesis focuses on developing a robust brain age prediction model to investigate brain ageing in HIV-positive individuals. We utilise the UK Biobank, CamCAN, and ENIGMA-HIV datasets for this task and train a convolutional neural network in two stages. First, we pre-train the model on the large UK Biobank dataset (N=21366) which contains individuals in the age range of 45-82 years. To this end, we achieve a mean absolute error (MAE) of 2.57 ± 1.94 years. Next, we fine-tune the pre-trained model on a smaller dataset, with a wider age range, aligned with that of our testing dataset from ENIGMA-HIV. We select the CamCAN dataset (N=484) for this, with individuals spanning the age range of 18-88 years. We obtain an MAE of 3.54 ± 2.59 years on the holdout CamCAN test set, substantially improving upon the 6.38 ± 5.30 years MAE achieved without pre-training. We then apply the trained model to the multi-site ENIGMA-HIV testing dataset which we have harmonised to remove inter-site variation. Following testing, we apply a fixed-effects model to analyse whether the brain age deltas are significantly higher in HIV-positive individuals compared to HIV-negative controls. Although no statistically significant difference is found in the brain age deltas due to HIV status, further analysis reveals significant correlations between the brain age deltas and specific HIV clinical measures, in particular, nadir CD4 count and current CD4 count. This thesis's findings contribute to understanding the impact of HIV on brain ageing and associated factors of significance, and highlights the value of DL techniques in medical research.

Acknowledgements

This research received funding from the University of Cape Town PGFO for the years 2021 and 2022, and from ETDP SETA for the year 2022.

The algorithm HD-BET was used for skull stripping [1]. The linear registration tool part of FSL's [2] package, *FLIRT* [3–5] was used to linearly register the brain MRI scans to MNI152 space with the ICBM 2009a nonlinear symmetric $1 \times 1 \times 1$mm template [6–8].

This research has been conducted using the UK Biobank Resource under application number 11559.

Data collection and sharing for this project was also provided by the Cambridge Centre for Ageing and Neuroscience (CamCAN). CamCAN funding was provided by the UK Biotechnology and Biological Sciences Research Council (grant number BB/H008217/1), together with support from the UK Medical Research Council and University of Cambridge, UK.

I would like to extend a very sincere thank you to Associate Professor Jonathan Shock, without whom this master's degree would not have been possible. I am immensely appreciative of the regular guidance, support, and invaluable knowledge provided by him.

I would also like to extend gratitude to my co-supervisor Associate Professor Deshen Moodley for providing me with advice on how to produce a well-constructed thesis and robust experimental pipeline.

Furthermore, thank you to Professor Dan Stein for assisting in getting the research running, to the ENIGMA-HIV working group for providing this research with access to the ENIGMA-HIV dataset, and the University of Southern California for providing this research with access to the UK Biobank and CamCAN datasets. Thank you to Dr Talia Nir and Alyssa Zhu for their time and guidance in advising me on statistical analysis methods, and thank you to my friends and lab mates for their support and advice.

Lastly, I would like to extend a huge thank you to my parents for all their support and care over the years. Thank you for making this journey a lot smoother, and more pleasant than it may otherwise have been.

Contents

List of Figures

List of Tables

List of Abbreviations

Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
AIDS	Acquired Immunodeficiency Syndrome
AP	Average Pooling
ART	Antiretroviral Therapy
BBB	Blood Brain Barrier
BN	Batch Normalisation
BRN	Batch Re-Normalisation
cART	Combined Antiretroviral Therapy
CNN	Convolutional Neural Network
Concat	Concatenation
CSF	Cerebrospinal Fluid
DA	Data Augmentation
DL	Deep Learning
DNP	Distal Neuropathic Pain
DP	Dropout
DWI	Diffusion-Weighted Imaging
EB	Empirical Bayes
ES	Early Stopping
FC	Fully Connected
fMRI	Functional MRI
FMRIB	Functional MRI of the Brain
FSL	FMRIB Software Library
GAM	Generalised Additive Model
GC	Gradient Clipping
GM	Grey Matter
GPR	Gaussian Process Regression
HIV	Human Immunodeficiency Virus
HIVE	HIV Encephalitis
HIVL	HIV Leukoencephalopathy
HIV−	HIV-negative
HIV+	HIV-positive
KLD	Kullback-Leibler Divergence
LReLU	Leaky Rectified Linear Unit
MAE	Mean Absolute Error
ML	Machine Learning
MNI	Montreal Neurological Institute
MP	Max Pooling

MP-RAGE	Magnetisation Prepared Rapid Gradient Echo
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
ResNet	Residual Neural Network
RF	Radiofrequency
SC	Skip Connection
SFCN	Simple Fully Convolutional Network
SGD	Stochastic Gradient Descent
SM	Softmax
SOTA	State-of-the-Art
SVR	Support Vector Regression
VL	Viral Load
WD	With Decay
WM	White Matter

Chapter 1

Introduction

1.1 Background

Tracking whether one is ageing along a healthy trajectory is of importance for monitoring wellness and promoting longevity [9]. One measure by which a healthy ageing trajectory can be tracked is through the analysis of an individual's brain health, specifically their 'brain age'. By assuming that in healthy individuals, brain age aligns with chronological age, we are able to estimate the brain age of individuals with unknown health status based on the appearance of their brain. Many afflictions, such as Alzheimer's [10], cardiometabolic disorders [11], mental illnesses such as depression [12] and even certain lifestyle factors [13], have a neurological impact. These factors may lead to the brain ageing faster than the individual's chronological age. Being able to determine and monitor the progression of brain ageing can provide a means of assessing the severity of the disease and efficacy of treatment. Furthermore, the ability to assess whether an individual is deviating from a healthy brain ageing trajectory can provide an early diagnostic tool for diseases such as Alzheimer's [14]. The brain age delta measures the difference between brain age and chronological age and is given by

$$\delta = \text{age}_{\text{brain}} - \text{age}_{\text{chron}} \tag{1.1}$$

where $\text{age}_{\text{brain}}$ is the apparent age of the brain as measured by a health professional, or a predictive model, such as a deep learning (DL) model. This brain age delta has been shown to be a reliable biomarker of an individual's brain health [9], measuring if, and how much, an individual is deviating from a healthy ageing trajectory. From Equation 1.1, individuals whose brain age is greater than their chronological age will have a positive delta, and vice versa. It is expected that healthy individuals will have a delta close to zero.

As the field of machine learning (ML) develops, ML models have been frequently utilised in the health care field and put to the task of assessing brain ageing. These models include Gaussian process regression (GPR) [9], support vector regression (SVR) [10], and more commonly, DL techniques; in particular, convolutional neural networks (CNNs) have proven useful [9, 15–17]. A variety of different physical features have been used as input to these models, as a means of determining brain age. These features include cortical thickness [18], functional connectivity [19], grey matter (GM) and white matter (WM) volumetric maps [9], as well as T_1-weighted brain MRI scans [15–17].

In addition to the aforementioned conditions associated with a positive brain age delta, a positive brain age delta has also been seen to be a feature of human immunodeficiency virus (HIV) [20]. Despite advancements in antiretroviral therapy (ART), HIV still impacts approximately 39 million people around the world [21]. HIV has also been linked to neurological effects [22, 23] which prompts the question of its impact on brain ageing. A gap in current research is evident; as of this study, no literature on the applications of CNNs to investigating brain ageing in HIV-positive (HIV+) individuals could be found. This gap reveals a substantial research opportunity. Thus, in this work we shall focus on researching applications of CNNs to brain ageing, with a focus on applying a CNN to investigate brain ageing in HIV+ individuals.

A fundamental challenge in neuroimaging research is the limited availability of diverse, large datasets. This limited availability poses an issue when developing ML models for tasks such as brain age prediction, which not only require extensive datasets for effective training, but also require diverse datasets to ensure generalisability to new, often smaller datasets, obtained through different scanners and scanning protocols. We see that often high performing models are trained with one dataset and the generalisation across datasets from other sources is unknown [24]. Addressing this challenge, our work not only focuses on exploring brain ageing in HIV+ individuals but also aims to develop a model that demonstrates high accuracy and adaptability when applied to smaller, unseen datasets, which will be the case for the dataset of HIV+ individuals.

1.2 Problem Statement

Despite advances in treatment, HIV continues to have a major global health impact [21]. Given the established association between HIV and neurological implications, we are prompted to investigate whether HIV contributes to accelerated brain ageing. Thus, this thesis aims to develop a CNN model and utilise transfer learning to explore the effects of HIV on brain age. This investigation extends to identifying which factors, particularly HIV-related clinical measures, may exacerbate this effect. By conducting this research, we aim to fill a current gap in the literature and contribute to a deeper understanding of the neurological impact of HIV and related factors of significance.

1.3 Research Questions

There are three notable research questions for this work, outlined below in the order that they will be addressed.

1. To what extent does model pre-training and transfer learning allow for application of the model to smaller, unseen datasets?
2. What are the effects of HIV on an individual's brain age, as predicted by an ML model?
3. What HIV-related clinical measures, such as an individual's CD4 count, nadir CD4 count, ART status and acquired immunodeficiency syndrome (AIDS) status, is the brain ageing effect correlated to?

1.4 Hypotheses

Our hypotheses are as follows:

1. Model pre-training and transfer learning substantially improves the model performance, thus allowing for improved application to smaller, unseen datasets.

2. HIV leads to accelerated brain ageing, resulting in a positive brain age delta. Specifically, the brain age delta obtained on HIV+ individuals will be significantly larger than the brain age delta obtained on HIV-negative (HIV−) individuals.

3. There will be a negative correlation between the individual's CD4 count and nadir CD4 count and their calculated brain age delta. The use of ARTs will be correlated to a smaller brain age delta and a status of AIDS will be correlated to a larger brain age delta. These correlations would demonstrate that a lower CD4 count, nadir CD4 count and status of AIDS are linked to accelerated brain ageing while the use of ARTs lessens the acceleration of brain ageing.

1.5 Aims and Objectives

To address the research questions of this work, we define our aims and objectives as follows:

1. Carry out an extensive literature review to build a foundational understanding of existing methodologies in brain age prediction, informing the design of an appropriate dataset pre-processing pipeline, CNN model and training strategy for this task.

2. Design, pre-train and evaluate the CNN model, and benchmark its performance against relevant models in the literature.

3. Investigate the efficacy of transfer learning on a pre-trained model compared to independent training, particularly focusing on the model's performance and adaptability when applied to smaller, unseen datasets.

4. Apply the developed model to a dataset comprising HIV+ individuals and conduct detailed analyses to investigate the impact of HIV on brain ageing, and identify correlations between brain ageing and HIV-related clinical measures.

1.6 Thesis Layout

1. **Chapter 1: Introduction** sets the stage by outlining the research problem, and identifying the research questions, hypotheses, and the aims and objectives of the thesis.

2. **Chapter 2: Literature Review** covers relevant literature, beginning with the development of ML and its progression to DL and CNNs. It then contextualises the use of these networks within medical imaging and brain age prediction and explores brain ageing, the neurological impact of HIV and applications of ML to investigating this impact - thus forming a foundation for experimental decisions made.

3. **Chapter 3: Experimental Design** uses the literature review to develop a dataset pre-processing pipeline, CNN model, and training and transfer learning strategy for application to HIV+ individuals, thus allowing for the research questions to be investigated.
4. **Chapter 4: Experimental Results** presents the results and analyses from each experimental phase, including model pre-training, benchmarking, transfer learning, and the investigation of HIV's impact on brain ageing, as well as the correlations with HIV-related clinical measures.
5. **Chapter 5: Discussion** interprets the results and analyses and highlights significant findings, discussing their alignment with existing literature.
6. **Chapter 6: Conclusion** summarises the research, highlighting the extent to which the aims and objectives were fulfiled, and details the contributions, limitations, and future work.
7. **Appendix** contains additional tables and figures relevant to the research.

Chapter 2

Literature Review

In this review, we explore the landscape of neural network applications to brain age prediction with an extension to individuals with HIV. Through this review, we build a set of tools to address the research questions of this work.

We first provide an overview of the development of ML and DL and focus on the constructions of the convolutional neural network (CNN) [25–27] and residual neural network (ResNet) [28] - highly successful tools in the field of computer vision. Next we discuss brain imaging and elaborate on the phenomenon of brain ageing, as well as the neurological consequences of HIV. We then explore previous CNN applications of brain age prediction using brain MRI scans and ML applications to the study of HIV on the brain. In doing so, we motivate why HIV may accelerate brain ageing and how DL is suited to this study.

2.1 An Overview of Machine Learning

Within the field of ML, there are three main categories: supervised, unsupervised and reinforcement learning. Supervised learning trains on a dataset with input and label pairs. After being trained, it can predict outputs for new, unlabelled data using the learned mapping from inputs to outputs. In contrast, unsupervised learning finds patterns or structures within the inputs, without labelled data. Reinforcement learning involves learning by trial-and-error and features an 'agent' exploring an environment, with the aim of maximising a goal. Supervised learning is most relevant to the research aims of this work. We shall first discuss the origins of ML before diving into DL and relevant supervised learning model architectures.

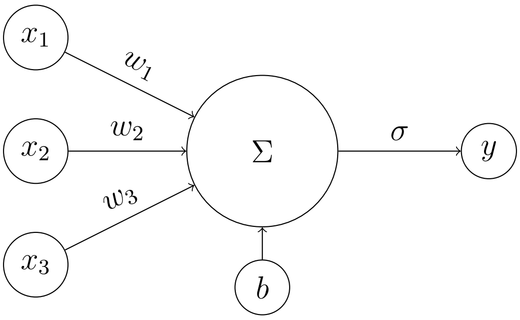
The first artificial neural networks were introduced in 1943 by Warren McCulloch and Walter Pitts [29]. They reasoned that due to the 'all-or-none' behaviour of nervous activity, propositional logic is applicable to neural events. Following this work, in 1957, Frank Rosenblatt [30] introduced the perceptron as an algorithm for pattern recognition. These perceptrons, inspired by biological neurons, came to be known as artificial neurons, contributing to the evolution of artificial intelligence (AI).

A perceptron is represented in Figure 2.1. Here the inputs x_1, x_2 and x_3, are multiplied with corresponding weights w_1, w_2 and w_3, added and passed to an activation function. The equation below the perceptron provides a mathematical representation of how the input is mapped to output. The activation function σ could be any non-linear activation function such as the rectified linear unit (ReLU). The ReLU activation function is defined as

$$f(z) \equiv \max(0, z). \tag{2.1}$$

and acts by setting negative values to zero.

For the case of the 'all or nothing' perceptron, binary inputs are mapped to binary outputs with the Heaviside step activation function. The binary output y is given by Equation 2.2 and corresponds to whether the value of this dot product of weights and inputs, summed with a bias, is less than or equal to, or greater than zero.



$$y = \sigma\left(w_1 x_1 + w_2 x_2 + w_3 x_3 + b\right)$$

Figure 2.1: Diagram illustrating a perceptron. The perceptron, in this case, receives three input values, x_1, x_2, and x_3. Each of these inputs is multiplied by its respective weight, w_i. In addition to these weighted inputs, a bias term, b, is added. This sum is passed through an activation function to produce the output y.

$$\mathrm{y} = \begin{cases} 0 \text{ if } \vec{w} \cdot \vec{x} + b & \leq 0 \\ 1 \text{ if } \vec{w} \cdot \vec{x} + b & > 0 \end{cases} \tag{2.2}$$

As described by Nielsen [31], the bias can be seen as a measure of how easily a perceptron may fire - like a biological neuron. By building larger networks of these perceptrons, one can create more complex decision-making systems, imitating the human decision-making system at a basic level [31]. Such a network is termed an artificial neural network.

Further key developments occurred in the 1970s, with the formulation of the backpropagation algorithm - a core component in the training of modern DL models. Precursors to the algorithm were first seen in control theory in the early 1960s [32, 33]. In the 1970s and 1980s, the backpropagation algorithm was described [34] and applied [35] to multi-layer perceptron training by Paul Werbos. This algorithm provided a way for neural networks to adjust their weights such that the prediction error was minimised. The backpropagation algorithm was later popularised in the 1980s by Rumelhart et al. [36], which was instrumental to the training of deeper networks.

ML algorithms have been applied to various tasks over the years and have demonstrated their potential on many occasions. For example, in 1952, an IBM computer program was written by Arthur Samuel whereby the machine was able to play checkers and learn from its experiences [37]. Furthermore, in 1996, the Deep Blue IBM chess program [38] was developed and successfully beat the reigning chess world champion Garry Kasparov. Power tools for image classification were proposed in 2012 by Krizhevsky et al.

[39]. Even more recently, AlphaGo [40] was released. This program was developed by DeepMind, with the aim of mastering the game of Go. Due to the high level of complexity of Go, the development of a program which could rival top Go players had been sought after and viewed as a worthy goal of developers. Many attempts were made to engineer algorithms which could accomplish this task, although none had been as successful as AlphaGo, with AlphaGo being the first to win against Go world champions. In other fields, complex models have taken the form of works such as DeepMind's AlphaFold [41] - an algorithm which solved the 50 year old protein-folding problem in 2020. There have also been models which allow for the colourisation of black and white images [42], and models allowing for the development of self driving cars, dating back to 1988 [43].

2.1.1 Neural Networks and Deep Learning

While a simple neural network consists of an input and an output layer, networks may have additional inner layers known as hidden layers. With time, more complex network structures have been developed, some consisting of hundreds of hidden layers [28]. These networks with hidden layers fall into the category of DL, a subset of ML. An example of a fully connected deep neural network is given in Figure 2.2.

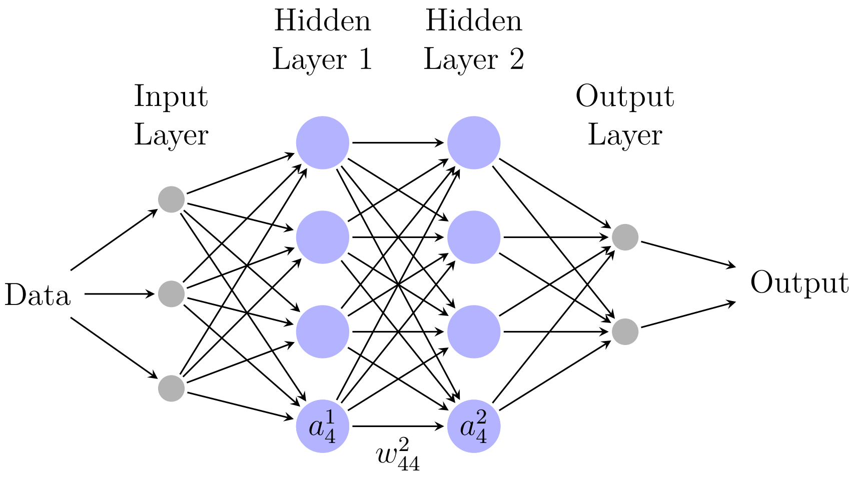


Figure 2.2: A diagram of a fully connected neural network with three input nodes, two hidden layers containing four nodes each, and two output nodes. This is an example of a multi-layer perceptron (MLP) mentioned in the previous section. We include a subset of network labels for illustrative purposes.

An essential component of deep neural network training is the back propagation algorithm. We closely follow the derivations and notations layed out in Nielsen [31], and give a detailed overview of a training **epoch** of a standard fully connected network, such as the network illustrated in Figure 2.2.

1. A training epoch commences with a 'batch' of inputs being fed through the network.
2. The **feed forward** step through each layer is dictated by an equation of the form

$$a_j^l = \sigma\Big(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\Big) \tag{2.3}$$

where,

(a) a_j^l corresponds to the activation of the j^{th} neuron in the l^{th} layer (see Figure 2.2),

(b) w^l is the weight matrix with entries being the weights connecting to the l^{th} layer of neurons,

(c) w_{jk}^l is the weight entry in the j^{th} row and k^{th} column, and thus connects from the k^{th} neuron in the $l-1^{\text{th}}$ layer to the j^{th} neuron in the l^{th} layer (see Figure 2.2),

(d) b_j^l corresponds to the bias of the j^{th} neuron in the l^{th} layer.

Hence, the activation, a_j^l, is determined by taking the weighted sum of activations from all neurons k in the preceding layer, adding the respective bias, and then applying the activation function σ. We see Equation 2.3 is similar to that of Equation 2.2, however the outputs will not be constrained to binary values of zero and one in general - the values are dependent on the activation function.

3. Once the input has been fed through the network, the **loss function** $\mathcal{L}$ can be calculated. Example of such loss functions include the mean absolute error (MAE) loss function

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{2.4}$$

and mean squared error (MSE) loss function

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{2.5}$$

where y denotes the network predictions and $\hat{y}$ denotes the associated label. This **loss** is a measure of how close the network predictions are to the desired output label. The minimisation of network loss through adjusting the weights is depicted in Figure 2.4.

4. This loss is then used to calculate a measure of error which is propagated backwards to previous layers via the **backpropagation** algorithm. Through the backpropagation algorithm, we are able to determine how changes to the weights and biases impact the loss function, and thus take steps to **minimise** the loss.

5. Updates to the weights and biases are performed by a chosen **gradient descent** algorithm, first proposed in 1847 by Cauchy [44]. Popular algorithms include stochastic gradient descent (SGD) [45] and Adaptive Moment Estimation (Adam) [46]. The gradient descent updates are given by

$$w'^l_{jk} = w^l_{jk} - \eta\frac{\partial\mathcal{L}}{\partial w^l_{jk}} \tag{2.6}$$

$$b'^l_j = b^l_j - \eta\frac{\partial\mathcal{L}}{\partial b^l_j}. \tag{2.7}$$

where, η is the learning rate which moderates how large the changes to the weights and biases are. In Figure 2.3, we give a simplified illustration of backpropagation and the update of weights.

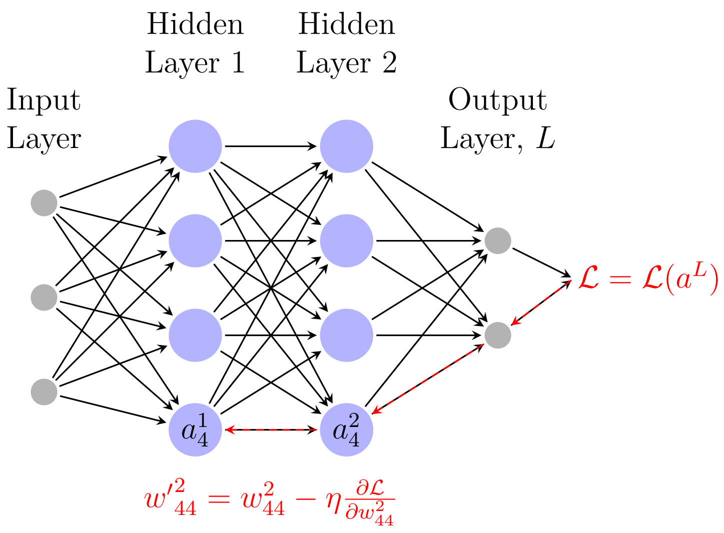


Figure 2.3: A diagram of a fully connected neural network with backpropagation illustrated. The red arrows represent the backwards flow of error from the output layer. For readability, we illustrate the update of a single weight.

Multiple epochs are typically utilised during the training of a neural network. Tracking the loss across these epochs provides insight into the training's progression. One of the challenges faced during this process is the potential for overfitting or underfitting. **Overfitting** occurs when the network 'learns' the training dataset too well, leading to poor generalisation on unseen data. On the other hand, **underfitting** arises when the model is too simplistic or is over-regularised and fails to capture underlying patterns in the data, thus performing poorly.

To mitigate these challenges, and to assess the model's performance, it's common practice to divide the dataset into training, validation, and testing datasets. Through calculating the loss on the validation dataset, the models's performance at the end of each epoch can be evaluated, offering insights into how the model is likely to perform on unseen data. The primary objective when training a robust model is to minimise the validation loss. By monitoring this loss, one can determine if adjustments are needed to prevent overfitting or underfitting. Finally, the test set, which the model hasn't seen during training or validation, provides a final assessment of the model's accuracy.

Neural network tasks can be broadly categorised into **classification** and **regression** based on the nature of their output. Classification models are designed to predict discrete labels or classes. For example, in medical imaging, a model might be trained to classify images into categories like 'tumour' or 'no tumour', or perhaps classify the type of tumour. In contrast, regression models predict continuous values. Such models are more suitable for tasks where the output can lie anywhere on a continuous scale. An example of such a task is the prediction of brain age. Since age doesn't fall into distinct classes but rather spans a continuous range, the brain age prediction problem is typically approached as a regression task [17, 48, 49].

2.1.2 The Convolutional Neural Network

A widely-used, successful neural network variant in the field of computer vision is the convolutional neural network (CNN). The initial form of the CNN was put forth by

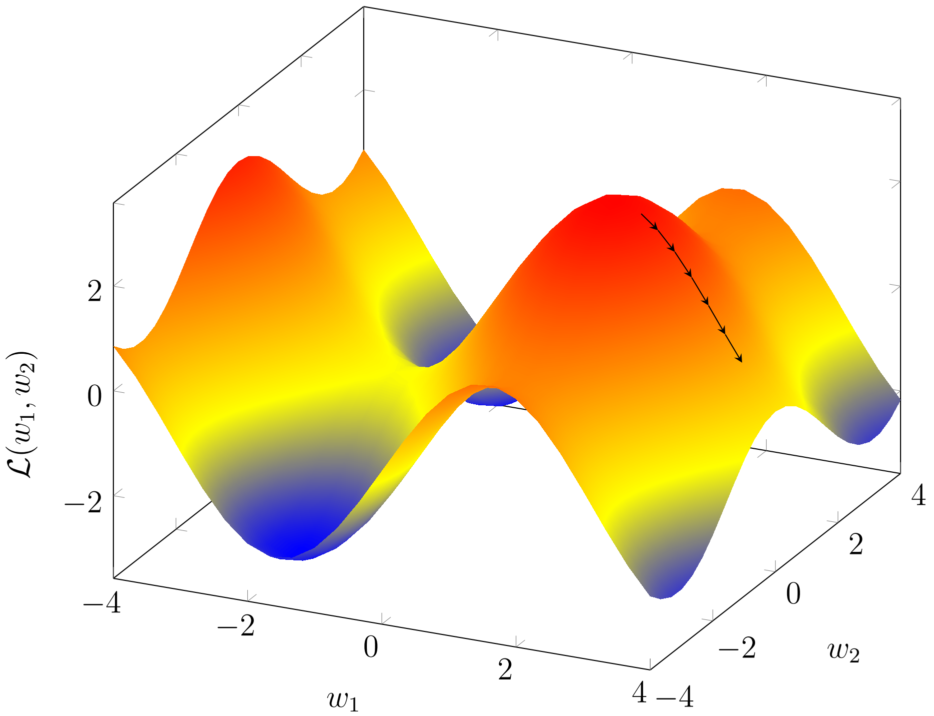


Figure 2.4: Illustration, adapted from [47], of a simplified loss landscape for a neural network. Here, w_1 and w_2 illustrate two weights within the network. This landscape illustrates that the network has been initialised with weights corresponding to a loss local maximum. The path illustrates the gradient descent optimisation technique, where the weights are updated to converge to a lower-loss region, representing the training process of the network.

Fukushima [25] in 1980. This pioneering work details a 'neocognition' model inspired by the hierarchical model of the visual nervous system proposed by Hubel and Wiesel [50]. Fukushima [25] aimed to develop a neural network with pattern recognition capacity on par with that of humans. To achieve this, they developed an extended hierarchical model consisting of complex components emulating the function of cells in the visual system.

The modern CNN architecture is largely based on three concepts, as described by Nielsen [31]. These concepts are local receptive fields, shared weights, and pooling layers. Whereas in a fully connected network, such as Figure 2.2, a vector of inputs is fed in to the model, in a CNN, the input is a matrix, or a tensor. The matrix entries, or input neurons, correspond to pixel intensity values. Unlike for the fully connected network, in the CNN, regions of input are connected to neurons in the following hidden layers. These input regions are termed **local receptive fields**. A **convolutional kernel**, or filter, sweeps over the local receptive fields, performing a convolutional operation. These convolutional kernels consist of shared weights, where the same weights are applied to the different local receptive fields. As a result of this, the convolutional kernels pick up on the same patterns in the local receptive fields. An example of a 2D convolution is illustrated in Figure 2.5. From this figure, one can see that the dimension of the convolutional kernel, or filter, determines the dimension of the local receptive field and the 'stride' determines the number of pixels the kernel moves after each convolution.

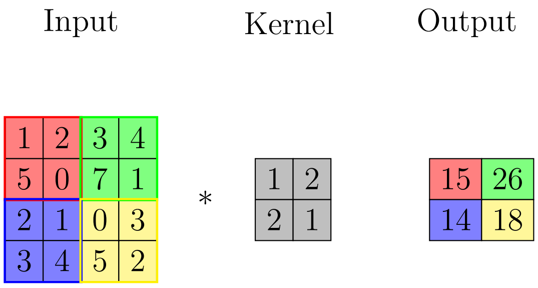


Figure 2.5: Illustration of a 2D convolution with a 2×2 convolutional kernel with stride 2. The '$*$' represents the convolutional operation where the convolutional kernel is applied to an input matrix (left) and results in an output matrix (right). The highlighted blocks of the input are the local receptive fields.

The mathematical formulation of this operation is given by

$$a^l_{j,k} = \sigma\Big(\sum_{m,n} w^l_{m,n} a^{l-1}_{j+m,k+n} + b^l\Big) \tag{2.8}$$

where Equation 2.3 is extended to handle a higher dimensional input, and convolutional kernel operations. The shared weights are given by $w^l_{m,n}$ while the shared bias is given by b^l.

These convolutional operations are often followed by **pooling layers**, which summarise and reduce the spatial information of the layer input. A regularly used form of pooling is max pooling which consists of a filter that preserves the highest intensity value in a local receptive field. Another common form of pooling, average pooling, calculates and outputs the average intensity value in a local receptive field. Max pooling is depicted in Figure 2.6.

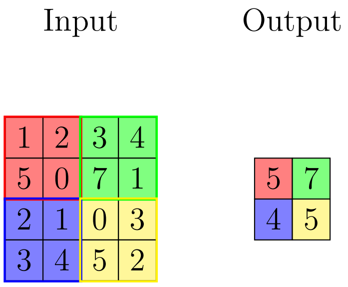


Figure 2.6: Illustration of a 2×2 max pooling with stride 2.

Other components include activation functions such as the ReLU. The ReLU activation function is commonly used in CNNs [39] and is a means of introducing non-linearity to the network. Since convolutional layers are linear this activation function aids the network in learning more complex, non-linear mappings [51]. The use of rectifying nonlinearities, such as ReLU, have been shown to be a key influential factor in accuracy of feature extraction tasks [52]. Other common activation functions include the sigmoid, softmax and hyperbolic tangent activation functions, to name a few.

The CNN architecture is often concluded with fully connected layers where the high-level abstracted features are combined to allow for classification or regression to be performed. In summary, through a combination of convolutions, pooling, activations and fully connected layers, CNNs are a successful tool for abstracting information from 'grid-like' input and performing classification and regression tasks.

There are a wide range of loss functions used for training CNNs. Commonly used loss functions include the MAE and MSE loss functions, given in Equation 2.4 and Equation 2.5, respectively. These loss functions can be interpreted to penalise deviations in the prediction from the desired output. For the case of the MAE loss function, in Equation 2.4, deviations are penalised equally for over and under-predictions, while the MSE loss function, in Equation 2.5, penalises larger deviations more harshly, as the loss is squared.

In the field of computer vision, CNNs have steadily gained recognition, with 'LeNet' by LeCun et al. [27] being one of the early and influential models. Additionally, LeCun et al. [27] introduced the Modified NIST (MNIST) dataset - a dataset of handwritten digits, which has become a common benchmark for evaluating the performance of models. Over the years, several prominent works have introduced variations and improvements on the CNN structure, yielding impressive results. These include the models ResNet [28], Inception [53], DenseNet [54], VGGNet [55] and EfficientNet [56], for example. CNNs have not only been put to the task of image classification [39], but also object detection [57] and semantic segmentation [58] with applications stretching from facial recognition [59] to medical image analysis [60] and art [61].

A technique known as transfer learning is widely used in the field of computer vision, as well as in ML in general. This method allows for a model to be pre-trained on a large initial dataset, which serves as a foundation for future specialisation. These pre-trained models can then be fine-tuned for more specific tasks using new, often smaller, datasets. This pre-training and fine-tuning through transfer learning has been shown to lead to higher performance [62]. The ImageNet dataset consisting of millions of annotated images is a well known pre-training dataset [63].

2.1.3 The Residual Neural Network

It is well established that as the depth of a CNN increases, hierarchical features are learnt from the input data. As shown by Zeiler and Fergus [64], initial layers largely detect simple patterns, for example, edges, while deeper layers recognize increasingly complex features. This ability to learn hierarchical features and capture intricate patterns naturally prompts the exploration of accuracy gains from increasing depth. While research to this end demonstrated improved accuracy [55], a number of training challenges arose.

It was discovered that increasing the network depth did not always guarantee improved performance. With the increase in depth, issues such as the exploding/vanishing gradient problem presented. These problems entailed gradients getting too large (exploding) or too small (vanishing), hindering the learning process. Ioffe and Szegedy [65] mitigated these issues with the introduction of **Batch Normalisation** (BN) which normalises the distribution of the input to the different layers. Another issue present was the degradation problem. It was seen that past a certain depth, degradation would occur whereby the accuracy of the network saturated and subsequently decreased [28]. He et al. [28] reasoned that this degradation is an optimisation problem and proposed a deep residual learning framework to overcome it.

To further understand the motivation behind residual networks, as discussed by He et al. [28], it is helpful to visualise two networks - one shallow and one deep, whereby the additional layers in the deep network are identity mapping and the other layers are taken from the shallow network. Thus, in theory, a deeper network should, at worst, obtain the same level of accuracy as the shallow network. This was not found to be the case. He et al. [28] thus proposed a method that assists deeper networks in learning this identity mapping, if the added layer would result in training accuracy degradation.

Consider the network component - a residual block - illustrated in Figure 2.7. Keeping with the notation used in He et al. [28], we can denote the ideal mapping through these layers to be given by $\mathcal{H}(x)$, where x is the input to the first block. As explained above, the network struggles to learn the mapping $\mathcal{H}(x) = x$ for the case where alternative mappings would degrade the accuracy. Hence, these layers are pushed to instead fit a residual mapping given by $\mathcal{F}(x) := \mathcal{H}(x) - x$, where the ideal mapping $\mathcal{H}(x)$ is explicitly given by the addition of a mapping $\mathcal{F}(x)$ and x, taken from the input of the block. By fitting this residual mapping, we can see that if the identity mapping is optimal the mapping $\mathcal{F}(x)$ would go to zero and the ideal underlying mapping of $\mathcal{H}(x) = x$ would be learnt. This connection of x from a previous to later layer is known as a **'shortcut'**, or **'skip connection'**. For the case where these shortcut connections are identity connections, no additional parameters (nor computational complexity) is added [28].

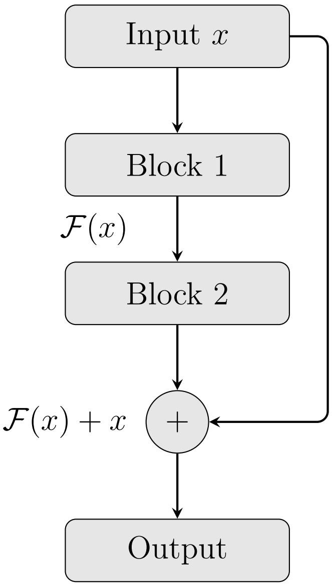


Figure 2.7: Depiction of a residual block proposed by He et al. [28].

As seen in the addition $\mathcal{H}(x) = \mathcal{F}(x) + x$, it is required that $\mathcal{F}(x)$ and x be of the same dimensions. For cases where channel dimensions have increased/decreased between layers, a linear projection (**'projection shortcut'**) may be applied to x to appropriately match channel dimensions.

In summary, if the added layers were to decrease the accuracy of the network, the shortcut connections should instead allow the network to push the weights of the residual mapping to zero so that the ideal underlying mapping may be learnt. These networks with residual connections should either benefit from the increased depth and train to a higher level of accuracy, or remain the same accuracy as the shallower version, and not degrade.

The residual networks proposed by He et al. [28] are known as ResNets and typically start with a 7×7 convolution and subsequent 3×3 max pooling layer. Following this, a series of residual blocks occur. The residual blocks consist of 3×3 convolutions. BN occurs after each convolution and is followed by ReLU activation. The depth varies according to the specific architecture, with ResNet-18 (18 layers) being the smallest and

ResNet-152 (152 layers) being the largest initially proposed. The ResNets conclude with an average pooling, fully connected and softmax layer.

These ResNets have proven to be successful with an ensemble being applied to the ImageNet test set, achieving an error of 3.57%, and winning first place on the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) classification task [66]. First place was also achieved in the ImageNet detection and localisation and Common Objects in Context (COCO) detection and segmentation tasks [67]. Aside from increasing accuracy, residual connections have been shown to increase network convergence speed [68]. Work by Zagoruyko and Komodakis [69] further demonstrates that these residual connections are useful for increasing performance of shallower networks. The work shows that shallow networks with increased width (number of channels/feature maps) outperform deeper and narrower networks. Similarly, Tan and Le [56] show that the balance of depth, width and resolution is key for performance.

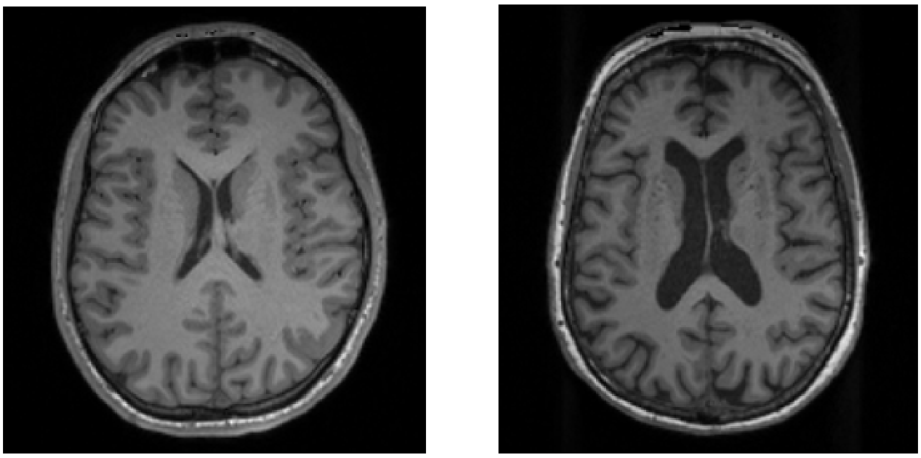
In summary, ResNets, as proposed by He et al. [28], have been instrumental in the training of deep neural networks. By introducing residual connections, these networks effectively mitigate the degradation problem, allowing for deeper architectures without a loss in performance. This is evident in their ability to consistently achieve high accuracy, as demonstrated in various competitions like the ILSVRC and COCO detection and segmentation tasks [28]. ResNets are suited for complex inputs, such as medical images and have been successfully applied to tasks such as tumor segmentation [70] and brain age prediction [11, 49]. We shall expand on this in Section 2.5.

2.2 Brain Ageing

As one ages, one undergoes numerous physical changes, both external and internal. In this section, we outline the brain changes that occur along a healthy brain ageing trajectory. One measure by which a healthy ageing trajectory can be tracked is through the analysis of an individuals brain health, or rather brain age. In a healthy individual, their physical - or chronological age - is matched to the apparent age of their brain. Mathematically, this is defined as

$$\delta_{\text{theory}} = \text{age}_{\text{brain}} - \text{age}_{\text{chron}} \approx 0. \tag{2.9}$$

Particularly in individuals over 60, the healthy ageing trajectory is linked to overall decreases in brain volume and weight, and increases in ventricular volumes, and other cerebrospinal fluid (CSF) filled spaces [71]. In more detail, as one ages, substantial brain volume decreases can be found in the prefrontal cortex [72, 73], followed by the striatum. Regions such as the temporal lobe, cerebellar vermis, cerebellar hemispheres and hippocampus also experience decreases in volume with age [71, 73], with studies showing the occipital cortex to be least affected [72]. Lesions [73, 74] and impairment of WM integrity [74–76] have also been found to become more prevalent with increased age. Furthermore, the myelin sheath has been seen to deteriorate after the age of 40, with the frontal lobes being most affected [73]. Over the age span of 30 to 90 years, Jernigan et al. [77] finds the cerebral cortex volume to decrease by 14%, the hippocampus by 35% and the cerebral WM by 26%. Due to these structural brain changes, there are cognitive changes that arise. Peters [73] discusses memory to be the most prominent cognitive change that occurs with ageing. Figure 2.8 demonstrates the change in brain structure that occurs with age. While the scans originate from different individuals, they are suitable to illustrate the overall changes that occur. The decrease in WM and GM volumes, as well as the increase in CSF in the ventricles are evident.



(a) 18 years (b) 87 years

Figure 2.8: Illustration of changes in the brain with age [78–80]

Studies have shown that certain diseases, such as Alzheimer's [81], type 2 diabetes [82], as well as mental illnesses such as depression [12], may cause individuals to deviate from a healthy ageing trajectory and cause an increase in brain ageing. In Section 2.4, we discuss the impact of HIV on the brain and motivate how this impact may lead to accelerated brain ageing.

2.3 Medical Imaging

As mentioned previously, CNNs and ResNets have had useful applications to medical imaging analysis. These applications include detection of disease [83], biomedical image segmentation [84], and most relevant to the research questions of this work, brain age prediction [9]. Medical imaging techniques for brain imaging include CT scans, PET scans and magnetic resonance imaging (MRI).

MRI is a medical imaging method which uses magnetic fields, as opposed to ionising radiation (such as positrons in Positron Emission Tomography scans), and has been valuable in producing highly detailed brain imagery. As a result, this non-invasive technique has become a highly useful tool in the healthcare industry, and is regularly used for diagnosing and monitoring health conditions. Due to the 'grid-like' structure of these MRI scans, they are suitable input for CNNs and are regularly used in the literature (see Section 2.5). For these reasons, we shall focus on MRIs in this section.

There are a number of magnetic resonance imaging sequences, such as T_1-weighted and T_2-weighted, functional MRI (fMRI), diffusion-weighted imaging (DWI) and GM/WM segmentation maps (see Figure 2.9). As seen in Section 2.5, T_1-weighted structural MRI brain scans are a commonly used modality for input to models tasked with brain age prediction.

MRIs work through applying an external magnetic field which results in nuclei in one's body aligning with the direction of the applied magnetic field [85]. When a second radiofrequency (RF) magnetic field is applied perpendicular to the initial external field, the nuclei may be excited. Once this occurs the nuclei may release energy and drop into a lower energy state. This energy is emitted from the body, detected by the machine as

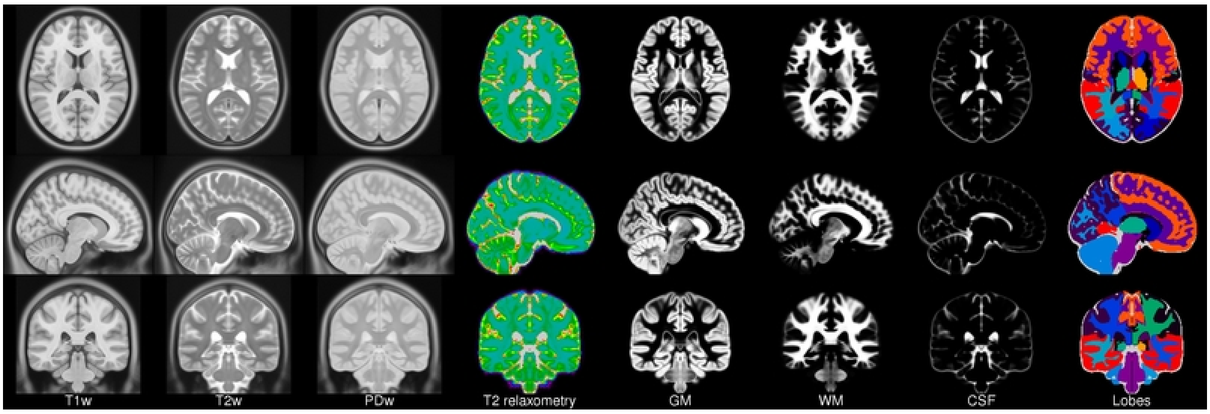


Figure 2.9: Example of axial (top), sagittal (middle) and coronal (bottom) slices of brain MRI scans of different modalities and processing [6–8]. Copyright (C) 1993-2004 Louis Collins, McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University.

an electrical current and transformed to an image by a computer [86]. Different types of nuclei types emit distinct frequency signals, allowing for the extraction of the MRI. Relaxation time - the time taken for the nuclei to return to thermal equilibrium - is an important factor in the MRI outcome. Two main types of relaxation are observed: longitudinal (denoted as T_1) and transverse (denoted as T_2) [85]. We discuss T_1 relaxation as T_1-weighted scans are the scan modality used in this thesis.

The properties of different tissues in the body determine their T_1 relaxation times, which in turn affect how they appear in MRI scans. Tissues with longer T_1 relaxation times appear darker in T_1-weighted images. An example is CSF, which has longer T_1 times, leading to its darker appearance in T_1-weighted scans [86]. Conversely, tissues with shorter T_1 times, like fat, exhibit the opposite behaviour. They appear bright in T_1-weighted images. In addition, WM, which has shorter T_1 times compared to GM, appears lighter in T_1-weighted scans [87].

To produce a T_1-weighted image, a short repetition time between RF pulses and a short signal recovery time is required. Contrasts in the T_1-weighted scans are desirable for the task of brain age prediction where a distinct anatomical contrast is beneficial [86].

The Magnetisation Prepared Rapid Gradient Echo (MP-RAGE) sequence is a widely used technique in T_1-weighted structural imaging in MRI. This sequence is efficient and produces high tissue contrast and high spatial resolution images [88]. Furthermore, in MR imaging, different resolutions are used, including the commonly seen 1mm isotropic resolution. This resolution refers to the size of the voxels in the scan, where each voxel is a cube measuring 1mm on each side. We see the use of these sequences and resolutions in Section 3.1.

Brain scans are an informative medium to investigate the anatomical and functional changes associated with disease and ageing in the brain. Given that manual analysis of these images by clinicians can be labour-intensive and time consuming, CNNs provide an efficient approach to medical image analysis.

2.3.1 Challenges with Medical Imaging

Medical imaging, particularly MRI, plays an important role in modern healthcare. However, the use of these images in ML, especially in DL applications, poses several challenges.

One common issue with MRI scans is the presence of motion artifacts [89]. These artifacts, often appearing as blurs or distortions in the images, are primarily due to patient movement during the scanning process. Such motion artifacts can obscure critical anatomical details, potentially leading to misinterpretations or reduced accuracy in diagnostic and predictive models. Another challenge in leveraging medical imaging datasets is the access restrictions due to the sensitive nature of the data. Ensuring patient privacy and adhering to data protection regulations often limits the availability of large-scale datasets for research and development. Furthermore, the processing and utilisation of high-resolution MRIs with deep CNNs prove to be a challenge computationally. The high dimensionality of MRI data requires substantial computational resources for storage, processing, and analysis, which can be a limiting factor, especially in resource-constrained environments.

2.3.2 Pre-Processing

It is common practice to pre-process brain MRI scans prior to inputting them to the model [9, 15, 17]. In this section, we give an overview of the common pre-processing steps used in the literature.

Pre-processing of brain MRI inputs for models includes steps such as reorientation, skull-stripping and registration (e.g. Cole et al. [9]). Certain studies [15, 17] use more complex pre-processing pipelines such as that engineered by the UK Biobank and detailed in Alfaro-Almagro et al. [90].

Re-orientation, usually the first pre-processing step, is performed to orientate the scans with a standard template image, such as the Montreal Neurological Institute (MNI) template. This ensures that the scans are orientated in the same way, and is also necessary for later pre-processing steps like registration. The *fslreorient2std* tool from Functional MRI of the Brain (FMRIB's) Software Library (FSL) [2] can be used for reorientation.

Brain extraction, commonly known as **skull stripping**, often follows re-orientation. Skull stripping is the process whereby the non-brain matter components, such as skull and meninges, are removed from a brain scan. Since these are non-brain matter components, we do not want them to factor in to the brain age prediction and thus it is typical to remove them. There are a number of brain extraction tools available for brain extraction. Isensee et al. [1] introduce a model geared towards brain extraction, known as HD-BET. They compare different softwares used for brain extraction and find HD-BET to be robust. Figure 2.10 demonstrates skull stripping on a template brain scan.

A recent study by Wood et al. [91] provides evidence that skull stripping may not be a necessary pre-processing step, particularly for the case of larger datasets. The paper showed, by inspecting the brain age saliency maps, that the model was able to learn that the non-brain matter was not relevant to the brain age calculation.

Registration is used to align images to ensure correspondence in spatial features between different images [92]. This step usually follows skull-stripping and consists of either/both linear and non-linear registration of the brain scans to a specific model template. Linear registration usually involves registration with six, nine or twelve degrees of freedom. These degrees include transformations such as rotations, translation or scaling [92]. Non-linear registration has less constraints (more degrees of freedom) and is thus far more computationally costly. Common templates originate from scans within the MNI152 database. Brain atlases can be generated by taking unbiased non-linear averages of these scans. A number of different atlas templates are available through the McConnell Brain Imaging Centre ICBM study [6–8]. These templates ensure the brain

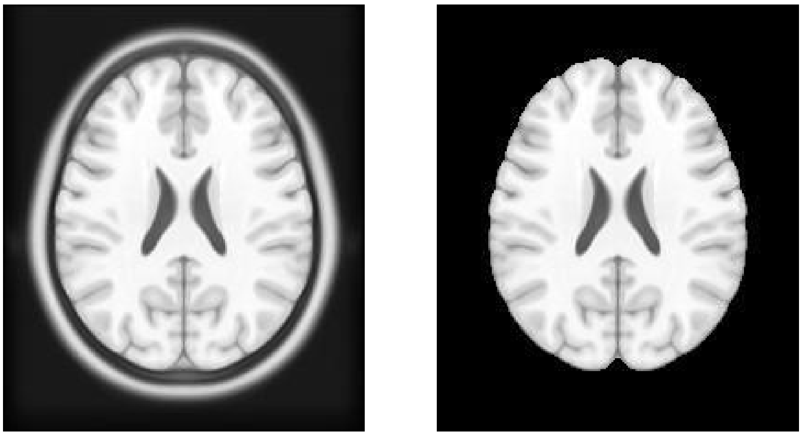

(a) Pre skull-strip (b) Post skull-strip

Figure 2.10: Template [6–8] example of the skull-stripping pre-processing step used to remove non-brain matter, such as skull and meninges, from a brain MRI scan.

positioning and size of the brains are uniform between scans.

Dinsdale et al. [17] discuss that non-linear registration suppresses morphological information, which is thought to be relevant to brain age prediction. Despite this, they achieve comparable MAEs on the linear and non-linearly registered data. The *FLIRT* [3–5] tool provided by FSL [2] may be used to linearly register the scans to standard MNI space. Similarly, the *FNIRT* tool may be used for non-linear registration.

2.3.3 Harmonisation of MRI Scans

When compiling a large, diverse dataset, it is often necessary to pool datasets from different sources [93]. While ensuring that this creates a larger dataset, this technique also introduces non-biological variance related to the sites at which the datasets were collected. This variance is a form of domain shift, where differences between training and testing data distributions may lead to a drop in ML model performance [94]. Bayer et al. [95] discuss and list causes of scanner variation which include scanner platforms, differing sequences and acquisition procedures.

A harmonisation method known as ComBat is a powerful batch adjustment tool. The first implementation of ComBat was proposed by Johnson et al. [96] as a means to correct for batch effects in microarray studies combining data from multiple experiments. Here batch effects take the form of non-biological variation as a result of a range of factors, such as differing experimental conditions. This variation reduces the amount of data that can be combined for analysis and results, thus limiting the statistical power of the experiment.

The proposed method implements a parametric and non-parametric empirical Bayes (EB) framework. Johnson et al. [96] discuss how EB methods are a useful tool in the field as they are adept at working with high-dimensional data for the case of small sample sizes. This is particularly relevant to neuroimaging data, where the dimensionality is high due to the large number of MRI voxels, while the number of available scans might be limited.

In more detail, their approach works through parametric shrinkage adjustment. The method of parametric shrinkage adjustment entails standardisation of data, estimation of batch effect parameters with EB parametric empirical priors and lastly application of estimated batch effects to the dataset. Standardisation of the data ensures that the batches have comparable overall means and variances. This step minimises bias in the EB estimates of the prior distribution of batch effect. Johnson et al. [96] detail how the estimation of batch effect parameters using EB parametric empirical priors is achieved by estimating the location and scale model parameters for batch effects. They use a 'pooling information' strategy across data within each batch, effectively 'shrinking' the batch effect parameter estimates towards the overall mean across the dataset. The parametric priors were found to not describe all data well, hence Johnson et al. [96] additionally proposed the non-parametric prior method. The final step involves the application of adjusted batch effect estimators, derived from the conditional posterior, to the dataset. This adjustment reduces the batch effects. The proposed methods prove to be effective for small datasets, which was previously a challenge in the field, while maintaining comparable performance with existing approaches for larger datasets.

There have been a number of extensions made for ComBat, some with neuroimaging support. We explore these extensions to find a tool able to reduce domain shift, such as shift due to scanner variation. For example, neuroComBat offers support for harmonisation of multi-site data [97]. Similarly, a package called neuroHarmonize extends neuroComBat with specific support for harmonisation of neuroimaging data [98]. The application of ComBat for harmonisation has been shown to effectively reduce variance due to site and scanners [99]. It has also been shown that ComBat is effective for harmonisation of small multi-site datasets, including those which do not have a balanced distribution of key covariates [97, 100]. It should be noted, that according to Fortin et al. [97], the efficacy of harmonisation for voxel-level whole-brain analyses might be limited. This limitation arises because ComBat works under the assumption that scanner effects are shared between voxels. However, this assumption may not hold true for whole-brain voxel analyses, where scanner effects could vary considerably across different voxels. This variation is due to voxel intensities varying depending on the tissue medium, and so Fortin et al. [97] propose a future expansion of ComBat which includes these varying effects.

Bayer et al. [95] discuss implications of the use of ComBat-GAM, which is offered within the neuroHarmonize package. ComBat-GAM, extends on ComBat, and features Generalised Additive Model (GAM) functionality. This statistical method makes use of EB for fitting prior distributions. The downside of this is that EB may not apply well to sites with varying sample sizes, with the smaller sites being over moderated. EB also draws on the assumption that the distributions of features are defined with a single mean and variance within each site. Due to the numerous factors (e.g. lifestyle, substance use, health conditions) that play a role in the brain ageing process, this is likely not to be the case for the datasets in question.

Recent research by Kushol et al. [94] indicates that ComBat harmonisation may be less effective for disease classification tasks. The study details the substantial classification accuracy drop of DL models when tested on data from different scanner manufacturers compared to the data used in training, highlighting an instance of domain shift. The study then focuses on the harmonisation of this multi-scanner dataset of individuals with neurodegenerative diseases and performs a classification task based on disease status. They make use of 3D structural MRIs and utilise deep learning for disease classification. They were unable to detect enhancements in classification ability due to the harmonisation indicating a need for further investigation into the efficacy of harmonisation techniques in DL models for disease classification tasks.

In the case of medical imaging, the behaviour of ComBat harmonisation may be assessed through analysis of the intensity histograms of the images. One is able to set a 'reference site', which is the site which the other medical images should be harmonised to. One is able to plot the histogram of intensity voxels per site, prior to, and following harmonisation. From these plots, one is able to see that the intensity histograms of the other sites are shifted and scaled to be of a similar distribution to the reference site. We demonstrate this further in Section 4. This is not a full verification of the harmonisation method though. Dinsdale et al. [101] discuss the challenges in validating output images that have been harmonised. As detailed in Moyer et al. [102], pairs of scans - one from the original scanner and one from the reference scanner - are required to validate harmonisation. In practice, however, such data is rarely available.

Liu et al. [103] tackle harmonisation as a style transfer problem, whereby the goal is to apply the style of one image (such as its colour) onto the content of another, while preserving the core elements of the target image. Liu et al. [103] note that statistical approaches have a tendency to remove relevant information. To achieve harmonisation, they implement an unsupervised DL model inspired by a generative adversarial network framework [103]. In essence the style information from a reference image is extracted and transferred to a wider set of images. To this end, they successfully removed inter-site related variance while maintaining anatomical and clinically valuable features.

2.4 HIV and the Brain

As of the end of 2022, an estimated 39 million people were living with HIV [21]. Of these 39 million, 1.5 million were children younger than 15 years old. Despite the rate of HIV infection decreasing over the years, over one million individuals were infected with HIV in 2022 [21]. Since the introduction of ART, the treated population of HIV+ individuals has been steadily increasing with approximately 76% of people living with HIV receiving treatment in 2022. However, HIV is still a critical health problem with over half a million people dying of AIDS-related conditions in 2022, of which approximately 16% were under the age of 20 [104]. In this section, we detail the mechanism through which HIV acts and discuss literature outlining the impact HIV has on an individual's health, with a focus on the neurological effect. This section motivates the research question of what the effects of HIV are on an individual's brain ageing.

Chen [105] discuss the mechanism of infection of the HIV-1 virus. The HIV-1 type is the more common of the two main HIV types: HIV-1 and HIV-2 [106]. Since HIV-2 generally has a less severe impact on the immune system, and is less prominent, we primarily discuss HIV-1, which we shall further refer to as HIV. The first stage of HIV entry to a host cell involves the binding of the HIV viral membrane, also referred to as the protein envelope (Env) [107], using the gp41 and gp120 receptors, to the CD4 receptors of immune system cells. These immune system cells include the helper T-cells and macrophages. The CD4 receptors play a role in initiating the activation of the helper T cells in the immune response [105]. Once bound to the CD4 receptors, the virus then binds to a coreceptor, such as the CCR5 or CXCR4 coreceptors. Once the virus is attached to the target, or host cell, it is able to pass through the cell membrane by fusion [105, 107]. The attachment process is illustrated in Figure 2.11.

Following the fusion of the viral cell membrane with the host/target cell membrane, the virus delivers its viral genome into the host cell cytoplasm. Since HIV is a retrovirus,

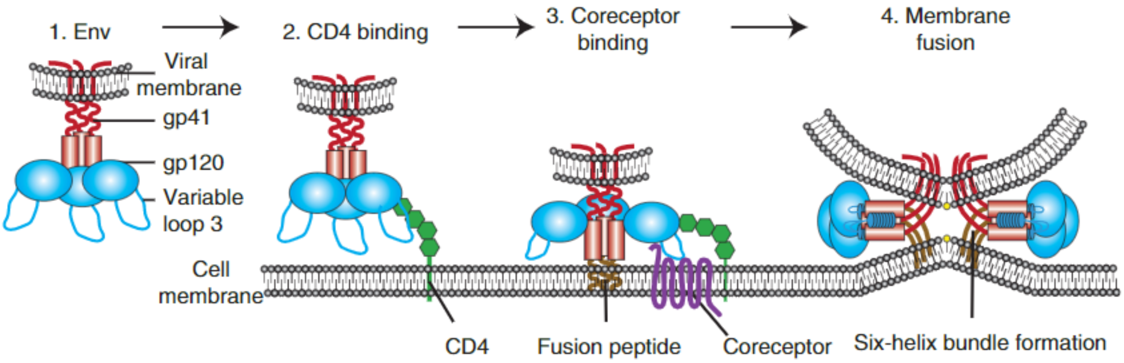


Figure 2.11: The mechanism of HIV entry as depicted by Wilen et al. [107].

reverse transcription is initiated on the delivered viral RNA [108]. The viral genome is then integrated into the hosts chromosome. Further transcription produces viral RNA which then leads to the production of HIV. The steps detailed above are illustrated in Figure 2.12.

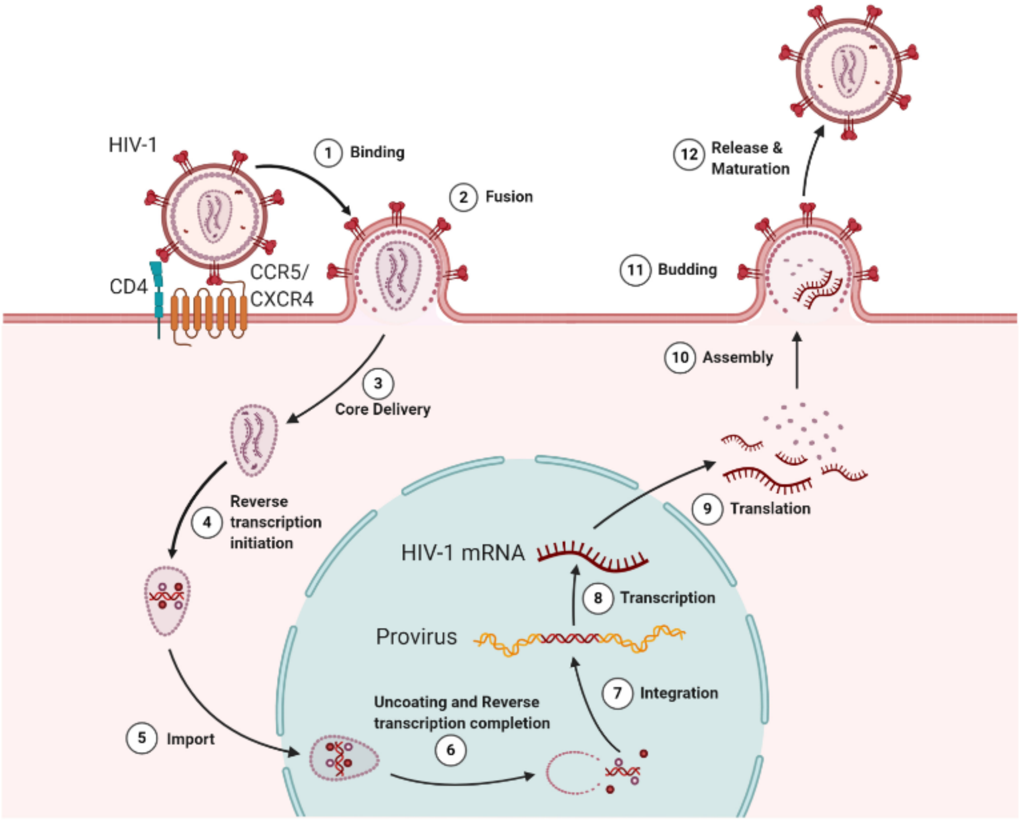


Figure 2.12: The mechanism of HIV entry, viral contents delivery, reverse transcription, integration and translation and release stages, as depicted by Ramdas et al. [108].

HIV leads to the destruction of CD4+ T-cells [106]. These CD4+ T-cells are a key component in a healthy immune system. As more HIV is produced throughout the body, more CD4+ T-cells are destroyed, thus weakening the individual's immune system. When the CD4+ T-cell count of an individual drops below 200 cells/mm^3 blood, they are said to have AIDS. We shall further refer to CD4+ T-cell counts as **CD4 count**.

Lower CD4 counts are associated with weakened immunity and increased susceptibility to opportunistic infections [109]. The CD4 count is a commonly used clinical measure reflecting the number of these cells per cubic millimetre of blood and is used for assessing immune system health in HIV+ individuals. The **nadir CD4 count** is an associated clinical measure, referring to the lowest recorded CD4 count of an individual. This measure is valuable for assessing the severity of previous immunodeficiency, for example, prior to treatment with ART. Another highly informative clinical measure [110] is the **HIV RNA level**, also known as **viral load**, which quantifies the amount of HIV virus in the blood.

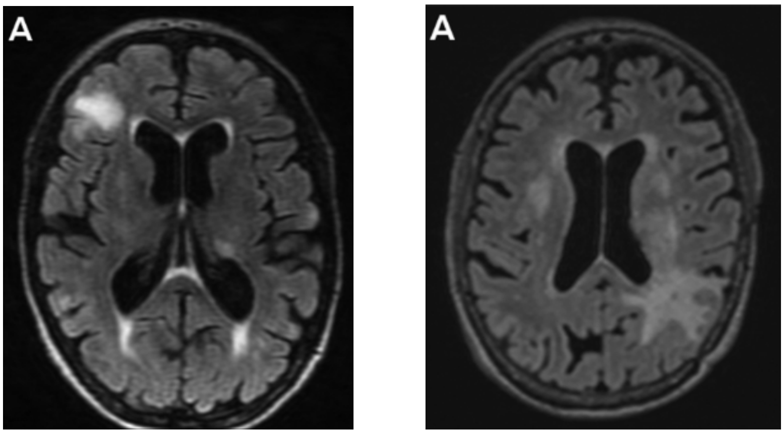
As a result of the widespread prevalence and serious health impact HIV has had on society, a great deal of research has been put towards eradicating this virus and treating those affected. It has also been of huge importance to understand the health impact HIV has.

HIV is known to have a large impact on an individual's health and can cause damage to various systems and organs within the body [111]. With the advent of ART, life expectancy for those living with HIV has substantially increased. Alongside this improvement, however, a rise in comorbidities, such as cardiovascular disease, cancer, diabetes, and liver disease, has been observed in HIV+ individuals compared to their HIV− counterparts [112, 113]. Additionally, ART has been associated with its own set of challenges, including toxicity and mitochondrial dysfunction, with these effects being particularly pronounced in those undergoing long-term treatment [114]. Mitochondria play a crucial role in cellular energy production, and impairment of their function can have detrimental effects. There is promising research that suggests that newer ART regimens may have less severe effects on mitochondrial function [115].

Aside from the health impact HIV may have on the body as a whole, HIV also has a considerable neurological impact. This neurological impact stems from HIV infecting monocytes and macrophages which cross the blood brain barrier (BBB) and enter the central nervous system (CNS) [116]. Once in the brain, further infection ensues. Despite combined antiretroviral therapy (cARTs) being beneficial to managing HIV and AIDS, early treatments were limited in the ability to traverse the BBB and thus the neurological effects of HIV persisted [117].

In 1991, Budka [22] discussed the neuropathological implication of infection with HIV. This includes HIV encephalitis (HIVE) and HIV leukoencephalopathy (HIVL), and are more prominent during infection with AIDS. In HIVL, the WM of the brain is often damaged [118]. This WM damage has links to brain ageing (see Section 2.2). Since the introduction of cART, HIVE and HIV-linked dementia have been more manageable [119], but not eradicated. A prominent, debilitating treatment-resistant HIV-related disease is distal neuropathic pain (DNP). Keltner et al. [120] show that the severity of this disease is correlated with cortical GM atrophy that occurs as a result of HIV. Their multi-site cohort study performed cross-sectional analysis on structural MRIs of HIV+ individuals with (n=66) and without (n=175) DNP.

Grill [23] further discuss the neurological impact HIV may have. Broadly, HIV related CNS conditions fall under the categories of cerebrovascular disease (e.g. stroke), meningitis (e.g. neurosyphilis), encephalitis (e.g. HIVE), intracranial mass/focal lesions (e.g. toxoplasmosis) and myelitis (e.g. HIV vacuolar myelopathy). As later discussed by Heaps et al. [121], HIV-associated neurocognitive disorders (HAND) are prevalent in HIV+ individuals. HAND ranges in severity from asymptomatic neurocognitive impairment to HIV-associated dementia. Grill [23] state the primary risk factors for HAND to be low nadir CD4 counts and the absence of ART treatment, with other factors such as older age, vascular risk factors (e.g. hypertension), and illegal drug use contributing to prevalence. Other infections, such as compartmentalised CNS infection, also known as CSF escape, can be present. This involves continued replication of the HIV virus in the CNS, while the individual displays relative viral suppression in the blood with ongoing ART treatment. Due to the subsequent weakening of the immune system which may feature in individuals with HIV, opportunistic infections and co-infections are also prominent. Since lower CD4 counts are linked to higher degrees of immunodeficiency, an individual with a lower CD4 count is more susceptible to opportunistic infections [23]. An example of such opportunistic infections includes toxoplasmosis and progressive multifocal leukoencephalopathy (PML), both of which contribute to lesions [23]. As stated by Schouten et al. [118], PML, or HIVL may cause damage to the WM. Toxoplasmosis is the primary cause of brain lesions in HIV+ individuals. The risk of toxoplasmosis is increased in individuals with CD4 counts < 200 cells/mm^3, and further so with CD4 counts < 100 cells/mm^3. These lesions are commonly found in the basal ganglia and cortical GM [23]. The symptoms of individuals with toxoplasmosis include headaches, fever, affected mental state and focal neurologic signs such as affected speech [23].



(a) Toxoplasmosis. Signal intensity changes can be seen in the left and right hemispheres.

(b) CNS lymphoma. Diffuse WM lesions can be seen in the left hemisphere (right side of scan).

Figure 2.13: Display, taken from Grill [23], of neurological effects of HIV using axial fluid-attenuated inversion recovery (FLAIR) brain MRI scans.

Heaps et al. [121] explore the relation between brain volumes and HAND in HIV+ individuals. The study uses a dataset of untreated HIV+ individuals with (HIV+ HAND+, n=37) and without HAND (HIV+ HAND−, n=37) and HIV− controls (n=29). Total and subcortical GM, total WM, caudate, corpus callosum and thalamus were studied using MRI scans to assess the relation in question. Neuropsychological tests were also performed. While a significant decrease in brain volume of total and subcortical GM was found for HIV+ HAND+ relative to HIV− controls, a statistically significant difference could not be identified between brain volumes of HIV+ HAND+ and HIV+ HAND− or between HIV+ HAND− and HIV− controls. The study cites that a larger sample size and longitudinal information may improve results.

Thompson et al. [122] study the cortical GM thickness in AIDS patients (n=26) and healthy controls (n=14). While a small study, they find that cortical GM tissue loss was prevalent in individuals with AIDS. Nir et al. [123] explore whether HIV+ individuals on cART, with successful viral suppression, still experience progressive brain atrophy and neurological consequences. Their dataset consists of n=155 HIV+ virologically suppressed individuals, with a nadir CD4 count less than 200 cells/mm^3, from the longitudinal HIV Neuroimaging Consortium (HIVNC). This longitudinal study allows for the rate of progression of the infection to be studied, as opposed to using a cross-sectional study. As mentioned in Heaps et al. [121], longitudinal information may assist in improving results. Selecting two time points from the longitudinal HIVNC study, with a mean interval of 1 ± 0.5 years, and applying tensor-based morphometry, Nir et al. [123] were able to analyse the rate of change of regional brain volumes in the HIV+ dataset. Furthermore, correlations between this rate of change of regional brain volumes (brain tissue loss) and measures such as nadir CD4 count, baseline CD4 count, baseline HIV plasma RNA concentration, duration of HIV infection, cART efficacy and AIDS Dementia Complex stage were analysed. Nir et al. [123] finds that HIV+ virologically suppressed individuals may still experience continual brain tissue atrophy. Furthermore, they found that measures of the disease severity, such as lower CD4 and nadir CD4 count and higher HIV RNA concentration, as well as age and neurocognitive decline, were linked with higher levels of atrophy. This study provides us with useful information on the severe impact of HIV, even in the case of treated individuals. Clifford et al. [124] compare n=24 age-matched healthy controls with n=38 HIV+ individuals. For both groups, they select individuals over 60 years to investigate the older HIV demographic which is a growing population since the introduction of ART. They analyse brain atrophy rates in longitudinal structural brain MRIs. To carry out this analysis, they make use of tensor-based morphometry. The study finds that the HIV+ group displayed higher levels of brain atrophy relative to the age-matched healthy controls.

The results from this section are summarised in Table 2.1. From Table 2.1, we can conclude that HIV can have a severe impact on the brain, once it crosses the BBB [116]. The effects present themselves in both untreated and treated HIV+ individuals and include cortical and subcortical GM atrophy [120–122], WM damage [22, 118] and lesions [23]. Overall brain atrophy [124] and subcortical atrophy [123] were seen in virally suppressed, treated individuals.

These neurological effects provide a strong motivation for investigating the brain ageing impact of HIV. It is important to note however, that there is evidence to suggest that MRI detection of HIV in individuals without neurological symptoms may not be useful [125], thus consideration of the severity of the HIV case is of importance when analysing predicted brain ages. In the following sections, we see the applications of DL to predicting brain age and the use of ML techniques to investigating the impact of HIV on the brain.

2.5 CNN Applications to Brain Ageing

In this section, we discuss previous CNN architectures tasked to predict brain age. In order to address the research questions of this work and investigate the impact of HIV on brain ageing, a model which can predict brain age is essential. This section will be instrumental in motivating the model design choices in Section 3.3.

MRI brain scans contain highly complex, spatially distributed features and 3D CNNs are a natural choice for the task of extracting information from them. The application of

CNNs to predict brain age began with Cole et al. [9] in 2017. Since their seminal work, numerous studies have further explored CNNs in this context of brain age prediction, some of which are highlighted below.

As of the writing of this thesis, the state-of-the-art (SOTA) for brain age prediction with a CNN is held by Peng et al. [15] with an MAE of 2.14y achieved on the UK Biobank dataset, spanning the age range 44-80y. Peng et al. [15] implement a simple fully convolutional network (SFCN) to predict the brain age of T_1-weighted structural brain MRI data. The SFCN has considerably less parameters than a majority of previous models (see Table 2.3) and is ideal for working with small datasets and 3D volumetric data. The work implements performance boosting techniques, such as data augmentation, model regularisation, model ensembles and prediction bias correction. We discuss this bias correction below in Section 2.5.1. The recent work of Yin et al. [16] also implement a lightweight 3D CNN, with their model containing fewer parameters than that of the SFCN. They combine datasets from multiple sources - the UK Biobank, CamCAN, Alzheimer's Disease Neuroimaging Initiative (ADNI) and Human Connectome Project (HCP) - and train and test on a subset of cognitively healthy scans from the UK Biobank, ADNI, and HCP. Overall their dataset spans the age range 18-92y. They achieve an MAE of 2.23-2.41y which is competitive with SOTA. This MAE range is due to Yin et al. [16] obtaining distinct MAEs for males and females. Furthermore, comparing the testing of the same unseen data on the SFCN of Peng et al. [15] and the 3D-CNN of Yin et al. [16], Yin et al. [16] obtain lower MAEs. Dinsdale et al. [17] also make use of the UK Biobank dataset (age span stated as 44.6-80.6y) and implement an ensemble of 3D CNNs to predict brain age. They explore the effects of training on linearly and non-linearly registered images and use a subset of axial slices to conserve memory during computations. Like Peng et al. [15] and Yin et al. [16], Dinsdale et al. [17] use a low number of convolutional layers (twelve) and achieve an MAE of 2.71-3.09y. The MAE age range is a result of training separately on male and female groups of linearly and separately, non-linearly registered images. Dinsdale et al. [17] monitor the MSE on the validation set to decide whether the learning rate should be decayed or if training should be stopped. Like Peng et al. [15], Dinsdale et al. [17] also follow Smith et al. [126] to remove brain age bias. Levakov et al. [48] train a CNN ensemble on T_1-weighted brain MRI data (age range 4-94y) and obtain an MAE of 3.07y. Like Peng et al. [15], Levakov et al. [48] implement dropout and data augmentation. To obtain the MAE, they implement a simple linear regression to find an optimal linear combination between each model output in the ensemble. Hofmann et al. [127] also explores the use of combined ensembles of CNNs to determine brain age and generate saliency maps to aid in the interpretability of how different brain regions contribute to brain ageing. Saliency maps to aid model interpretability are also used by Yin et al. [16], and in other brain ageing related research by Taylor et al. [128]. Hofmann et al. [127] makes use of population-based data (ages 18-82y) to train the neural network, and obtains accurate predictions with an MAE of 3.37-3.86y for both single and multiple modalities, and regionally restricted and whole-brain images. Jonsson et al. [49], Kolbeinsson et al. [11] and Bashyam et al. [93] make use of residual blocks, inspired by the ResNet architecture, in their models. They achieve MAEs of 3.39y, 2.87-3.42y and 3.702y, respectively. Note that the Kolbeinsson et al. [11] MAE range derives from them reporting an MAE of 2.87y on an N=2057 health stratified subset of UK Biobank and an MAE of 3.42y on a subset not stratified for health with N=12196. Jonsson et al. [49] and Bashyam et al. [93] span age ranges of 18-82y and 3-95y, respectively. While the age range of Kolbeinsson et al. [11] is unknown, they make use of the UK Biobank dataset which is commonly quoted to span an age range of approximately 44-80y. Feng et al. [129] and Cole et al. [9] follow very similar model structures and achieve MAE's of 4.06y and 4.16-4.66y, respectively. Their datasets span the age range 18-97y and 18-90y, respectively.

In Table 2.2 a summary of the datasets, input modalities, dataset age ranges, and dataset sizes of previous works are detailed. We include an in-depth summary of these works with details of the architectures, regularisation techniques, loss functions, optimisers, learning rates and batch sizes in Table 2.3. For both tables, we order the rows in ascending order of MAE.

In Table 2.3, as expected from Section 2.1.2, we see convolutional layers, often combined with activation functions like ReLU or its variants (Leaky ReLU, exponential linear unit), and pooling operations (max pooling, average pooling). The complexity and depth of these architectures vary; some papers, such as Peng et al. [15] and Yin et al. [16], feature very lightweight models, while others, like Bashyam et al. [93], utilise much deeper models.

Regularisation plays a crucial role in ensuring generalisation in DL models. Dropout and L2 weight regularisation are the most commonly employed techniques across the selected studies. Dropout aids generalisation by randomly deactivating a specified percentage of the features in a layer during each training iteration. L2 regularisation penalises large weights by introducing a penalty term to the loss function. This encourages the model to learn smaller weights and reduce the risk of overfitting. The dropout percentages used vary, with a range between 0.2 (20%) to 0.8 (80%), indicating that different studies have found varying levels of regularisation effective. The choice of optimisers and learning rates also shows variability across the literature. The Adam optimiser is more commonly used, followed by SGD and RMSprop. Initial learning rates generally range between 2×10^{-5} to 10^{-2}, with most works employing learning rate decay. For the batch sizes, a size of eight appears most frequently, though some works opt for larger batches [9, 17].

From Table 2.2 and Table 2.3, it can be seen that studies leveraging the UK Biobank dataset [11, 15, 17, 130] typically observed MAEs ranging between 2.14y and 3.42y. These studies contain age ranges between the mid-40s to early 80s, utilising between 10000 to over 21000 UK Biobank scans. The variation of dataset size within the different studies utilising UK Biobank data can be put down to varying quality control procedures, study selection criteria and date at which the study was carried out in relation to ongoing imaging collections, as well as participant withdrawals [131]. On the other hand, studies utilising datasets from multiple sources, such as Levakov et al. [48], Feng et al. [129] and Bashyam et al. [93], with the strong exception of Yin et al. [16], present MAEs between 3.07y and 4.06y. These multi-site studies span broader age ranges, from early childhood to late 90s, and feature dataset sizes ranging from just over 10000 to more than 30000 scans. It is worth noting that although studies utilising multiple sources and broader age ranges, on average, perform worse than the studies utilising the UK Biobank dataset, this difference may be attributed to the increased complexity that accompanies brain age prediction on datasets diverse in site and age. Yin et al. [16] diverge from this and obtain an MAE highly competitive with SOTA, despite utilising a smaller dataset that is more diverse (multiple sources) and spans a broader age range. As noted above, directly comparing the MAE across studies is arguably not fair since different studies often utilise different datasets and span different age ranges. A 'weighted' MAE could be a useful method for fair comparisons, however this was not seen in the literature.

2.5.1 The Brain Age Bias

Throughout the literature, unspecific to the type of ML technique used [132], a trend is found to be prevalent within the prediction of the brain age. This trend is a bias between the chronological and predicted age and presents itself with the over-prediction of age in younger individuals and under-prediction of age in older individuals. Hence, the bias results in a negative correlation between chronological age and the brain age delta, which is given by Equation 1.1. For unbiased results, in healthy individuals, there should be no correlation between chronological age and calculated brain age delta. We primarily discuss the works of Smith et al. [126] and Treder et al. [133] where this bias and proposed corrections are outlined. Causes of bias, as listed in Smith et al. [126], include regression towards the mean and a non-Gaussian distribution of ages of individuals in the study datasets. As seen in the above section, it is common practice for studies involved in the prediction of brain age to implement these bias correction methods [11, 15–17].

Smith et al. [126] discuss a number of methods of correcting for this bias. We closely follow their derivation for linear corrections below. To begin, they represent the vector of predicted brain ages Y_{pred} from imaging matrix X as

$$Y_{\text{pred}} = Y + \delta = f(X) \tag{2.10}$$

where Y is the vector of corresponding chronological age labels, δ is the vector of differences between predicted and chronological age and $f(X)$ is some vector function. They first postulate $f(X)$ to be approximated by a simple linear model acting on an input matrix, i.e. $f(X) = X\beta$ where β is a vector of parameters. Equation 2.10 can then be reformulated as

$$Y = X\beta_1 - \delta_1. \tag{2.11}$$

The subscripts are introduced for clearer notation in the following paragraphs. Equation 2.11 can be solved, for example by representing $\beta_1 = X^{\dagger}Y$, where $X^{\dagger} = (X^TX)^{-1}X^T$ is the pseudo-inverse of X, as Smith et al. [126] lays out. Substituting this in one gets

$$\delta_1 = (XX^{\dagger} - I)Y. \tag{2.12}$$

From this, one can see that the δ_1 vector will not be orthogonal to the chronological ages Y, and thus not independent. Note that since we use $f(X) = X\beta_1$, it is apparent from Equation 2.10 that $Y_{\text{pred}} = XX^{\dagger}Y$. The study corrects for the bias by removing the dependencies of δ_1 on chronological ages Y and achieves this first by assuming linear corrections only, later detailing the case of nonlinear corrections. We present the derivation by Smith et al. [126] for the simpler case of linear corrections.

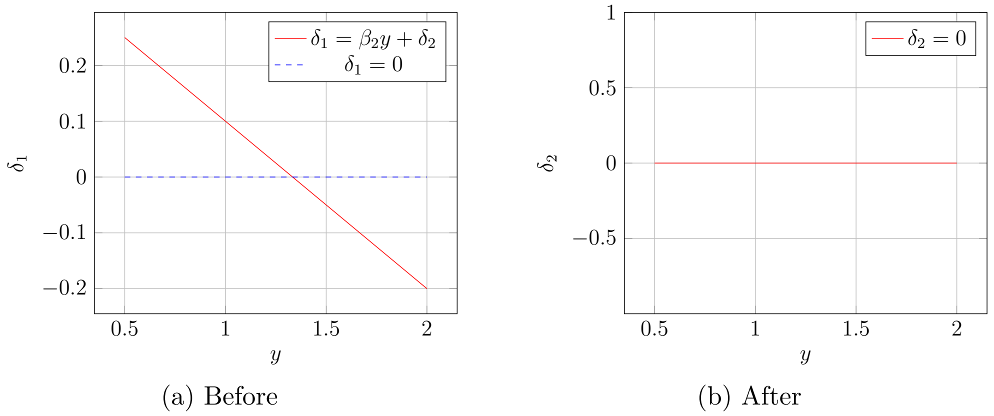
The dependency between chronological age Y and uncorrected brain age delta δ_1 can be given by the linear equation

$$\delta_1 = Y\beta_2 + \delta_2. \tag{2.13}$$

This equation represents the line of best fit of the uncorrected brain age deltas and chronological ages, and is illustrated by the red line in Figure 2.14a. By taking $\beta_2 = Y^{\dagger}\delta_1$, analogously to the procedure of deriving Equation 2.12, Smith et al. [126] then derive the equation

$$\delta_2 = M_yXX^{\dagger}Y \tag{2.14}$$

where $M_Y = I - YY^{\dagger}$ orthogonalises a vector with respect to the chronological age vector Y. As mentioned above, the work shows that $XX^{\dagger}Y$ can be identified as the predicted



(a) Before

(b) After

Figure 2.14: Line of best fit for the brain age delta and the chronological age, y, before and after linear correction proposed by Smith et al. [126]

brain age Y_{pred} and so we can interpret δ_2 as the corrected brain age delta. The line of best fit for corrected delta and chronological age is illustrated in Figure 2.14b.

Peng et al. [15], following Smith et al. [126], use a linear bias correction method to bias correct the brain age delta and fit a linear regression $x = ay + b$ to the validation set where y is the chronological age and x is the predicted age. The corrected predicted age is then given by $\hat{x} = (x - b)/a$. Using these assumed generalisable a, b, the corrected brain age delta is then obtained on the test set.

An alternative approach is proposed by Treder et al. [133] and addresses the bias by developing an analytical solution to a constrained optimisation problem. The constraints are added to the correlation between predicted age and calculated delta during model training. This solution is developed for linear, ridge and kernel ridge regression. Mathematically, the crux of the method can be laid out as the optimisation of the loss function

$$\mathcal{L}(y, y_{\text{pred}}) \tag{2.15}$$

subject to constraint

$$|\text{corr}(y, \delta)| \leq \rho \tag{2.16}$$

where $\rho \geq 0$ is the correlation bound.

The work finds that the least-squares model is the optimal model, allowing correlation constraints to be met and a best fit to be achieved, thus finding a model to minimise bias with threshold ρ while maximising accuracy (quantified by MAE). The work shows that for all models, increasing ρ (a looser constraint) caused the test set MAE to decrease - thus emphasising the use of ρ as a balancing parameter between the level of bias and predictive accuracy. Through using the Predictive Analytics Competition (PAC) 2019 data, the work shows that optimal unbiased predictive models are obtained. They state that their approach holds many advantages over previously established approaches. These advantages include correcting residuals on the training set and validating the corrections on a separate test set. This contrasts with the approach whereby the correction for the residuals is calculated on the training set and then applied to the test set which introduces dependencies within the the test set [133]. Furthermore, since this approach corrects the predictions (indirectly correcting residuals), the method is applicable to unlabeled datasets.

We keep these alternative approaches in mind when later applying a suitable correction to our experimental results in Chapter 4.

The deep neural networks discussed above are detailed in the context of predicting the brain age of healthy controls. We now dive deeper into the effects of HIV and applications of ML models to the case of HIV+ individuals.

2.6 Machine Learning Applications to HIV

In this section, we dive into the literature of previous ML applications to investigating the impact of HIV on individuals. In doing so, we improve our understanding of the task and complexities involved.

Guaraldi et al. [134] present evidence for the increased susceptibility of non-infectious comorbidities (NICMs) in HIV+ individuals with a history of ART. The NICMs investigated include diabetes mellitus type 2, renal failure, hypertension, bone fractures and cardiovascular disease. The study involved HIV+ individuals (n=2854) and age, race and sex matched HIV− controls (n=8562). By implementing logistic regression, the work shows that certain age-associated NICMs, and presence of multiple (≥ 2)NICMs, are more prevalent in HIV+ individuals than matched controls. Specifically, higher prevalence of diabetes mellitus, bone fractures and renal failure were found in the HIV+ group. As shown by Franke et al. [82], individuals with diabetes mellitus type 2 exhibit an increase in brain predicted age. This suggests increased predicted brain age is likely in HIV+ individuals with diabetes mellitus type 2. Guaraldi et al. [134] also identify correlations between multiple NICMs, lower nadir CD4 counts, and prolonged ART usage. El-Sadr et al. [135] provides insight to this correlation between multiple NICMs and prolonged ART use by studying the occurrence of opportunistic disease and death in HIV+ individuals. Their study comprises of a virologically suppressed group and an episodically (based on CD4 counts) treated group. Their findings indicate a higher risk of opportunistic disease or death in the episodically treated group, seemingly in contradiction to the associations found by Guaraldi et al. [134]. This discrepancy suggests that the occurrence of NICMs could potentially be even higher without a history of ART.

Focusing on neural alterations associated with ageing in HIV+ individuals, Jahanshad et al. [136] study disruptions in brain networks. They utilise high angular resolution diffusion tensor imaging (HARDTI) tractography, tracing WM connections between cortical regions based on MRI-derived cortical parcellations. Furthermore, they investigate the relationship between neurodegeneration in HIV+ individuals and carriers of the apolipoprotein E4 allele (ApoE4). This allele has been found to be a predictor of increased risk of late-onset Alzheimer's disease and, in individuals carrying two copies of ApoE4, dementia [137]. Jahanshad et al. [136] hypothesise that the neural networks of HIV+ individuals will experience impairment, with this impairment further increased in ApoE4 carriers. In both cases, this impairment may provide evidence for increased brain ageing. Jahanshad et al. [136] further hypothesise that the extent of this impairment may be correlated to the duration of infection with HIV. Their dataset includes HIV+ individuals (n=55) from the longitudinal study at the Memory and Aging Center, University of California in San Francisco (UCSF), and age and sex matched HIV− controls (n=30). At the time of assessment, all but three HIV+ individuals were on cART and only six did not have undetectable plasma HIV RNA (less than 400 copies/mL). Jahanshad et al. [136] implement a multilinear model and find that older HIV+ individuals exhibit disrupted brain connectivity and integrity, with this effect further heightened in those who carry the ApoE4 genotype. They also find evidence of the interaction between duration of HIV and presence of the ApoE4 genotype, with the longer HIV affected carriers exhibiting further deterioration of connectivity.

Cole et al. [20] investigate the link between HIV and increased levels of brain atrophy. Furthermore, the work explores whether increased predicted age is linked to HIV status, age, cognitive performance and HIV-related clinical parameters. As seen in Nir et al. [123], these correlations are informative to investigate. Cole et al. [20] uses neuroimaging data from virologically suppressed HIV+ individuals (n=162) and comparable HIV− controls (n=105). These individuals are part of the Comorbidity in Relation to AIDS (COBRA) collaboration. The work uses a GPR model trained on concatenated GM and WM tissue segmentation maps from independent healthy controls (n=2001). The coefficients from the validated GPR model are applied to the testing datasets, allowing for the brain age deltas for both HIV+ and HIV− groups to be derived. Cole et al. [20] reports a brain age delta of 2.15 ± 7.79 years for HIV+ individuals and -0.87 ± 8.40 years for HIV− individuals. The standard deviation observed in the HIV− control group can be attributed to the selection of HIV− controls to align closely with the characteristics of the HIV+ dataset, ensuring comparability between the two groups. In conclusion, the work finds evidence that HIV may accentuate brain ageing, even in virologically suppressed individuals, due to the positive brain age delta obtained. The findings suggest accentuated ageing rather than accelerated ageing as the calculated delta was not correlated to duration of infection.

Following on from the work of Cole et al. [20], Kuhn et al. [138] implement a SVR model to predict brain age on a HIV+ dataset. The study trains the model to predict age on DTI metrics from the CamCAN (n=765) and University of Oslo (n=177) datasets. Once trained, the model is applied to the HIV+ group (n=70) and HIV− comparable controls (n=34). The findings indicate that HIV is linked to WM ageing, revealing a significantly larger brain age delta in the HIV+ group in comparison to the HIV− controls. Furthermore, the study identifies a significant positive correlation between the viral load of HIV RNA and the brain age delta, and a negative correlation between domain-level cognitive function and brain age delta.

Continuing the exploration of HIV's impact on cognitive function and brain ageing, Underwood et al. [139] investigate longitudinal changes in cognitive function of HIV+ individuals. They utilise T_1-weighted MRI scans of virologically suppressed HIV+ participants (n=129) from the CNS HIV Anti-Retroviral Therapy Effects Research (CHARTER) cohort. They employ structural imaging techniques and segment the scans into WM and GM maps, followed by spatial normalisation, to prepare them for analysis. For training, they implement a multivariate model to predict age and assess cognitive function. The results display a noticeable difference in predicted age between the HIV+ individuals and the HIV− controls. On average, the predicted age for HIV+ individuals was found to be 1.17 years higher than that of the HIV− controls. This age discrepancy was even more pronounced in specific subsets of the HIV+ group: a 5.87 years increase was observed in individuals with HIV and additional comorbidities, and a 3.03 years increase was noted in those with a history of AIDS. Accurate prediction of cognitive impairment through the ML models was not achieved.

More recently, Petersen et al. [140] investigate accelerated WM ageing in individuals with HIV, with a focus on those with detectable viral load (VL). They train a GPR model to predict age on DTIs of healthy controls from the CamCAN dataset. They apply the trained model to HIV+ individuals (n=290, 85% with undetectable VL) and HIV− controls (n=165). Following this, they calculate the brain age delta and model it as an age-VL interaction. They also use measures of fractional anisotropy (FA) and mean diffusivity (MD) to quantify WM microstructure. FA is useful at deducing axonal degeneration and altered myelination whereas MD quantifies membrane density, necrosis and oedema [140]. They carry out multivariate regression analysis to assess correlations and find age and detectable VL to have a significant interaction effect. In more detail, HIV+ individuals with detectable VL were found to gain delta at a rate of 1.5 years δ/decade compared to HIV− controls (p=0.018), and HIV+ individuals with undetectable VL were found to gain delta at a rate of 0.86 years δ/decade relative to HIV− controls (p=0.052). Furthermore, the study finds a correlation between brain age delta and lower levels of overall cognition for individuals with detectable VL. Additionally, Petersen et al. [140] utilise DeepBrainNet [93] with T_1-weighted images of the same individuals for multi-modal comparison and find the predicted deltas to be significantly correlated, thus emphasising the reliability of the results. The results from this section are summarised in Table 2.4.

From Table 2.4, we can see how ML techniques have been applied to investigating the physical impact of HIV. The effects of HIV include NICMs [134], with evidence indicating that HIV causes brain damage [136, 140] and increases brain ageing [20, 138]. Together with Section 2.4, we motivate the importance of investigating brain ageing in HIV+ individuals.

2.7 Summary

This literature review provides an overview of DL, and key neural network variants. The review focuses on medical applications of DL, specifically the task of investigating brain ageing in HIV+ individuals. Thus, throughout this chapter we study medical imaging, including available training and testing datasets and pre-processing steps. We also discuss brain ageing and study previous CNN applications to predicting brain age. Following this, we explore the impact of HIV on the brain and previous ML applications to investigating the effect of HIV on individuals.

The information detailed in the literature review paves the way for further research to be performed. As seen in Section 2.4 and Section 2.6, there is substantial motivation supporting the research questions of this work. Through the analysis of the architectures detailed in Section 2.5, the foundations of a starting point model to conduct the experiment and address the research questions can be constructed. Finally, from Sections 2.5 and 2.3.2, we note suitable datasets and how to pre-process them accordingly for the task at hand.

Paper	**Methodology**	**Results**
Budka et al. [22] (1991)	Review of neuropathological and other data.	HIV linked to increased prevalence of HIVE and HIVL which may cause brain WM damage.
Thompson et al. [122] (2005)	Study of cortical GM thickness in AIDS patients and healthy controls.	Found prevalent cortical GM tissue loss in individuals with AIDS.
Heaps et al. [121] (2014)	Total and subcortical GM, total WM, caudate, corpus callosum, and thalamus analyses, along with neuropsychological tests.	Significant decrease in brain volume of total and subcortical GM found for HIV+ HAND+ relative to HIV− controls.
Keltner et al. [120] (2014)	Cross-sectional analysis on structural MRIs of HIV+ individuals with and without DNP.	Severity of DNP correlated with cortical GM atrophy in HIV+ individuals.
Clifford et al. [124] (2018)	Analysis of brain atrophy rates in longitudinal structural brain MRIs of older HIV+ individuals and age-matched healthy controls.	HIV+ group displayed higher levels of brain atrophy relative to age-matched healthy controls.
Nir et al. [123] (2019)	Tensor-based morphometry between selected time points.	Significant brain tissue loss found; predominantly in subcortical regions. Correlation between disease severity, age, cognitive decline, and greater atrophy.
Grill et al. [23] (2021)	Review of clinical syndromes and neurological conditions related to HIV infection.	Susceptibility to range of neurological illnesses such as HAND (e.g. HIV-associated dementia), opportunistic infections (e.g. toxoplasmosis) and other CNS conditions.

Table 2.1: Summary of previous key HIV research.

Paper	Dataset	Data Type	Age Range (y)	Size
Peng et al. [15]	UK Biobank	T_1-weighted	44-80	14503
Yin et al. [16]	UK Biobank, HCP, ADNI, CamCAN	T_1-weighted	18-92	7211
Dinsdale et al. [17]	UK Biobank	T_1-weighted	44.6-80.6	19687
Kolbeinsson et al. [11]	UK Biobank	T_1-weighted		21382
Levakov et al. [48]	Multiple sources	T_1-weighted	4-94	10176
Hofmann et al. [127]	LIFE-Adult	T_1-weighted, fluid-attenuated inversion recovery, susceptibility weighted imaging	18-82	2637
Jonsson et al. [49]	Icelanders, IXI, UK Biobank	T_1-weighted	18-82	>12378
Bashyam et al. [93]	LifespanCN	T_1-weighted	3-95	14468
Feng et al. [129]	Multiple sources	T_1-weighted	18-97	>30000
Cole et al. [9]	Brain-Age Healthy Control	GM, WM, T_1-weighted	18-90	2001

Table 2.2: Summary of the studies and corresponding datasets, input modalities, dataset age ranges, and dataset sizes. We order in ascending order of performance achieved.

Paper	Structure	Regularisation Techniques	Loss Function	Optimiser	Learning Rate	Batch Size	MAE (y)
Peng et al. [15]	(3x3x3 Conv, BN, 2x2x2 MP, ReLU)x5, 1x1x1 Conv, BN, ReLU, AP, DP, 1x1x1 Conv, SM	DA, DP (p=0.5), L2 Reg	KLD	SGD	10^{-2} (WD)	8	2.14
Yin et al. [16]	Conv, ReLU, MP, BN, Conv, ReLU, MP, BN, DP, Conv, ReLU, MP, BN, AP, FC, ReLU, DP, FC, ReLU	DP (p=0.3)	MSE	Adam	10^{-3}		2.23-2.41
Dinsdale et al. [17]	((3x3x3 Conv, ReLU, BN)x2, 2x2x2 MP)x3, ((3x3x3 Conv, ReLU, BN)x3, 2x2x2 MP)x2, BN, FCx2	ES	MSE	RMSprop	10^{-2} (WD)	16	2.71-3.09
Kolbeinsson et al. [11]	(3x3x3 Conv, BN, ReLU, 3x3x3 MP)x N, (3x3x3 Conv, BN, ReLU, SC, 3x3x3 Conv, BN, ReLU)x4, AP, FC	L2 Reg	MSE	Adam	10^{-4} (WD)	8	2.87-3.42
Levakov et al. [48]	(BN, (3x3x3 Conv, ReLU)x2, 2x2x2 MP)x2, BN, DP, (FC, ReLU)x2, DP, FC	DA, DP (p=0.3)	MSE				3.07
Hofmann et al. [127]	(3x3x3 Conv, LReLU, 3x3x3 MP)x4, 1x1x1 Conv, LReLU, 3x3x3 MP, DP, FC	DP (p=0.5), L2 Reg	MSE	Adam	5×10^{-4}		3.37-3.86
Jonsson et al. [49]	(3x3x3 Conv, BRN, ELU, 3x3x3 Conv, BRN, SC, ELU , 3x3x3 MP)x5, FC, ELU, DP, Concat, FC	DA, DP (p=0.2), ES, L2 Reg	MAE	Adam	10^{-3} (WD)	4	3.39
Bashyam et al. [93]	DeepBrainNet	DA, DP (p=0.8)	MSE	Adam	10^{-4} (WD)		3.702
Feng et al. [129]	(3x3x3 Conv, ReLU, 3x3x3 Conv, BN, ReLU, MP)x5, FC	L2 Reg	MAE	Adam	2×10^{-5}	5	4.06
Cole et al. [9]	(3x3x3 Conv, ReLU, 3x3x3 Conv, BN, ReLU, 2x2x2 MP)x5, FC	DA, L2 Reg	MAE	SGD	10^{-2} (WD)	28	4.16-4.66

Table 2.3: Summary of model architectures and related details from CNN brain age prediction literature.

Paper	**Dataset Size**	**Methodology**	**Results**
Guaraldi et al. [134] (2011)	HIV+ n=2854, HIV− n=8562	Logistic regression model	Certain age-associated NICMs, and presence of multiple (≥ 2)NICMs, more prevalent in HIV+ than matched HIV− controls.
Jahanshad et al. [136] (2012)	HIV+ n=55, HIV− n=30	Multilinear model	Older HIV+ exhibit disrupted brain connectivity, further deterioration in ApoE4 carriers.
Cole et al. [20] (2017)	Independent healthy n=2001, HIV+ n=161, HIV− n=102	Gaussian process regression model	Brain age delta calculated to be 2.15 ± 7.79 years in HIV+ individuals and -0.87 ± 8.40 years in HIV− individuals.
Kuhn et al. [138] (2018)	Independent healthy n=942, HIV+ n=70, HIV− n=34	Support vector regression model	Larger delta for HIV+, positive correlation between HIV RNA viral load and delta, and negative correlation between domain-level cognitive function and delta.
Underwood et al. [139] (2018)	HIV+ n=139	Multivariate model	Predicted age on average 1.17 years higher in HIV+; 5.87 years higher in those with confounding comorbidities; 3.03 years higher in those with a history of AIDS.
Petersen et al. [140] (2022)	HIV+ n=290, HIV− n=165	Gaussian process regression model	Significant interaction between age and detectable VL on brain ageing. Detectable VL gain delta at 1.5 years/decade compared to HIV− (p=0.018). Undetectable VL gain delta at 0.86 years/decade compared to HIV− (p=0.052). Correlation between delta and lower global cognition in detectable VL.

Table 2.4: Summary of previous ML applications investigating effects of HIV in individuals.

Chapter 3

Experimental Design

In this chapter, we outline the pipeline and methods used to conduct the investigation of the research questions outlined in Section 1.3. The pipeline begins with dataset acquisition and dataset pre-processing. In the model design phase, we build on previous brain age prediction models and construct a novel architecture suited to this task. The model then undergoes training, tuning and optimisations to improve the accuracy of results and computational performance. For generalisability to broader datasets and age ranges, transfer learning is then applied. Lastly we carry out testing, harmonisation and analysis with the ENIGMA-HIV dataset. This pipeline is illustrated in Figure 3.1.

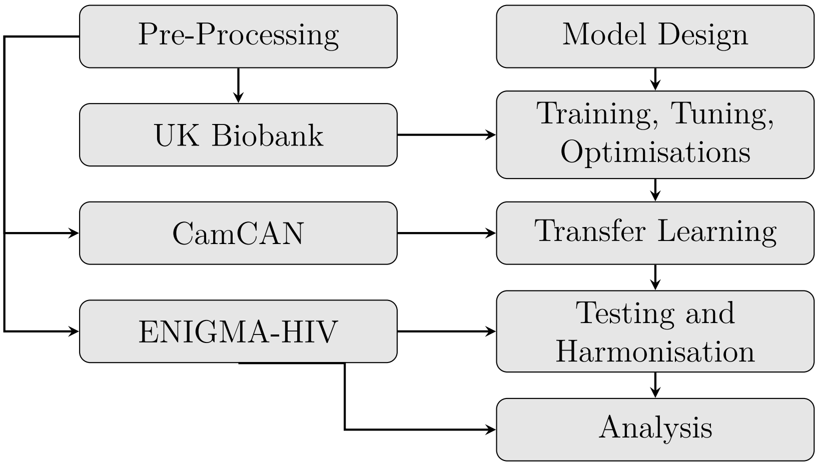


Figure 3.1: Experimental pipeline implemented for investigating brain ageing in HIV+ individuals.

3.1 Datasets

Selecting a suitable dataset is a crucial component in carrying out a successful experiment. In this section, we discuss the selection of training and testing datasets, and the key factors we look for in these datasets.

3.1.1 Training Datasets

We first require a dataset to train a neural network to predict brain age. We select T_1-weighted brain MRI scans for the input, as is commonly seen in the literature (see

Table 2.2). A suitable dataset is one which consists of scans of healthy individuals for which chronological age is matched to apparent brain age. Recall from Chapter 1 that the apparent age of the brain is defined to be the age value as measured by a health professional, or a predictive model. Through training a model on these T_1-weighted brain MRI scans, the model learns a mapping between input brain scan and predicted age value. Furthermore, to develop a robust model, it is important to ensure a suitably large training dataset is used, as well as one which has a wide distribution of ages [17]. Table 3.1 details training datasets available (and used in Section 2.5) and features details on the age range and size of the datasets.

Dataset	Number of Scans (N)	Age Range (Years)	Relevant Studies
UK Biobank	21369	45-82	[11, 15–17, 49, 130]
CamCAN	652	18-88	[16, 129, 141]
IXI	544	20-86	[48, 49, 129]
LifespanCN	14468	3-95	[93]
Brain-Age Healthy Control	2001	18-90	[9]

Table 3.1: Relevant brain MRI datasets with details on number of scans, age ranges and brain age prediction studies that they have previously been applied in.

As seen from Table 3.1, the UK Biobank [142] is a large dataset of healthy controls. We note that while the LifespanCN dataset is also large and spans a wide range, the access is restricted [93]. We found the UK Biobank dataset to be more accessible, through the University of Southern California (USC). Furthermore, the UK Biobank is widely used and has been shown to lead to successful results [15, 17, 130] - in particular, Peng et al. [15] is SOTA for brain age prediction at the time this work was carried out. A limitation of the UK Biobank is that it spans a narrow age range. When developing a robust model it is necessary for training and testing datasets to have comparable age ranges and distributions. We do not want to restrict the applicability of the model to individuals within the range 45-82 years, especially since HIV affects individuals of all ages. Hence, we are motivated to obtain a second dataset, with a wide age range, for further training of the model through transfer learning. Training the model on a larger sample of datasets will not only allow for a greater range of age inputs to be used, but it will also allow for the model to become more generalisable [48, 129], as different datasets often originate from different scanners and scanning protocols. An accessible dataset which satisfies the age range requirement is the CamCAN dataset [78, 79].

Additional Imaging Study Details for Selected Datasets

UK Biobank imaging data. The UK Biobank [142] is a large biomedical database that contains genetic, lifestyle and health information of approximately half a million participants from the UK population, recruited between 2006 and 2010. At the time of first recruitment, the participants spanned the age range of 40-69 years. A follow up study, commencing in 2016, involved the acquisition of medical images. The participants, at the time of imaging, spanned the age range of 45-82 years. This study was first detailed in Miller et al [131]. These multimodal medical images were acquired between three identical scanning centres using Siemens Skyra 3T scanners. The T_1-weighted structural data was acquired at a 1mm isotropic resolution using a MP-RAGE sequence.

CamCAN imaging data. The Cambridge Centre for Ageing and Neuroscience (CamCAN) [78, 79] is a collaborative research project, started in October 2010. The CamCAN dataset (available at http://www.mrc-cbu.cam.ac.uk/datasets/camcan/) contains brain MRI data from the UK population and investigates healthy cognitive ageing. CamCAN participants undergo screening to ensure healthy cognitive function. The participants, at the time of imaging, spanned the age range of 18-88 years. The imaging was carried out on a 3T Siemens TIM Trio System. As for the UK Biobank dataset, the T_1-weighted structural images were acquired at 1mm isotropic resolution using MP-RAGE sequence [79].

The age distributions for the UK Biobank and CamCAN datasets are illustrated in Figure 3.2. Similarly, we see the sex distributions of the UK Biobank and CamCAN dataset in Figure A.1. In Figure 3.2 we see the higher frequency of UK Biobank scans, particularly within the middle of the age range, and the broader age range of the CamCAN scans. It is important to note that we have modified the original CamCAN dataset to ensure that its age distribution aligns more closely with that of the testing dataset, which will be discussed in the following section. This modification involved defining age bins for both the CamCAN and testing dataset, and then removing scans from the CamCAN dataset in bins where its frequency was higher than that of the testing dataset. When removing scans we remove a proportional amount of male and female scans. Following the removal of scans, N=484 scans remained in the CamCAN dataset.

3.1.2 Testing Dataset

To fulfil the research aims of this work, we require a dataset of T_1-weighted brain MRI scans of individuals with HIV, from which the effects of HIV on the brain can be investigated.

As seen in Section 2.6, there are a variety of datasets available for the study of HIV. However, these datasets are fairly small and as seen in [121, 139, 143] the acquisition of a sufficiently large dataset is beneficial to obtaining accurate results. It is also of importance to have comparable HIV$-$ controls [20, 143]. Comparable controls are often matched in terms of age, sex, and/or lifestyles and allow for more accurate comparison of predicted ages. When selecting a dataset, we aim to obtain a large, diverse dataset paired with HIV$-$ controls.

A large group involved in the study of HIV, as well as various other neuroimaging and genetic studies, is the Enhancing NeuroImaging Genetics through Meta Analysis (ENIGMA) consortium [144]. This group, launched in 2009, is a wide-scale collaboration of over 1400 scientists from around the world. ENIGMA works to answer research questions in neuroscience, psychiatry, neurology, and genetics by combining datasets from sources around the world. One of these working groups within the ENIGMA consortium is the ENIGMA-HIV working group, formed in May 2013, which holds a diverse dataset of HIV+ individuals. This diverse dataset is extremely valuable and features individuals from South Africa, the USA, Asia, and Europe [145]. This dataset is thus a better representative of HIV impacted individuals than a dataset from a single site would be. The works of Jahanshad et al. [136], Underwood et al. [139] and Nir et al. [123], detailed in Section 2.6, use datasets from sites contributing to the ENIGMA-HIV dataset.

Through ENIGMA-HIV, this investigation on brain ageing in HIV+ individuals, was

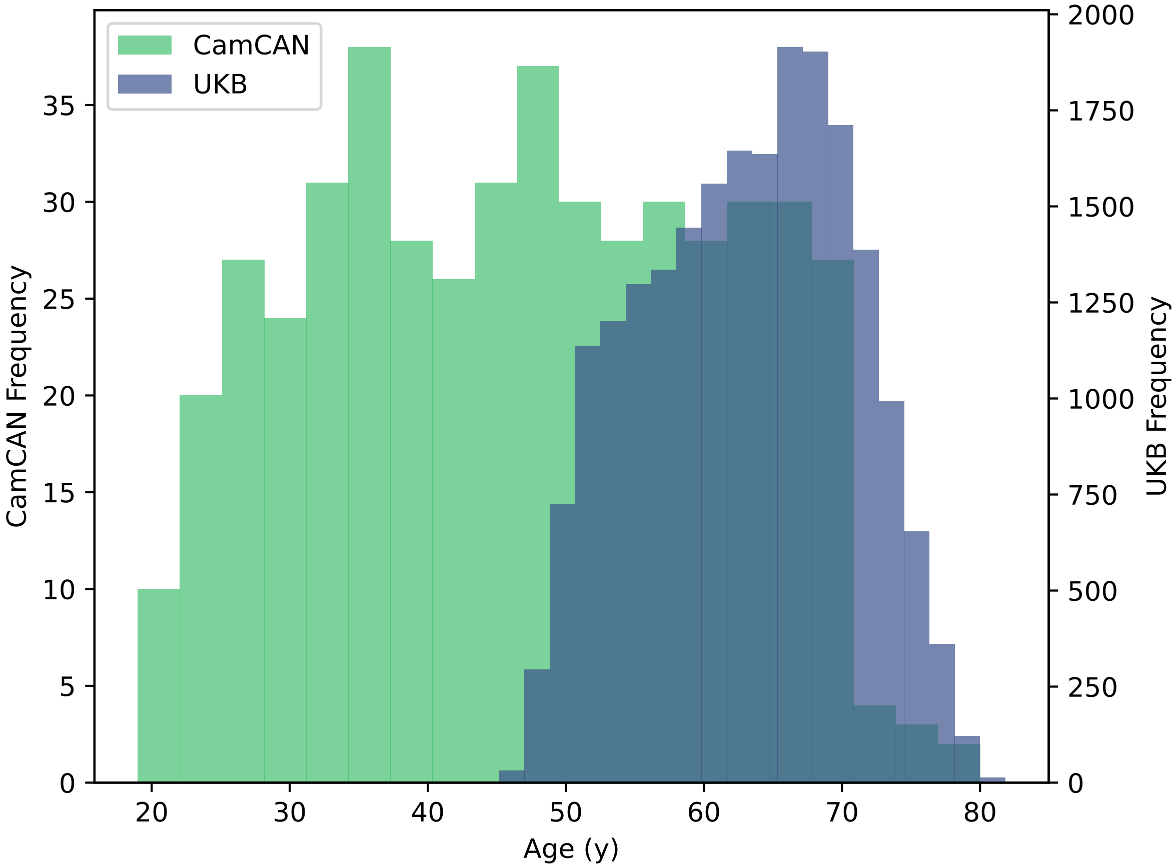


Figure 3.2: Age frequency histograms of the UK Biobank (UKB) and CamCAN training datasets.

given access to the data from twelve sites within the group. Table 3.2 contains details on the number of scans and age ranges each site contains, as well as the case (HIV+) to control and male to female proportions. Figure 3.3 represents these details graphically, and additionally displays the age distributions within the sites. Six of the twelve sites contain controls. In Table 3.3 we detail the key clinical measures provided by the sites within ENIGMA-HIV. Due to sparsity of information, we primarily focus on the clinical measures CD4, nadir CD4, duration of infection, treatment status and HIV plasma RNA (plRNA) detectability status (see Section 2.4 for definitions). Note we exclude one of the twelve sites due to the scan acquisition protocol which produced scans of very large intensities and biased results. Following this exclusion, six sites with controls and five sites with purely HIV+ cases remain. Figure C.1, C.2, C.3 illustrate the mean CD4, nadir CD4 and HIV durations per site, while Figure C.4 and Figure C.5 illustrate the percentage of HIV+ cases on ART, and with detectable levels of HIV plasma RNA, respectively.

To ensure that the age distributions of the training and testing datasets are representative, we excluded N=168 older scans from the CamCAN dataset that fell outside the ENIGMA-HIV age range. Leaving these scans in may introduce bias whereby the model may tend to over-predict the ages of the individuals in the ENIGMA-HIV dataset and thus produce misleading results. We see a comparison between the age and sex distributions of the CamCAN dataset and the ENIGMA-HIV dataset in Figure 3.4 and Figure A.2, respectively. To assess the comparability of age distributions and sex frequencies

Site	Scans (N)	Age Range (Yrs)	HIV+ % (N)	Male % (N)
Arch	134	21-69	**59.0 (79)**	61.2 (82)
Boban	157	17-76	**51.6 (81)**	81.5 (128)
Brew	32	39-75	100.0 (32)	96.9 (31)
Brown	131	23-79	**61.8 (81)**	64.9 (85)
Chang	191	20-74	100.0 (191)	91.1 (174)
Cysique	122	44-69	**67.2 (82)**	100.0 (122)
Hinkin	65	26-67	**20.0 (13)**	75.4 (49)
Navia	49	25-62	100.0 (49)	65.3 (32)
Paul	147	22-45	100.0 (147)	12.2 (18)
Search	62	22-56	100.0 (62)	41.9 (26)
Valcour	83	60-69	**63.9 (53)**	92.8 (77)

Table 3.2: Details of number of scans, age ranges, proportions of cases to controls and males to females for sites forming part of the ENIGMA-HIV dataset. We highlight the cohorts with controls in bold.

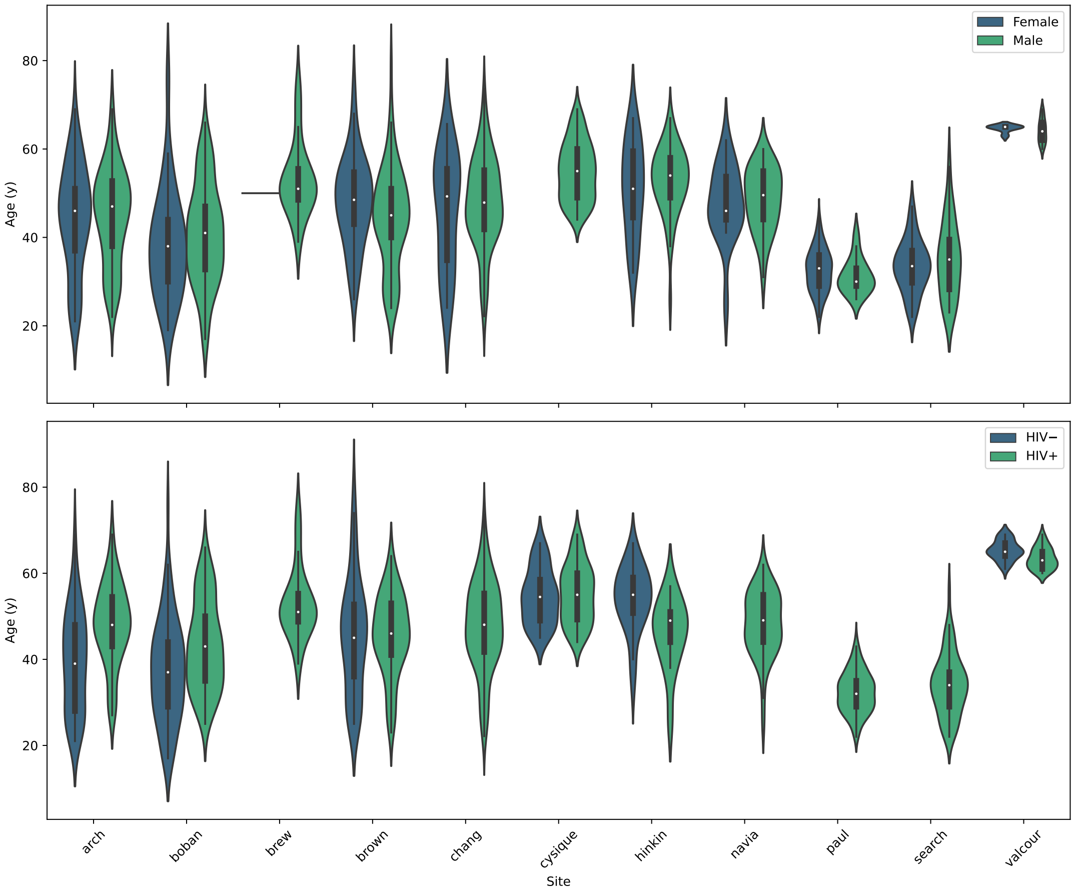


Figure 3.3: Age distributions of scans for the sites in the ENIGMA-HIV dataset with sex (top) and HIV status (bottom) splits.

between the CamCAN training dataset and the ENIGMA-HIV testing dataset, we conducted Kolmogorov-Smirnov (KS) tests for the continuous age data and Chi-squared tests for the categorical sex data. For these tests, the null hypothesis is that the datasets are drawn from the same distribution. The Chi-squared test results indicated that the null

Site	CD4, cells/mm^3, Mean (Std)	CD4 (N)	Nadir CD4, cells/mm^3, Mean (Std)	Nadir CD4 (N)	HIV Duration, Mean (Std)	HIV Duration (N)	ART Ratio (%) [N]	plRNA Ratio (%) [N]
Arch	574.9 (273.6)	75	214.3 (161.5)	76	18.4 (8.3)	76	96.00 [75]	24.36 [78]
Boban	586.9 (337.9)	79	303.1 (193.2)	81		0	100.00 [72]	25.93 [81]
Brew	585.2 (282.5)	32		0		0	[0]	[0]
Brown	474.3 (250.6)	79	195.5 (176.5)	80	12.6 (6.9)	81	85.19 [81]	30.77 [78]
Chang	477.8 (276.3)	185	220.5 (194.2)	175		0	93.68 [190]	[0]
Cysique	552.8 (276.5)	82	176.5 (126.3)	82		0	[0]	[0]
Hinkin	604.1 (289.1)	12		0		0	[0]	[0]
Navia	361.8 (204.0)	49	62.4 (43.5)	49	11.6 (5.8)	49	67.35 [49]	[0]
Paul	224.6 (138.4)	147		0		0	[0]	[0]
Search	236.0 (139.0)	61		0		0	[0]	[0]
Valcour	529.3 (218.5)	50	209.8 (190.5)	52	20.1 (6.4)	51	[0]	[0]

Table 3.3: Details of key clinical measures for the sites forming part of the ENIGMA-HIV dataset. Note some clinical measures are unavailable for certain cohorts.

hypothesis could not be rejected, while the KS test rejected the null hypothesis. From Figure 3.4 it is evident that there are age bins where the ENIGMA-HIV dataset has a higher frequency of scans, and vice versa. Hence, without excluding further training and testing data, we will not be able to have matched age distributions. However, the critical step of removing older scans from the CamCAN dataset, which could introduce age prediction bias, was performed. Since the UK Biobank dataset is purely for pre-training the model, the distribution of ages relative to the ENIGMA-HIV dataset is not of high importance.

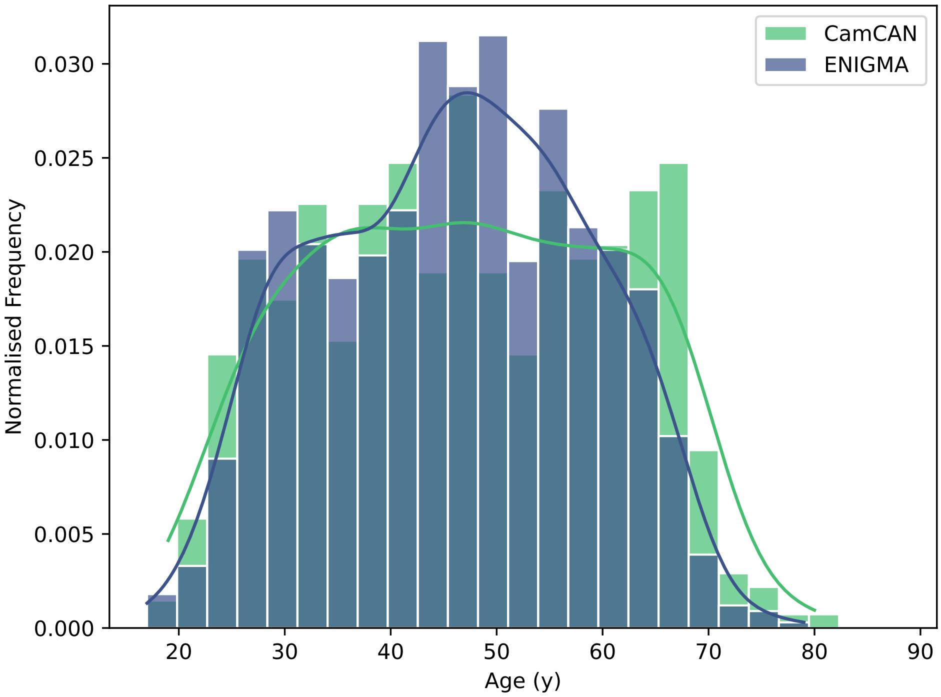


Figure 3.4: Comparison plot of the normalised age distributions of the CamCAN transfer learning dataset and the ENIGMA-HIV testing dataset.

It is important to note that it is not viable to solely train and test on the ENIGMA-HIV dataset. It is imperative that the model learns a mapping between a brain scan and an age value, where the mapping is defined on healthy controls. As detailed in Section 2.2, the apparent brain age of healthy controls is matched to the individual's chronological age. Healthy controls are defined here as HIV− individuals who have been screened for health factors; for example, the controls in the UCSF Valcour cohort (see below). However, within the ENIGMA-HIV dataset, the majority of controls are not classified as healthy, or their health status was unavailable, hence they are not suitable for teaching the model the appropriate mapping. Furthermore, as the ENIGMA-HIV dataset originates from multiple sites, of which different sites contribute different numbers and distributions of scans, a large dataset imbalance would be present.

Details on controls

From the distributions in Figure 3.3 we see, visually, that the controls appear to be generally well matched by age and sex, with the exception of Hinkin, where there were fewer HIV+ cases than controls. To numerically validate the similarity of age distributions and sex frequencies across these groups, we applied KS tests for the age data and

Chi-squared tests for the sex data, as before. The KS test results revealed significant differences in age distributions at the Arch, Hinkin, Valcour, and Boban sites between HIV+ and HIV− groups. Conversely, the Brown and Cysique sites, did not have the null hypothesis (defined as before) rejected. Regarding sex frequencies, the Chi-squared tests indicated differences in the frequencies within the Arch and Boban sites, but not for the Brown, Hinkin, and Valcour sites. The Cysique site contains only males. We were able to obtain additional information about the controls from four sites [1].

Arch and Brown. The health of the cases and controls were largely similar. The majority of controls were recruited from the same population as the cases, many with similar comorbidities. There were no major medical issues, but multiple risk factors were present.

Cysique. The controls were selected to be age matched, with comparable demographic and lifestyle criteria.

UCSF Valcour. The controls form part of a healthy ageing cohort at UCSF and have been extensively screened for factors linked to brain health, hence they are healthy controls [123].

Although our training datasets from the UK Biobank and CamCAN were screened for health, variations in the strictness of these criteria may affect the model's performance. Despite this, we believe that training a model on these health-screened datasets and applying it to the ENIGMA-HIV dataset is feasible. In particular, ensuring that the model is generalisable is important for improving the accuracy of results on the ENIGMA-HIV dataset. Furthermore, we shall utilise statistical methods (see Section 3.5.3) to assess differences between the HIV− controls and HIV+ cases, thus allowing for deeper insights into the results to be obtained.

3.2 Pre-Processing

As detailed in Section 2.3.2, pre-processing is a standard feature in medical imaging studies. The UK Biobank pre-training dataset was pre-processed by the USC. The first of these pre-processing steps applied by the USC involved the reorientation of the brain MRI scans to correspond with that of the standard template (MNI152) orientation. To do this, FSL's [2] *fslreorient2std* tool was used. For skull-stripping, the brain extraction tool HD-BET was implemented [1]. A linear registration tool part of FSL's [2] package, *FLIRT* [3–5] was used to linearly register the scans to MNI152 space with the ICBM 2009a nonlinear symmetric $1 \times 1 \times 1$mm template [6–8]. The scans were registered with nine degrees of freedom and the mutual information cost function. As discussed by Sengupta et al. [146], the mutual information cost function aims to maximise the aligned voxel information between images and has been shown to be very successful in image registration. Following registration, the scans were of dimensions $197 \times 233 \times 189$. Dinsdale et al. [17] discusses the potential loss of morphological information due to the registration process. We justify the need for this pre-processing step since we are pooling data from multiple sites [9] where the sites use varied scanners and scanning protocols and produce scans of different dimensions. The USC performed visual quality control and excluded scans with motion or Gibbs artifacts [2].

---

[1] Private correspondence

[2] Private correspondence

To avoid introducing unnecessary variation to the model, it is sensible to be consistent with pre-processing steps for the different datasets. Hence, the same pre-processing pipeline, as used for the UK Biobank dataset by the USC, was applied to the CamCAN and ENIGMA-HIV datasets. Since HD-BET was created to be robust to brain tumours, lesions and resection cavities [1] it is also a suitable choice for the ENIGMA-HIV dataset where lesions, and other neurological pathology, may be present as a result of progressed HIV [23].

We choose to zero pad (in x and y dimensions) and crop (in y dimension) the scans in the UK Biobank, CamCAN and ENIGMA-HIV datasets to have final dimension $225 \times 225 \times 225$, similar to He et al. [28]. Since the brain scans feature an outer coating of black (zero) voxels it is justifiable to further add or remove zero voxels as they do not contribute to brain structure.

Normalisation is typically the final step in the input pre-processing pipeline. This process involves standardisation of the scans so that the scan voxels have intensity values of mean $\mu = 0$ and standard deviation $\sigma = 1$. The PyTorch *torchvision transform* library was used to perform normalisation, with the mean and standard deviation being calculated for each scan. The pre-processing pipeline utilised in this work is illustrated in Figure 3.5.

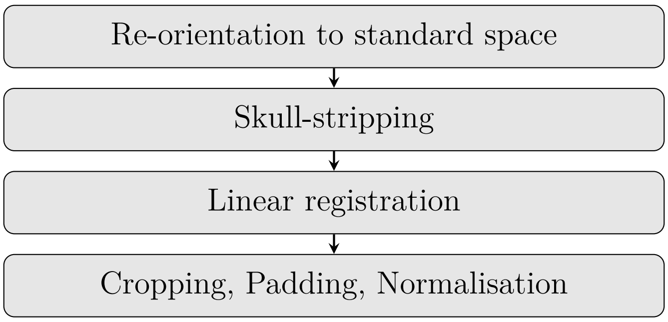


Figure 3.5: Pre-processing pipeline for T_1-weighted brain MRI scans used as input to the model.

Once the initial pre-processing steps are complete, the MRI scans are in a suitable format to input to the model for training, validation and testing. We transpose the tensor of scan intensity values such that the input tensor to the model represents axial slices [17, 93]. An example of an axial slice can be seen in Figure 2.10.

3.3 Model Design

To construct a novel brain age prediction model, we utilise aspects of the literature detailed in Section 2.5 and other successful architectures in the field.

We draw on the architecture of the ResNet [28], a highly successful computer vision architecture, and implement a 3D CNN with residual skip connections. In Table 2.3, we see Kolbeinsson et al. [11] implement a ResNet and achieve the fourth ranking MAE in the literature. Our model design begins with a 3D batch normalisation layer and ReLU activation function. This is followed by a $7 \times 7 \times 7$ convolution operation with stride 2 and a $3 \times 3 \times 3$ max pooling layer to down-sample the spatial dimensions. This format of $7 \times 7 \times 7$ convolution and max pooling is the standard opening to the model proposed in He et al. [28]. We differ from the usual sequence by implementing batch normalisation and ReLU prior to the convolution which has been shown to improve the accuracy of results by Zagoruyko and Komodakis [69].

This initial block is followed by a series of four residual blocks, each composed of two sets of batch normalisation, ReLU activation and $3 \times 3 \times 3$ convolution operations. After every residual block, a $2 \times 2 \times 2$ max pooling layer is applied to progressively reduce the spatial dimensions, similarly to Dinsdale et al. [17]. The residual connections implement a $1 \times 1 \times 1$ convolution (projection) on the convolved input tensor when feature map dimensions have increased (or decreased) between layers. These residual connections help mitigate the degradation of model learning and have been successful in brain age prediction tasks [11, 49].

The final residual block replaces the usual max pooling layer with a $7 \times 7 \times 7$ average pooling layer. This average pooling is frequently used to summarise the abstracted spatial dimensions of the tensor to a single value per feature map [11, 15, 28]. Next the model comprises of a flattening operation to prepare the tensor for the fully connected layer and a dropout layer to prevent overfitting. The model culminates in a fully connected layer that outputs the predicted brain age. Throughout the model we zero pad to ensure edge information is not discounted. The selected model architecture is illustrated in Figure 3.6.

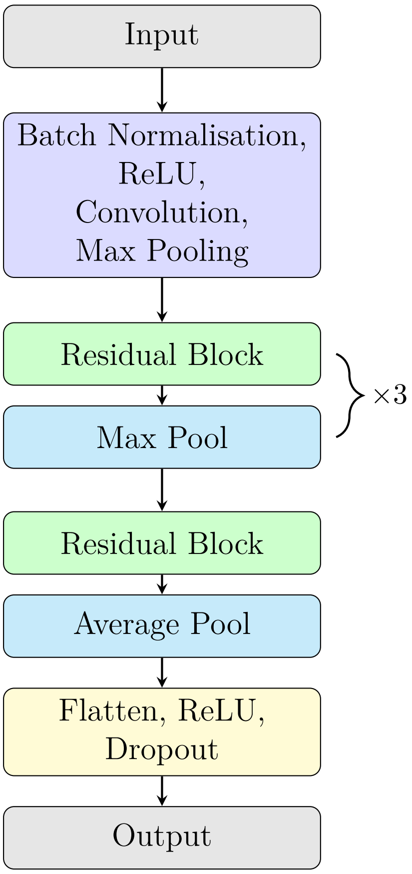


Figure 3.6: Model designed for brain age prediction task. The residual block contains two sequences of batch normalisation, ReLU activation and a $3 \times 3 \times 3$ convolution, all enclosed by a skip connection.

In total, the model consists of nine convolutional layers. As seen in Peng et al. [15], the current SOTA for CNN brain age prediction, a model with as few as seven convolutional layers has been shown to achieve high performance. The selection of nine convolutional layers is motivated by the dimension of the input scans, which is $225 \times 225 \times 225$ - similar to what is used for the ResNet by He et al. [28]. The sequence of convolutional and max pooling layers reduce the spatial dimensions of the scans to $7 \times 7 \times 7$ which is appropriate for average pooling [11, 28]. As seen in Section 2.1.3, Zagoruyko and Komodakis [69] show that residual connections are beneficial for performance in models with fewer layers, such as ours. Throughout the model, we use filters of sizes $[32, 32, 64, 128, 256]$. He et al. [28] use a maximum filter size of 512 for the ResNet-18, and so, bearing in mind computational constraints, we justify selecting a maximum filter size of 256 for our smaller model.

3.4 Pre-Training, Transfer Learning and Tuning

In this section, we provide a detailed overview of the pre-training, transfer learning, tuning and optimisation. We use PyTorch [147] for the computational tasks of this work.

Since the UK Biobank dataset is far larger, and has a substantially different age distribution than the CamCAN dataset, we decide against pooling the two datasets into one. Furthermore, the ENIGMA-HIV dataset also consists of a considerably different age distribution than the UK Biobank dataset. It is important to train a model on datasets with similar age distributions to prevent biasing the model. It is also essential for the model to be sufficiently exposed to these younger (and older) individuals for accurate brain age prediction. We thus motivate using the UK Biobank dataset as a pre-training dataset, followed by applying transfer learning with the CamCAN dataset to develop a robust model, trained on a large range of ages. As seen in Figure 3.4, the distributions of the CamCAN and ENIGMA-HIV datasets are more closely aligned, justifying the use of the CamCAN dataset for transfer learning. To this end, we first pre-train on the UK Biobank dataset and then perform transfer learning on the UK Biobank pre-trained model with the CamCAN dataset.

3.4.1 Pre-Training with UK Biobank Dataset

We split the UK Biobank pre-training dataset into training, validation and testing subsets using a split ratio of 70%-15%-15%. The age and sex proportions across the training, validation and testing subsets are illustrated in Figure 3.7. Similarly, the breakdown of number of scans and sex proportions is given in Table 3.4. We ensure a proportional split across ages and sex within each subset of the UK Biobank dataset. It is not uncommon for brain age studies to stratify/split data based on sex [16, 17]. The inclusion of a test subset from the UK Biobank is motivated by the need to benchmark our model's performance against previous studies utilising the UK Biobank dataset, as detailed in Section 2.5.

	Training	**Validation**	**Testing**
Total Scans (N)	14956	3205	3205
Male (%)	46.04	46.02	45.99

Table 3.4: Summary of number of scans and sex proportions within the UK Biobank training, validation and testing subsets.

For all training, we utilise an NVIDIA Tesla V100-SXM2-32GB GPU accessed through

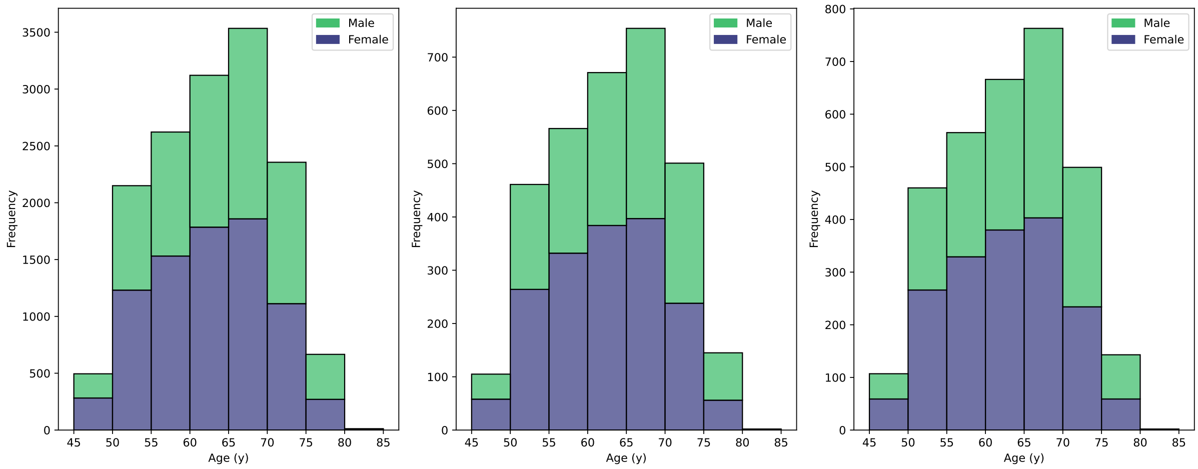


Figure 3.7: Histogram displaying the number of scans and proportions of males and females for each age bracket for training (left), validation (middle) and testing (right) subsets of the UK Biobank.

the USC. The UK Biobank dataset size, following pre-processing, is approximately 1TB in size and thus poses a prominent computational challenge. We write a custom DataLoader class, which inherits from the classic PyTorch [147] DataLoader class, that loads the required batches of input into memory as needed. This avoids the requirement for the dataset to be loaded into memory in its entirety, for which we do not have the compute. While this method of data loading increases the computational time required for the training procedure, the memory burden was greatly reduced and allowed for a large pre-training dataset to be used.

3.4.2 Transfer Learning with CamCAN Dataset

Following the same procedure as for the UK Biobank, we split the CamCAN dataset into training, validation and testing subsets using a split ratio of 70%-15%-15%, again stratifying by age and sex. The age and sex proportions across the training, validation and testing subsets are illustrated in Figure 3.8, and the breakdown of the number of scans and sex proportions is given in Table 3.5. To validate the similarity of age and sex distributions across these subsets, we again applied KS tests for the continuous age data and Chi-squared tests for the categorical sex data. The null hypothesis, that the subsets are drawn from the same distribution, was not rejected for either test, verifying the proportionality of our splits.

	Training	**Validation**	**Testing**
Total Scans (N)	338	73	73
Male (%)	50.30	47.95	47.95

Table 3.5: Summary of number of scans and sex proportions within the CamCAN training, validation and testing subsets.

To further motivate the use of first pre-training with the UK Biobank dataset, we compare the accuracy of the model on the CamCAN dataset with and without pre-training. For transfer learning, we employ the strategy of loading the pre-trained model

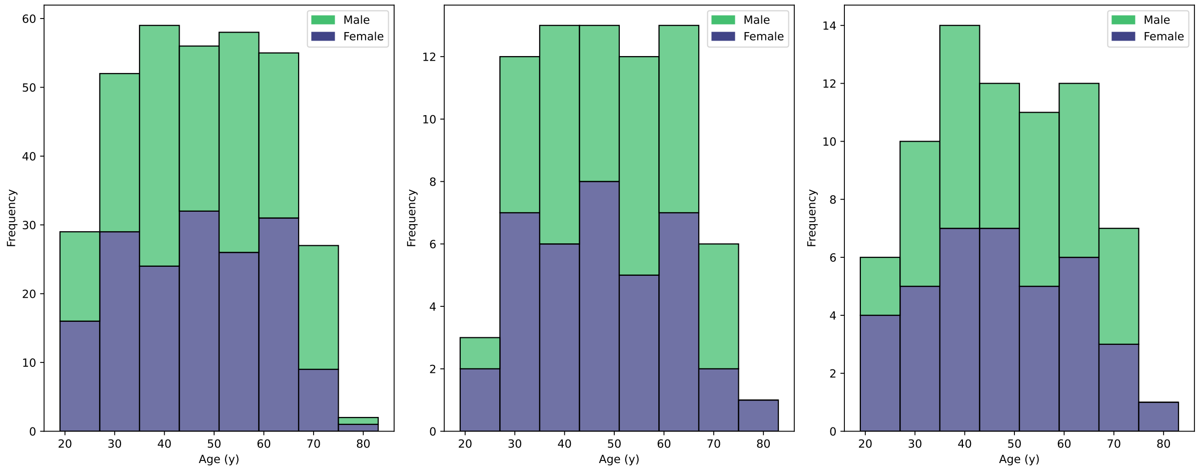


Figure 3.8: Histogram displaying the number of scans and proportions of males and females for each age bracket for training (left), validation (middle) and testing (right) subsets of CamCAN.

saved at the epoch with the lowest validation loss. Once the model weights and biases are initialised, we commence training as usual.

Bashyam et al. [93] and Jonsson et al. [49], mentioned in Section 2.5, both make use of transfer learning. Bashyam et al. [93] transfer learns to apply their model to datasets featuring disease, while Jonsson et al. [49] uses transfer learning as a means of training the model to accurately predict brain age on a new site, similar to the transfer learning use case of this work.

3.4.3 Hyperparameter Tuning

We perform hyperparameter sweeps using the functionality provided by Weights and Biases [148]. We note commonly used learning rates, loss functions, optimisers and regularisation techniques in Table 2.3. We perform three separate hyperparameter sweeps, for pre-training on the UK Biobank, transfer learning on CamCAN, and independent training on CamCAN. The sweeps are executed by a grid search with validation loss as the metric to be minimised. We sweep over different values of learning rate, learning rate decay, L2 weight decay and dropout, as well as the MAE and MSE loss functions. To ensure the model is robust to different initialisations, we assess its performance across three random seed values, using the previously found optimal hyperparameters.

3.4.4 Reproducibility

When searching for the optimal hyperparameter configuration, it is important to eliminate sources of randomness and non-determinism within the model. This allows one to draw conclusions about how the choice of hyperparameters is linked to model accuracy, without confounding factors in play. For a given library, fixing the seed sets a constant starting point for the random number generator. For example, setting the PyTorch manual seed to a fixed value ensures the convolutional weights within the model are initialised with the same values for each subsequent instantiation of the model. During training we experiment with different fixed seeds. By ensuring the weights and biases are instantiated to the same values, it is clearer which hyperparameter changes impact results.

3.4.5 Optimisations

Due to the large volume of data, the data loading process was computationally costly. As mentioned previously, due to memory constraints we create a custom DataLoader class which, while solving the memory issues, results in longer run times. Optimisation of data loading was achieved by increasing the number of workers in the PyTorch DataLoader function. For number of workers > 1, multi-process data loading is enabled. We experiment with varying the number of workers within the range $[1, 2, 4, 6, 8]$ and find four workers to be optimal. The PyTorch DataLoader function also provides the option to 'pin memory'. Further optimisation of data loading was obtained by setting this parameter to true, thus moving the called data to pinned memory and allowing for more efficient data transfer from memory to GPU [147].

3.5 Testing

Since not all HIV$-$ controls can be classified as 'healthy', we decide against further transfer learning with these controls. Their primary use shall be for comparison of results from the HIV+ cases. To obtain predicted brain ages, and thus calculate brain age deltas on the ENIGMA-HIV dataset, we load the model saved at the lowest validation loss epoch during transfer learning with the CamCAN dataset.

3.5.1 Harmonisation

A notable concern when pooling data from multiple sites is the additional variation in scans that is introduced. The ENIGMA-HIV [145] dataset utilised in this work consists of scans originating from eleven sites. These sites make use of different scanners and scanning protocols. It is important to ensure that the model does not learn to associate this variation as a factor contributing to the predicted brain age. A technique known as harmonisation (see Section 2.3.3) has been engineered to detect and remove these differences. The ENIGMA-HIV Working Group regularly incorporates this technique in their analysis pipeline. We also note that two of the sites within ENIGMA-HIV use two scanning protocols to obtain the MRIs. Hence, we specify these scans obtained through differing protocols as distinct sites. Figure A.3 illustrates the age distributions each new distinct site contains, as well as the male to female and HIV+ case to control proportions.

We use an algorithm known as neuroHarmonize [98] to implement harmonisation. This software is an extension of neuroCombat [97] and ComBat [96], with additional neuroimaging processing support. The method aims to harmonise the post-registration NIFTI images, a common file format for neuroimaging data. We use post-registration images as inputs to the harmonisation algorithm since the inputs are required to have the same dimensions. The algorithm treats each voxel as a feature and computes appropriate scale and shift parameters for each distinct site input. Through applying these parameters, harmonisation corrects for the inter-site and scanner differences and preserves the effects of specified covariates. In this work the covariates we specify for preservation are age, sex and HIV status. Age and sex are commonly specified as covariates for harmonisation [98]. We select the CamCAN test set as the reference site that the ENIGMA-HIV scans should be harmonised to. The functionality to specify a reference site was accessed through a fork of neuroHarmonize available at GitHub [149]. The specification of a reference site scales and shifts the voxel intensities of the NIFTI images within each site to align with that of the reference site, while preserving the effects of the specified covariates.

Due to time constraints, we were unable to explore the method proposed by Liu et al. [103] (see Section 2.3.3) but believe it is a worthwhile future endeavour.

3.5.2 Brain Age Bias Correction

We implement a brain age bias correction (see Section 2.5.1) at the end of each stage of pre-training, transfer learning and testing. We follow the brain age bias correction detailed in Peng et al. [15] whereby a linear regression is fit to the age predictions of the validation set. The estimated fit parameters are then used to correct the brain age delta predictions on the test set. Hence, we follow this procedure for pre-training on the UK Biobank and independent training and transfer learning with the CamCAN dataset. For the ENIGMA-HIV dataset, following harmonisation to the CamCAN test set, we use the correction parameters calculated on the transfer learning CamCAN validation set and apply it to the ENIGMA-HIV testing results. Since the controls in the ENIGMA-HIV dataset are of mixed health, it would not be accurate to calculate the correction on the controls.

3.5.3 Analysis Overview

We analyse and compare the results obtained on the testing dataset with and without the use of harmonisation. Pre-harmonisation, we perform two scopes of analysis, one using a mixed-effects model on the full dataset and another with fixed-effects models for within site analyses. Post-harmonisation, we use a fixed-effects model on the full harmonised dataset.

For the analysis, we aim to determine the overall effect of HIV. As a first scope of analysis for the pre-harmonisation dataset, we implement a mixed-effects model and specify delta (the brain age delta) as the outcome variable and HIV, sex and age, as the predictor variables. We group the input by site, which is the random effect, and allow for random intercepts. This has the corresponding linear mixed-effects formula

$$\text{Delta} \sim \text{HIV} + \text{Sex} + \text{Age} + (1 \mid \text{Site}), \tag{3.1}$$

where ‘$(1 \mid \text{Site})'$ denotes that site is the random effect. We choose to include sex as a predictor variable instead of stratifying by sex. This approach allows for performing analyses on a larger subset of data.

To harmonise, we train the harmonisation model on the whole dataset and specify the CamCAN test set as the reference site. For the first scope of analysis post-harmonisation, we implement the fixed-effects model and specify delta as the outcome variable and HIV, sex and age, as the predictor variables, i.e.,

$$\text{Delta} \sim \text{HIV} + \text{Sex} + \text{Age}. \tag{3.2}$$

Furthermore, pre- and post-harmonisation, we perform additional analysis to investigate correlations between the brain age deltas and clinical measures such as CD4 count, nadir CD4 count, HIV duration, ART status and HIV plasma RNA detectability status.

For the CD4 counts, we create four distinct bins; controls, counts > 500 cells/mm^3, $200 \leq$ counts ≤ 500 cells/mm^3 and counts < 200 cells/mm^3. These bins establish whether an individual is a HIV$-$ control, HIV+ in good health, moderate health or has AIDS [150]. Similarly, for the nadir CD4 count, we create three distinct bins; for HIV− controls, counts ≥ 200 cells/mm^3 and counts < 200 cells/mm^3. These bins categorise the individuals into HIV− controls and HIV+ without, and with, a history of AIDS, respectively. We then perform the aforementioned analyses using the bin values, which are treated as categorical variables. We choose to perform the analysis on discrete bins as opposed to continuous values since we do not have CD4 and nadir CD4 information for the HIV− controls. When defining the bins, we follow the general HIV CD4 classification system [151]. We also conduct analyses exclusively on the HIV+ cases. As seen in Section 4.5.1, the mixed-effects model, which includes age, sex, CD4 count, and nadir CD4 count as predictor variables, given by

$$\text{Delta} \sim \text{Age} + \text{Sex} + \text{CD4 Bin} + \text{Nadir CD4 Bin} + (1 \mid \text{Site}), \tag{3.3}$$

fails to converge when controls are included. This is due to the high complexity of the model and limited data available. In these analyses, the HIV− control bin, or the 'healthiest' bin in the case of analyses without controls, serves as the reference bin for the corresponding clinical measure in the analyses. Table 3.6 and Table 3.7 display the frequency, means and standard deviations of the CD4 counts and nadir CD4 counts, respectively, in the different count ranges.

CD4 Count (cells/mm^3)	Frequency (N)	Mean (cells/mm^3)	Std. Deviation (cells/mm^3)
> 500	310	742.70	219.66
$200 \leq$ count ≤ 500	379	338.82	83.80
< 200	162	120.99	53.46

Table 3.6: Frequency, mean and standard deviation of CD4 counts in different category ranges.

Nadir CD4 Count (cells/mm^3)	Frequency (N)	Mean (cells/mm^3)	Std. Deviation (cells/mm^3)
≥ 200	273	360.44	150.20
< 200	322	78.02	60.51

Table 3.7: Frequency, mean and standard deviation of nadir CD4 counts in different category ranges.

We opt to not report the MAEs for the HIV+ cases and HIV− controls since we believe that the mixed and fixed-effects models provide a more comprehensive means of assessing the differences between the groups. These models account for multiple influencing factors and capture the direction of the differences between the predicted and chronological ages of the individuals.

3.5.4 Noteworthy Limitations

It is important to address limitations associated with the testing dataset. As evidenced in Section 3.1.2, we see not all controls may be classified as 'healthy'. While we aim to determine brain age differences solely due to HIV, HIV independent factors present, such as substance use [143], mental health [12, 152], and cardiovascular health [153] play a role in brain ageing and may confound results. HIV dependent factors, such as incomplete data on the duration of infection and treatment, further complicate the analysis. While we aim to carry out the investigation with precision, it is necessary to be aware of these constraints. Further limitations arise from the imbalanced distributions of age and sex across cohorts, as well as between cases and controls within ENIGMA-HIV, which could introduce biases in subsequent analyses. Also, due to differences in scanners and scanning protocols, processing steps such as harmonisation are required. Despite implementing post-harmonisation scan checks, this step adds further processing that might result in the loss of HIV-relevant information.

3.6 Summary

In this chapter, we outlined the methodology employed to predict brain age using a 3D CNN, largely inspired by the ResNet [28]. We leverage the UK Biobank and CamCAN datasets for training, and partition them separately into training, validation, and testing subsets, ensuring a proportional distribution across ages and sexes.

Due to the disparity in age distributions across the UK Biobank dataset versus the CamCAN and ENIGMA-HIV datasets, we adopted a strategy of pre-training on the UK Biobank dataset, followed by transfer learning on the CamCAN dataset. We hyperparameter tune and optimise to improve the accuracy and efficiency of the model. As the datasets originated from multiple sites, harmonisation is necessary; this will be addressed using the neuroHarmonize algorithm and contrasted with that of a mixed-effects model. The model's predictions will also be analysed through a fixed-effects model, with particular emphasis on the differentiation between HIV+ and HIV− brain age deltas, and correlations with clinical measures, such as CD4 count, nadir CD4 count, HIV duration, ART status and HIV plasma RNA detectability status.

Chapter 4

Experimental Results

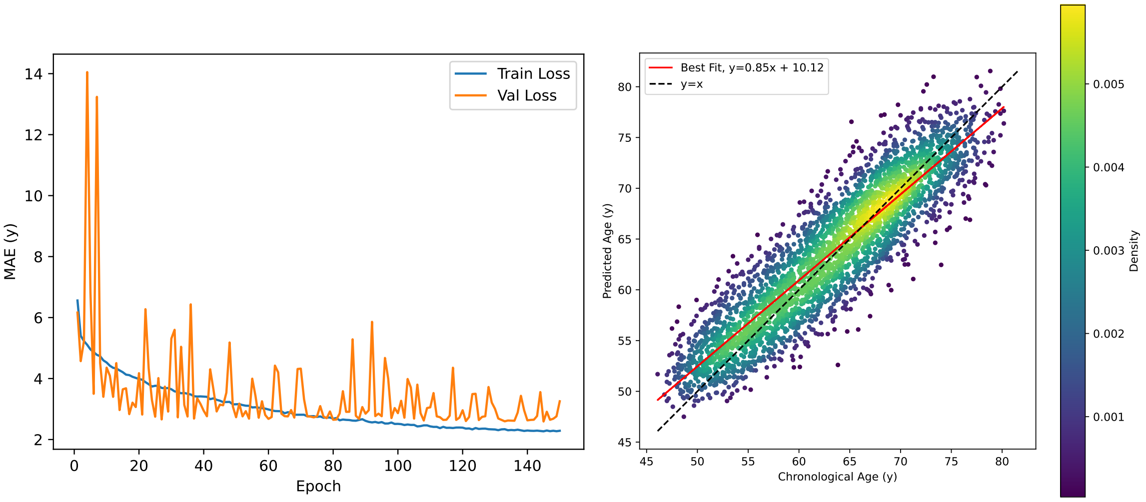
In this chapter we present the experimental results obtained. These results include model pre-training with the UK Biobank dataset, independent training and transfer learning with the CamCAN dataset, and pre- and post-harmonisation results and analyses on the ENIGMA-HIV testing dataset.

4.1 Pre-Training with the UK Biobank Dataset

We first train, validate and test the model, developed in Section 3.3, on the UK Biobank dataset (N=21366). We use the dataset split detailed in Table 3.4 whereby the training, validation and testing datasets are stratified by age and sex.

As mentioned in Section 3.4.3, we perform hyperparameter tuning on the UK Biobank and find suitable parameters to include a batch size of 8, initial learning rate of 10^{-4}, L2 weight decay of 10^{-5} and a dropout value of 0.25. Likely due to the large volume of data within the UK Biobank training dataset, we found minor accuracy differences when sweeping neighbouring learning rates and weight decays. This was also seen in Peng et al. [15]. Furthermore, we find the MAE loss function to result in a lower validation loss compared to the MSE loss function, and we use the Adam optimiser, which is a common choice (Table 2.3). We apply learning rate decay by halving the learning rate on epochs 20, 40, 60 and 80. The model is trained for 150 epochs and saved at the epoch with the lowest validation loss.

Applying the testing dataset to this model, we achieve an MAE of 2.57 ± 1.94y and Pearson correlation coefficient of 0.90. The training, validation, and testing results are shown in Figure 4.1, and detailed in Table 4.1. The mean brain age delta is found to be 0.46y. While less indicative of model accuracy than the MAE, it is useful for assessing whether the model is over-predicting or under-predicting ages.



(a) Training and validation loss

(b) Testing dataset correlation

Figure 4.1: Training, validation and testing results obtained on the UK Biobank dataset.

Metric	**Full Dataset**
Number of Scans	21366
MAE (y) - Training	2.27
MAE (y) - Validation	2.59
MAE (y) - Test	2.57 ± 1.94
Correlation (r)	0.90
Delta, δ (y)	0.46
Corrected MAE (y) - Test	2.83 ± 2.13

Table 4.1: Performance metrics of the model from pre-training and testing on the UK Biobank dataset. The MAE metrics correspond to the epoch with the lowest validation loss.

We follow the brain age bias correction inspired by Smith et al. [126] and implemented by Peng et al. [15] whereby correction parameters are calculated on the validation set and applied to the test set. Post-correction, the MAE increases to 2.83 ± 2.13y. It is not unusual for the correction to increase the MAE [15]. The histograms displaying the brain age delta distributions, pre- and post-correction, are shown in Figure 4.2, while Figure 4.3 illustrates the correlation between the deltas and chronological age, also pre- and post-correction. In Figure 4.2 we see that the distribution of deltas is shifted to become more centred around zero after correction, with the mean delta value reducing from 0.46y to -0.03y. In Figure 4.3, we see that the dependency between delta and chronological age is mostly removed, with the gradient of the line of best fit increasing from -0.15 to 0.01. Furthermore, the intercept of the line of best fit decreases from 10.12 to -0.53, which is a considerable improvement.

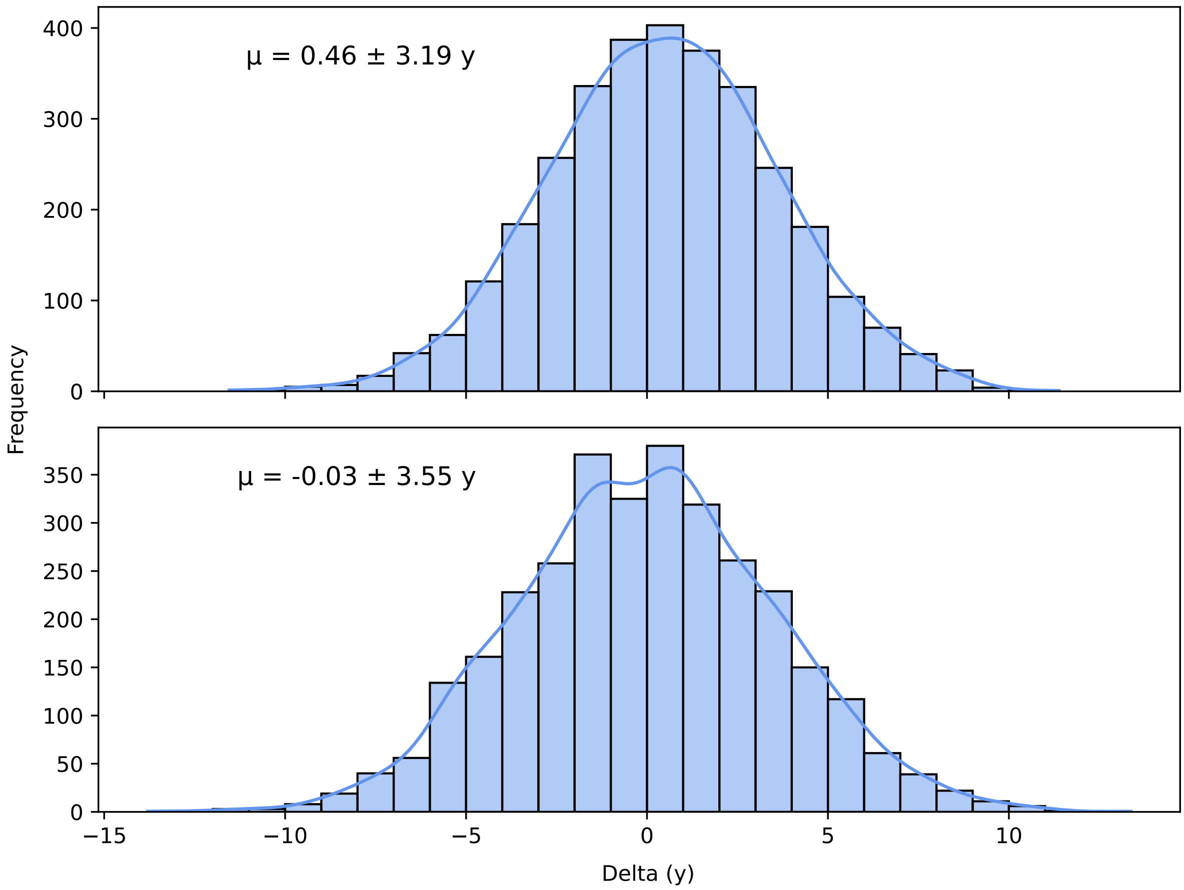


Figure 4.2: Histogram depicting the distribution of brain age delta values for the UK Biobank test set, shown both before (top) and after (bottom) applying brain age bias correction. We use a bin size of one year.

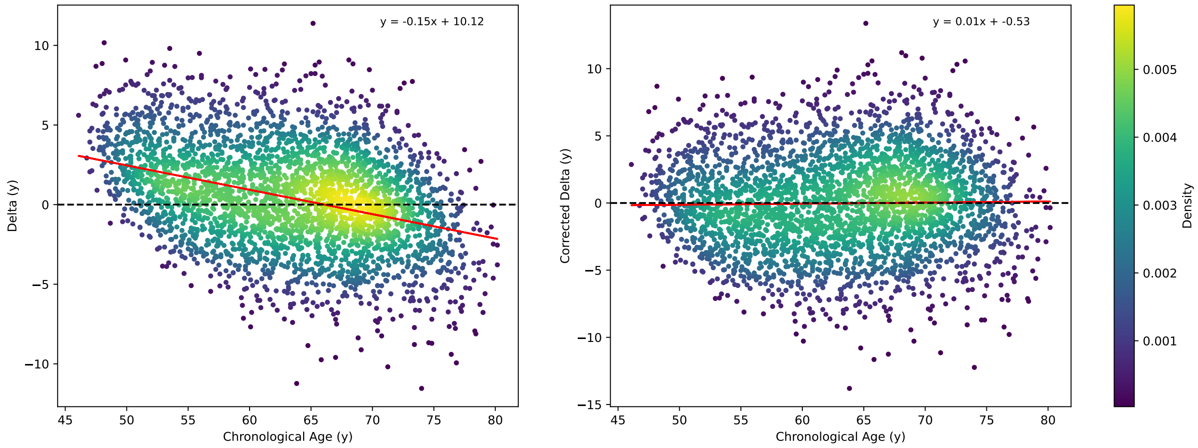


Figure 4.3: Plot of brain age delta values verse chronological age of UK Biobank test set calculated prior to (left) and post (right) the brain age bias correction.

4.2 UK Biobank Size and Model Performance

While rich in features, 3D T_1-weighted scans are costly. We explore the impact of dataset size on the training, validation and testing accuracy. We use the large UK Biobank dataset, and the MAE metric to assess accuracy. From Figure 3.2 we see the distribution of ages in the UK Biobank. Note that the tail ends of the distribution feature fewer scans than the midsection. When successfully training a model for application to a testing dataset, having training, validation and testing splits with equivalent distributions is important. For this reason we explore the impact of dataset size by implementing a cut off on the number of scans per age bin, where age bins span one year. Through this method we preserve the population of scans in the least represented age bins. We implement this experiment with bin frequency maxima of 50, 100, 200 and 400. We use the same hyperparameters for the different subsets of data. Since the subsets have comparable distributions, we justify that the hyperparameters are suitable throughout. We fix the seed for consistent weight initialisation between the runs.

The training and validation loss trends for each run are displayed in Figure B.1, with their respective dataset sizes and minimum validation loss epoch metrics summarised in Table 4.2. The models at the epochs with the minimum validation loss are saved for testing. The correlation plots of chronological and predicted age on the test set are shown in Figure B.2. Corresponding metrics for each run, including MAE, standard deviation, Pearson correlation coefficient r, and brain age delta δ, are also detailed in Table 4.2.

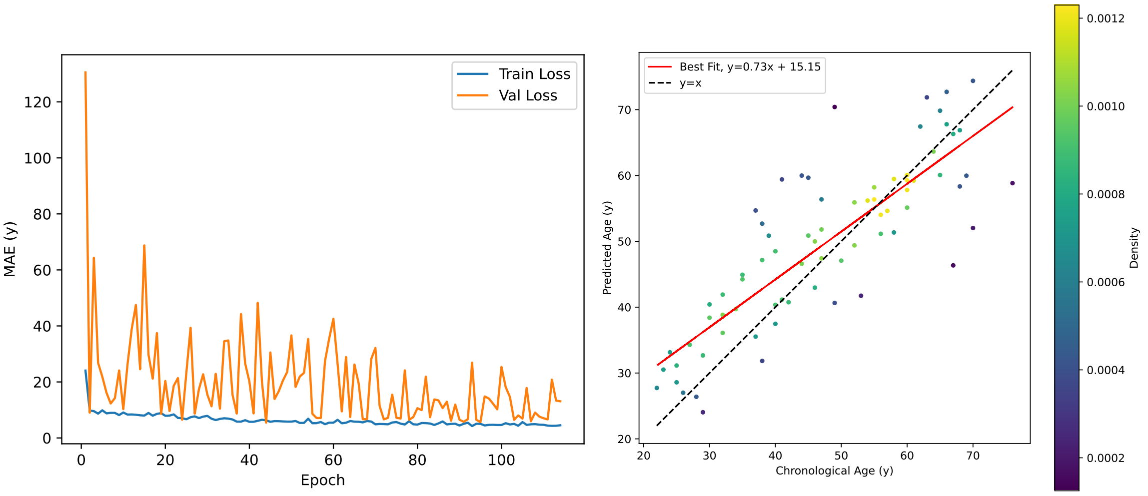
Bin Maxima	**50**	**100**	**200**	**400**	**Full**
Number of Scans	1690	3277	6245	11678	21366
MAE (y) - Training	3.64	3.27	2.67	2.51	**2.27**
MAE (y) - Validation	3.74	3.43	3.10	2.84	**2.59**
MAE (y) - Test	4.06 ± 3.13	3.47 ± 2.69	3.10 ± 2.34	2.78 ± 2.12	$\mathbf{2.57 \pm 1.94}$
Correlation, r	0.86	0.89	0.90	**0.91**	0.90
Delta, δ (y)	-1.33	-0.42	**0.11**	-0.38	0.46

Table 4.2: Summary of the effects of dataset size on the training, validation, and test MAE, correlation coefficient (r), and delta (δ). The best results in each category are highlighted in bold.

4.3 Independent Training on CamCAN Dataset

We performed extensive hyperparameter tuning on the CamCAN dataset. This entailed running a sweep over 80 epochs with various configurations including MAE and MSE loss functions, learning rates of $[10^{-4}, 10^{-3}]$, dropout values of [0.2, 0.3, 0.4], L2 weight decay of $[10^{-4}, 10^{-5}]$ and learning rate decays of [0.25, 0.5, 1], where 1 is no decay. We decayed the learning rate on epochs 20 and 50 which were seen to be suitable decay points from inspection of the loss.

From the hyperparameter sweep, we obtain the optimal hyperparameters to be an initial learning rate of 10^{-4}, L2 weight decay of 10^{-5} and a dropout value of 0.3. We run this configuration for 300 epochs with the early stopping callback occurring after the 114th epoch by which the learning rate had decreased to 2.5×10^{-5}. For the independent CamCAN training, early stopping was triggered if there was no decrease in validation loss (by more than 0.01y) in 70 epochs. We also implemented learning rate decay on plateau whereby if the validation loss had not decreased in 30 epochs, the learning rate was decayed by a factor of 2. The model was saved at the epoch with the lowest validation loss. This minimum occurred at epoch 44. The training, validation, and testing results are shown in Figure 4.4, and detailed in Table 4.3. The test set MAE was 6.38 ± 5.30y.



(a) Training and validation loss

(b) Testing dataset correlation

Figure 4.4: Training, validation and testing results obtained on the CamCAN dataset.

Metric	**Full Dataset**
Number of Scans	484
MAE (y) - Training	6.27
MAE (y) - Validation	5.46
MAE (y) - Test	6.38 ± 5.30
Correlation (r)	0.83
Delta, δ (y)	2.18
Corrected MAE (y) - Test	6.97 ± 5.87

Table 4.3: Performance metrics of the model from independent training and testing on the CamCAN dataset. The MAE metrics correspond to the epoch with the lowest validation loss, which was epoch 44.

We follow the same brain age bias correction procedure as performed on the UK Biobank pre-training stage. Post-correction, the MAE increases to 6.97 ± 5.87y. We see the histogram of brain age deltas in Figure 4.5 and the plots of delta verse chronological age in Figure 4.6. Similar to the results of the correction on the UK Biobank dataset, in Figure 4.5 we see that the distribution of CamCAN test set deltas shifts to become more centred around zero, with the mean delta reducing from 2.18y to -0.38y. In Figure 4.6 we see that the dependency between delta and chronological age is reduced, with the gradient of the line of best fit increasing from -0.27 to -0.05. The intercept of the line of best fit decreases from 15.15 to 2.22.

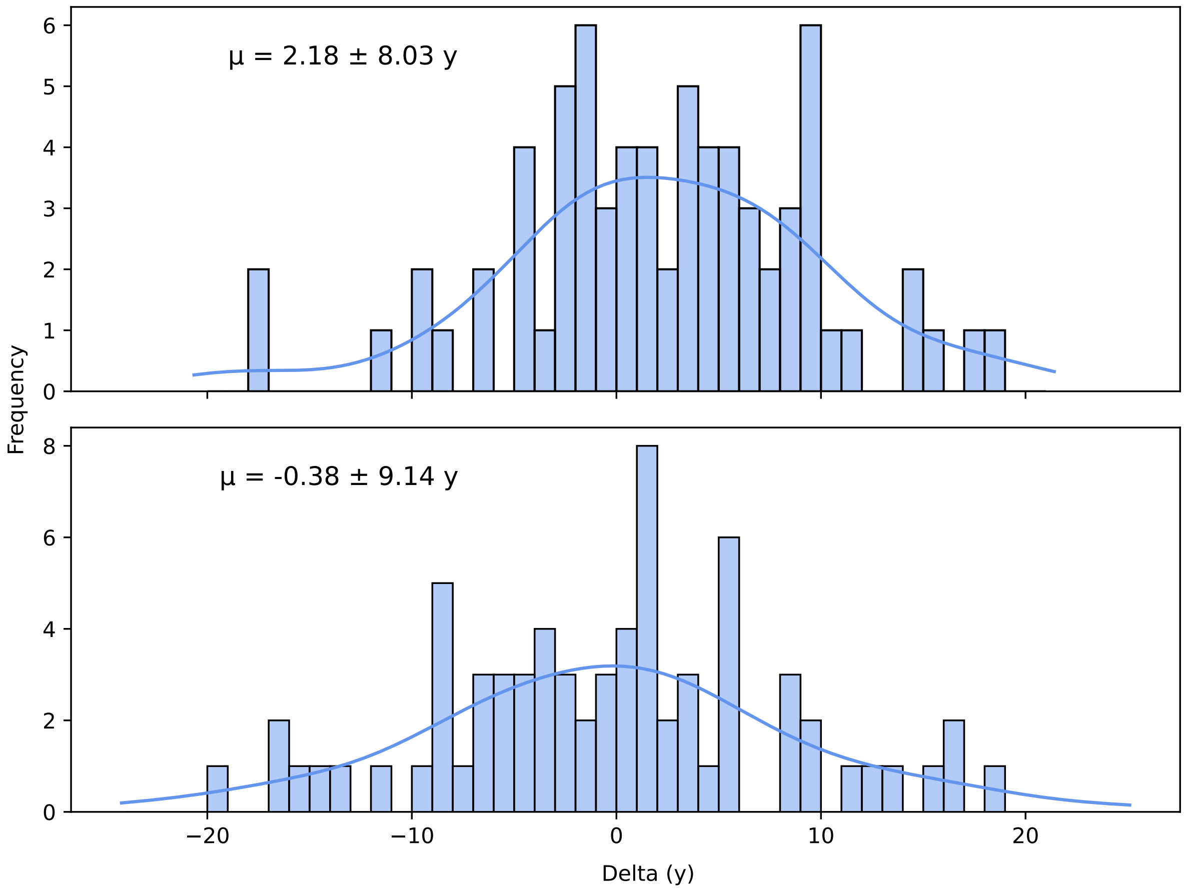


Figure 4.5: Histogram depicting the distribution of brain age delta values for the CamCAN test set following independent training, shown both before (top) and after (bottom) applying brain age bias correction. We use a bin size of one year.

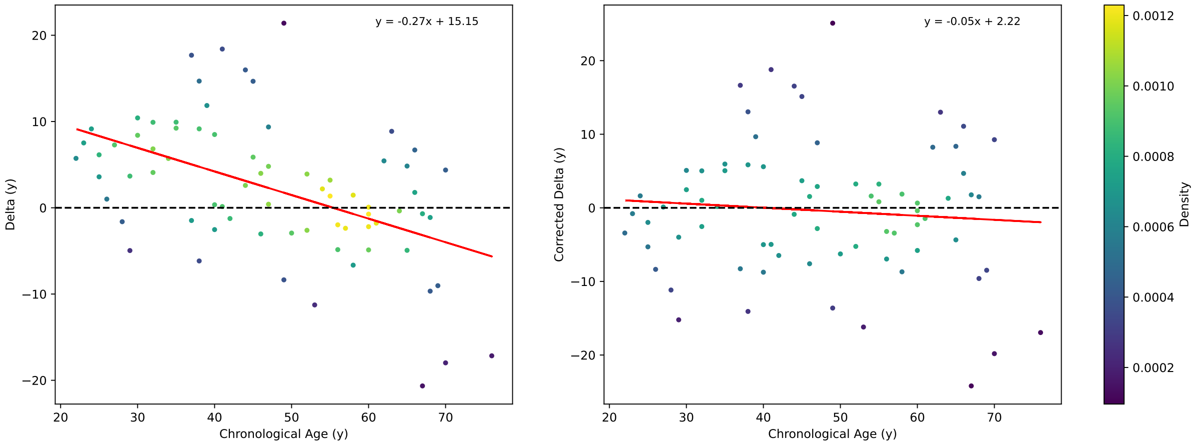


Figure 4.6: Plot of brain age delta values verse chronological age of CamCAN test set following independent training, calculated prior to (left) and post (right) the brain age bias correction.

Using the tuned hyperparameters, we swept over three randomly generated seeds and ran the model for 80 epochs to ensure that the results are consistent and not a feature of the model initialisation. We display the training and validation loss plots in Figure 4.7. It is evident that the loss trends follow similar trajectories.

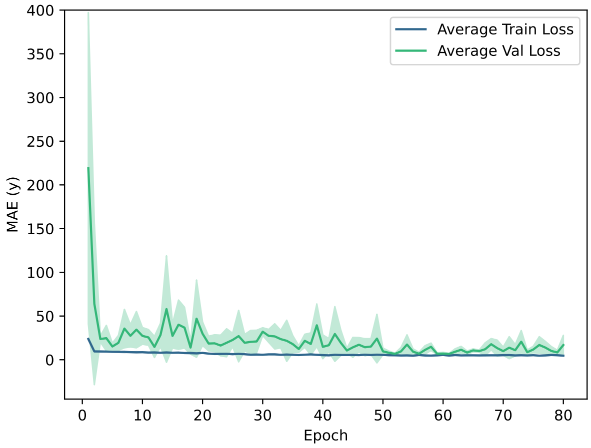


Figure 4.7: Average training and validation loss curves obtained on the CamCAN dataset through independent training for three random seeds. Note we do not use a log scale on the y-axis due to the confidence interval entering negative values.

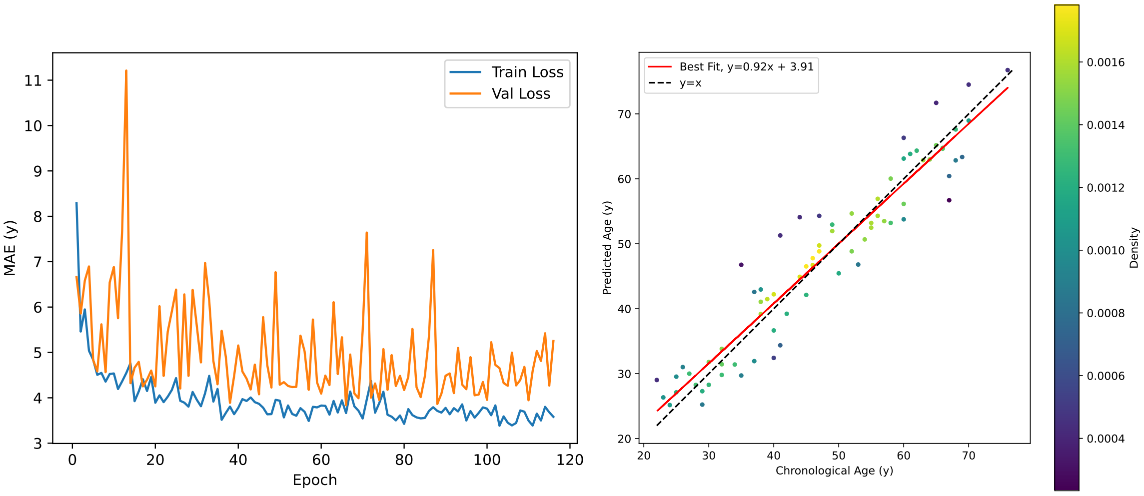
4.4 Transfer Learning on CamCAN Dataset

To train a model applicable to the ENIGMA-HIV testing set, we perform transfer learning on the pre-trained UK Biobank model. We use the CamCAN dataset for this transfer learning. While considerably smaller, this dataset has age and sex distributions more comparable to the ENIGMA-HIV testing dataset. To transfer learn, we load the pre-trained model and commence training, validation and testing.

We performed extensive hyperparameter tuning on the CamCAN dataset. This entailed running a sweep over 80 epochs with various configurations including MAE and MSE loss functions, learning rates of $[10^{-4}, 10^{-3}]$, dropout values of [0.45, 0.55, 0.65], L2 weight decay of $[0,\ 10^{-4}, 10^{-5}]$ and learning rate decays of [0.5, 1], where 1 is no decay. Again, we decayed the learning rate on epochs 20 and 50 which were seen to be suitable decay points from inspection of the loss. We found the fine-tuning with transfer learning to require higher rates of dropout to prevent overfitting, hence the differing dropout sweep values.

We find the MAE loss function, learning rate of 10^{-4} and learning rate decay of 0.5 to consistently perform better. An L2 weight decay of 10^{-5} was also found to be appropriate. From the performance of dropout we conclude that a value of 0.65 is preferable. For reproducibility we fix the seed. As with the independent CamCAN training, we implement learning rate decay on plateau whereby if the validation loss had not decreased in 30 epochs, we decayed the learning rate by half. We implement early stopping and run for 300 epochs. The early stopping callback occurred after the 116th epoch by which the learning rate had decreased to 5×10^{-5}. The minimum validation loss occurred at epoch 59. The training, validation, and testing results are shown in Figure 4.8, and detailed in Table 4.4. The test set MAE was 3.54 ± 2.59y.

We follow the standard procedure for brain age bias correction. Post-correction, the MAE slightly increases to 3.72 ± 2.82y. We see the histogram of brain age deltas in Figure 4.9 and the plots of delta verse chronological age in Figure 4.10. Similar to previous



(a) Training and validation loss

(b) Testing dataset correlation

Figure 4.8: Training, validation and testing results obtained from transfer learning on the CamCAN dataset.

Metric	**Full Dataset**
Number of Scans	484
MAE (y) - Training	3.66
MAE (y) - Validation	3.84
MAE (y) - Test	3.54 ± 2.59
Correlation (r)	0.95
Delta, δ (y)	0.24
Corrected MAE (y) - Test	3.72 ± 2.82

Table 4.4: Performance metrics of the model from transfer learning on the CamCAN dataset. The MAE metrics correspond to the losses at the epoch with the lowest validation loss, which was epoch 59.

results, in Figure 4.9 we see that the distribution of deltas shift to become more centred around zero, with the mean delta value reducing from 1.35y to -0.30y. In Figure 4.10 we see that the dependency between delta and chronological age is also reduced, with the gradient of the line of best fit increasing from -0.15 to 0.00. The intercept of the line of best fit decreases from 8.58 to -0.53.

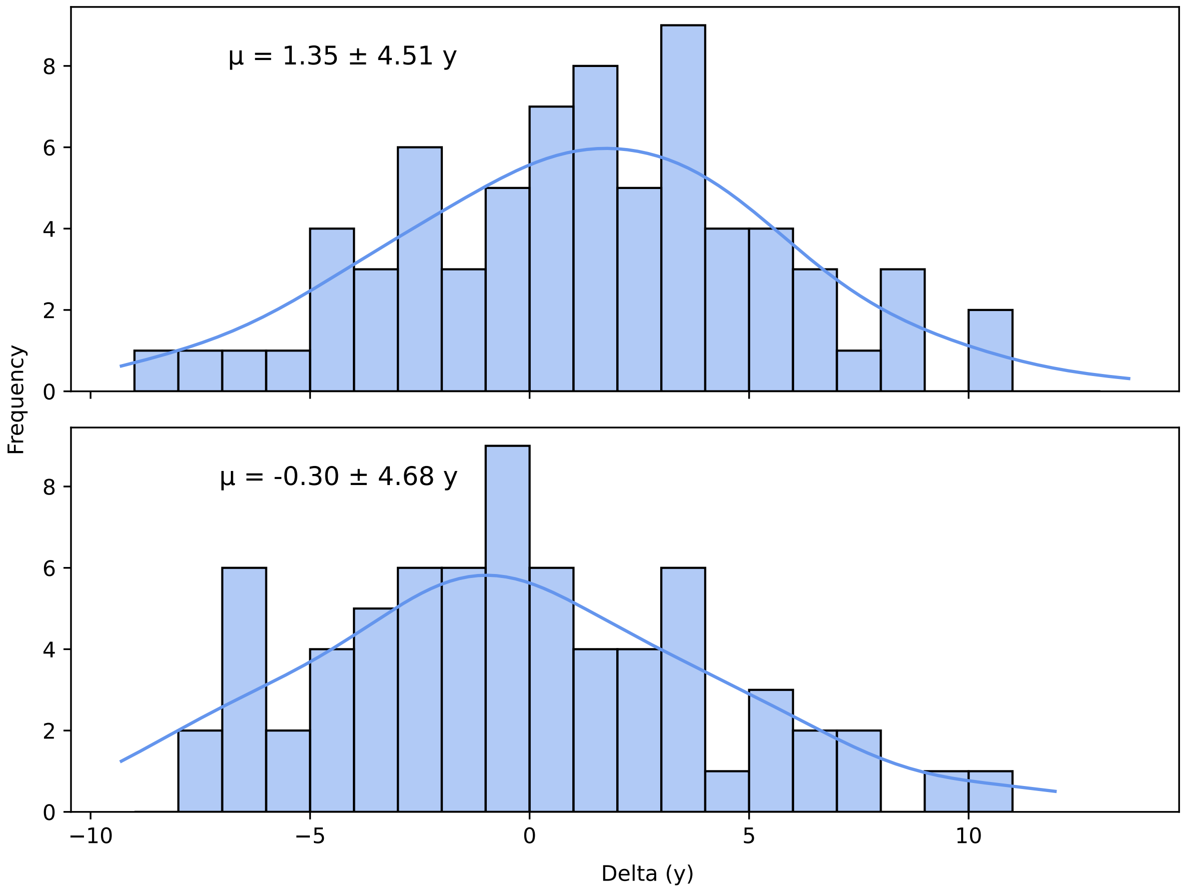


Figure 4.9: Histogram depicting the distribution of brain age delta values for the CamCAN test set after transfer learning, shown both before (top) and after (bottom) applying brain age bias correction. We use a bin size of one year.

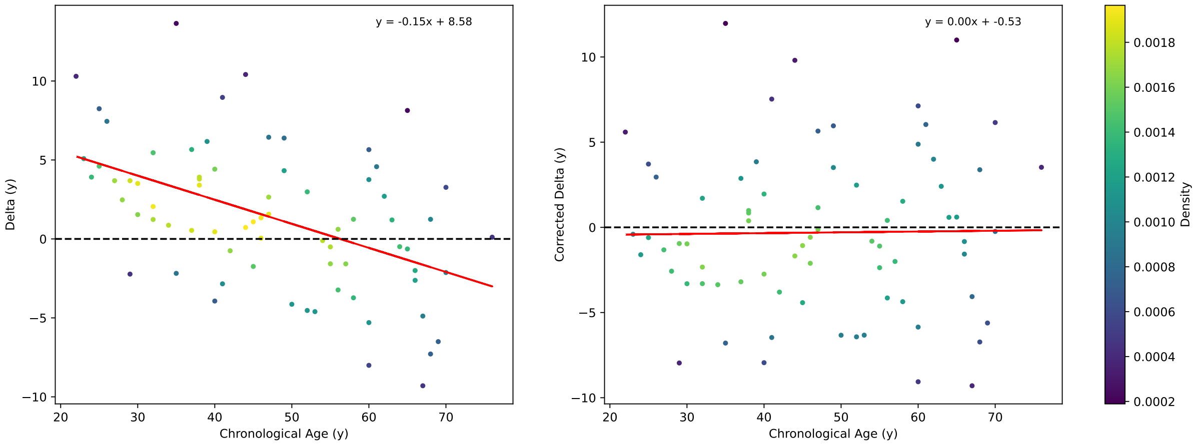


Figure 4.10: Plot of brain age delta values verse chronological age of CamCAN test set calculated prior to (left) and post (right) the brain age bias correction.

As for the independent CamCAN training runs, using the tuned hyperparameters, we swept over three randomly generated seeds and ran the model for 80 epochs. We display the training and validation loss plots in Figure 4.11. Again, we can see that the loss trends for different seeds follow similar trajectories.

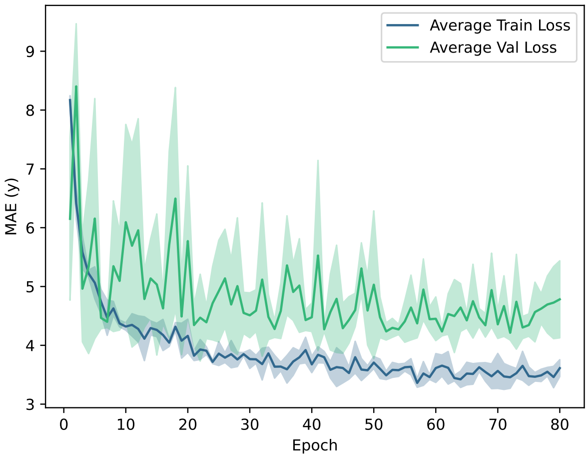


Figure 4.11: Average training and validation loss curves obtained on the CamCAN dataset through transfer learning for three random seeds.

4.5 Pre-Harmonisation Testing Results

In this section, we lay out the findings of the pre-harmonisation testing and analysis conducted on the ENIGMA-HIV dataset. For testing, we use the model that has been pre-trained and fine-tuned through transfer learning. Note that despite the differing age and sex distributions between the CamCAN and ENIGMA-HIV datasets (Figures 3.4 and A.2, respectively), we argue that our model, trained on age and sex-balanced datasets like the UK Biobank and CamCAN, should robustly handle the variations in testing dataset age and sex distributions. This variation is discussed further in Section 3.1.2.

In Sections 4.5 and 4.7 we use the legend defined in Table 4.5 for the analysis tables. We set our significance threshold at $\alpha = 0.05$ and use this threshold throughout the analysis. When conducting analyses, we are interested in the directional influence of HIV (and associated clinical measures CD4, nadir CD4, HIV duration, ART status and plasma RNA detectability) on the brain age delta, hence we convert the two-tailed p-values to one-tailed p-values, following standard statistical practice when testing for directional hypotheses. We are able to use the reported mixed or fixed-effects model coefficients to do this. The reported coefficients indicate the estimated effect of each predictor variable on the dependent variable. More specifically, for continuous variables, the coefficients indicate how much the dependent variable, in this case, delta, is expected to change when the predictor variable increases by one unit, keeping all other variables constant. For categorical variables, the coefficients represent the difference in delta when compared to a baseline or reference category, assuming other variables are held constant. Since we hypothesise the measures of HIV, and above mentioned associated clinical measures, will lead to an increase in delta, if the reported coefficients are positive the one-tailed p-values are calculated as

$$p_{\text{1-tail}} = \frac{p_{\text{2-tail}}}{2} \tag{4.1}$$

else, if the coefficient is negative,

$$p_{\text{1-tail}} = 1 - \frac{p_{\text{2-tail}}}{2}. \tag{4.2}$$

Note that all p-values in the reported mixed and fixed effects tables are given as two-tailed p-values. We highlight, in bold, p-values of significance, accounting for their appropriately calculated one-tailed p-value. We also note that the reported intercept values, while standard practice to include in the mixed and fixed-effects model outputs, are not meaningful to our analyses and so we do not highlight, nor discuss them.

As we are conducting multiple hypothesis tests within our analyses, it is important to correct for the increased risk of false positives. To address this, we use the Bonferroni correction for adjusting p-values. This correction is conservative, adjusting the p-value threshold for significance in proportion to the number of tests performed. The Bonferroni method of correction has been widely used in the literature [11, 17, 154]. We do not include the intercept in the p-value corrections. We follow the aforementioned procedures throughout Sections 4.5 and 4.7.

Variable	Description
Sex	1: Male, 2: Female
HIV Status	0: Control, 1: Case
CD4 Bin	0: Control, 1-Mild: $>$ 500 cells/mm^3, 2-Medium: $\leq$ 500 cells/mm^3, $\geq$ 200 cells/mm^3, 3-Severe: $<$ 200 cells/mm^3
Nadir CD4 Bin	0: Control, 1-Mild/Medium: $\geq$ 200 cells/mm^3, 2-Severe: $<$ 200 cells/mm^3
ART	0: Untreated, 1: Treated, 2: Control
plRNA Detectable	0: Undetectable, 1: HIV plasma RNA Detectable

Table 4.5: Definition of variables used in analysis tables.

4.5.1 Mixed-Effects Model

In this section we apply a mixed-effects model to the test results from the ENIGMA-HIV dataset. Within the testing dataset there is data pooled from eleven sites. Within two sites, two distinct scanning protocols were used to obtain the data. As outlined in Section 2.3.3, data pooled from different sites and scanning protocols contains site and protocol related variation. This variation should not contribute to the model age estimates, therefore when conducting analysis, the variation should be factored in. As discussed in Section 3.5.3, a mixed-effects model is a means of accounting for this variation.

The plot of chronological and predicted age for the cohorts within ENIGMA-HIV, split by scanner protocol, is given in Figure 4.12. We run a mixed-effects model on the data within this plot, first specifying brain age delta as the outcome variable, and HIV, age and sex as the predictor variables and grouping by site, given by Equation 3.1. The results from the mixed-effects model are given in Table 4.6. Applying Bonferroni correction, our significance threshold becomes $\alpha/3 = 0.017$. The one-tailed p-value for HIV status is calculated as 0.773, which does not indicate statistical significance. Similarly, sex does not demonstrates a significant contribution to delta, with a p-value of 0.097. We note that age is statistically significant, with a negative reported coefficient. This is likely attributable to the brain age bias phenomenon discussed in Section 2.5.1. This feature with age is present throughout all analyses, hence we do not discuss it each time. Due to the data being obtained from different sites and scanner protocols, we are not able to perform a brain age bias correction for this stage.

The distribution of residuals of the mixed effect fit is given in Figure D.1. The normal distribution fit has mean $\mu = -0.00 \pm 4.98$ which indicates it is centred around zero. This

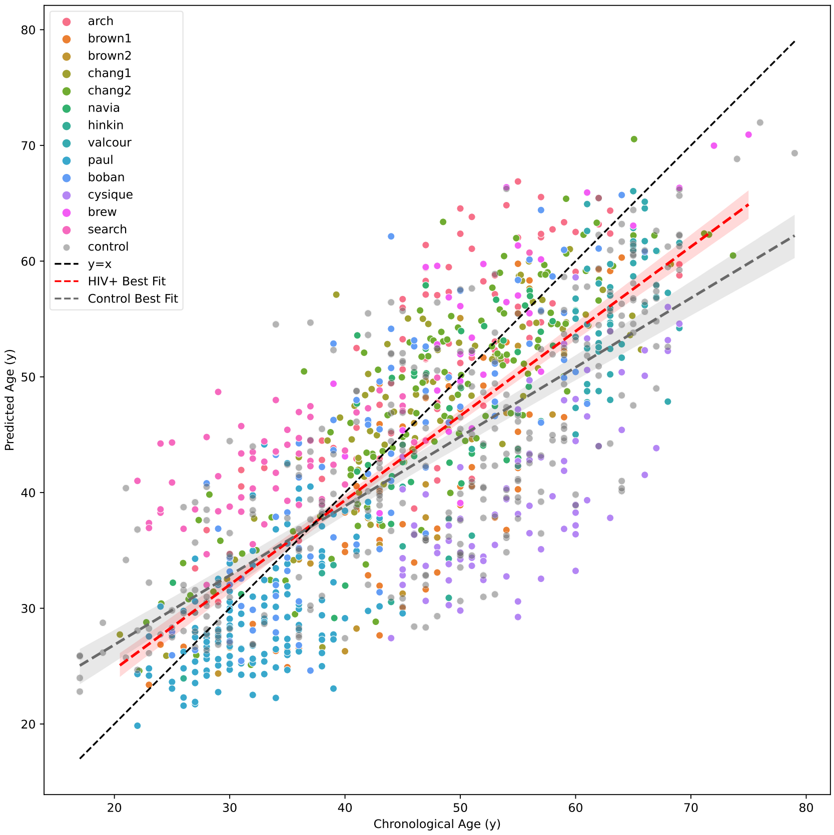


Figure 4.12: Comparison between predicted and chronological age of cases and controls for the cohorts within ENIGMA-HIV.

Variable	Coefficient	Std. Error	z-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	8.769	1.755	4.996	0.000	5.329	12.209
HIV=1	-0.303	0.405	-0.749	0.454	-1.097	0.491
SEX=2	-0.660	0.398	-1.658	0.097	-1.441	0.120
AGE	-0.232	0.016	-14.251	**0.000**	-0.264	-0.200

Table 4.6: Statistical results from mixed-effects model for predictor variables HIV, sex and age with site (split by scanner protocol) as the random effect.

is a typical characteristic of well-fitted residuals in a mixed-effects model, suggesting an appropriate model fit.

We then implement a mixed-effects model, focussing purely on the HIV cases and specify age, sex, CD4 count and nadir CD4 count as predictor variables, and delta as the outcome variable. See Table 3.3 for the sites with available covariate information. As before, we specify site as the random effect. This mixed-effects model formula is given in Equation 3.3. Note that for the analysis including the HIV$-$ controls, and specifying the above predictor variables, the mixed-effects model fails to converge. This is likely attributable to the high complexity of the model and insufficient data. The results of the analysis with the HIV cases and specifying age, sex, CD4 count and nadir CD4 count as predictor variables are given in Table 4.7. This analysis consisted of N=588 individuals. The significance threshold, after Bonferroni correction, becomes $\alpha/5 = 0.010$. The nadir CD4 bin 2 was found to have a one-tailed p-value of 0.010 which is just within statistical significance. As before, age is statistically significant. The residuals from the fit are given in Figure D.2. The normal distribution fit has mean $\mu = 0.00 \pm 5.25$.

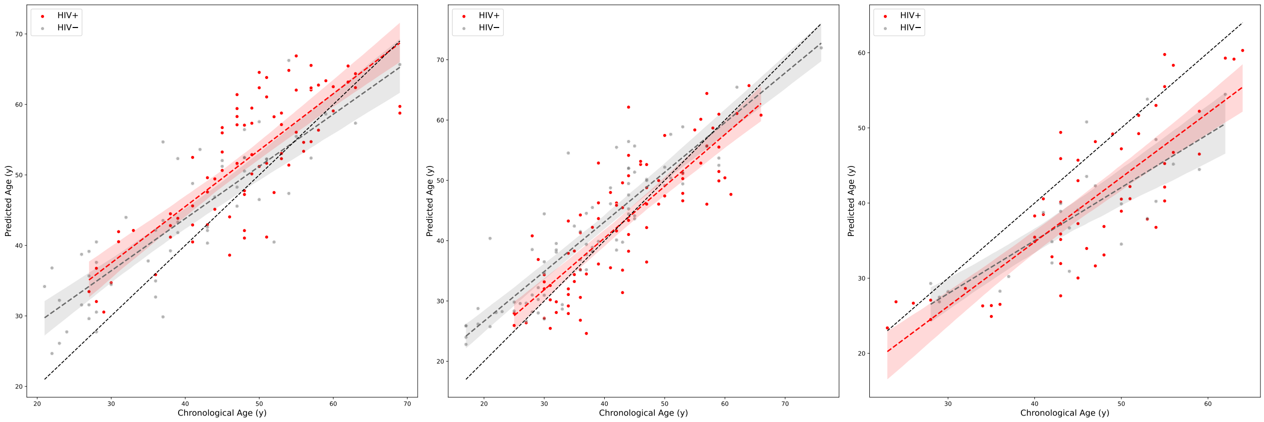
Variable	Coefficient	Std. Error	z-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	6.520	2.120	3.076	0.002	2.365	10.675
SEX=2	0.595	0.662	0.899	0.369	-0.702	1.891
CD4 Bin=2	0.257	0.491	0.522	0.601	-0.706	1.219
CD4 Bin=3	1.057	0.840	1.258	0.208	-0.590	2.704
nadCD4 Bin=2	1.178	0.505	2.333	**0.020**	0.188	2.169
AGE	-0.224	0.024	-9.158	**0.000**	-0.272	-0.176

Table 4.7: Statistical results from mixed-effects model for predictor variables age, sex, CD4 bin and nadir CD4 bin, for HIV cases only.

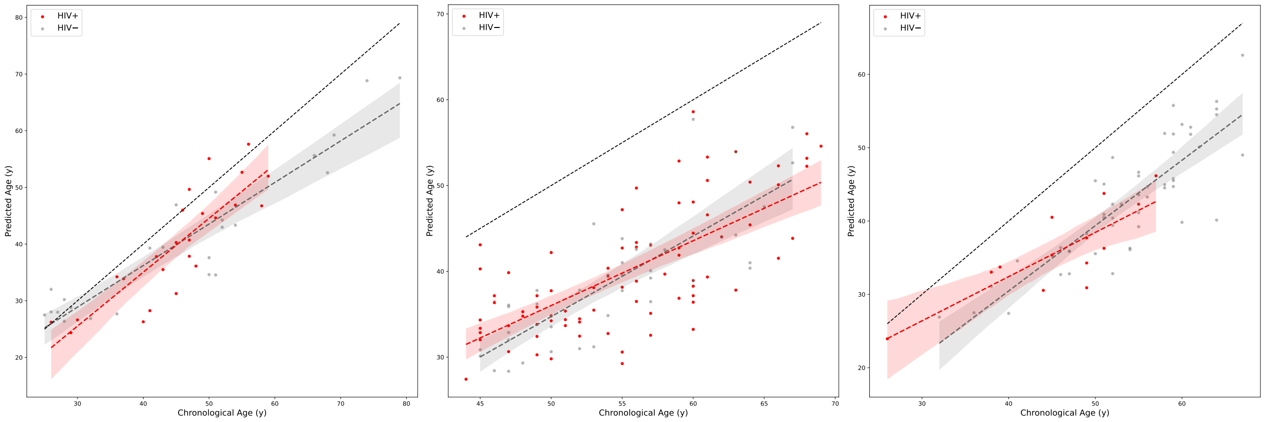
We carry out analyses for the combined cases and controls for CD4 bins and nadir CD4 bins separately, but do not detect significant results for the bins corresponding to the lowest counts. We did not detect the significance of ART, plasma RNA detectability status or HIV duration on delta, for both the HIV only and combined HIV case and control analyses.

4.5.2 Fixed-Effects Model

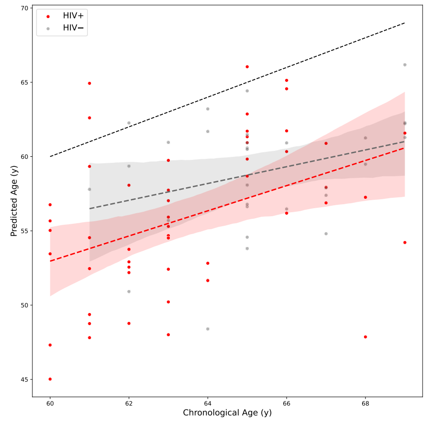
In Figure 4.13 we plot the chronological and predicted age comparison for the sites containing controls.



(a) Arch (b) Boban (c) Brown 1



(d) Brown 2 (e) Cysique (f) Hinkin



(g) Valcour

Figure 4.13: Comparison between predicted and chronological age of cases and controls for HIV cohorts with controls.

We apply a fixed-effects model to each of the cohorts with controls. As with the mixed-effects model, we specify delta as the outcome variable and HIV, sex and age as predictor variables, as in Equation 3.2. The significance threshold, after Bonferroni correction, becomes $\alpha/3 = 0.017$. The results are displayed in Table 4.8 and Table D.1 to D.5 in Appendix D.

From Table 4.8, we find the one-tailed p-value for HIV in Arch to be 0.023 which is

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	12.7337	2.000	6.366	0.000	8.776	16.691
HIV=1	2.1928	1.084	2.023	0.045	0.048	4.337
SEX=2	0.8137	1.012	0.804	0.423	-1.189	2.816
AGE	-0.2341	0.046	-5.079	**0.000**	-0.325	-0.143

Table 4.8: Arch cohort statistical results from fixed-effects model for predictor variables age, sex and HIV.

statistically significant for the significance threshold $\alpha = 0.05$, but outside of statistical significance after Bonferroni correction.

Next we specify age, sex, CD4 count and nadir CD4 count as the predictor variables. Arch is the sole cohort to reveal results of significance. We present these results in Table 4.9. Note that the sites with these predictor variables are detailed in Table 3.3. The significance threshold, after Bonferroni correction, becomes $\alpha/7 = 0.007$.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	12.9660	2.047	6.333	0.000	8.914	17.018
SEX=2	0.6557	1.073	0.611	0.542	-1.468	2.779
CD4 Bin=1	0.2720	1.039	0.262	0.794	-1.786	2.330
CD4 Bin=2	1.0987	1.082	1.015	0.312	-1.044	3.241
CD4 Bin=3	1.8703	2.141	0.873	0.384	-2.368	6.109
nadCD4 Bin=1	1.3861	1.213	1.142	0.255	-1.016	3.788
nadCD4 Bin=2	1.8549	0.963	1.925	0.057	-0.052	3.762
AGE	-0.2378	0.047	-5.029	**0.000**	-0.331	-0.144

Table 4.9: Arch cohort statistical results from fixed-effects model for predictor variables age, sex and CD4 bin and nadir CD4 bin.

From Table 4.9, we find the one-tailed p-value for nadir CD4 bin 2 in Arch to be 0.029 which is statistically significant for the significance threshold $\alpha = 0.05$, but again outside of significance after Bonferroni correction.

Performing analysis with age, sex and CD4, again solely Arch has significant results, before, but not after, Bonferroni correction. Similarly Arch shows significant results for the analysis with age, sex and nadir CD4, before, but not after, Bonferroni correction. These analyses are displayed in Tables D.6 and D.8, respectively.

4.6 Harmonisation

We applied harmonisation to the NIFTI post-registration scans and re-ran the testing. The cohorts Brown and Chang, acquired using two different scanner protocols each, were divided into sub-cohorts for harmonisation. Figures E.1 and E.2 illustrate the histograms of average intensity for a sampled axial slice across each site, before and after harmonisation, respectively. This process aligns the intensity distributions with the CamCAN reference dataset. The neuroHarmonize algorithm, using the CamCAN test set as the reference set, computes scale and shift parameters for each unique site to mitigate variations in scans unrelated to the preserved parameters (age, sex, and HIV status). The site Navia was excluded from this analysis due to harmonisation biases, attributed to the small number of younger scans in this site.

4.7 Post-Harmonisation Testing Results

In this section, we lay out the findings of the post-harmonisation testing and analysis conducted on the ENIGMA-HIV dataset.

We first implement the brain age bias correction using the correction parameters calculated on the CamCAN validation set as outlined in Section 3.5.2. Post brain age bias correction, a change in gradient and intercept of chronological age and delta is evident in Figure 4.15. We note that the negative correlation between chronological age and delta is aligned with the brain age bias phenomenon discussed in Section 2.5.1. In Figure 4.16 we see the comparison of predicted and chronological ages for cases and controls, following harmonisation and brain age bias correction.

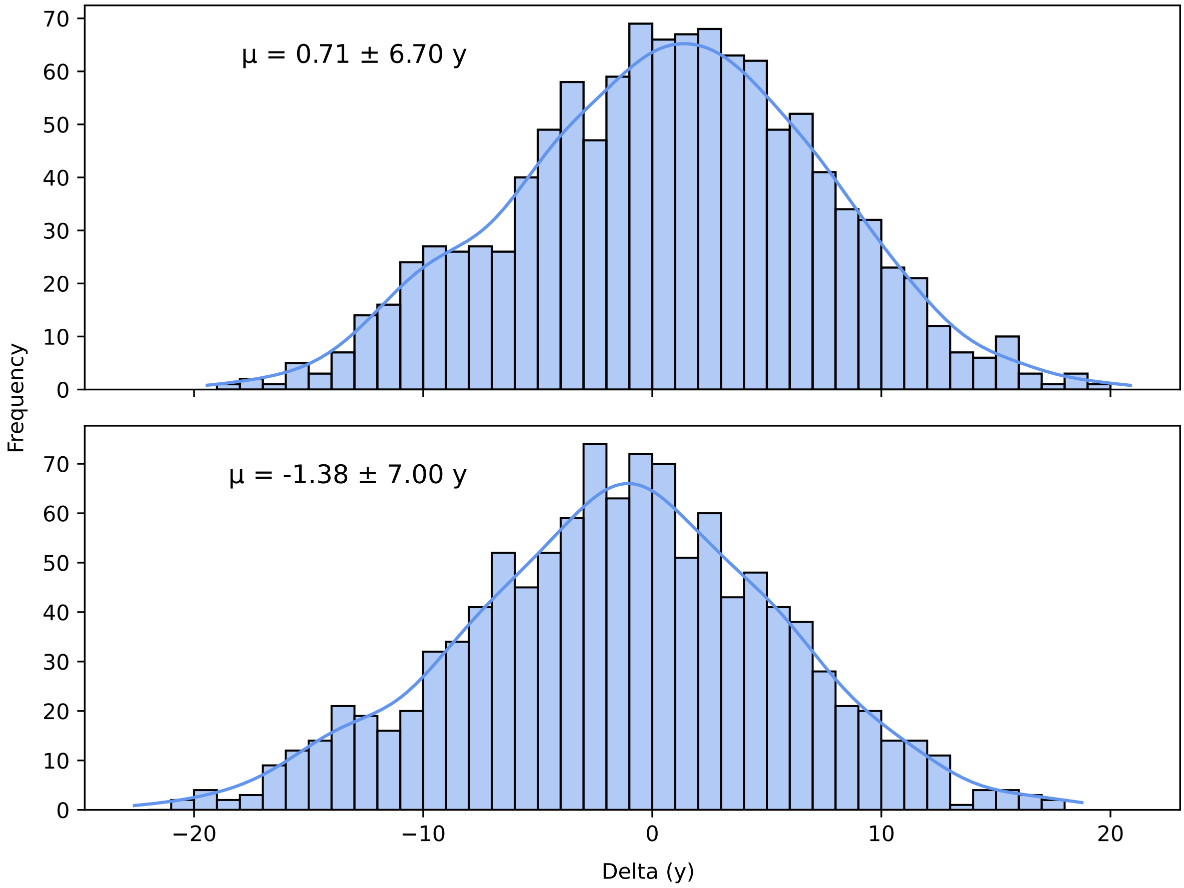


Figure 4.14: Histogram depicting the distribution of brain age delta values for the ENIGMA-HIV test set (Navia and Shikuma excluded), shown both before (top) and after (bottom) applying brain age bias correction. We use a bin size of one year.

We perform all further analyses using the brain age bias corrected results. The fixed effects results for the analysis with age, sex and HIV as predictor variables, and delta as the outcome variable (see Equation 3.2), are displayed in Table 4.10. We note that the

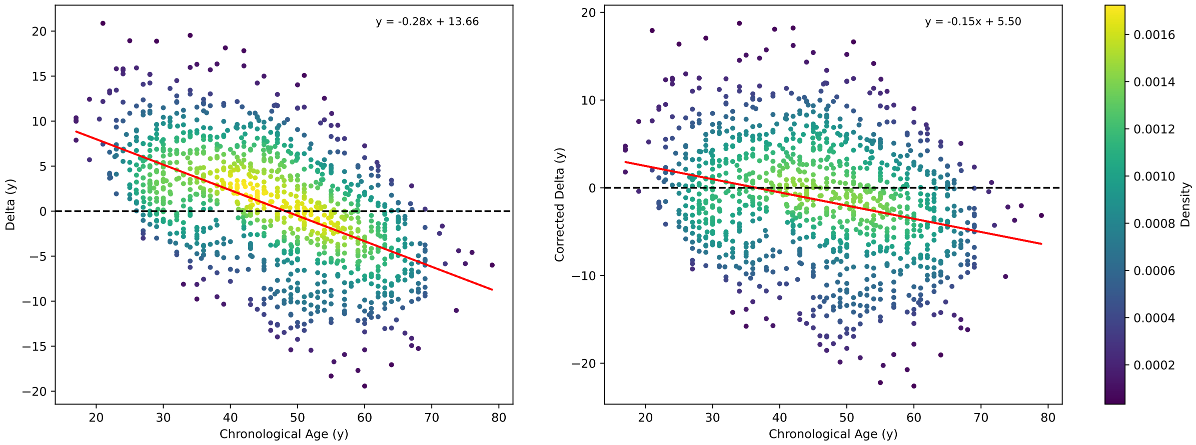


Figure 4.15: Plot of brain age delta values against chronological age of ENIGMA-HIV test set (Navia and Shikuma excluded) calculated prior to (left) and post (right) the brain age bias correction.

distribution of deltas was shifted slightly away from normal. The significance threshold, after Bonferroni correction, becomes $\alpha/3 = 0.017$. From Table 4.10, we see that the one-tailed p-value for HIV is 0.613 which is not statistically significant.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	5.8510	0.941	6.215	0.000	4.004	7.698
SEX=2	-0.2884	0.467	-0.617	0.537	-1.206	0.629
HIV=1	-0.1310	0.454	-0.288	0.773	-1.022	0.760
AGE	-0.1543	0.017	-9.035	**0.000**	-0.188	-0.121

Table 4.10: Statistical results from fixed-effects model for predictor variables age, sex and HIV.

We then implement a fixed-effects model and specify age, sex, CD4 count and nadir CD4 count as predictor variables. This fixed-effects model formula is given by,

$$\text{Delta} \sim \text{Age} + \text{Sex} + \text{CD4 Bin} + \text{Nadir CD4 Bin}. \tag{4.3}$$

Figure 4.17 displays the comparison between chronological and predicted age, stratified by CD4 bin and nadir CD4 bin, respectively. The results of the analysis are given in Table 4.11. This analysis consisted of N=790 individuals. As in Section 4.5, the p-values in the table correspond to a two-tail test and since we are investigating if the clinical measures correspond to a statistically significant positive increase in delta, the sign of the corresponding coefficients allow for calculation of the one-tailed p-value. The significance threshold, after Bonferroni correction, becomes $\alpha/7 = 0.007$.

The results indicate that the statistically significant predictor variables corresponding to an increase in delta are the most severe CD4 bin (bin 3, CD4 $<$ 200 cells/mm^3) and nadir CD4 bin (bin 2, nadir CD4 $<$ 200 cells/mm^3). The residuals from the fit are given in Figure F.13 and are well-distributed. The normal distribution fit has mean $\mu = -0.00 \pm 6.70$.

To include more data, we then implement a fixed-effects model and specify age, sex and CD4 count as predictor variables - Delta $\sim$ Age + Sex + CD4 Bin. In Figure 4.18

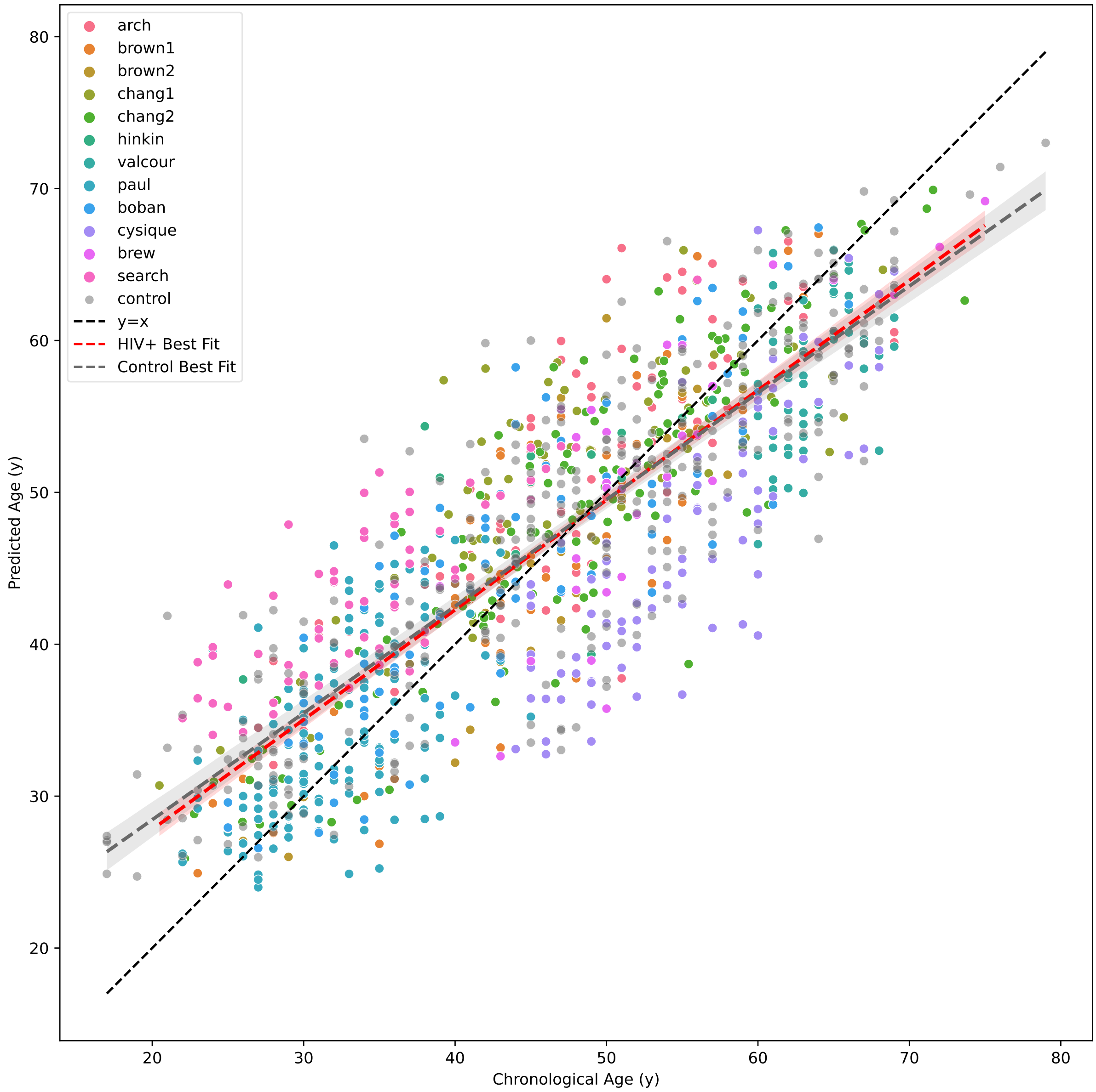


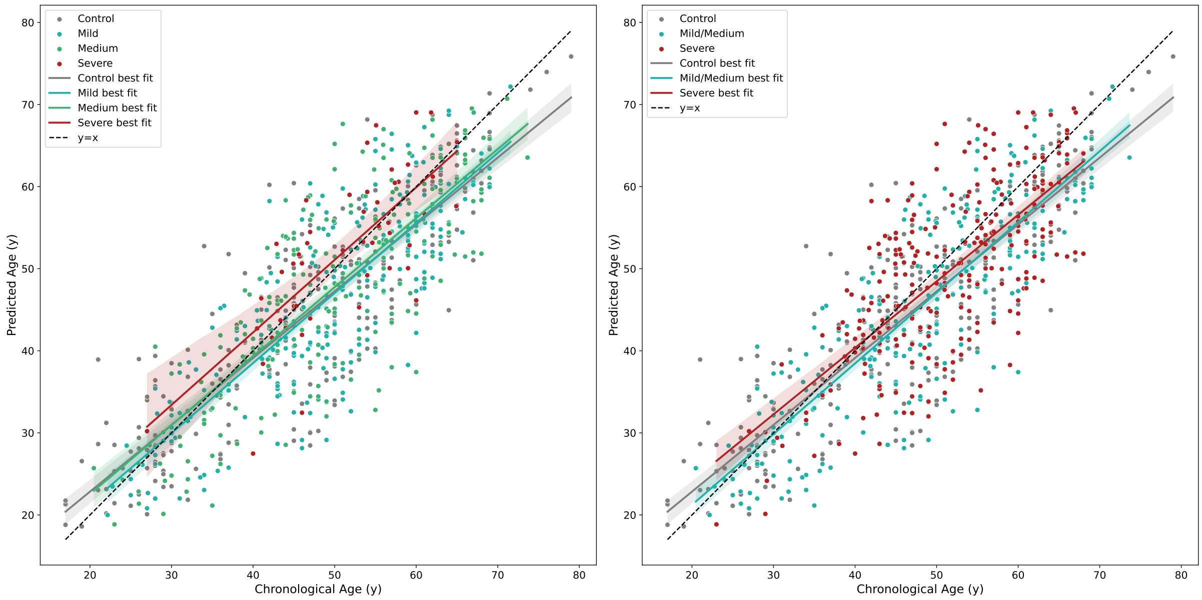
Figure 4.16: Comparison between predicted and chronological age of cases and controls for the cohorts within ENIGMA-HIV following harmonisation and brain age bias correction.

the comparison between chronological and predicted age stratified by CD4 bin is plotted. The results of the analysis are given in Table 4.12. This analysis consisted of N=1105 individuals. The significance threshold, after Bonferroni correction, becomes $\alpha/5 = 0.010$.

The results indicate that the statistically significant predictor variables corresponding to an increase in delta is the most severe CD4 bin (bin 3, CD4 < 200 cells/mm^3), but only prior to Bonferroni correction. The residuals from the fit are given in Figure F.14 and are well-distributed. The normal distribution fit has mean $\mu = 0.00 \pm 6.72$.

The fixed-effects models specifying HIV duration, CD4 count and nadir CD4 count, as well as the model with ART status, CD4 count and nadir CD4 count, did not find statistically significant effects due to HIV duration and ART, respectively.

Conducting analysis on the HIV cases only, we still do not find a statistically significant effect for HIV duration or ART. The results of the fixed-effects model with predictor



(a) CD4

(b) Nadir CD4

Figure 4.17: Plots of chronological and predicted age with lines of best fit plotted according to CD4 (left) and nadir CD4 (right) bin values. These plot only includes data points where both CD4 and nadir CD4 data was present, relevant to Table 4.11. Refer to Table 4.5 for the bin values.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	5.1765	1.046	4.950	0.0000	3.124	7.229
SEX=2.0	1.4660	0.634	2.313	0.0210	0.222	2.710
CD4 Bin=1.0	-0.8208	0.464	-1.769	0.0772	-1.732	0.090
CD4 Bin=2.0	-0.3501	0.446	-0.786	0.4322	-1.225	0.524
CD4 Bin=3.0	2.9422	0.831	3.539	**0.0004**	1.310	4.574
nadCD4 Bin=1.0	0.5585	0.505	1.107	0.2688	-0.432	1.549
nadCD4 Bin=2.0	1.2128	0.446	2.719	**0.0067**	0.337	2.088
AGE	-0.1653	0.020	-8.169	**0.0000**	-0.205	-0.126

Table 4.11: Statistical results from fixed-effects model for predictor variables age, sex, CD4 bin and nadir CD4 bin.

variables age, sex, CD4 count, nadir CD4 count and plasma RNA detectability, given by

$$\text{Delta} \sim \text{Age} + \text{Sex} + \text{CD4 Bin} + \text{Nadir CD4 Bin} + \text{plasma RNA Bin}, \tag{4.4}$$

are displayed in Table 4.13. For this analysis, the significance threshold, after Bonferroni correction, becomes $\alpha/6 = 0.008$. The most severe nadir CD4 bin (bin 2, nadir CD4 $<$ 200 cells/mm^3) and a positive plRNA detectability status (bin 1) are within statistical significance, prior to, but not following, Bonferroni correction. We see that the most severe CD4 bin (bin 3, CD4 $<$ 200 cells/mm^3) is no longer within statistical significance. Note that only N=232 data points, from three sites (four when split by scanner protocol) were able to be used for this analysis. Recall that for harmonisation, it is necessary to classify the data by scanner and protocol due to the variation in scans introduced by

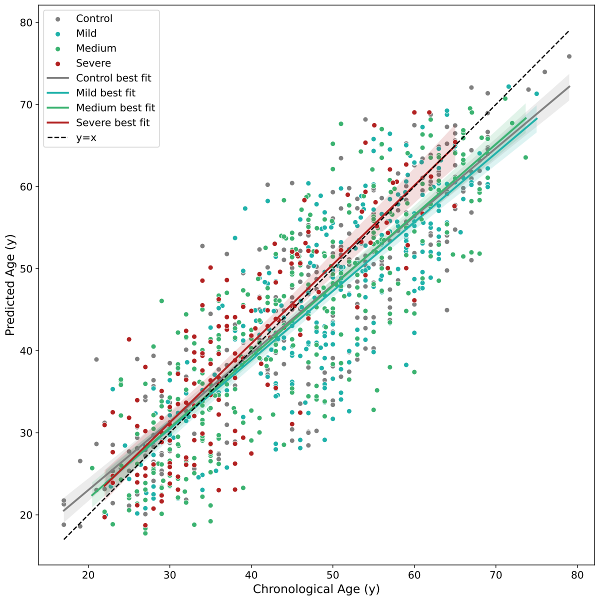


Figure 4.18: Plots of chronological and predicted age with lines of best fit plotted according to CD4 bins. Refer to Table 4.5 for the bin values.

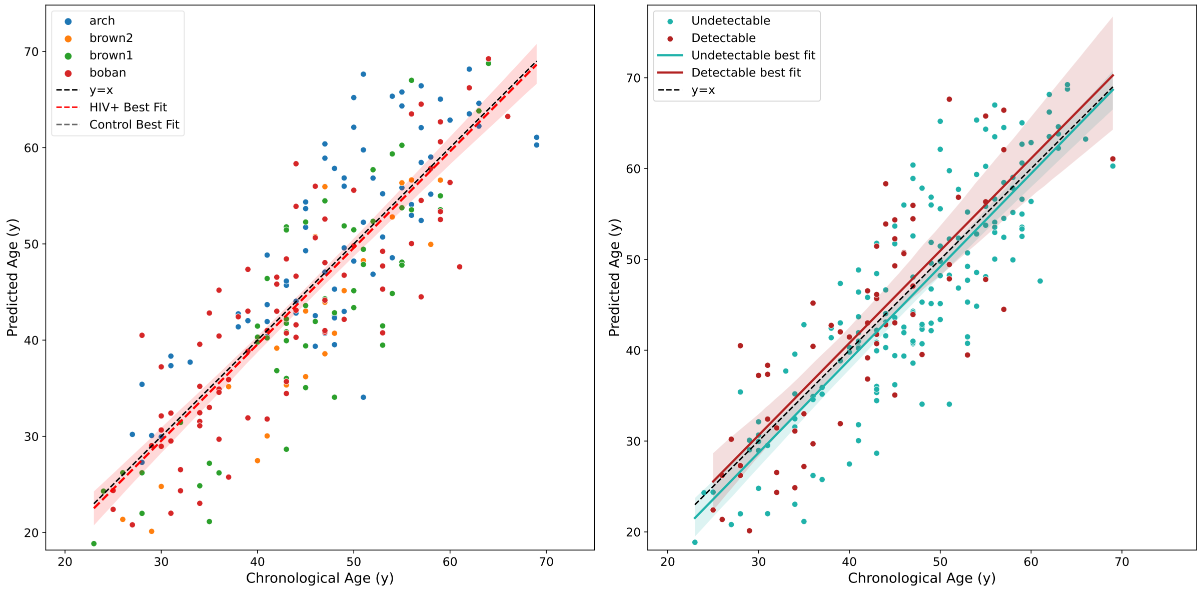
Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	5.6782	0.952	5.965	0.000	3.811	7.546
SEX=2	-0.6100	0.482	-1.267	0.205	-1.555	0.335
CD4 Bin=1	-0.8901	0.555	-1.604	0.109	-1.979	0.199
CD4 Bin=2	-0.2194	0.530	-0.414	0.679	-1.259	0.821
CD4 Bin=3	1.3797	0.684	2.017	0.044	0.038	2.722
AGE	-0.1485	0.017	-8.533	**0.000**	-0.183	-0.114

Table 4.12: Statistical results from fixed-effects model for predictor variables age, sex and CD4 bin.

different scanners and protocols. In Figure 4.19 the comparison between chronological and predicted age, both without and with stratification of plRNA detectability status, is plotted. We show that the fixed-effects model residuals are well-distributed in Figure F.15.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	-2.0406	2.156	-0.946	0.345	-6.290	2.209
SEX=2.0	0.1432	1.023	0.140	0.889	-1.873	2.159
CD4 Bin=2.0	-0.2893	0.958	-0.302	0.763	-2.178	1.599
CD4 Bin=3.0	0.4660	1.967	0.237	0.813	-3.411	4.343
nadCD4 Bin=2.0	1.8516	0.960	1.928	0.055	-0.041	3.744
plRNA detectable=1.0	1.9722	1.036	1.904	0.058	-0.069	4.014
AGE	0.0071	0.045	0.159	0.874	-0.081	0.095

Table 4.13: Statistical results from fixed-effects model for predictor variables age, sex, CD4 bin, nadir CD4 bin and HIV plasma RNA detectability.



(a) Chronological and predicted age

(b) Chronological and predicted age stratified by plasma RNA detectability

Figure 4.19: Plots of chronological and predicted age both without (left) and with (right) HIV plasma RNA detectability stratification.

Chapter 5

Discussion

In this chapter we discuss the experimental results outlined in Chapter 4 and to what extent our research questions were addressed.

5.1 Pre-Training with the UK Biobank Dataset

We obtained high performing results on the UK Biobank test set following pre-training, achieving an MAE of 2.57 ± 1.94y. Although the primary role of the UK Biobank dataset in this study is for model pre-training, including a UK Biobank test set is valuable. It provides a point of comparison of our model's performance with other studies in the literature. From Table 2.3, we see that this result ranks third in the literature surveyed. The leading two works both make use of the UK Biobank dataset, however, while the leading work of Peng et al. [15] solely uses the UK Biobank dataset, Yin et al. [16] uses three additional datasets. If we compare the results of this study with other works which solely use the UK Biobank dataset, we rank second. Peng et al. [15] leads with an MAE of 0.43y less, while Dinsdale et al. [17] follows our work with a reported MAE range of $2.71 - 3.09$y. Dinsdale et al. [17] provides differentiated results for female and male subjects, as well as for subjects with linearly and non-linearly registered scans. Lastly Kolbeinsson et al. [11] features with a reported MAE range of $2.87 - 3.42$y. The MAE of 2.87y was obtained on a health stratified test set, while the MAE of 3.42y was obtained on a test set not stratified for health.

See Table 5.1 for a summary of these results, including MAEs, standard deviations, and details of the model architectures.

Our work features smaller standard deviations in comparison to those reported by Dinsdale et al. [17] and Kolbeinsson et al. [11], indicative of more consistent model performance across test samples. It is important to note that while Peng et al. [15] reports a substantially smaller standard deviation, their calculation methodology differs. Therefore, a direct comparison of standard deviation values between our work and theirs is not a fair comparison.

It is also worth mentioning that our pre-trained model obtains a brain age delta of 0.46y on the test set. This result is indicative that the model is not greatly biased towards over-predicting or under-predicting ages. Studies often do not include their brain age delta and so we do not compare our delta with other works.

Looking at the complexity of the architectures, we see Peng et al. [15] contains less convolutional layers than our work while Dinsdale et al. [17] contains more. Kolbeinsson et al. [11] does not state the number of layers in their architecture. Like this work,

Study	MAE and Std. Dev. (y)	Architecture
Peng et al. [15]	2.14 ± 0.05	(3x3x3 Conv, BN, 2x2x2 MP, ReLU)x5, 1x1x1 Conv, BN, ReLU, AP, DP, 1x1x1 Conv, SM
Ours	2.57 ± 1.94	BN, ReLU, Conv, MP ((BN, ReLU, Conv)x2, SC, MP)x3, (BN, ReLU, Conv)x2, AP, Flatten, ReLU, DP, FC
Dinsdale et al. [17]	2.86 ± 2.22, 3.09 ± 2.37, 2.71 ± 2.10, 2.91 ± 2.18	((3x3x3 Conv, ReLU, BN)x2, 2x2x2 MP)x3, ((3x3x3 Conv, ReLU, BN)x3, 2x2x2 MP)x2, BN, FCx2
Kolbeinsson et al. [11]	$2.87 - 3.42 \pm 3.72$	(3x3x3 Conv, BN, ReLU, 3x3x3 MP)x N, (3x3x3 Conv, BN, ReLU, SC, 3x3x3 Conv, BN, ReLU)x4, AP, FC

Table 5.1: This table presents a comparative analysis of our study against other research works that exclusively utilised the UK Biobank dataset. It is worth noting that Peng et al. [15] derived the test set standard deviation by bootstrapping the data 1000 times, resulting in a considerably smaller standard deviation value.

Kolbeinsson et al. [11] implements a ResNet, and obtains an MAE larger by 0.3y. Thus, when compared to the study with the most similar architecture, our model achieves an improved MAE.

5.2 Experimentation with Dataset Size

From Table 4.2 we see that as the size of the dataset increased, there was a noticeable improvement in performance across the majority of metrics. More specifically, the test MAE showed consistent improvements as the number of scans increased. The improvements were approximately 14.53% when increasing the size of the dataset from 1690 (bin 50) to 3277 (bin 100), 10.66% from 3277 (bin 100) to 6245 scans (bin 200), 10.32% from 6245 (bin 200) to 11678 (bin 400) scans, and 7.55% from 11678 (bin 400) to 21366 scans (full dataset). Overall, comparing the runs with the smallest and largest dataset sizes, we observe a considerable reduction in the MAE, of 36.70%. It is clear that although the accuracy is improving, these returns are diminishing. The Pearson correlation coefficient also gets stronger, increasing by 5.81%, as the dataset size increases from 1690 (bin 50) to 11678 (bin 400), peaking at 11678 (bin 400) with $r = 0.91$. The brain age delta closest to zero corresponds to the run with dataset size 6245 (bin 200).

These results suggest that maximising dataset size is not always essential for successful DL applications. Instead, depending on available resources, one can aim for an optimal balance between computational costs and the accuracy of the results, and achieve a successful outcome. It is useful to identify the threshold where the increase in computational load outweighs the gains in accuracy, such as the test MAE. Unfortunately, accurate and reproducible runtimes for each experiment could not be provided as the computational resources for this work were shared, which influenced the durations of the different runs.

5.3 Comparison of Independent Training and Transfer Learning

Through transfer learning with the CamCAN dataset on the pre-trained UK Biobank model, we achieve substantially improved performance. The MAE on the test set is reduced to 3.54 ± 2.59y, demonstrating a 44.51% improvement compared to the MAE of 6.38 ± 5.30y obtained when independently training the model on the CamCAN dataset. Similarly, the Pearson correlation coefficient increases from $r = 0.83$ to $r = 0.95$. These improvements, along with information on the training and validation losses, are detailed in Table 5.2.

Metric	**Independent Training**	**Transfer Learning**
MAE (y) - Training	6.27	3.66
MAE (y) - Validation	5.46	3.84
MAE (y) - Test	6.38 ± 5.30	3.54 ± 2.59
Correlation (r)	0.83	0.95
Delta, δ (y)	2.18	0.24

Table 5.2: Comparison of performance metrics between independent training and transfer learning on the CamCAN dataset.

Figures 4.4 and 4.8 further highlight the differences in loss scales during training and validation. It is apparent that pre-training has considerably stabilised and reduced the validation loss and training loss magnitudes. Moreover, in the transfer learning results, predicted ages align more closely with the ideal line, the line along which theoretical brain age values fall for healthy individuals. Following brain age bias correction, we observe a more effective removal of chronological age dependency on delta in transfer learning results than in independent training. This effectiveness is evidenced by the gradient and intercept of chronological age and delta being closer to zero.

Since our objective is to apply the trained model to investigate brain age in HIV+ individuals, achieving a minimal MAE is important. A lower MAE allows for more accurate identification of the specific effects of HIV on brain ageing.

We could only find two other works using CNNs for brain age prediction on the CamCAN dataset. Taylor et al. [128] utilised a ResNet trained solely on the CamCAN dataset and obtained an MAE of 6.55y. Yin et al. [16], which we previously discussed, trained a CNN on an aggregate dataset. When tested on the CamCAN dataset, they obtained an MAE of 4.71 y for males and 3.01 y for females. Following transfer learning, we obtained an MAE of 3.54 ± 2.59y. This is notably better than the result of Taylor et al. [128]. In comparison with Yin et al. [16], our MAE is slightly better than the average of their reported MAEs, which is 3.86 y, although they do not train on any CamCAN data. To our knowledge, our work is the first to achieve such a high level of accuracy on the CamCAN dataset through transfer learning.

5.4 Pre-Harmonisation Results

Section 4.5 presented a comprehensive analysis of the pre-harmonisation testing on the ENIGMA-HIV dataset, utilising mixed and fixed-effects models to assess the influence of various predictor variables on the brain age delta. The variables used in the analysis are outlined in Table 4.5. We apply Bonferroni correction to our significance threshold of $\alpha = 0.05$, for each analysis, to assess significance in results.

5.4.1 Mixed-Effects Model Analysis

The mixed-effects model, as detailed in Section 4.5.1, considers HIV status, sex and age as predictor variables, with site (stratified by scanning protocol) as the random effect. This method is a means of mitigating the variation in scans across different sites and scanner protocols. From the mixed-effects model analysis (see Table 4.6), we find that HIV does not show a statistically significant influence on brain age delta, with $p = 0.773$. In all our analyses, age consistently shows a significant negative correlation with brain age delta. This trend primarily stems from the bias in brain age prediction models. Brain age bias correction, based on the CamCAN parameters, mitigates but does not entirely remove the age-related bias, maintaining age as a significant factor in our model's predictions. From the residual distribution of the mixed-effects model (Figure D.1), we validate the reliability of the analyses.

Analysis with HIV Clinical Measures

Further analyses incorporating HIV clinical measures reveal additional insights. Individuals with lower nadir CD4 counts (nadir CD4 bin 2, count < 200 cells/mm^3) show a significant increase in brain age delta ($p = 0.010$), with a coefficient of 1.18, indicating an increase in delta of 1.18y, if other predictor variables are held constant. This suggests that individuals who have experienced severe immunosuppression, with a history of AIDS, tend to have higher brain age deltas, potentially reflecting accelerated brain ageing. The residuals of the model were well-distributed (Figure D.2). We did not find statistically significant correlations between the other HIV-related clinical measures (i.e., HIV duration, ART status and HIV plasma RNA detectability) on the brain age delta.

Our findings, that individuals with nadir CD4 counts less than 200 cells/mm^3 exhibit increased brain age deltas, are consistent with previous research detailed in Sections 2.4 and 2.6. For example, Underwood et al. [139] found larger brain age deltas in individuals with nadir CD4 counts less than 200 cells/mm^3. Our results also align with the observations made by Thompson et al. [122], which linked increased cortical GM loss to AIDS, and Nir et al. [123], which found continued brain tissue atrophy in virologically suppressed individuals with a history of AIDS.

5.4.2 Fixed-Effects Model Analysis

The fixed-effects model applied to cohorts with controls (Figure 4.13) allows for a direct comparison of cases and controls within each cohort, allowing for distinct within-site effects to be determined. We also use fixed-effects models for within-site analyses of clinical measures.

Cohort-Specific Observations

In general, we did not find correlations between HIV status, clinical measures and brain age delta within cohorts, with the exception of the Arch cohort where HIV status was significant. This is aligned with the trends in Figure 4.13 where it appears Arch is the only cohort with controls where the effect of HIV has resulted in higher predicted ages overall.

Here, the fixed-effects model HIV+ variable exhibited an increase in delta of 2.19y, with other variables held constant ($p = 0.023$). However, this result did not remain significant following Bonferroni correction. Due to the conservative nature of this correction, we report both the uncorrected and corrected findings. This result suggests the potential of accelerated brain ageing occurring in the Arch cohort, although further analysis is required.

We conduct further analysis, applying a fixed-effects model to each site with CD4 or nadir CD4 data. We do not find results of significance for the sites within this analysis, with the exception of the Arch cohort, but only prior to correction. The plots of chronological age and predicted age, stratified by CD4 bin, and nadir CD4 bin, can be seen in Figures D.3 and D.5, respectively.

An examination of Figures C.1 to C.5, and Table 3.3, suggests that the Arch dataset is not the most significantly affected by HIV-related factors when compared with other cohorts with controls. In terms of average CD4 and nadir CD4 counts, duration of HIV infection, ART ratio and plasma RNA detectability, Arch ranks consistently lower in severity. Specifically, it holds the fourth position for current CD4 count, third for nadir CD4 count, third for HIV duration, second for ART ratio, and is the least severe in terms of plasma RNA detectability, ranking by most healthy for all measures. These rankings are out of eleven, seven, four, five and three sites, respectively.

The cohorts with the lowest CD4 counts and nadir CD4 counts, are Paul and Navia, respectively. However, we did not find results of significance from their corresponding fixed-effects models, investigating the effects of CD4 and nadir CD4 count on brain age delta. In Table D.7, we see that the coefficient of the lowest CD4 bin in the Paul cohort is more positive than that of the middle CD4 bin, but these results are not significant relative to the highest CD4 bin. However, from Figure D.4 we see that there are very few data points for the highest CD4 bin, the reference for the fixed-effects model, hence this could confound results. In Table D.9, we see a positive coefficient for the effect of the lowest nadir CD4 bin on brain age delta in the Navia cohort, but this effect is not significant relative to the higher nadir CD4 bin. From Figure D.6, we see that there is only one data point in the higher nadir CD4 bin, the reference for the fixed-effects model, hence the results are not meaningful. It is evident that the smaller individual cohort sizes, and imbalanced spread of data within each site impacts the analysis. Methods such as the mixed-effects model, where the site data can be combined, are useful.

5.4.3 Summary

In the pre-harmonisation mixed-effects model analysis, where age, sex, CD4 count, and nadir CD4 count were included as predictor variables, nadir CD4 count emerged as a significant factor. Specifically, individuals in the lowest, most severe, nadir CD4 count bin exhibited an increase of 1.18 years in brain age delta, holding other variables constant. While the coefficient for the lowest current CD4 bin indicated a delta increase of 1.06 years, this was not statistically significant ($p = 0.104$). **Our findings align with our hypothesis that there is a negative correlation between an individual's nadir CD4 count and their brain age delta. The hypothesis that ART use correlates with a smaller delta and that AIDS status correlates with a larger delta was partially validated: AIDS status (not ART use) was linked to an increased brain age delta.**

5.5 Harmonisation Considerations

The harmonisation of the ENIGMA-HIV dataset was a critical step in preparing the multi-site neuroimaging data for consistent testing and analysis. This process aimed to align intensity distributions across the different sites and scanning protocols, and reduce non-biological variation in the scans. It also allowed for larger scale, and more meaningful analysis to be conducted. The majority of sites within ENIGMA-HIV are quite small (N<150), with only six sites containing controls. A majority of analyses would be lost if we were not able to pool the scans for comparison.

We used the neuroHarmonize algorithm [98], which adjusts the data from each site and scanner protocol using calculated scale and shift parameters. The scale and shift parameters are crucial for adjusting each site's data to match the reference dataset. As one of the checks of the efficacy of the harmonisation, we visually inspected the scale and shift maps within each site to ensure that the scales and shifts were reasonable. Figures E.1 and E.2 show the effectiveness of this harmonisation, displaying histograms of intensities for a selected axial slice before and after the process. Post-harmonisation, we observed that the intensity distributions closely matched the mean and variance of the CamCAN reference dataset, indicative of the harmonisation's efficacy.

We encountered a challenge with the Navia site, which was ultimately excluded from the analysis. In Figure F.9 we see the massive over-prediction of the ages of the younger individuals. This was likely due to an imbalanced age distribution and a low number of younger scans, skewing the harmonisation results, but further investigation is required.

While ComBat harmonisation has been shown to be effective for small multi-site datasets, and for those with imbalanced covariate distributions [97, 100] such as ENIGMA-HIV, its efficacy for whole-brain voxel analyses may be limited. According to Fortin et al. [97], ComBat works under the assumption that scanner effects are consistent across voxels. However, this assumption may not be true in whole-brain voxel analyses, where scanner effects depend on the tissue medium. The software we utilised, neuroHarmonize [98], builds upon ComBat, offering additional features. However, to the best of our knowledge, it has not yet addressed this specific limitation in voxel-level analysis.

In summary, while harmonisation effectively standardised the majority of the datasets, it also highlighted the challenges posed by imbalanced dataset distributions within individual sites. Therefore, conducting both pre- and post-harmonisation analyses is good practice, although one must be aware of the limitations inherent in the harmonisation process [97], and the difficulties in validating the output scans [101].

5.6 Post-Harmonisation Results

In this section, we discuss the key findings and implications of the post-harmonisation testing and fixed-effects model analyses conducted on the ENIGMA-HIV dataset as detailed in Section 4.7.

5.6.1 Fixed-Effects Model Analysis

The fixed-effects model results (Table 4.10) indicate that post-harmonisation, with brain age bias correction, HIV status does not significantly influence the brain age delta ($p = 0.613$). Throughout the post-harmonisation analysis, age continues to be significantly correlated with a decrease in brain age delta due to the age bias in brain age prediction models.

Analysis with HIV Clinical Measures

Further analysis incorporating HIV clinical measures (Table 4.11) reveals that individuals with lower CD4 counts (CD4 bin 3, count < 200 cells/mm^3) and lower nadir CD4 counts (nadir CD4 bin 2, count < 200 cells/mm^3) show significant increases in brain age delta ($p = 0.0002$ and $p = 0.0034$, respectively). The corresponding increases in delta are 2.94y and 1.21y, respectively, if all other variables are held constant each time. From Figure 4.17 the increased predicted ages, corresponding to individuals in the lowest CD4 and nadir CD4 bins, are evident from the respective lines of best fit.

Similarly to our pre-harmonisation results, our findings that individuals with nadir CD4 counts less than 200 cells/mm^3 display increased brain age deltas align with that of previous works [122, 123, 139]. Additionally, we found significance for a current status of AIDS ($p = 0.0002$). These findings also align with the work of Grill [23], where lower CD4 counts are linked to increased risk of opportunistic infections. As detailed in Section 2.4, these opportunistic infections may have neurological impacts, for example, causing lesions.

In the analysis of HIV+ cases only, plasma RNA detectability emerges as a significant factor ($p = 0.029$, see Table 4.13) with a corresponding increase in delta of 1.97y, if all other variables are held constant. While the significance is lost post-Bonferroni correction, this result highlights the potential influence of HIV plasma RNA presence on brain ageing. This result aligns with the findings of Petersen et al. [140], who found an increased positive delta in those with detectable viral loads. Similarly, Kuhn et al. [138] found a correlation between HIV RNA viral load (related to HIV plasma RNA detectability) and delta. The non-significance of the lowest CD4 bin in this fixed-effects model could be attributed to the smaller sample size (four sites/scanners, N=232) and the specific clinical measures of the included cohorts, as seen in Figure 4.19. We also find the lowest nadir CD4 count bin to be of significance, prior to but not following correction ($p = 0.028$), with a delta increase of 1.85y. In all above analyses we find the residual distributions to be normal and validate the reliability of the analyses.

We do not find significance of HIV duration and ART status for the fixed-effects models specifying HIV duration, CD4 count and nadir CD4 count, as well as the model with ART status, CD4 count and nadir CD4 count. This may indicate that the predictor variables of HIV duration and treatment status are less influential to accelerated brain ageing than the state of the immune system (CD4 count and nadir CD4 count) and amount of virus present in the blood (plasma RNA detectability). As aforementioned, Nir et al. [123] found that virologically suppressed individuals with a history of AIDS still experience brain atrophy. This may be linked to why we do not detect correlations between delta and ART status, as nadir CD4 count confounds this result.

5.6.2 Summary

In the post-harmonisation fixed-effects model analysis with age, sex, CD4 count, and nadir CD4 count as predictor variables, both CD4 count and nadir CD4 count significantly affected brain age delta. The lowest CD4 count bin was associated with an increase in delta of 2.94 years, while the lowest nadir CD4 count bin was associated with an increase in delta of 1.21 years, assuming other variables remained constant. **This supports our hypothesis that there is a negative correlation between current and nadir CD4 counts and brain age delta**. For the fixed-effects model including age, sex, CD4 count, nadir CD4 count, and HIV plasma RNA detectability, significant influences on brain age delta were observed for nadir CD4 count and HIV plasma RNA detectability, but only prior to Bonferroni correction. Here, the lowest nadir CD4 count bin showed an increase in delta of 1.85 years, and HIV plasma RNA detectability showed an increase in delta of 1.97 years, again holding other predictor variables constant. There is research to suggest that MRI screening is not useful for HIV+ individuals without neurological symptoms [125]. **It is therefore reasonable that we have only detected increased brain age deltas in individuals with lower current CD4 counts, a history of AIDS (lowest nadir CD4 count bin) and a detectable HIV plasma RNA status**. These findings contribute to a better understanding of the neurological impact of HIV and the crucial clinical factors to monitor.

Chapter 6

Conclusion

In this chapter, we revisit the aims and objectives outlined in Section 1.5 and the research questions from Section 1.3, discussing the extent to which they have been met. This chapter also highlights the contributions of our work to brain age prediction, and primarily to investigating brain ageing in HIV+ individuals. Lastly, we detail the limitations of the research and conclude with recommendations for future work.

Objective 1: Conduct an extensive literature review

Through the comprehensive literature review in Chapter 2, we established a thorough understanding of the methodologies and common practices in the field of brain age prediction. This review encompassed an overview of ML, with a specific focus on CNNs, including ResNets. Additionally, the review delved into the phenomenon of brain ageing, detailing the changes that occur, as well as medical imaging background relevant to brain scans, focussing on pre-processing steps and the role of harmonisation of MRI scans - elements vital to this work. Furthermore, the neurological impact of HIV and relevant ML applications were explored, providing motivation for our second and third research questions and corresponding hypotheses. The review also offered an in-depth analysis of specific applications of CNNs in brain ageing, which, along with the background on CNNs and ResNets, allowed for the formulation of our model. In summary, the insights gained from the literature review not only aided the design of our CNN model but also highlighted the novelty of our research in the context of existing literature, laying a foundation for our experimental approach.

Objective 2: Develop an accurate pre-trained model

We designed, pre-trained and hyperparameter tuned the CNN model on the UK Biobank dataset (N=21366). To this end, we achieved an MAE of 2.57 ± 1.94 years and Pearson correlation coefficient $r = 0.90$ on the test set. This MAE ranked third in the literature surveyed, following Peng et al. [15] and Yin et al. [16], and second within studies using the UK Biobank dataset, following Peng et al. [15]. From these results, we can conclude that our model design and pre-training was highly successful.

Objective 3: Investigate the use of transfer learning for improving model performance and application to small datasets

Our third objective, directly addressing our first research question, *'To what extent does model pre-training and transfer learning allow for application of the model to smaller, unseen datasets?'*, focused on exploring the effectiveness of model pre-training and transfer learning in adapting our model for use with smaller, unseen datasets like the ENIGMA-HIV dataset. We fine-tuned our model through transfer learning with the CamCAN dataset and compared the results with those from independent training on the same dataset. This comparison, particularly focusing on the MAE and Pearson correlation coefficient, demonstrated the effectiveness of transfer learning, which notably led to a 44.51% improvement in the test MAE. This large improvement in MAE results in a robust model able to obtain more accurate results on smaller, unseen datasets, including those of wider age ranges. However, it is important to note that for the ENIGMA-HIV testing dataset, which comprised multi-site data, harmonisation was an essential step in ensuring reliable and accurate results. In summary, while transfer learning substantially improved model performance, the implementation of harmonisation was crucial for its effective application to smaller, unseen datasets.

Objective 4: Apply the model to the testing dataset and analyse correlations between HIV-related clinical measures and brain age delta

Our fourth objective addressed our second and third research questions, *'What are the effects of HIV on an individual's brain ageing?'* and *'What HIV-related clinical measures, such as an individual's CD4 count, nadir CD4 count, ART status and AIDS status, is the brain ageing effect correlated to?'*, respectively. Regarding our second research question, we did not find a direct correlation between HIV status and brain age delta, both pre- and post-harmonisation. This finding suggests that HIV status alone may not be a direct indicator of accelerated brain ageing. However, due to the variation in scan quality and case-control, age, and sex distributions within the ENIGMA-HIV dataset, we cannot definitively conclude this and further investigation is required.

For our third research question, our post-harmonisation analysis led to valuable insights regarding HIV-related clinical measures and brain age delta. We found statistically significant correlations between lower CD4 counts and nadir CD4 counts, and increased positive brain age delta. These findings suggest a potential link between weakened previous and current immunity (as indicated by lower nadir CD4 and current CD4 counts, respectively) and accelerated brain ageing. Additionally, HIV plasma RNA detectability showed a correlation with increased positive brain age delta, but only prior to applying Bonferroni correction. This is indicative of a potential link between HIV plasma RNA detectability and accelerated brain ageing, but requires further investigation.

Pre-harmonisation, we found a statistically significant correlation between nadir CD4 count and increased positive brain age delta. It is worthwhile noting that we found nadir CD4 count to be a significant factor both before and after harmonisation, indicating its reliability. No significant correlations were observed with HIV duration or ART status, either pre- or post-harmonisation. Throughout our analyses, we found age to be statistically significant, however this was largely attributable to the brain age bias present in predictive models, despite correction efforts.

In summary, while HIV status alone did not directly correlate with accelerated brain ageing, our findings post-harmonisation revealed links between accelerated brain ageing and lower CD4 and nadir CD4 counts, with nadir CD4 count emerging as a consistent factor both pre- and post-harmonisation. However, no significant correlations were found with HIV duration or ART status, and the influence of age remained significant, likely due to inherent model bias. We note that since clinical covariate information such as CD4 count, nadir CD4 count, HIV duration, ART status, and plasma RNA detectability status was available only for HIV+ cases, our conclusions regarding accelerated brain ageing are applicable solely to the HIV+ cases, not the HIV− controls.

6.1 Contributions

This thesis makes several notable contributions to the field of brain ageing research. Firstly, we developed a pre-trained model, which achieved high accuracy in brain age prediction. The model code will be made publicly available on GitHub. As of this thesis submission, we are investigating the permissions needed for the public release of the pre-trained and fine-tuned model weights. It is important to note, however, that our findings suggest the necessity of harmonisation as a step prior to applying these models to unseen datasets.

Furthermore, through testing and analyses on the ENIGMA-HIV dataset, we have contributed to improving the understanding of the neurological impact of HIV. In particular we have identified significant factors, largely CD4 count and nadir CD4 count, associated with the acceleration of brain ageing in HIV+ individuals. These findings highlight the need for regular monitoring of these measures to effectively discern HIV's impact on brain health. Our findings also emphasise the vital role of ART in improving immune health and mitigating these neurological effects. Additionally, the development of a DL model capable of estimating 'brain age' from neuroimaging data provides a valuable first step towards creating practical and efficient DL tools for assessing the severity of HIV-related neurological changes.

6.2 Limitations

Our study offers valuable insights into the correlation between brain age delta and HIV-associated clinical measures, but it is not without its limitations. Firstly, our analysis did not include all possible influencing factors, for example, other HIV related clinical measures like treatment duration, nor factors such as substance use or mental health conditions which may also impact brain ageing [12, 155]. The challenge of incorporating these factors stems from the limited availability of subjects with comprehensive data and the complexities involved in such extensive analyses.

Additionally, the lack of control groups in some sites limited the extent to which we could conduct case-control comparisons, leading us to rely on methods like mixed-effects models and harmonisation. The registration pre-processing step, involving the alignment of scans to a common template, might have also contributed to the loss of age-predictive information. The necessity for harmonisation, due to the multi-site and multi-scanner nature of the ENIGMA-HIV dataset, introduced additional complexity to the study. Although we investigated the reliability of harmonisation, this process could inadvertently introduce additional variations. A further limitation is the dependence on the CamCAN reference test set for harmonisation, necessitating future studies using our models to access the CamCAN dataset.

Computational constraints also limited further in-depth pre-training and exploration of model architecture variations, particularly regarding depth and width. These limitations also affected the type of harmonisation strategies we could implement. For example, we considered using ComBat-GAM [98], known for effectively accounting for nonlinear covariate effects like age. However, the high computational demands of this approach made it infeasible.

As shown by Kushol et al. [94], in-depth evaluations of scan quality are crucial. Due to the extensive volume of data in our study, individual scan quality assessments, both pre- and post-harmonisation, were not feasible within the time constraints.

Lastly, seemingly healthy individuals included in the study may also feature undetected underlying conditions which could impact their brain health and thus the accuracy at which the model could be trained.

6.3 Future Work

Considering the limitations identified in this study, future research should include the utilisation of minimally pre-processed datasets to preserve informative features potentially lost during pre-processing such as registration. Due to the variability in the initial dimensions of the scans, resulting from the use of different scanners and scanning protocols, registration was necessary. Investigating the effects of registration with fewer degrees of freedom could provide insights into the preservation of critical information. Bias field correction is also worth exploring for removing variation due to scanner effect[1]. Further investigation into the effects of imbalanced datasets on neuroHarmonize output is suggested. Exploring alternative harmonisation methods is also recommended. Approaches like ComBat-GAM, given sufficient computational resources, or the style transfer method proposed by Liu et al. [103], could offer valuable insights. Another direction for future work involves domain shift. While we have aimed to mitigate the effects of domain shift through transfer learning and harmonisation, future work could include extensive verification of whether the impact of domain shift was alleviated[2]. In addition to this, future work should involve further investigations into the efficacy of harmonisation for whole-brain voxel images and their subsequent use in DL models.

While we have aimed to develop a pre-trained model that minimises the MAE and brain age delta, ensuring model generalisability is highly important for accurate performance on the HIV dataset. Thus, an important extension of this work would be to expose the model to a more diverse range of datasets, featuring individuals from varying demographics. For example, there is demographic shift present in the datasets used whereby the UK Biobank and CamCAN training datasets were collected in the UK, while the ENIGMA-HIV dataset is international. This may introduce a bias in the model[3]. The goal would be to train a model on multiple diverse datasets, enhancing its generalisability and reducing the need for harmonisation. Furthermore, having a training dataset of a more similar age distribution to the testing dataset is a point for future work. Alternatively, strategies such as unsupervised pre-training could be explored, potentially elim-

---

[1] Private correspondence

[2] Private correspondence

[3] Private correspondence

inating the need for large training datasets with matched distributions[4]. Using model ensembles, as suggested by Dinsdale et al. [17], could also lead to enhanced accuracy. Moreover, implementing SOTA models such as the SFCN introduced by Peng et al. [15], and applying the ENIGMA-HIV testing dataset could provide a comparative analysis of results. Longitudinal studies could also offer further understanding, correlating the progression of HIV with clinical measures. Exploring additional analyses, such as using continuous values for the CD4 and nadir CD4 counts in HIV+ individuals instead of discrete bins, would be interesting.

As discussed in Section 6.2, time constraints made in-depth scan inspection infeasible, thus future studies could include such evaluations.

Further exploration could also involve the implementation of techniques like saliency mapping to better understand the aspects of brain scans that the network prioritises in ageing analysis, such as that done by Taylor et al. [128]. This would be particularly informative in the context of investigating brain ageing in individuals with HIV.

Finally, the application of transformers, known for their success in various fields, to brain age prediction in the context of HIV, is a promising avenue for exploration.

6.4 Concluding Remarks

In summary, this thesis contributes an investigation of brain ageing in HIV+ individuals, utilising CNNs and transfer learning. The development of a robust model, along with harmonisation, allowed for application to smaller, diverse datasets like ENIGMA-HIV and proved valuable in understanding HIV's impact on the brain. We gained insights into the roles of current CD4 and nadir CD4 counts in the brain ageing process among HIV+ individuals, highlighting the importance of regular monitoring of these measures and effective treatment. While there is future work to be done, this thesis contributes to both the fields of brain age prediction and HIV-related neurological research, being the first work to apply a CNN to the investigation of brain ageing in HIV+ individuals, and thus aiding in laying down a foundation for the development of further DL clinical tools and strategies.

---

[4] Private correspondence

Appendix A

Dataset Distribution Plots

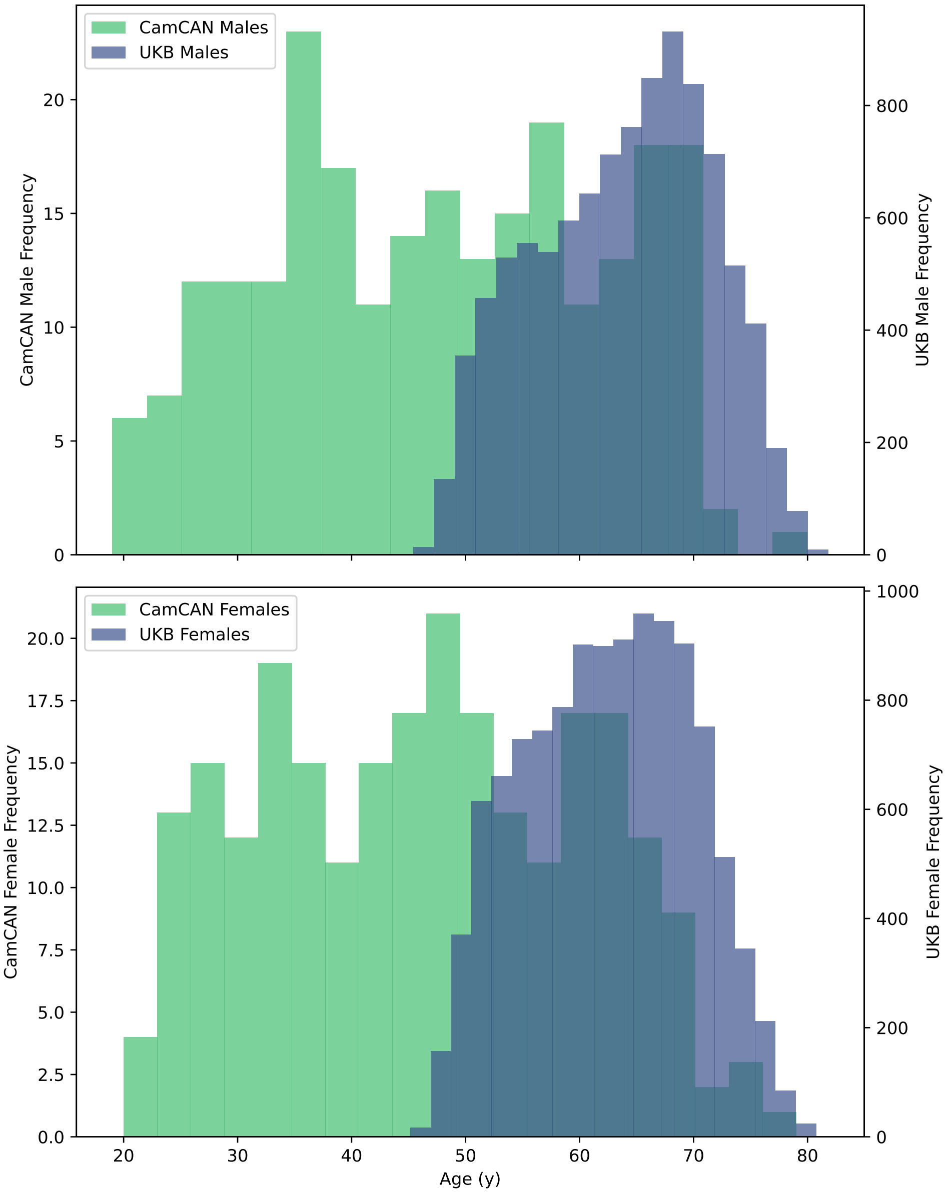


Figure A.1: Comparison of sex distributions of the UK Biobank and CamCAN training datasets.

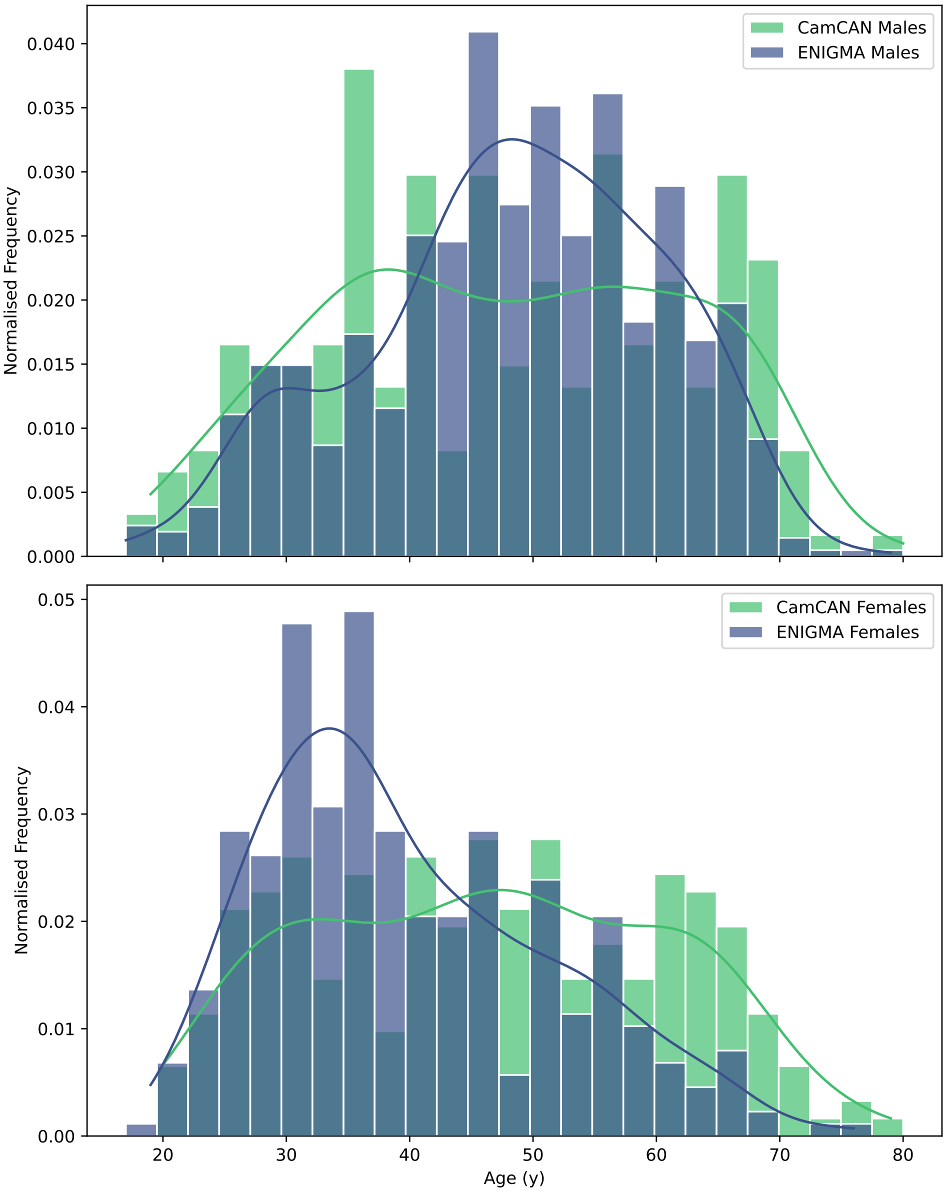


Figure A.2: Comparison plot of the normalised sex distributions of the CamCAN dataset and the ENIGMA-HIV testing dataset.

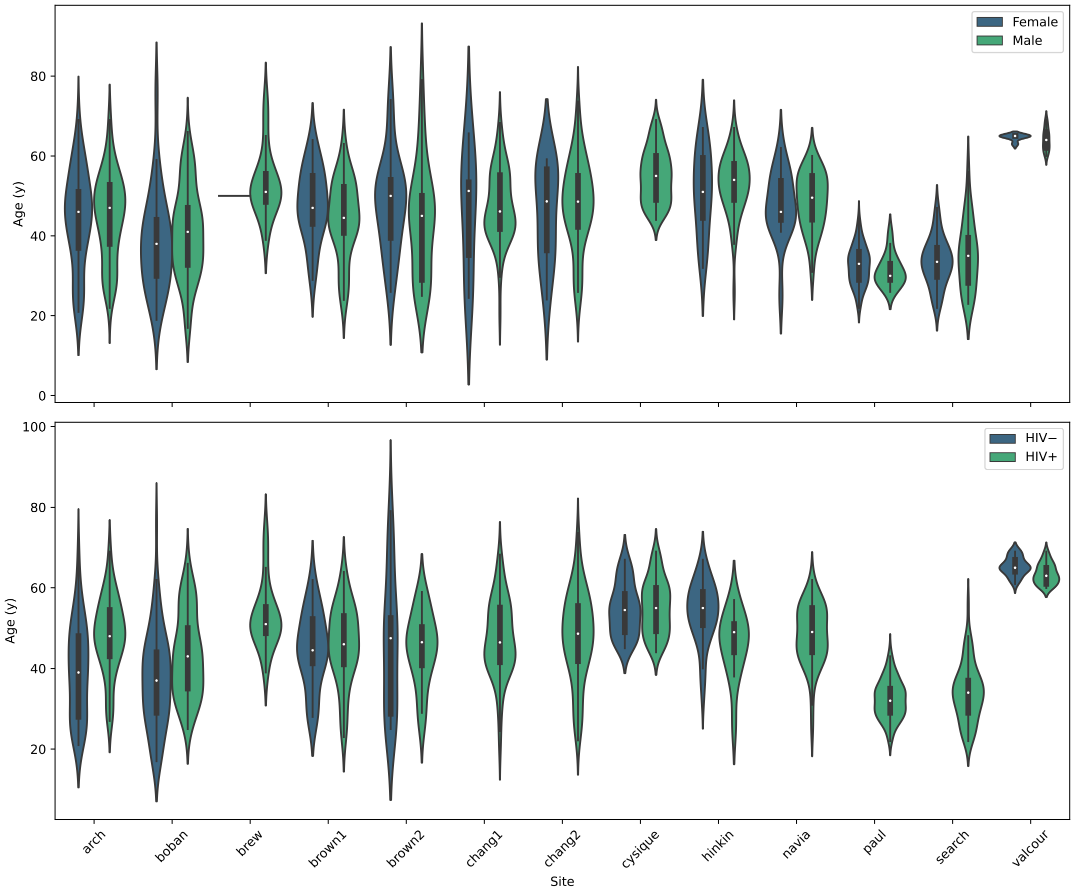
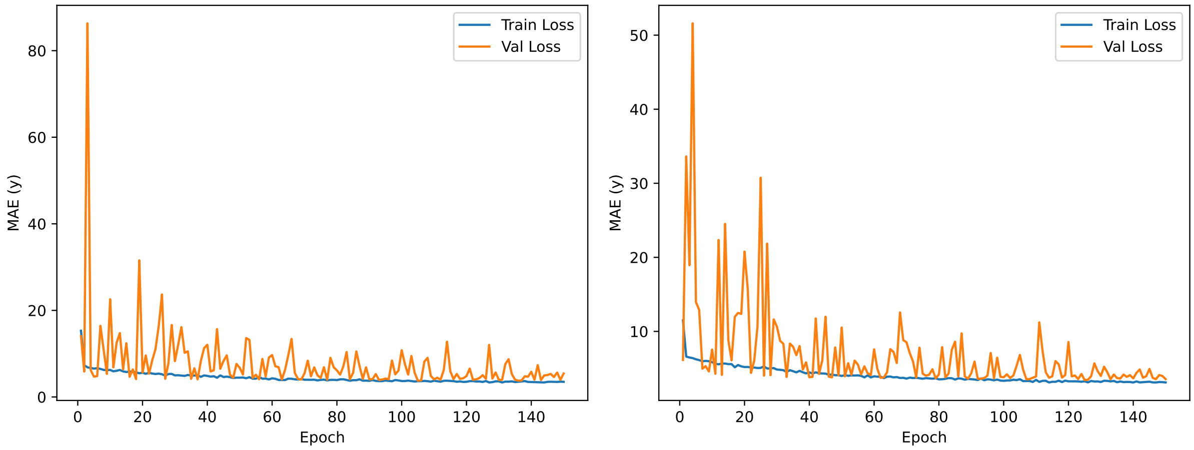


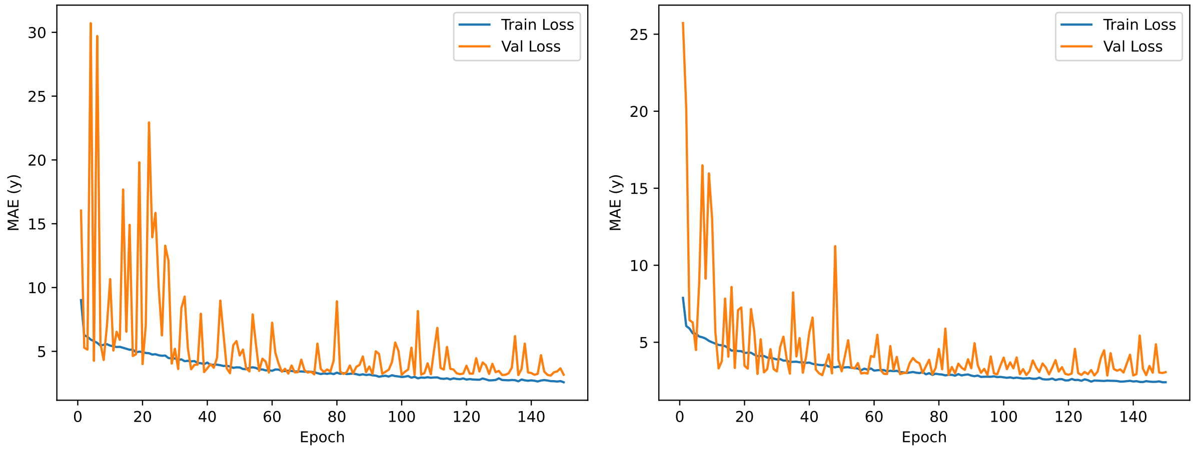
Figure A.3: Age distributions of scans for the sites in the ENIGMA-HIV dataset with sex (top) and HIV status (bottom) splits. Note that the sites Brown and Chang have been split into two sub-sites to account for the two protocols used within each site.

Appendix B

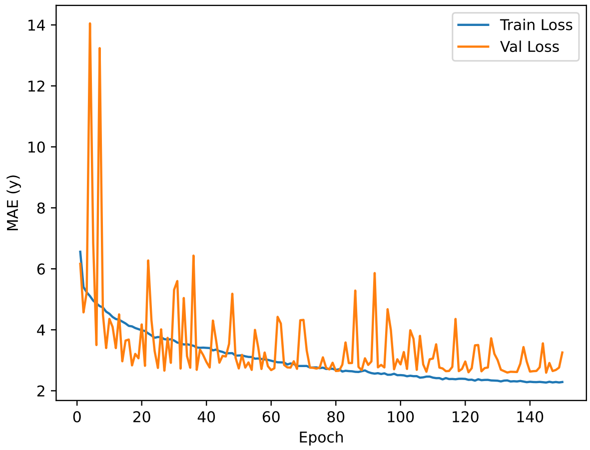
Training Dataset Size Effects Plots



(a) Scan bin maximum 50. Dataset size N=1690. (b) Scan bin maximum 100. Dataset size N=3277.

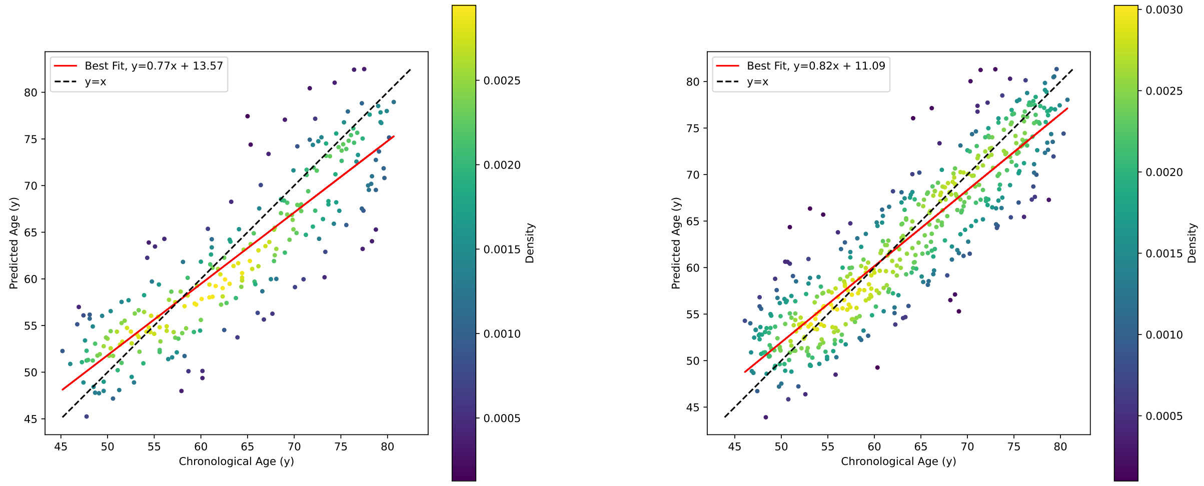


(c) Scan bin maximum 200. Dataset size N=6245. (d) Scan bin maximum 400. Dataset size N=11678.



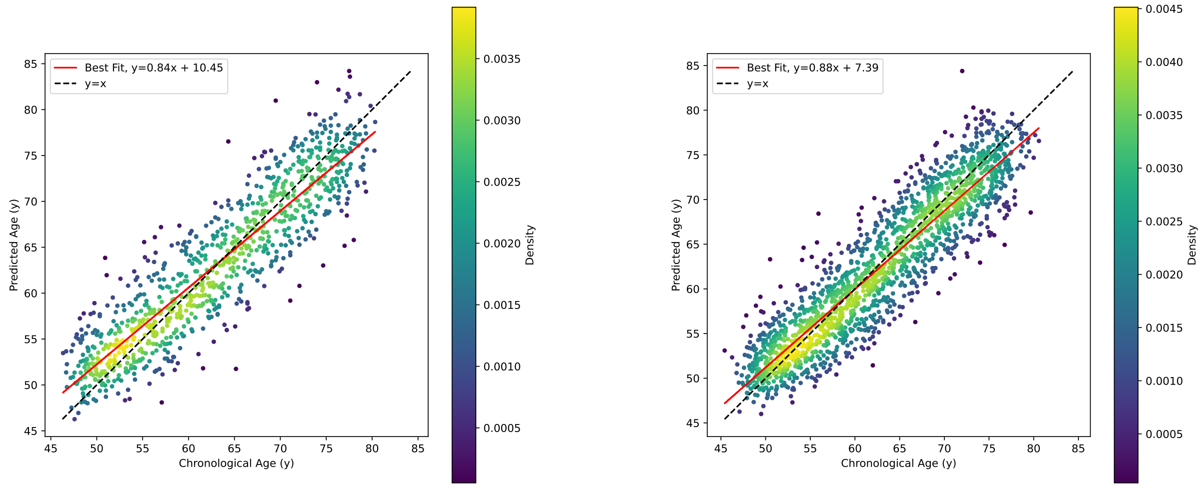
(e) Full dataset. Dataset size N=21366.

Figure B.1: Training and validation loss curves for different initial UK Biobank dataset sizes.



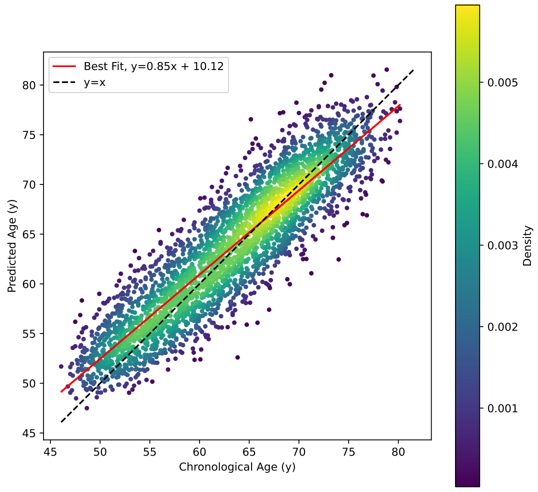
(a) Scan bin maximum 50. Dataset size N=1690.

(b) Scan bin maximum 100. Dataset size N=3277.



(c) Scan bin maximum 200. Dataset size N=6245.

(d) Scan bin maximum 400. Dataset size N=11678.



(e) Full dataset. Dataset size N=21366.

Figure B.2: Chronological and predicted age comparison plots for different initial UK Biobank dataset sizes.

Appendix C

Testing Dataset Clinical Variable Plots

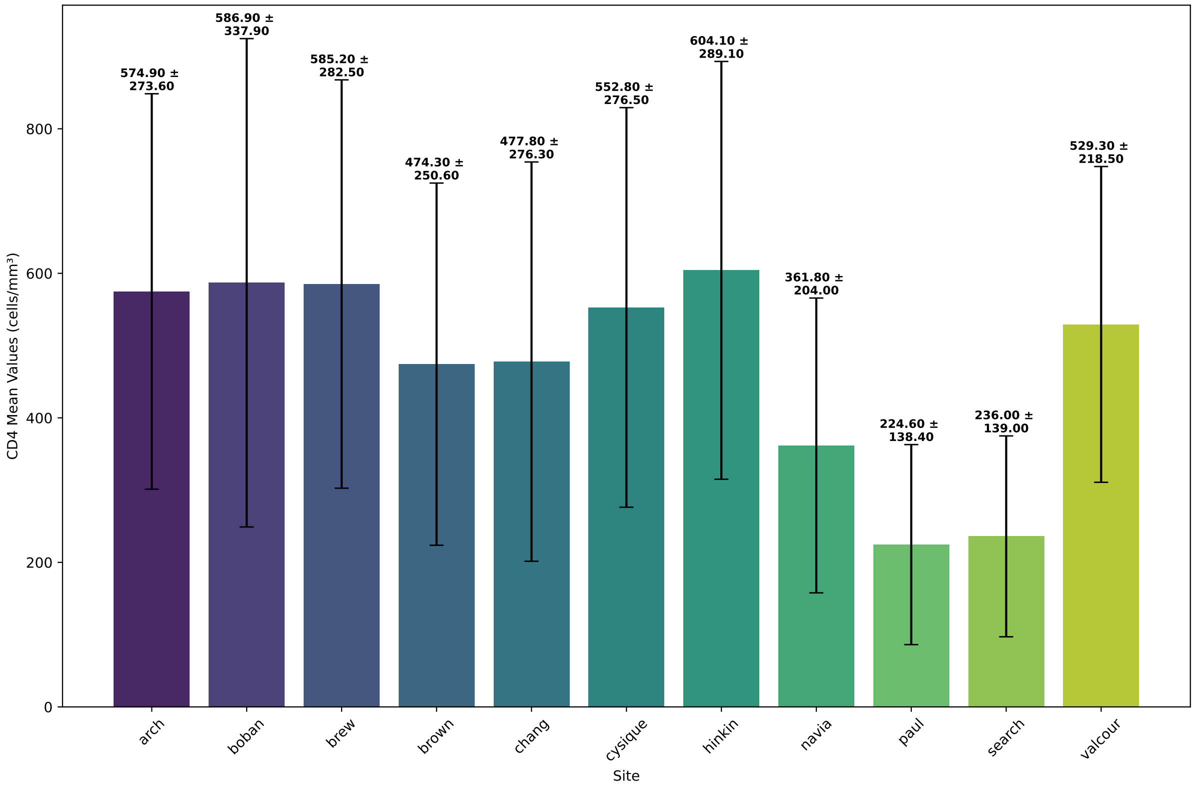


Figure C.1: Plot of the mean values of CD4 cell counts across different sites.

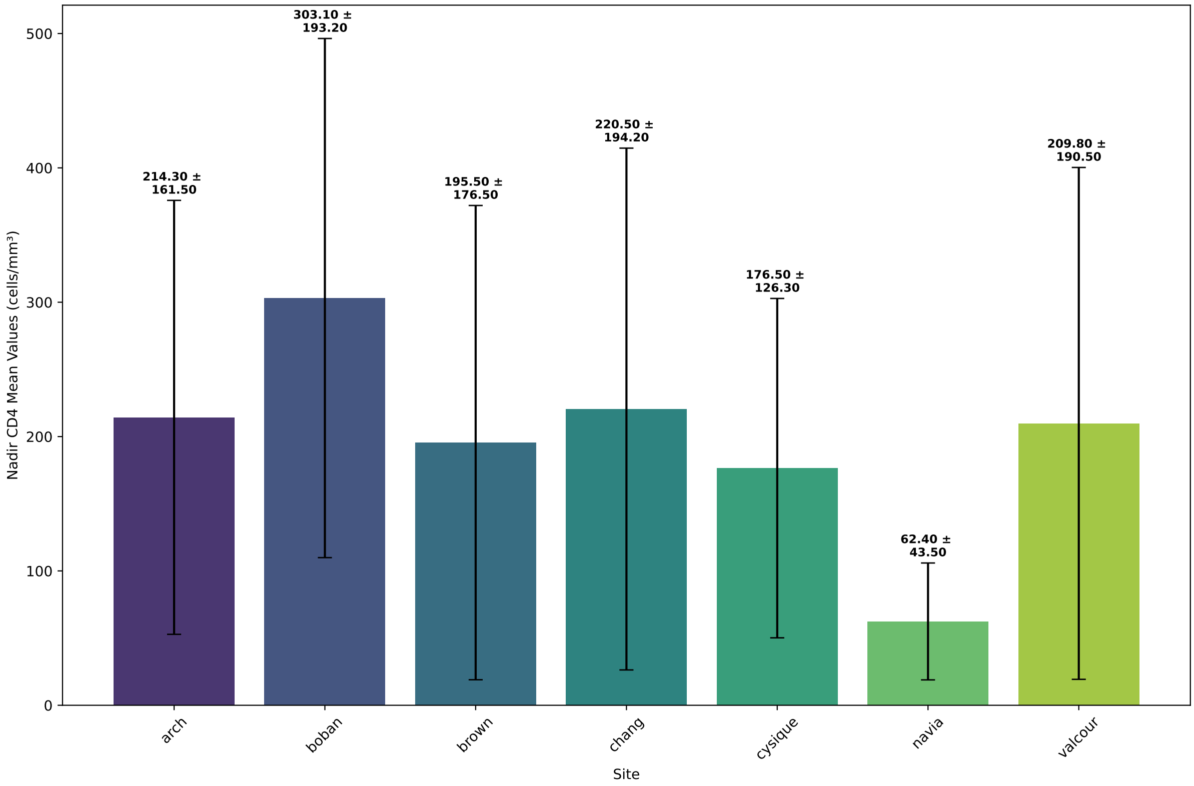


Figure C.2: Plot of the mean values of nadir CD4 cell counts across various sites.

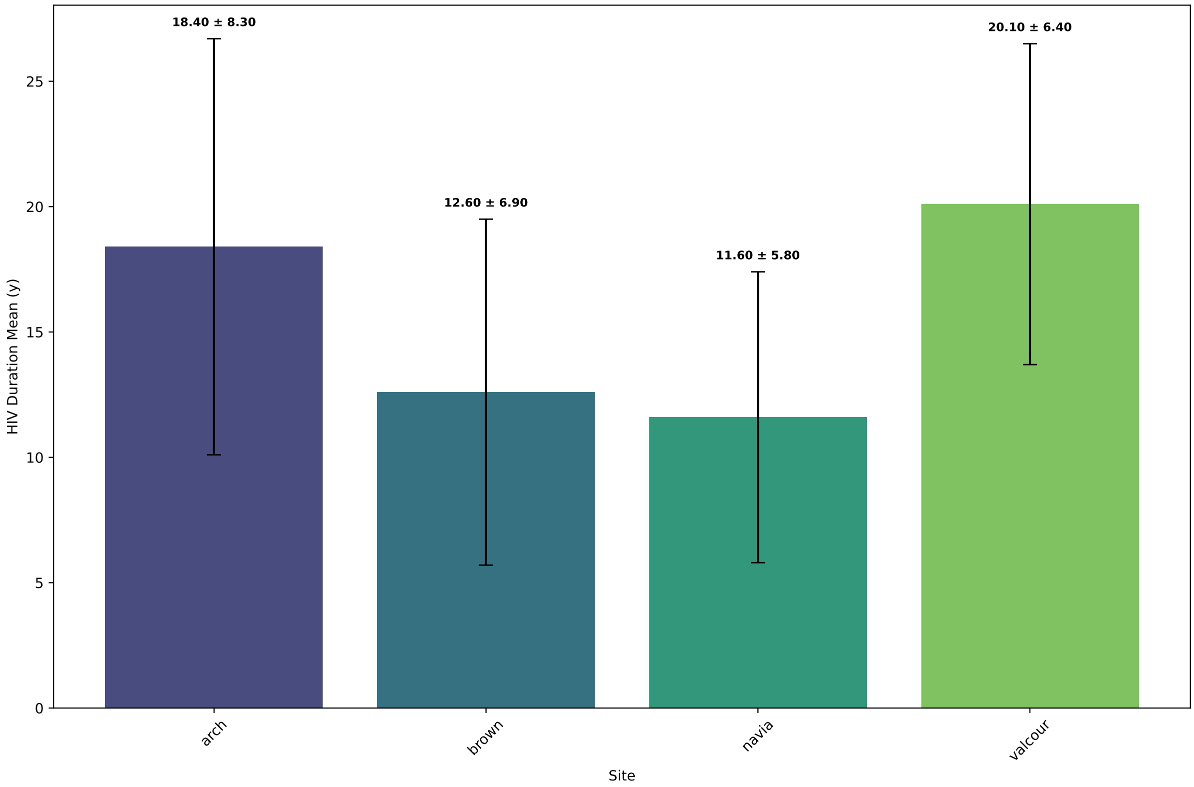


Figure C.3: Plot of mean duration of HIV across different sites.

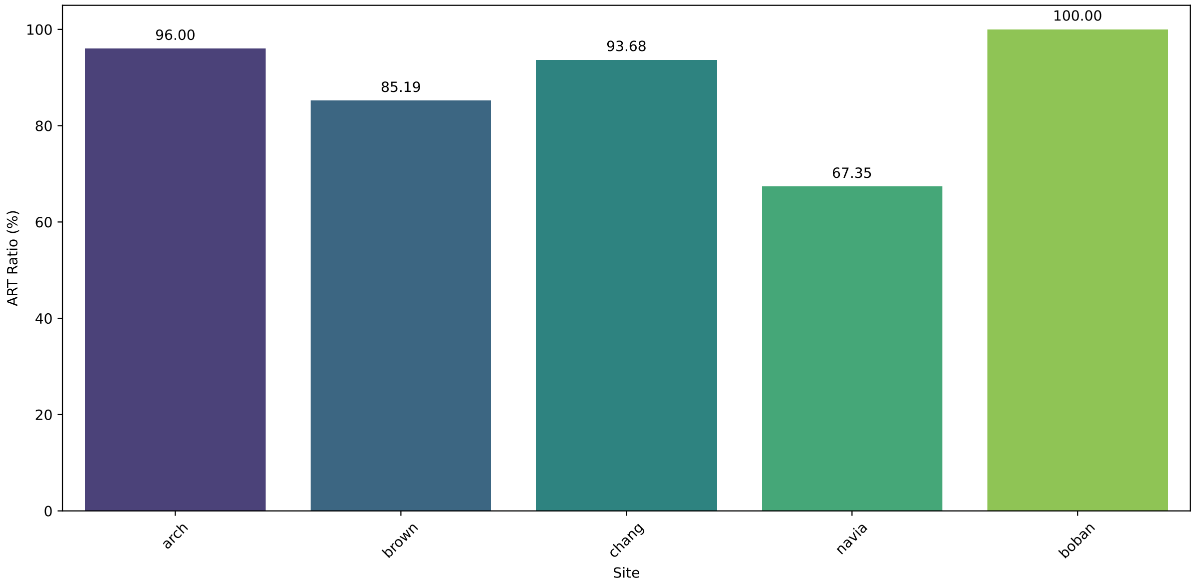


Figure C.4: Plot of the ART ratio percentages across sites.

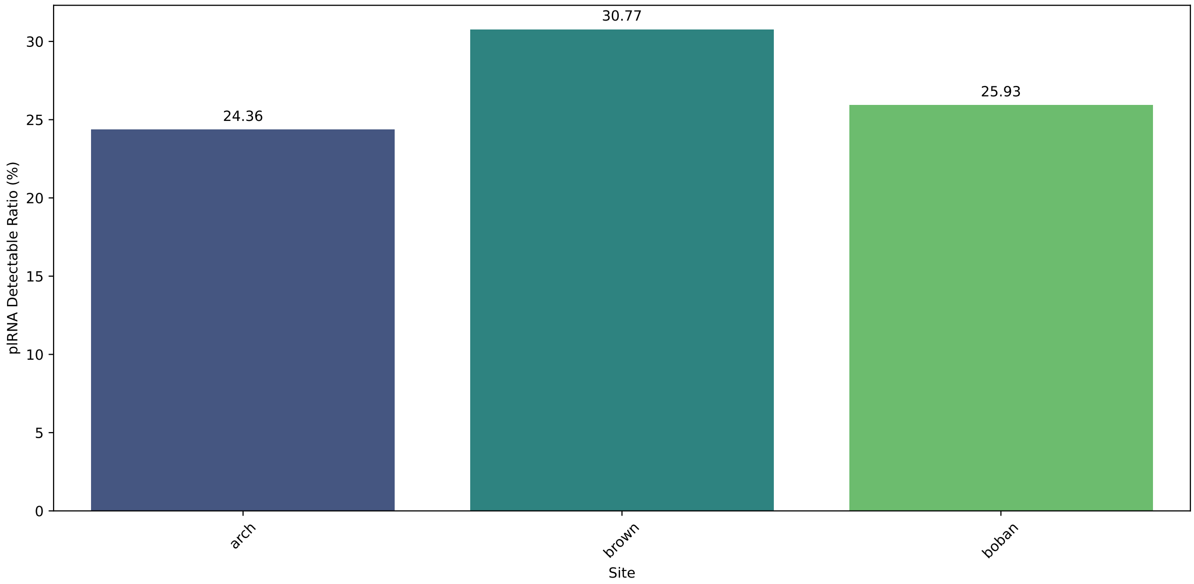


Figure C.5: Plot of the HIV plasma RNA detectable ratios across sites.

Appendix D

Pre-Harmonisation Fixed-Effects Model Results

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	10.0732	1.661	6.065	0.000	6.792	13.354
HIV	-2.9781	0.949	-3.139	0.002	-4.852	-1.104
SEX	-1.5933	1.184	-1.346	0.180	-3.933	0.746
AGE	-0.1615	0.040	-4.031	**0.000**	-0.241	-0.082

Table D.1: Boban cohort statistical results from fixed-effects model for predictor variables age, sex and HIV.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	3.2775	2.102	1.559	0.121	-0.881	7.436
HIV	0.1999	0.972	0.206	0.837	-1.723	2.123
SEX	0.1277	1.003	0.127	0.899	-1.856	2.112
AGE	-0.2036	0.044	-4.646	**0.000**	-0.290	-0.117

Table D.2: Brown cohort statistical results from fixed-effects model for predictor variables age, sex and HIV.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	-5.2841	3.810	-1.387	0.168	-12.828	2.259
HIV	0.4226	0.961	0.440	0.661	-1.480	2.325
AGE	-0.1882	0.068	-2.755	**0.007**	-0.324	-0.053

Table D.3: Cysique cohort statistical results from fixed-effects model for predictor variables age, sex and HIV.

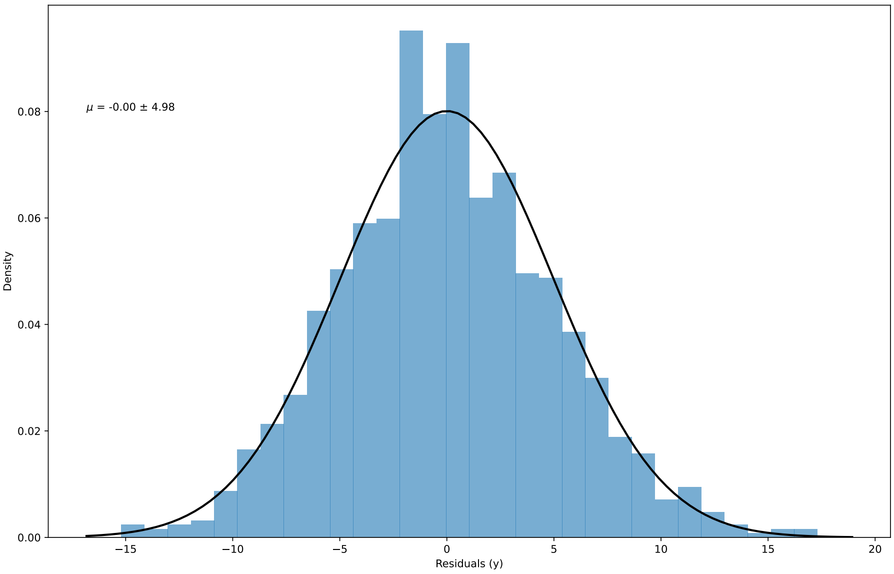


Figure D.1: Histogram of residuals from mixed-effects model, for predictor variables HIV, sex and age with site (split by scanner protocol) as the random effect, with normal distribution fitted.

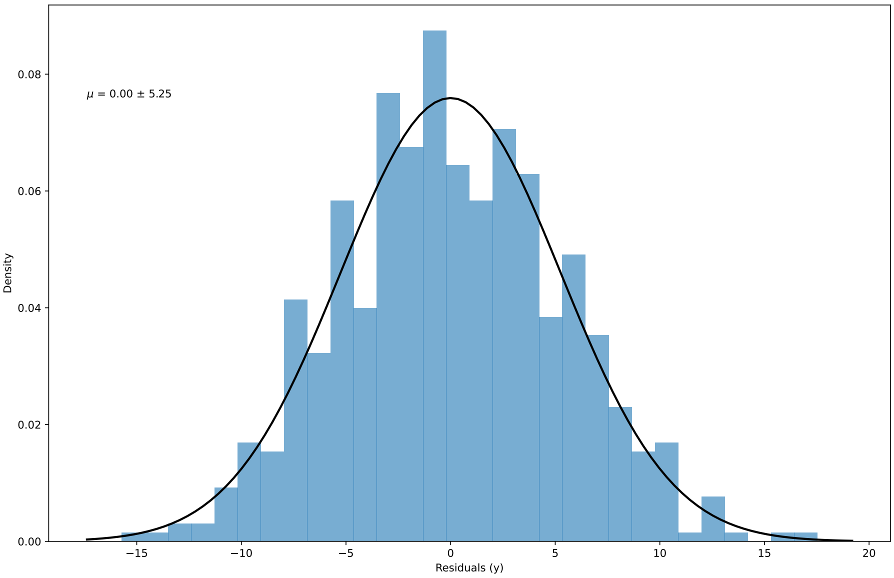


Figure D.2: Histogram of residuals from mixed-effects model with predictor variables age, sex, CD4 bin and nadir CD4 bin, with normal distribution fitted.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	0.1749	4.003	0.044	0.965	-7.830	8.180
HIV	-1.1064	1.528	-0.724	0.472	-4.163	1.950
SEX	-2.0516	1.317	-1.558	0.124	-4.684	0.581
AGE	-0.1963	0.071	-2.759	**0.008**	-0.339	-0.054

Table D.4: Hinkin cohort statistical results from fixed-effects model for predictor variables age, sex and HIV.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	8.9267	13.278	0.672	0.503	-17.502	35.355
HIV	-1.4105	1.093	-1.291	0.200	-3.585	0.764
SEX	2.2169	1.901	1.166	0.247	-1.567	6.001
AGE	-0.2390	0.203	-1.179	0.242	-0.643	0.165

Table D.5: Valcour cohort statistical results from fixed-effects model for predictor variables age, sex and HIV.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	12.9364	2.038	6.349	0.000	8.903	16.969
SEX=2	0.7243	1.047	0.692	0.490	-1.347	2.796
CD4 Bin=1	1.8203	1.260	1.445	0.151	-0.673	4.314
CD4 Bin=2	2.8219	1.389	2.032	0.044	0.073	5.571
CD4 Bin=3	3.7503	2.640	1.420	0.158	-1.475	8.976
AGE	-0.2379	0.047	-5.052	**0.000**	-0.331	-0.145

Table D.6: Arch cohort statistical results from fixed-effects model for predictor variables age, sex and CD4 bin.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	16.6506	2.591	6.427	0.000	11.529	21.772
SEX=2	-3.1835	0.997	-3.194	**0.002**	-5.154	-1.213
CD4 Bin=2	-1.7923	1.581	-1.133	0.259	-4.918	1.334
CD4 Bin=3	-0.3602	1.586	-0.227	0.821	-3.495	2.775
AGE	-0.4969	0.069	-7.157	**0.000**	-0.634	-0.360

Table D.7: Paul cohort statistical results from fixed-effects model for predictor variables age, sex and CD4 bin.

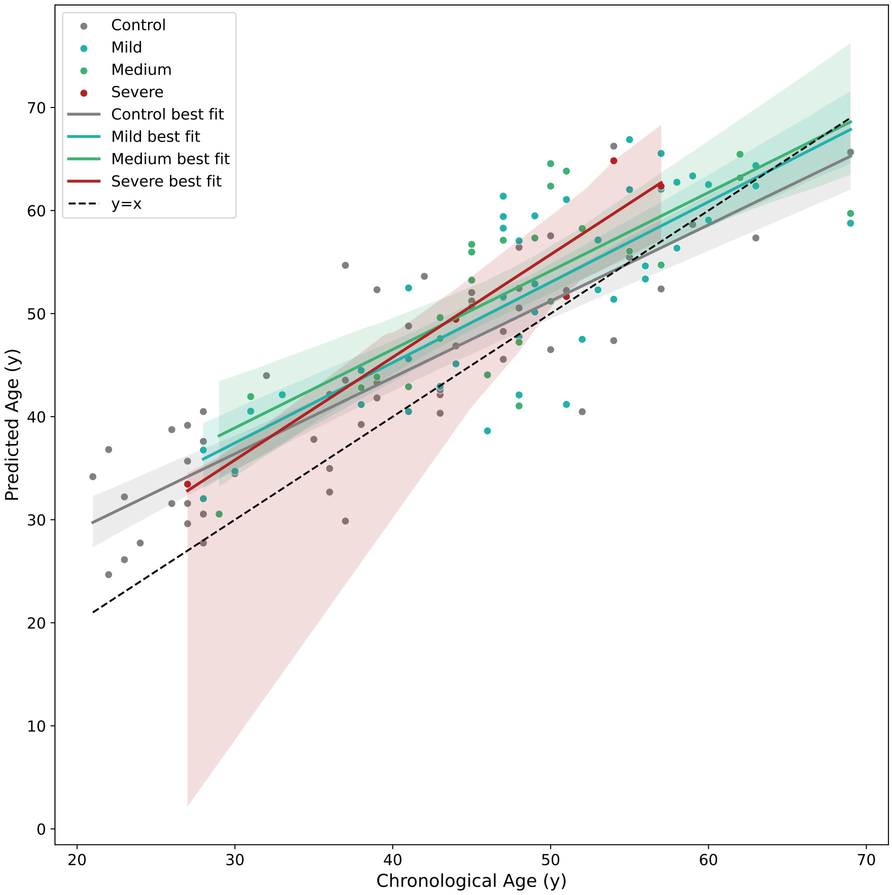


Figure D.3: Arch cohort plot of predicted and chronological age stratified by CD4 bin.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	13.0595	2.026	6.447	0.000	9.051	17.068
SEX=2	0.6819	1.054	0.647	0.519	-1.404	2.767
nadCD4 Bin=1	1.8624	1.308	1.424	0.157	-0.726	4.451
nadCD4 Bin=2	2.7061	1.253	2.160	0.033	0.226	5.186
AGE	-0.2405	0.047	-5.143	**0.000**	-0.333	-0.148

Table D.8: Arch cohort statistical results from fixed-effects model for predictor variables age, sex and nadir CD4 bin.

Variable	Coefficient	Std. Error	t-Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	7.3103	7.390	0.989	0.328	-7.573	22.194
SEX=2	1.0062	1.681	0.598	0.553	-2.380	4.392
nadCD4 Bin=2	6.0542	5.679	1.066	0.292	-5.383	17.492
AGE	-0.3191	0.107	-2.994	**0.004**	-0.534	-0.104

Table D.9: Navia cohort statistical results from fixed-effects model for predictor variables age, sex and nadir CD4 bin.

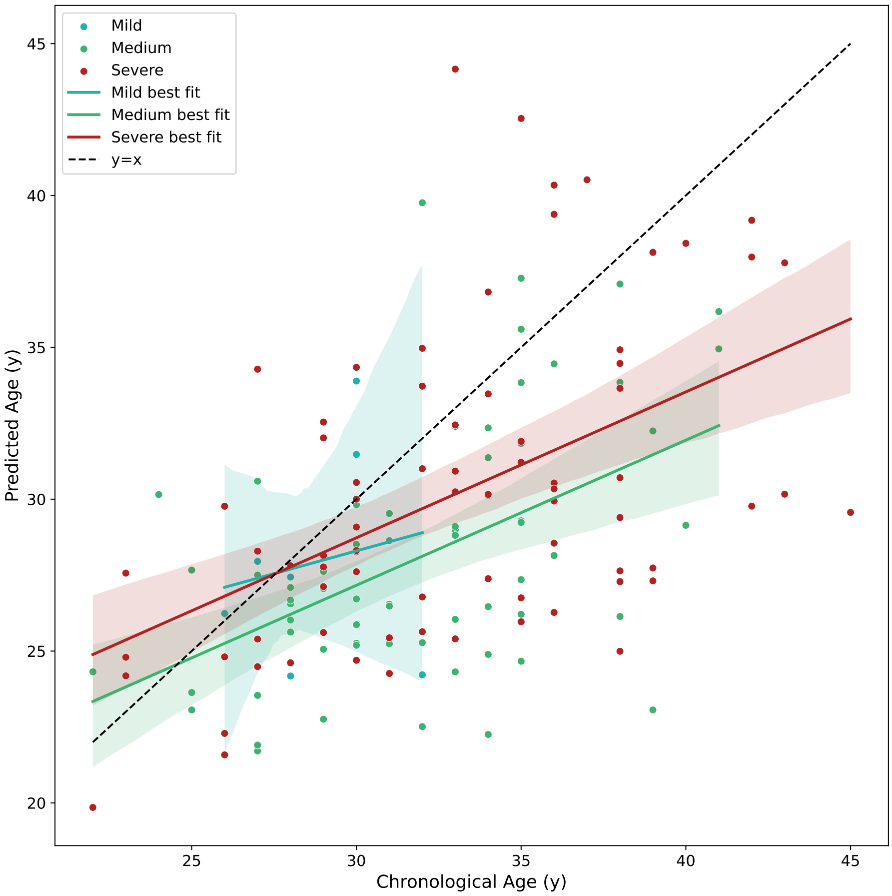


Figure D.4: Paul cohort plot of predicted and chronological age stratified by CD4 bin.

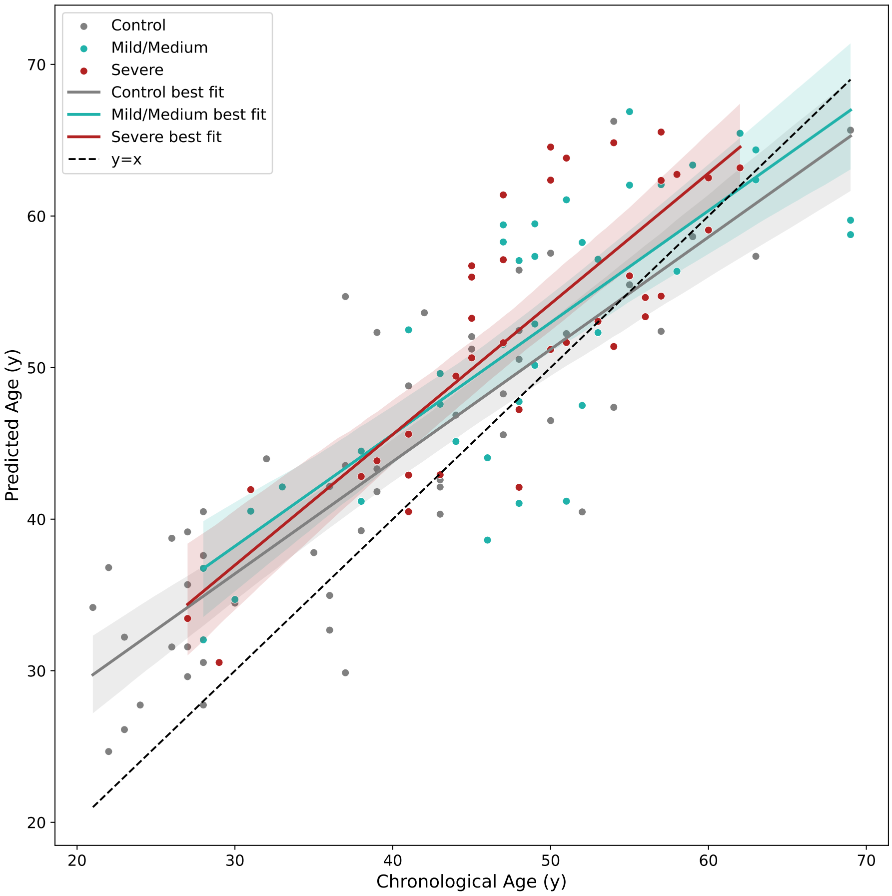


Figure D.5: Arch cohort plot of predicted and chronological age stratified by nadir CD4 bin.

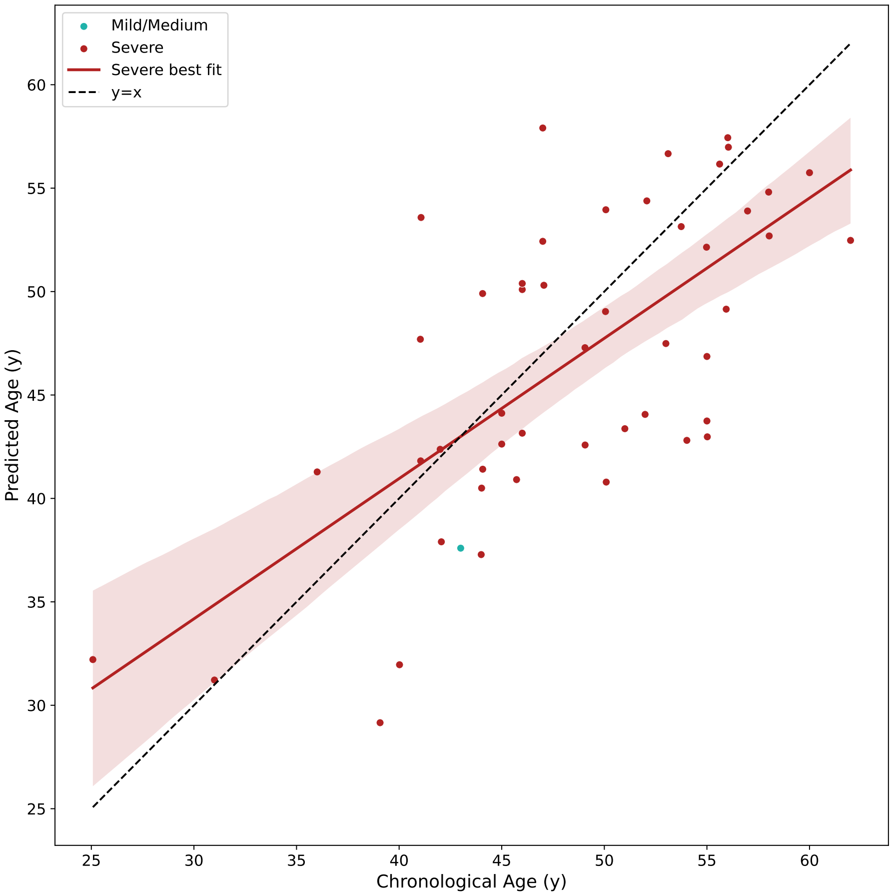


Figure D.6: Navia cohort plot of predicted and chronological age stratified by nadir CD4 bin.

Appendix E

Harmonisation Intensity Histograms

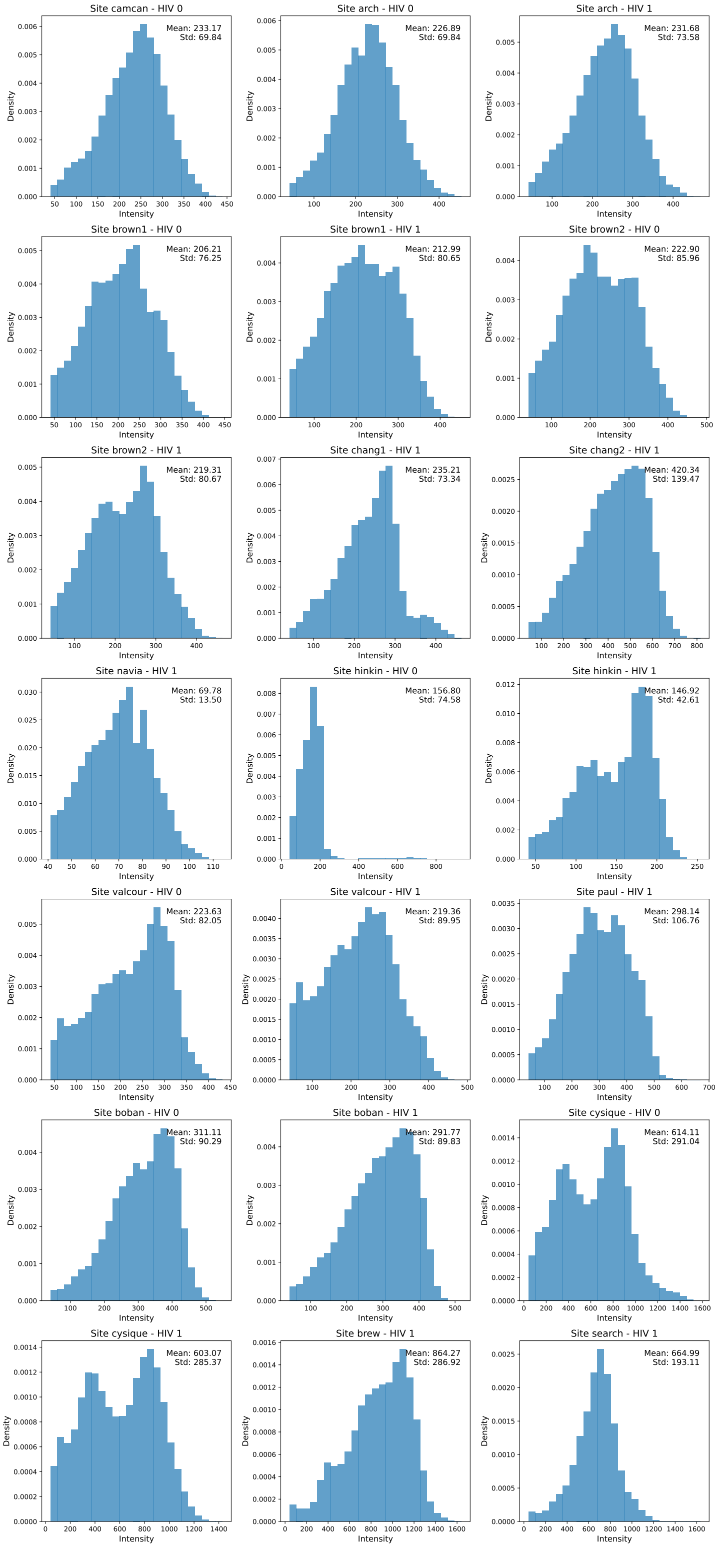


Figure E.1: Average intensity distribution of sampled axial slice (100) pre-harmonisation. Intensity values below a threshold of 40 are not displayed to avoid the high frequency of low intensity background voxels distorting the distribution.

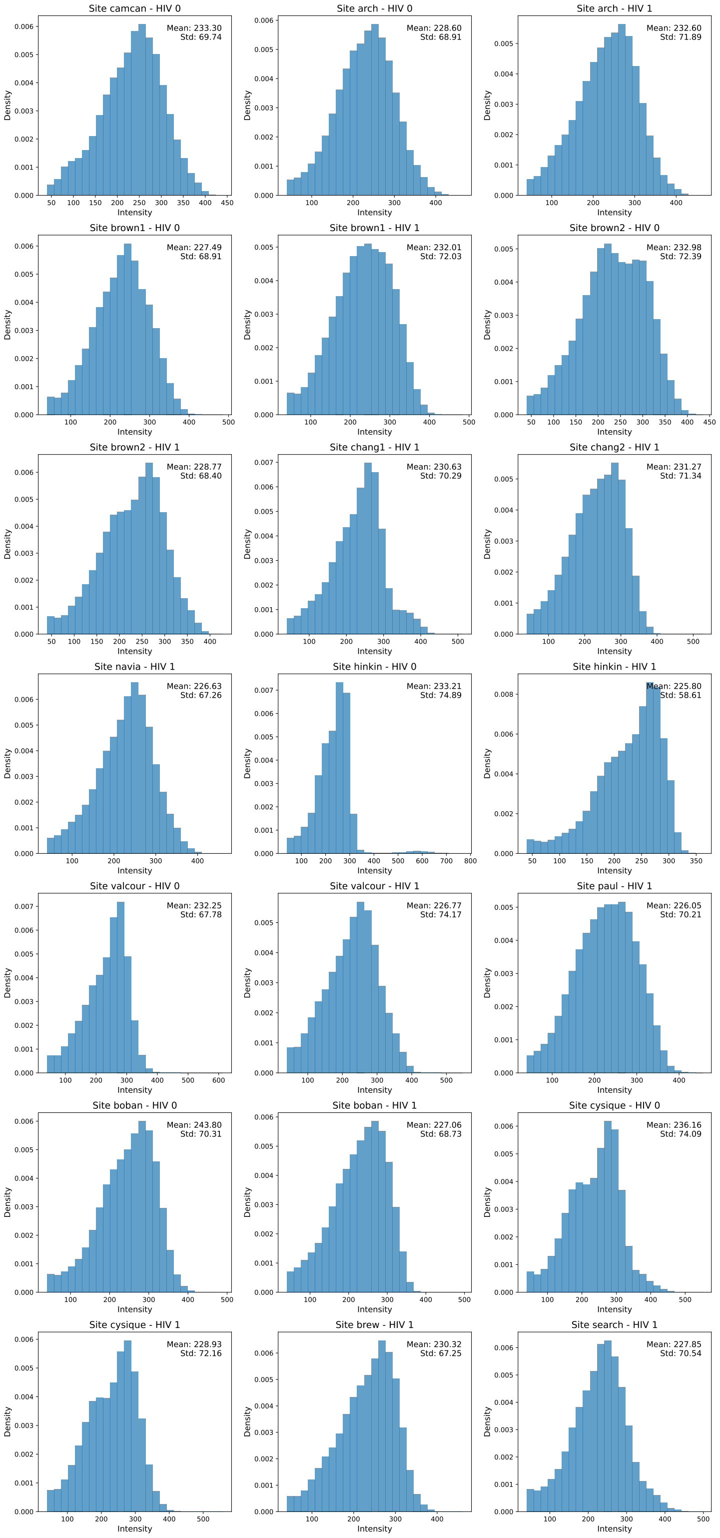


Figure E.2: Average intensity distribution of sampled axial slice (100) post-harmonisation. Intensity values below a threshold of 40 are not displayed to avoid the high frequency of low intensity background voxels distorting the distribution.

Appendix F

Post-Harmonisation Comparison Plots and Results

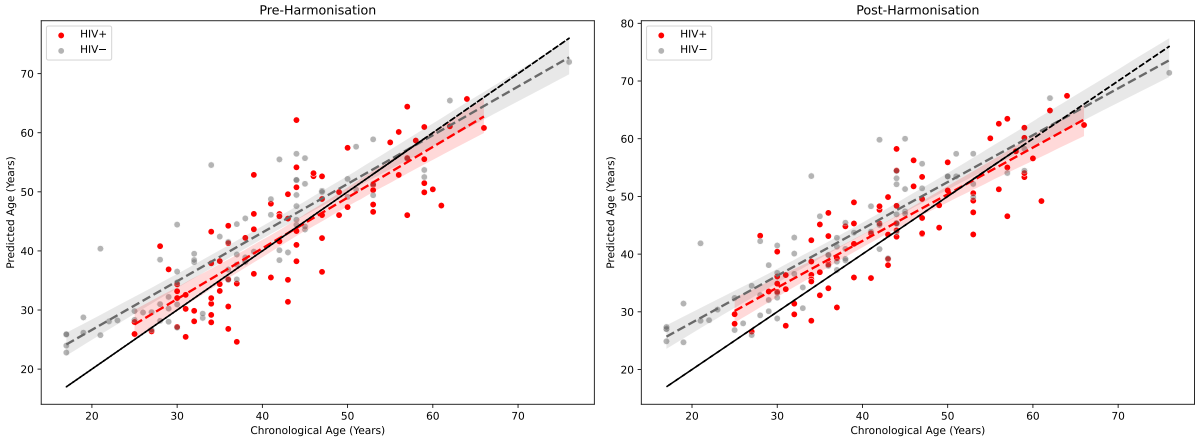


Figure F.1: Boban cohort comparison between predicted and chronological age of cases and controls, pre- and post-harmonisation.

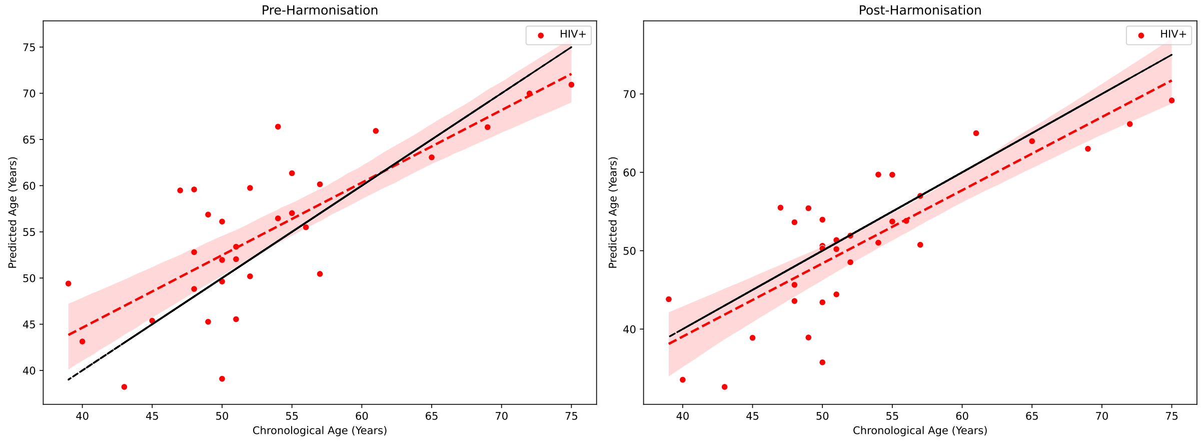


Figure F.2: Brew cohort comparison between predicted and chronological age, pre- and post-harmonisation.

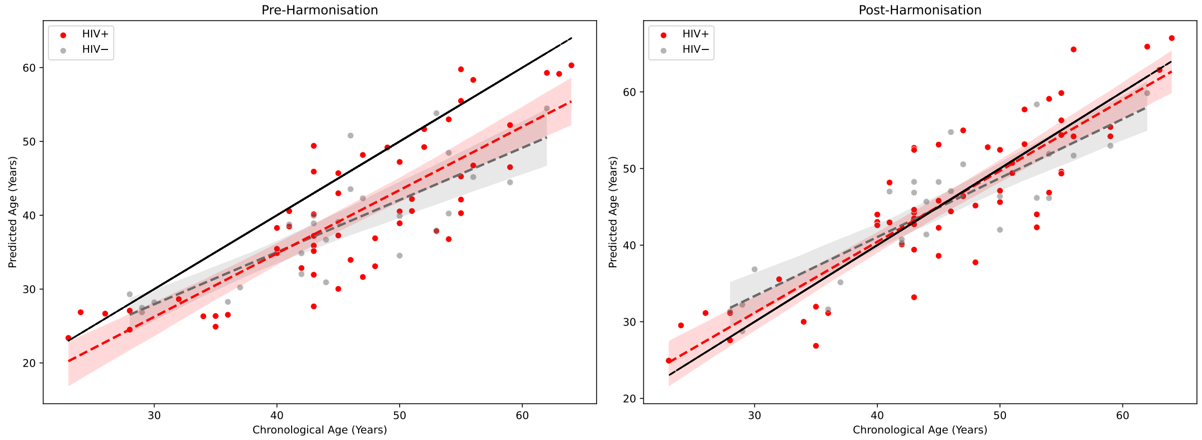


Figure F.3: Brown 1 cohort comparison between predicted and chronological age of cases and controls, pre- and post-harmonisation.

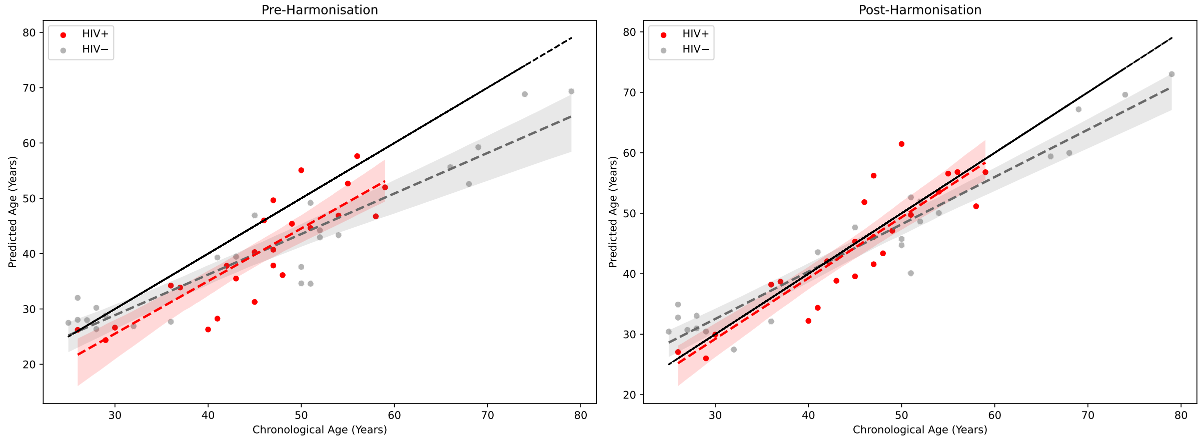


Figure F.4: Brown 2 cohort comparison between predicted and chronological age of cases and controls, pre- and post-harmonisation.

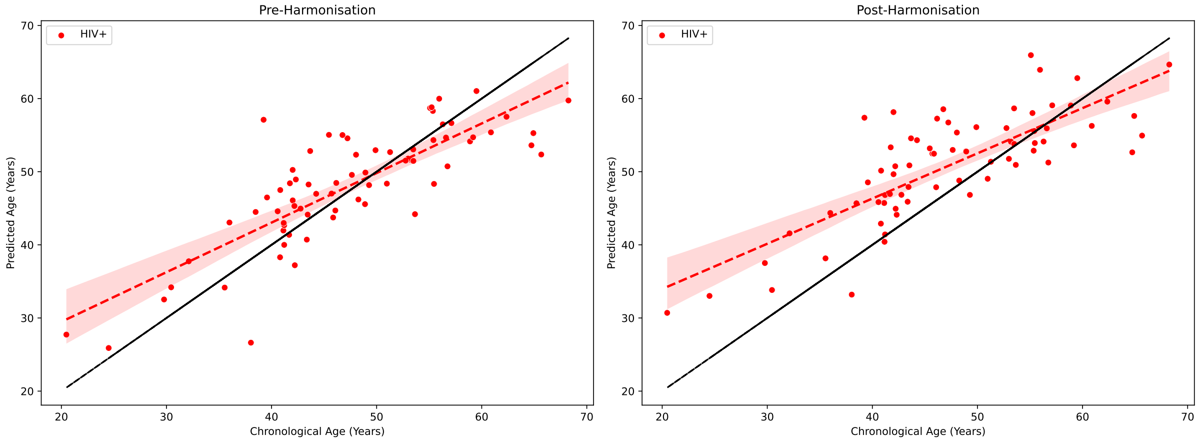


Figure F.5: Chang 1 cohort comparison between predicted and chronological age, pre- and post-harmonisation.

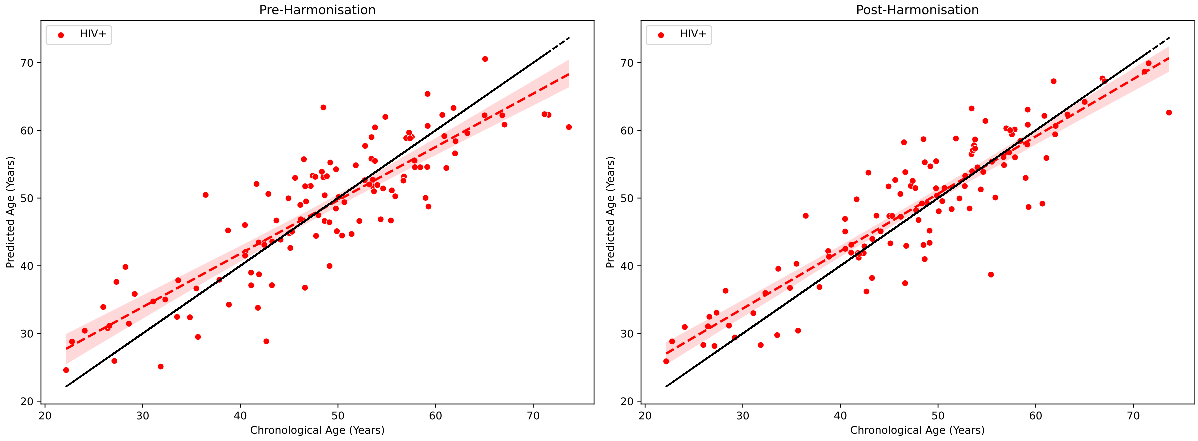


Figure F.6: Chang 2 cohort comparison between predicted and chronological age, pre- and post-harmonisation.

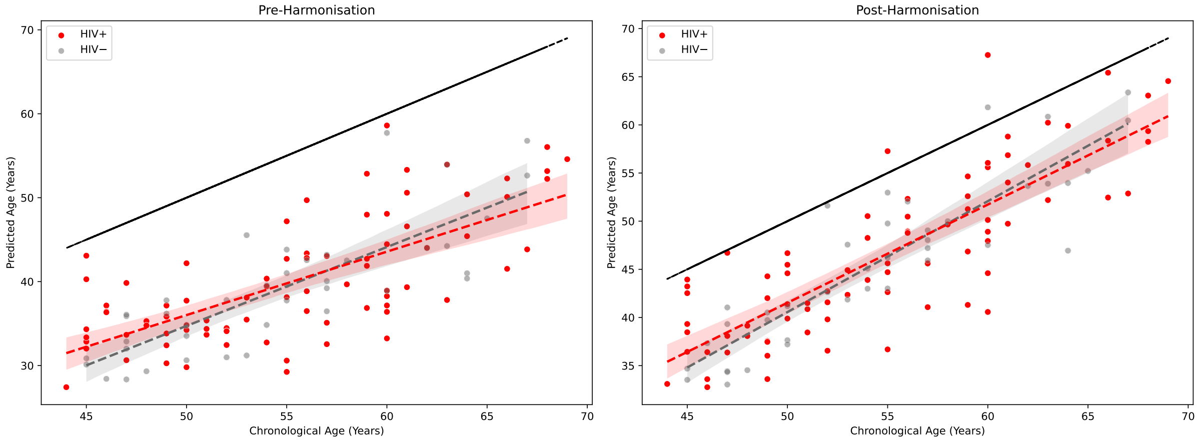


Figure F.7: Cysique cohort comparison between predicted and chronological age of cases and controls, pre- and post-harmonisation.

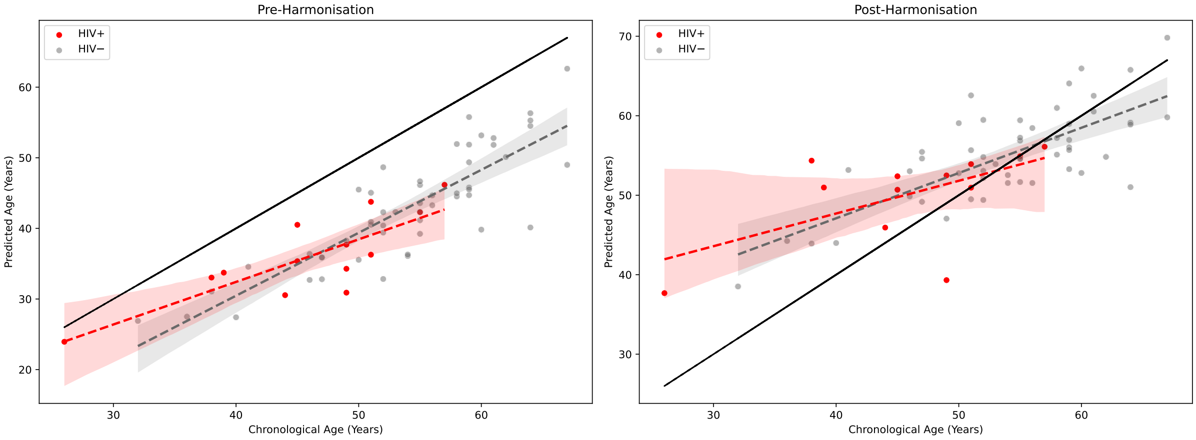


Figure F.8: Hinkin cohort comparison between predicted and chronological age of cases and controls, pre- and post-harmonisation.

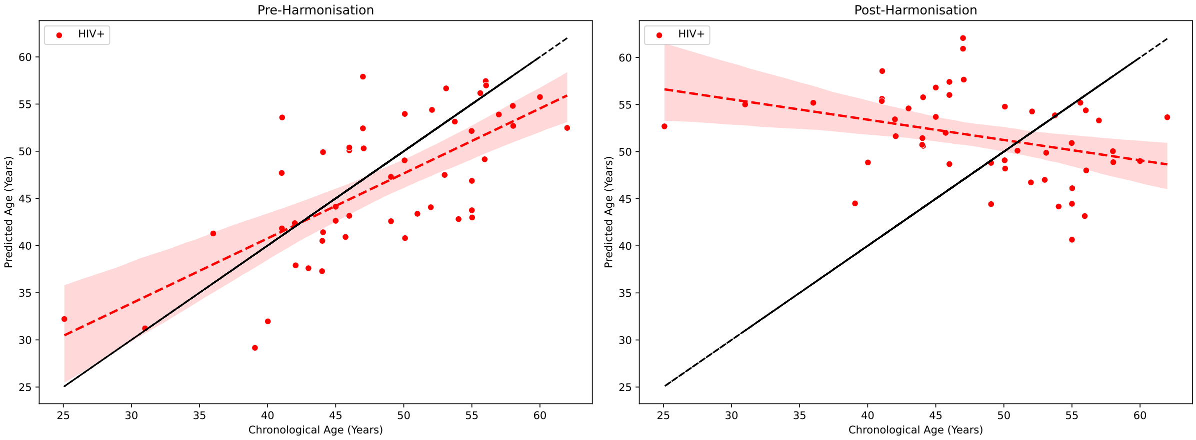


Figure F.9: Navia cohort comparison between predicted and chronological age, pre- and post-harmonisation.

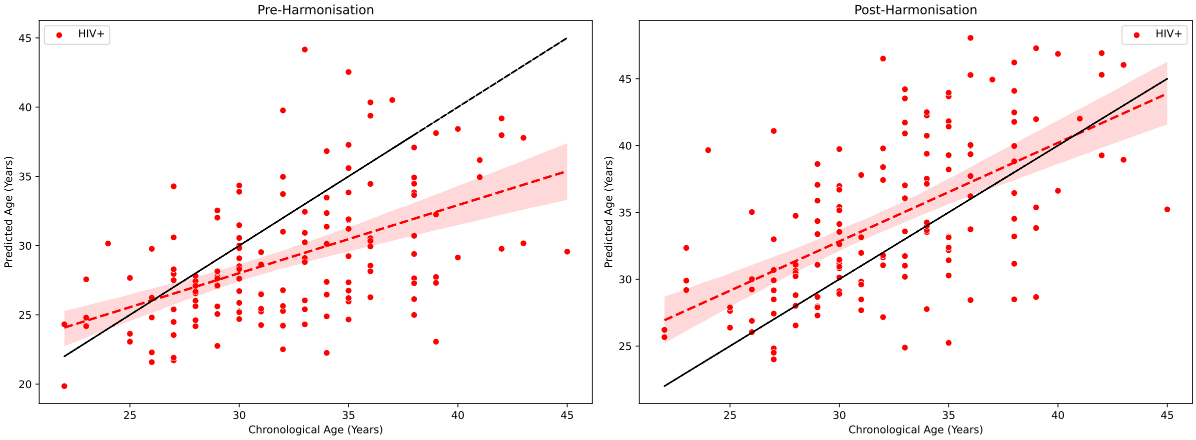


Figure F.10: Paul cohort comparison between predicted and chronological age, pre- and post-harmonisation.

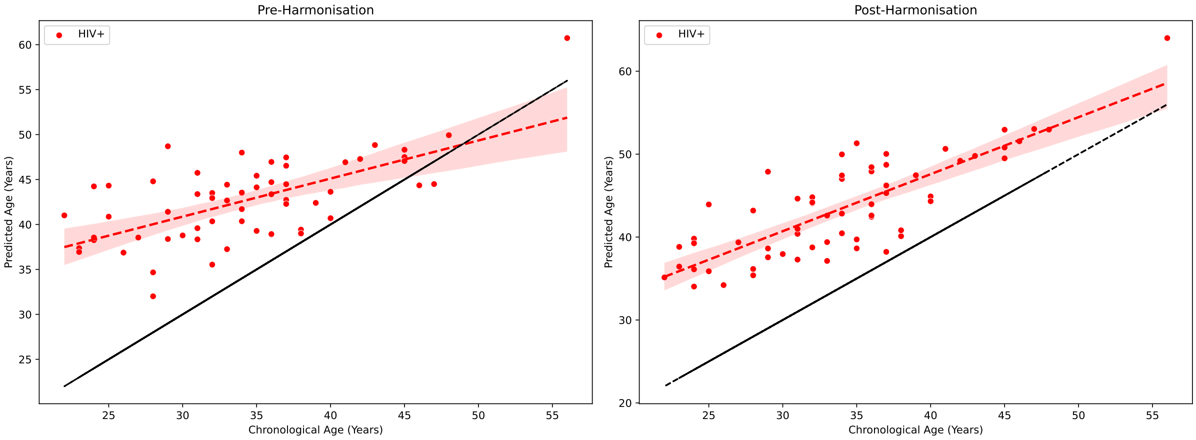


Figure F.11: Search cohort comparison between predicted and chronological age, pre- and post-harmonisation.

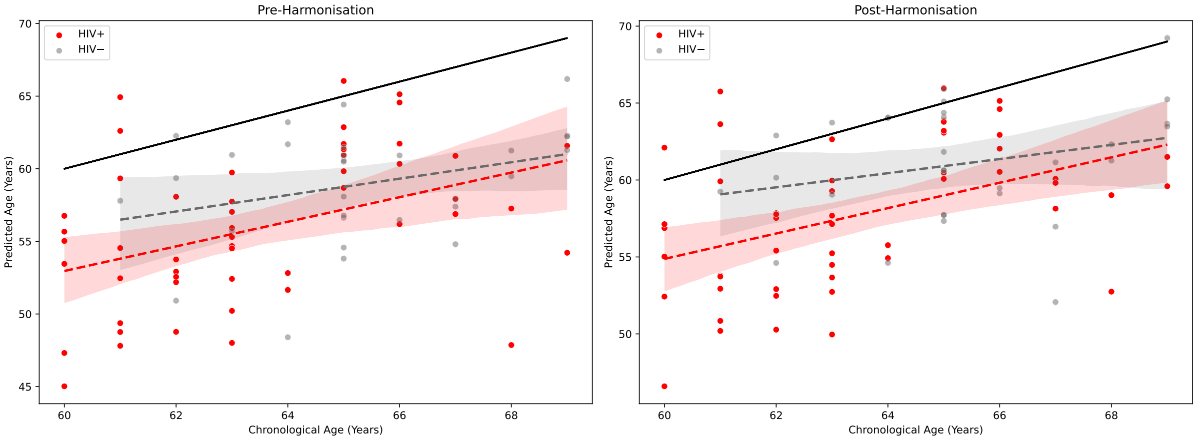


Figure F.12: Valcour cohort comparison between predicted and chronological age of cases and controls, pre- and post-harmonisation.

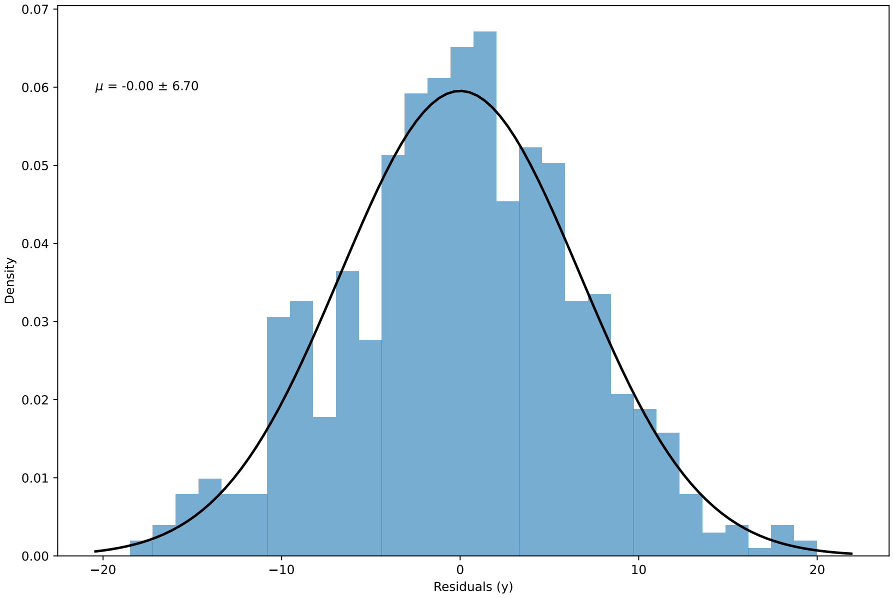


Figure F.13: Histogram of residuals from fixed-effects model with predictor variables age, sex, CD4 and nadir CD4, with normal distribution fitted.

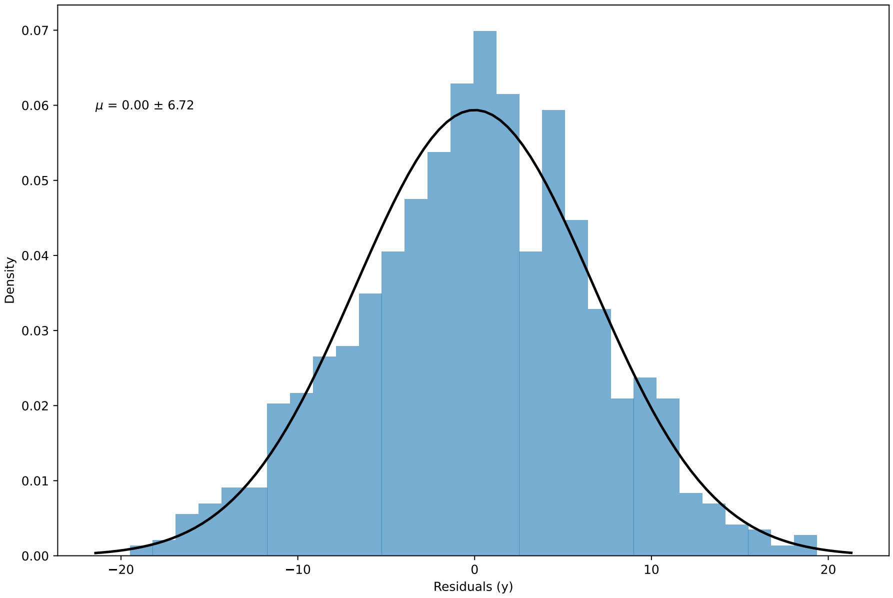


Figure F.14: Histogram of residuals from fixed-effects model with predictor variables age, sex and CD4, with normal distribution fitted.

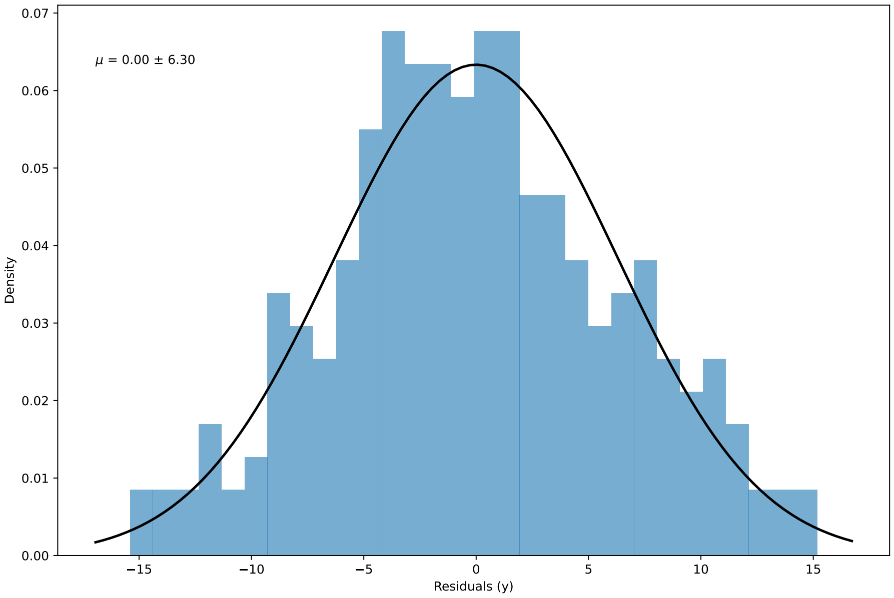


Figure F.15: Histogram of residuals from fixed-effects model with predictor variables age, sex, CD4, nadir CD4 and HIV plasma RNA detectability, with normal distribution fitted.

Bibliography

[1] Fabian Isensee, Marianne Schell, Irada Pflueger, Gianluca Brugnara, David Bonekamp, Ulf Neuberger, Antje Wick, H. P. Schlemmer, Sabine Heiland, Wolfgang Wick, Martin Bendszus, Klaus Maier-Hein, and Philipp Kickingereder. Automated brain extraction of multisequence mri using artificial neural networks. *Human Brain Mapping*, 40:4952 – 4964, 2019. iii, 17, 43, 44

[2] Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy Edward John Behrens, Heidi Johansen-Berg, Peter R. Bannister, M. De Luca, Ivana Drobnjak, David Flitney, Rami K. Niazy, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, Joanne Brady, and Paul M. Matthews. Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage*, 23:S208–S219, 2004. iii, 17, 18, 43

[3] Mark Jenkinson, Peter R. Bannister, Michael Brady, and Stephen M. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17:825–841, 2002. iii, 18, 43

[4] Mark Jenkinson and Stephen M. Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5 2:143–56, 2001.

[5] Douglas N. Greve and Bruce R. Fischl. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48:63–72, 2009. iii, 18, 43

[6] Vladimir S. Fonov, Alan C. Evans, Kelly N. Botteron, C. Robert Almli, Robert C. McKinstry, and D. Louis Collins. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54:313–327, 2011. iii, vii, 16, 17, 18, 43

[7] V. S. Fonov, A. Evans, RC McKinstry, C. Robert Almli, and DL Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, 2009.

[8] D. Louis Collins, Alex P. Zijdenbos, Wim F. C. Baaré, and Alan C. Evans. Animal+insect: Improved cortical structure segmentation. In *IPMI*, 1999. iii, vii, 16, 17, 18, 43

[9] J. Cole, Rudra P. K. Poudel, Dimosthenis Tsagkrasoulis, M. Caan, C. Steves, T. Spector, and G. Montana. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163:115–124, 2017. 1, 15, 17, 25, 26, 33, 34, 37, 43

[10] Katja Franke, Gabriel Ziegler, Stefan Klöppel, and Christian Gaser. Estimating the age of healthy subjects from t_1-weighted mri scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, 50:883–892, 2010. 1

[11] Arinbjörn Kolbeinsson, S. Filippi, Yannis Panagakis, P. Matthews, P. Elliott, A. Dehghan, and I. Tzoulaki. Accelerated MRI-predicted brain ageing and its associations with cardiometabolic and brain disorders. *Scientific Reports*, 10, 2020. 1, 14, 25, 26, 27, 33, 34, 37, 44, 45, 46, 63, 74, 75

[12] Yu Luo, Wen-Xiu Chen, Jiang Qiu, and Tao Jia. Accelerated functional brain aging in major depressive disorder: evidence from a large scale fmri analysis of chinese participants. *ArXiv*, abs/2205.04871, 2022. 1, 15, 51, 84

[13] James H. Cole. Multimodality neuroimaging brain-age in uk biobank: relationship to biomedical, lifestyle, and cognitive factors. *Neurobiology of Aging*, 92:34 – 42, 2020. 1

[14] Jennifer H. Barnett, Lily Lewis, Andrew D. Blackwell, and Matthew Taylor. Early intervention in Alzheimer's disease: a health economic study of the effects of diagnostic timing. *BMC Neurology*, 14(1):101, May 2014. ISSN 1471-2377. doi: 10.1186/1471-2377-14-101. URL `https://doi.org/10.1186/1471-2377-14-101`. 1

[15] Han Peng, Weikang Gong, Christian F. Beckmann, Andrea Vedaldi, and Stephen M. Smith. Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, 68:101871, 2021. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2020.101871. URL `https://www.sciencedirect.com/science/article/pii/S1361841520302358`. xii, 1, 17, 25, 26, 27, 28, 33, 34, 37, 45, 50, 53, 54, 74, 75, 82, 86

[16] Chenzhong Yin, Phoebe Imms, Mingxi Cheng, Anar Amgalan, Nahian F. Chowdhury, Roy J. Massett, Nikhil N. Chaudhari, Xinghe Chen, Paul M. Thompson, Paul Bogdan, Andrei Irimia, and the Alzheimer's Disease Neuroimaging Initiative. Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment. *Proceedings of the National Academy of Sciences*, 120(2):e2214634120, March 2023. doi: 10.1073/pnas.2214634120. URL `https://www.pnas.org/doi/10.1073/pnas.2214634120`. Publisher: Proceedings of the National Academy of Sciences. 25, 26, 33, 34, 37, 46, 74, 76, 82

[17] Nicola K. Dinsdale, Emma Bluemke, Stephen M. Smith, Zobair Arya, D. Vidaurre, M. Jenkinson, and A. Namburete. Learning patterns of the ageing brain in mri using deep convolutional networks. *bioRxiv*, 2021. 1, 9, 17, 18, 25, 26, 27, 33, 34, 37, 43, 44, 45, 46, 63, 74, 75, 86

[18] Habtamu Minassie Aycheh, Joon-Kyung Seong, Jeong-Hyeon Shin, Duk L. Na, Byungkon Kang, Sang Won Seo, and Kyung ah Sohn. Biological brain age prediction using cortical thickness data: A large scale cohort study. *Frontiers in Aging Neuroscience*, 10, 2018. 1

[19] Jian wei Zhai and Ke Li. Predicting brain age based on spatial and temporal features of human brain functional networks. *Frontiers in Human Neuroscience*, 13, 2019. 1

[20] James H. Cole, Jonathan Underwood, Matthan W. A. Caan, Davide De Francesco, Rosan A. van Zoest, Robert Leech, F. W. Wit, Peter Portegies, Gert J. Geurtsen,

Ben A. Schmand, Maarten Franciscus Schim van der Loeff, Claudio Franceschi, Caroline A. Sabin, Charles B. L. M. Majoie, Alan Winston, Peter Reiss, and David James Sharp. Increased brain-predicted aging in treated HIV disease. *Neurology*, 88:1349 – 1357, 2017. 2, 30, 31, 35, 38

[21] Global hiv programme, 2023. URL `https://www.who.int/teams/global-HIV$-$hepatitis-and-stis-programmes/hiv/strategic-information/HIV$-$data-and-statistics`. Accessed: 2023-10-18. 2, 20

[22] H Budka. Neuropathology of human immunodeficiency virus infection. *Brain Pathology*, 1, 1991. 2, 22, 24, 32

[23] Marie F. Grill. Neurologic complications of human immunodeficiency virus. *CONTINUUM: Lifelong Learning in Neurology*, 27:963 – 991, 2021. vii, 2, 22, 23, 24, 32, 44, 80

[24] M. Tanveer, M.A. Ganaie, Iman Beheshti, Tripti Goel, Nehal Ahmad, Kuan-Ting Lai, Kaizhu Huang, Yu-Dong Zhang, Javier Del Ser, and Chin-Teng Lin. Deep learning for brain age estimation: A systematic review. *Information Fusion*, 96: 130–143, August 2023. ISSN 15662535. doi: 10.1016/j.inffus.2023.03.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S156625352300088X`. 2

[25] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 2004. 5, 10

[26] Yann le Cun. A theoretical framework for back-propagation. *Proceedings of the 1988 Connectionist Models Summer School*, 1988.

[27] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86:2278–2324, 1998. 5, 12

[28] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. vii, 5, 7, 12, 13, 14, 44, 45, 46, 52

[29] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 52:99–115, 1990. 5

[30] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65 6:386–408, 1958. 5

[31] Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. 6, 7, 10

[32] Henry J. Kelley. Gradient theory of optimal flight paths. *ARS Journal*, 30:947–954, 1960. URL `https://api.semanticscholar.org/CorpusID:121072881`. 6

[33] Arthur E. Bryson and Walter F. Denham. A steepest-ascent method for solving optimum programming problems. *Journal of Applied Mechanics*, 29:247–257, 1962. 6

[34] Paul J. Werbos. *Beyond Regression : "New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, 1974. 6

[35] Paul J. Werbos. Applications of advances in nonlinear sensitivity analysis. In R. F. Drenick and F. Kozin, editors, *System Modeling and Optimization*, pages 762–770, Berlin, Heidelberg, 1982. Springer Berlin Heidelberg. ISBN 978-3-540-39459-4. 6

[36] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. URL `https://api.semanticscholar.org/CorpusID:205001834`. 6

[37] Arthur L. Samuel. Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.*, 44:206–227, 1959. 6

[38] Feng hsiung Hsu. Ibm's deep blue chess grandmaster chips. *IEEE Micro*, 19:70–81, 1999. 6

[39] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60:84 – 90, 2012. URL `https://api.semanticscholar.org/CorpusID:195908774`. 7, 11, 12

[40] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, L. Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Vedavyas Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy P. Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489, 2016. 7

[41] John M. Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Andy Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David A. Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583 – 589, 2021. 7

[42] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. In *ECCV*, 2016. 7

[43] Dean A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In *NIPS*, 1988. 7

[44] Augustin-Louis Cauchy. Méthode générale pour la résolution des systemes d'équations simultanées. *Comptes rendus de l'Académie des sciences*, 1847. 8

[45] Herbert E. Robbins. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 2007. 8

[46] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015. 8

[47] Plot gradient descent, 2020. URL `https://tex.stackexchange.com/questions/544796/plot-gradient-descent`. Accessed: 2023-10-10. vii, 10

[48] Gidon Levakov, Gideon Rosenthal, Ilan Shelef, Tammy Riklin Raviv, and Galia Avidan. From a deep learning model back to the brain—identifying regional predictors and their relation to aging. *Human Brain Mapping*, 41:3235 – 3252, 2020. 9, 25, 26, 33, 34, 37

[49] Benedikt A. Jonsson, Gyda Bjornsdottir, Thorgeir E. Thorgeirsson, L.M. Ellingsen, G. Bragi Walters, Daniel Fannar Gudbjartsson, Hreinn Stefánsson, Kári Stefánsson, and Magnus Orn Ulfarsson. Brain age prediction using deep learning uncovers associated sequence variants. *Nature Communications*, 10, 2019. 9, 14, 25, 33, 34, 37, 45, 48

[50] David H. Hubel and Torsten N. Wiesel. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, 28:229–89, 1965. 10

[51] Yiyang Jiang, Fan Yang, Hengliang Zhu, Dian Zhou, and Xuan Zeng. Nonlinear cnn: improving cnns with quadratic convolutions. *Neural Computing and Applications*, 32:8507–8516, 2019. 11

[52] Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ranzato, and Yann LeCun. What is the best multi-stage architecture for object recognition? *2009 IEEE 12th International Conference on Computer Vision*, pages 2146–2153, 2009. 11

[53] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, D. Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. 12

[54] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 12

[55] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2015. 12

[56] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *ArXiv*, abs/1905.11946, 2019. 12, 14

[57] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2015. URL `https://api.semanticscholar.org/CorpusID:206594738`. 12

[58] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2014. URL `https://api.semanticscholar.org/CorpusID:1629541`. 12

[59] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015. URL https://api.semanticscholar.org/CorpusID:206592766. 12

[60] Geert J. S. Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017. URL https://api.semanticscholar.org/CorpusID:2088679. 12

[61] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016. URL https://api.semanticscholar.org/CorpusID:206593710. 12

[62] Maxime Oquab, Léon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1717–1724, 2014. URL https://api.semanticscholar.org/CorpusID:206592191. 12

[63] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, K. Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. URL https://api.semanticscholar.org/CorpusID:57246310. 12

[64] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *ArXiv*, abs/1311.2901, 2013. URL https://api.semanticscholar.org/CorpusID:3960646. 12

[65] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ArXiv*, abs/1502.03167, 2015. 12

[66] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y. 14

[67] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, 2014. URL https://api.semanticscholar.org/CorpusID:14113767. 14

[68] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), February 2017. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v31i1.11231. URL https://ojs.aaai.org/index.php/AAAI/article/view/11231. 14

[69] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *ArXiv*, abs/1605.07146, 2016. 14, 45, 46

[70] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. *CoRR*, abs/1709.00382, 2017. URL `http://arxiv.org/abs/1709.00382`. 14

[71] Brian H. Anderton. Ageing of the brain. *Mechanisms of Ageing and Development*, 123:811–817, 2002. URL `https://api.semanticscholar.org/CorpusID:43097130`. 14

[72] Naftali Raz and Karen M. Rodrigue. Differential aging of the brain: Patterns, cognitive correlates and modifiers. *Neuroscience & Biobehavioral Reviews*, 30:730–748, 2006. 14

[73] Riccarda Peters. Ageing and the brain. *Postgraduate Medical Journal*, 82:84 – 88, 2006. 14

[74] Hui Guo, William Siu, Ryan C. N. D'Arcy, Sandra E. Black, Lukas A. Grajauskas, Sonia Singh, Yunting Zhang, Kenneth Rockwood, and Xiaowei Song. Mri assessment of whole-brain structural changes in aging. *Clinical Interventions in Aging*, 12:1251 – 1270, 2017. 14

[75] Denise Head, Randy L. Buckner, Joshua S. Shimony, Laura E Williams, Erbil Akbudak, Thomas Edward Conturo, Mark P. McAvoy, John C. Morris, and Abraham Z. Snyder. Differential vulnerability of anterior white matter in nondemented aging with minimal acceleration in dementia of the alzheimer type: evidence from diffusion tensor imaging. *Cerebral cortex*, 14 4:410–23, 2004.

[76] Yunting Zhang, Chunhua Zhang, Jing Zhang, and Wei Li. Age-related changes of normal adult brain structure: analysed with diffusion tensor imaging. *Chinese medical journal*, 118 13:1059–65, 2005. 14

[77] Terry L. Jernigan, Sarah L. Archibald, Christine Fennema-Notestine, Anthony Collins Gamst, J C Stout, Julie C. Bonner, and J.R. Hesselink. Effects of age on tissues and regions of the cerebrum and cerebellum. *Neurobiology of Aging*, 22:581–594, 2001. URL `https://api.semanticscholar.org/CorpusID:9811454`. 14

[78] Meredith A. Shafto, Lorraine K. Tyler, Marie Dixon, Jason R. Taylor, James B. Rowe, Rhodri Cusack, Andrew J. Calder, William D. Marslen-Wilson, John Duncan, Tim Dalgleish, Richard N. Henson, Carol Brayne, Fiona E. Matthews, and CamCAN. 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):204, October 2014. ISSN 1471-2377. doi: 10.1186/s12883-014-0204-1. URL `https://doi.org/10.1186/s12883-014-0204-1`. vii, 15, 37, 38

[79] Jason R. Taylor, Nitin Williams, Rhodri Cusack, Tibor Auer, Meredith A. Shafto, Marie Dixon, Lorraine K. Tyler, CamCAN, and Richard N. Henson. 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*, 144:262–269, January 2017. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2015.09.018. URL `https://www.sciencedirect.com/science/article/pii/S1053811915008150`. 37, 38

[80] Daniel Taylor. Saliency mapping in convolutional neural networks to determine brain age trajectories. vii, 15

[81] Katja Franke and Christian Gaser. Longitudinal changes in individual brainage in healthy aging, mild cognitive impairment, and alzheimer's disease. *GeroPsych*, 2012. URL `https://api.semanticscholar.org/CorpusID:17157572`. 15

[82] Katja Franke, Christian Gaser, Brad Manor, and Vera Novak. Advanced brainage in older adults with type 2 diabetes mellitus. *Frontiers in Aging Neuroscience*, 5, 2013. 15, 29

[83] Linda Wang, Zhong Qiu Lin, and Alexander Wong. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10, 2020. URL `https://api.semanticscholar.org/CorpusID:215768886`. 15

[84] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597, 2015. URL `https://api.semanticscholar.org/CorpusID:3719281`. 15

[85] Vijay P. B. Grover, Joshua M Tognarelli, Mary M. E. Crossey, I Jane Cox, Simon D. Taylor-Robinson, and Mark J Mcphail. Magnetic resonance imaging: Principles and techniques: Lessons for clinicians. *Journal of clinical and experimental hepatology*, 5 3:246–55, 2015. 15, 16

[86] Girish Katti, Syeda Ara, and Dr Shireen. Magnetic resonance imaging (mri) - a review. *Intl J Dental Clin*, 3, 03 2011. 16

[87] Janaka P. Wansapura, Scott K Holland, R. Scott Dunn, and William S. Ball. Nmr relaxation times in the human brain at 3.0 tesla. *Journal of Magnetic Resonance Imaging*, 9, 1999. 16

[88] Jinghua Wang, Lili He, Hairong Zheng, and Zhong-Lin Lu. Optimizing the magnetization-prepared rapid gradient-echo (mp-rage) sequence. *PLoS ONE*, 9, 2014. URL `https://api.semanticscholar.org/CorpusID:9339020`. 16

[89] Maxim Zaitsev, Julian R. Maclaren, and Michael Herbst. Motion artifacts in mri: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42, 2015. URL `https://api.semanticscholar.org/CorpusID:33461916`. 17

[90] Fidel Alfaro-Almagro, Mark Jenkinson, Neal K. Bangerter, Jesper L. R. Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Stamatios N. Sotiropoulos, Saâd Jbabdi, Moisés Hernández-Fernández, Emmanuel Vallée, Diego Vidaurre, Matthew A.

Webster, Paul McCarthy, Chris Rorden, Alessandro Daducci, Daniel C. Alexander, Honghui Zhang, Iulius Dragonu, Paul M. Matthews, Karla L. Miller, and Stephen M. Smith. Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank. *Neuroimage*, 166:400 – 424, 2018. 17

[91] David A. Wood, Sina Kafiabadi, Ayisha Al Busaidi, Emily Guilhem, Antanas Montvila, Jeremy Lynch, Matthew Townend, Siddharth Agarwal, Asif Mazumder, Gareth J. Barker, Sebastian Ourselin, James H. Cole, and Thomas C. Booth. Accurate brain-age models for routine clinical mri examinations. *NeuroImage*, 249, 2022. 17

[92] Xinyuan Zhang, Yanqiu Feng, Wufan Chen, Xin Li, Andreia Vasconcellos Faria, Qianjin Feng, and Susumu Mori. Linear registration of brain mri using knowledge-based multiple intermediator libraries. *Frontiers in Neuroscience*, 13, 2019. 17

[93] Vishnu M Bashyam, Guray Erus, Jimit Doshi, Mohamad Habes, Ilya Nasralah, Monica Truelove-Hill, Dhivya Srinivasan, Liz Mamourian, Raymond Pomponio, Yong Fan, Lenore J. Launer, Colin L. Masters, Paul Maruff, Chuanjun Zhuo, Henry Völzke, Sterling C. Johnson, Jurgen Fripp, Nikolaos Koutsouleris, Theodore Daniel Satterthwaite, Daniel H. Wolf, Raquel E. Gur, Ruben C. Gur, John C. Morris, Marilyn S. Albert, Hans Jörgen Grabe, Susan M. Resnick, R. Nick Bryan, David A. Wolk, Haochang Shou, and Christos Davatzikos. Mri signatures of brain age and disease over the lifespan based on a deep brain network and 14468 individuals worldwide. *Brain : a journal of neurology*, 2020. 18, 25, 26, 31, 33, 34, 37, 44, 48

[94] Rafsanjany Kushol, Pedram Parnianpour, Alan H. Wilman, Sanjay Kalra, and Yee-Hong Yang. Effects of mri scanner manufacturers in classification tasks with deep learning models. *Scientific Reports*, 13, 2023. URL `https://api.semanticscholar.org/CorpusID:263705268`. 18, 19, 85

[95] Johanna Margarete Marianne Bayer, Paul M. Thompson, Christopher R. K. Ching, Mengting Liu, Andrew A. Chen, Alana Castro Panzenhagen, Neda Jahanshad, Andre F. Marquand, Lianne Schmaal, and Philipp G. Sämann. Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. *Frontiers in Neurology*, 13, 2022. URL `https://api.semanticscholar.org/CorpusID:253239971`. 18, 19

[96] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8 1:118–27, 2007. URL `https://api.semanticscholar.org/CorpusID:14297870`. 18, 19, 49

[97] Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fava, Patrick J. McGrath, Melvin McInnis, Mary L. Phillips, Madhukar H. Trivedi, Myrna M. Weissman, and Russell T. Shinohara. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167:104–120, February 2018. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2017.11.024. URL `https://www.sciencedirect.com/science/article/pii/S105381191730931X`. 19, 49, 79

[98] Raymond Pomponio, Guray Erus, Mohamad Habes, Jimit Doshi, Dhivya Srinivasan, Elizabeth Mamourian, Vishnu M Bashyam, Ilya M. Nasrallah,

Theodore Daniel Satterthwaite, Yong Fan, Lenore J. Launer, Colin L. Masters, Paul Maruff, Chuanjun Zhuo, Henry Völzke, Sterling C. Johnson, Jurgen Fripp, Nikolaos Koutsouleris, and Christos Davatzikos. Harmonization of large mri datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, 208, 2019. URL `https://api.semanticscholar.org/CorpusID:208878565`. 19, 49, 79, 85

[99] Tiffany K. Bell, Kate J. Godfrey, Ashley L. Ware, Keith Owen Yeates, and Ashley D. Harris. Harmonization of multi-site mrs data with combat. *NeuroImage*, 257, 2021. URL `https://api.semanticscholar.org/CorpusID:236953121`. 19

[100] Jean-Philippe Fortin, Drew Parker, Birkan Tunç, Takanori Watanabe, Mark A. Elliott, Kosha Ruparel, David R. Roalf, Theodore D. Satterthwaite, Ruben C. Gur, Raquel E. Gur, Robert T. Schultz, Ragini Verma, and Russell T. Shinohara. Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, 161:149–170, November 2017. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2017.08.047. URL `https://www.sciencedirect.com/science/article/pii/S1053811917306948`. 19, 79

[101] Nicola K. Dinsdale, Emma Bluemke, Vaanathi Sundaresan, Mark Jenkinson, Stephen Smith, and Ana I. L. Namburete. Challenges for machine learning in clinical translation of big data imaging studies. *Neuron*, 110:3866–3881, 2021. URL `https://api.semanticscholar.org/CorpusID:235829550`. 20, 79

[102] Daniel Moyer, Greg Ver Steeg, Chantal M. W. Tax, and Paul M. Thompson. Scanner invariant representations for diffusion mri harmonization. *Magnetic Resonance in Medicine*, 84:2174 – 2189, 2019. URL `https://api.semanticscholar.org/CorpusID:119313678`. 20

[103] Mengting Liu, Alyssa H. Zhu, Piyush Maiti, Sophia I. Thomopoulos, Shruti Gadewar, Yaqiong Chai, Hosung Kim, and Neda Jahanshad. Style transfer generative adversarial networks to harmonize multi-site mri to a single reference image to avoid over-correction. *bioRxiv*, 2022. 20, 50, 85

[104] Hiv statistics - global and regional trends, 2023. URL `https://data.unicef.org/topic/hivaids/global-regional-trends/`. Accessed: 2023-10-18. 20

[105] Bing Chen. Molecular mechanism of hiv-1 entry. *Trends in microbiology*, 2019. 20

[106] K. K. Vidya Vijayan, Krithika P. Karthigeyan, Srikant Tripathi, and Luke Elizabeth Hanna. Pathophysiology of cd4+ t-cell depletion in hiv-1 and hiv-2 infections. *Frontiers in Immunology*, 8, 2017. 20, 21

[107] Craig B. Wilen, John C. Tilton, and Robert W. Doms. Hiv: cell binding and entry. *Cold Spring Harbor perspectives in medicine*, 2 8, 2012. vii, 20, 21

[108] Pavitra Ramdas, Amit Kumar Sahu, Tarun Mishra, Vipin Bhardwaj, and Ajit Chande. From entry to egress: Strategic exploitation of the cellular processes by hiv-1. *Frontiers in Microbiology*, 11, 2020. vii, 21

[109] Jonathan E. Kaplan, Debra L. Hanson, Mark S. Dworkin, Toni Frederick, Jeanne Bertolli, Mary Lou Lindegren, Scott Holmberg, and Jeffrey L. Jones. Epidemiology of human immunodeficiency virus-associated opportunistic infections in the united

states in the era of highly active antiretroviral therapy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 30 Suppl 1: S5–14, 2000. URL https://api.semanticscholar.org/CorpusID:21981253. 22

[110] John W. Mellors, Charles R. Rinaldo, Phalguni Gupta, R. M. White, John A. Todd, and Lawrence A. Kingsley. Prognosis in hiv−1 infection predicted by the quantity of virus in plasma. *Science*, 272:1167 – 1170, 1996. URL https://api.semanticscholar.org/CorpusID:20955051. 22

[111] Honghua Shao and Sijun Li. A new perspective on HIV: effects of HIV on brain-heart axis. *Frontiers in Cardiovascular Medicine*, 10:1226782, August 2023. ISSN 2297-055X. doi: 10.3389/fcvm.2023.1226782. URL https://www.frontiersin.org/articles/10.3389/fcvm.2023.1226782/full. 22

[112] Sophia Pathai, Hendren Bajillan, Alan L. Landay, and Kevin P. High. Is hiv a model of accelerated or accentuated aging? *The journals of gerontology. Series A, Biological sciences and medical sciences*, 69 7:833–42, 2014. URL https://api.semanticscholar.org/CorpusID:15734672. 22

[113] Kaku A So-Armah, Laura A. Benjamin, Gerald S Bloomfield, Matthew J. Feinstein, Priscilla Y. Hsue, Benson Njuguna, and Matthew S. Freiberg. Hiv and cardiovascular disease. *The lancet. HIV*, 7 4:e279–e293, 2020. URL https://api.semanticscholar.org/CorpusID:214787195. 22

[114] Madison Schank, Juan Zhao, Jonathan P. Moorman, and Zhi Q. Yao. The impact of hiv− and art-induced mitochondrial dysfunction in cellular senescence and aging. *Cells*, 10, 2021. 22

[115] Ana Blas-García, Miriam Polo, Fernando Alegre, Haryes A. Funes, E. Herrero Martinez, Nadezda Apostolova, and Juan Vicente Esplugues. Lack of mitochondrial toxicity of darunavir, raltegravir and rilpivirine in neurons and hepatocytes: a comparison with efavirenz. *The Journal of antimicrobial chemotherapy*, 69 11: 2995–3000, 2014. 22

[116] Marianne Strazza, Vanessa Pirrone, Brian Wigdahl, and Michael R. Nonnemacher. Breaking down the barrier: The effects of hiv−1 on the blood-brain barrier. *Brain Research*, 1399:96–115, 2011. 22, 24

[117] Paul Thompson and Neda Jahanshad. Novel Neuroimaging Methods to Understand How HIV Affects the Brain. *Current HIV/AIDS reports*, 12(2):289–298, June 2015. ISSN 1548-3568. doi: 10.1007/s11904-015-0268-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4433315/. 22

[118] Judith Schouten, Paola Cinque, Magnus Gisslén, Peter Reiss, and Peter Portegies. Hiv−1 infection and cognitive impairment in the cart era: a review. *AIDS*, 25 5: 561–75, 2011. 22, 23, 24

[119] Beau M. Ances and Ronald J. Ellis. Dementia and neurocognitive disorders due to hiv−1 infection. *Seminars in neurology*, 27 1:86–92, 2007. 22

[120] John R. Keltner, Christine Fennema-Notestine, Florin Vaida, Dongzhe Wang, Donald R. Franklin, Robert H. Dworkin, Chelsea Sanders, John Allen McCutchan, Sarah L. Archibald, David J. Miller, George Kesidis, Clint Cushman, Sung Min Kim, Ian Abramson, Michael J. Taylor, Rebecca J. Theilmann, Michelle D. Julaton, R. Notestine, Stephanie Corkran, Mariana Cherner, Nichole A. Duarte, Terry J. Alexander, Jessica Robinson-Papp, Benjamin B. Gelman, David M. Simpson, Ann C Collier, Christina M. Marra, Susan Morgello, Greg G. Brown, Igor Grant, Joseph Hampton Atkinson, Terry L. Jernigan, Ronald J. Ellis, and for the Charter Group. Hiv−associated distal neuropathic pain is associated with smaller total cerebral cortical grey matter. *Journal of NeuroVirology*, 20:209 − 218, 2014. URL https://api.semanticscholar.org/CorpusID:8812476. 22, 24, 32

[121] Jodi M. Heaps, Pasiri Sithinamsuwan, Robert H. Paul, Sukalaya Lerdlum, Mantana Pothisri, David B. Clifford, Somporn Tipsuk, Stephanie Catella, Edgar Busovaca, James L. K. Fletcher, Benjamin Raudabaugh, Silvia Ratto-Kim, Victor G. Valcour, Jintanat Ananworanich, and on behalf of the French Jdrf Diabetic Nephropathy Col groups. Association between brain volumes and hand in cart-naïve hiv+ individuals from thailand. *Journal of NeuroVirology*, 21:105–112, 2014. 22, 23, 24, 32, 38

[122] Paul M. Thompson, Rebecca A. Dutton, Kiralee M. Hayashi, Arthur W. Toga, Oscar L. Lopez, Howard J. Aizenstein, and James T. Becker. Thinning of the cerebral cortex visualized in HIV/AIDS reflects CD4 $^+$ T lymphocyte decline. *Proceedings of the National Academy of Sciences*, 102(43):15647–15652, October 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0502548102. URL https://pnas.org/doi/full/10.1073/pnas.0502548102. 23, 24, 32, 77, 80

[123] Talia M. Nir, Neda Jahanshad, Christopher R. K. Ching, Ronald Cohen, Jaroslaw Harezlak, Giovanni Schifitto, Hei Y. Lam, Xue Hua, Jianhui Zhong, Tong Zhu, Michael J. Taylor, Thomas B. Campbell, Eric S Daar, Elyse J. Singer, Jeffry R. Alger, Paul M. Thompson, and Bradford Navia. Progressive brain atrophy in chronically infected and treated hiv+ individuals. *Journal of NeuroVirology*, 25:342–353, 2019. 24, 30, 32, 38, 43, 77, 80

[124] Katherine M Clifford, Vishal Samboju, Yann Cobigo, Benedetta Milanini, Gabriel A. Marx, Joanna Hellmuth, Howard J. Rosen, Joel H. Kramer, Isabel Elaine Allen, and Victor G. Valcour. Progressive brain atrophy despite persistent viral suppression in hiv patients older than 60 years. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 76:289–297, 2017. URL https://api.semanticscholar.org/CorpusID:3507403. 24, 32

[125] Takeshi Nishijima, Hiroyuki Gatanaga, Katsuji Teruya, Tsuyoshi Tajima, Yoshimi Kikuchi, Kanehiro Hasuo, and Shinichi Oka. Brain Magnetic Resonance Imaging Screening Is Not Useful for HIV-1-Infected Patients Without Neurological Symptoms. *AIDS Research and Human Retroviruses*, 30(10):970–974, October 2014. ISSN 0889-2229. doi: 10.1089/aid.2014.0123. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4180124/. 24, 81

[126] Stephen M. Smith, Diego Vidaurre, Fidel Alfaro-Almagro, Thomas E. Nichols, and Karla L. Miller. Estimation of brain age delta from brain imaging. *bioRxiv*, 2019. viii, 25, 27, 28, 54

[127] Simon M. Hofmann, F. Beyer, S. Lapuschkin, Markus Loeffler, K. Müller, A. Villringer, W. Samek, and A. Witte. Towards the interpretability of deep learning models for human neuroimaging. *bioRxiv*, 2021. 25, 33, 34

[128] Daniel Taylor, Jonathan Shock, Deshendran Moodley, Jonathan C Ipser, and Matthias Sebastian Treder. Brain structural saliency over the ages. In *International Conference on Machine Learning, Optimization, and Data Science*, 2022. URL `https://api.semanticscholar.org/CorpusID:247058642`. 25, 76, 86

[129] Xinyang Feng, Zachary Chase Lipton, Jie Yang, Scott A. Small, and Frontotemporal Lobar Degeneration Neuroimaging Initiative. Estimating brain age based on a uniform healthy population with deep learning and structural magnetic resonance imaging. *Neurobiology of Aging*, 91:15–25, 2020. 26, 33, 34, 37

[130] Pradeep K. Lam, Vigneshwaran Santhalingam, Parth Suresh, Rahul Baboota, Alyssa H. Zhu, Sophia I. Thomopoulos, Neda Jahanshad, and Paul M. Thompson. Accurate brain age prediction using recurrent slice-based networks. *bioRxiv*, 2020. 26, 37

[131] Karla L. Miller, Fidel Alfaro-Almagro, Neal Kepler Bangerter, David L. Thomas, Essa Yacoub, Junqian Xu, Andreas J. Bartsch, Saâd Jbabdi, Stamatios N. Sotiropoulos, Jesper L. R. Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Thomas W. Okell, Peter J. Weale, Iulius Dragonu, Steve Garratt, Sarah Hudson, Rory Collins, Mark Jenkinson, Paul M. Matthews, and Stephen M. Smith. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature neuroscience*, 19:1523 – 1536, 2016. URL `https://api.semanticscholar.org/CorpusID:1018393`. 26, 37

[132] Hualou Liang, Fengqing Zhang, and Xin Niu. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Human Brain Mapping*, 40:3143 – 3152, 2019. 27

[133] Matthias S. Treder, Jonathan P. Shock, Dan J. Stein, Stéfan du Plessis, Soraya Seedat, and Kamen A. Tsvetanov. Correlation constraints for regression models: Controlling bias in brain age prediction. *Frontiers in Psychiatry*, 12, 2021. ISSN 1664-0640. doi: 10.3389/fpsyt.2021.615754. URL `https://www.frontiersin.org/articles/10.3389/fpsyt.2021.615754`. 27, 28

[134] Giovanni Guaraldi, Gabriella Orlando, Stefano Zona, Marianna Menozzi, Federica Carli, Elisa Garlassi, Alessandra Berti, Elisa Rossi, Alberto Roverato, and Frank J Palella. Premature age-related comorbidities among hiv−infected persons compared with the general population. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 53 11:1120–6, 2011. 29, 31, 35

[135] Wafaa M El-Sadr, Jens D. Lundgren, James D. Neaton, Fred M. Gordin, Donald I. Abrams, Roberto C. Arduino, A G Babiker, William J. Burman, Nathan Clumeck, Calvin J. Cohen, David L. Cohn, David Cooper, Janet H. Darbyshire, Sean Emery, Gerd Fätkenheuer, B. G. Gazzard, Birgit Grund, Jennifer F Hoy, Karin L. Klingman, Marcelo Losso, Norman Markowitz, Jacquie Neuhaus, Andrew N Phillips, and Claire Rappoport. Cd4+ count-guided interruption of antiretroviral treatment. *The New England journal of medicine*, 355 22:2283–96, 2006. 29

[136] Neda Jahanshad, Victor G. Valcour, Talia M. Nir, Omid Kohannim, Edgar Busovaca, Krista Nicolas, and Paul M. Thompson. Disrupted brain networks in the aging hiv+ population. *Brain connectivity*, 2 6:335–44, 2012. 29, 31, 35, 38

[137] Eric M. Reiman, Richard J. Caselli, Lang Sheng Yun, K. Chen, Daniel Bandy, Satoshi Minoshima, Stephen N. Thibodeau, and David Osborne. Preclinical evidence of alzheimer's disease in persons homozygous for the epsilon 4 allele for apolipoprotein e. *The New England journal of medicine*, 334 12:752–8, 1996. 29

[138] Taylor P Kuhn, Tobias Kaufmann, Nhat Trung Doan, Lars Tjelta Westlye, Jacob D. Jones, Rodolfo A. Núñez, Susan Y. Bookheimer, Elyse J. Singer, Charles H. Hinkin, and April D. Thames. An augmented aging process in brain white matter in hiv. *bioRxiv*, 2018. 30, 31, 35, 80

[139] Jonathan Underwood, James H. Cole, Robert Leech, David James Sharp, and Alan Winston. Multivariate pattern analysis of volumetric neuroimaging data and its relationship with cognitive function in treated hiv disease. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 78:429–436, 2018. 30, 35, 38, 77, 80

[140] Kalen J Petersen, Jeremy Strain, Sarah Cooley, Florin Vaida, and Beau M Ances. Machine Learning Quantifies Accelerated White-Matter Aging in Persons With HIV. *The Journal of Infectious Diseases*, 226(1):49–58, April 2022. ISSN 0022-1899. doi: 10.1093/infdis/jiac156. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9890925/`. 30, 31, 35, 80

[141] Daniel Taylor, Jonathan Shock, Deshendran Moodley, Jonathan Ipser, and Matthias Treder. Brain Structural Saliency Over The Ages, July 2022. URL `http://arxiv.org/abs/2202.11690`. arXiv:2202.11690 [cs, eess, q-bio]. 37

[142] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, 12(3):e1001779, March 2015. ISSN 1549-1676. doi: 10.1371/journal.pmed.1001779. URL `https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001779`. Publisher: Public Library of Science. 37

[143] Jonathan C Ipser, John A. Joska, Tatum Sevenoaks, Hetta Gouse, Carla Patricia Freeman, Tobias Kaufmann, Ole Andreas Andreassen, Steven Shoptaw, and Dan J Stein. Evidence for a moderating effect of drinking severity but not hiv status on brain age in heavy episodic drinkers. *medRxiv*, 2021. 38, 51

[144] Paul M. Thompson et al. The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, 8:153–182, 2014. URL `https://api.semanticscholar.org/CorpusID:215520737`. 38

[145] Talia M. Nir, Jean-Paul Fouche, Jintanat Ananworanich, Beau M. Ances, Jasmina Boban, Bruce J. Brew, Linda Chang, Joga R. Chaganti, Christopher R. K. Ching, Lucette A. Cysique, Thomas Ernst, Joshua Faskowitz, Vikash Gupta, Jaroslaw Harezlak, Jodi M. Heaps-Woodruff, Charles H. Hinkin, Jacqueline Hoare, John A.

Joska, Kalpana J. Kallianpur, Taylor Kuhn, Hei Y. Lam, Meng Law, Christine Lebrun-Frenay, Andrew J. Levine, Lydiane Mondot, Beau K. Nakamoto, Bradford A. Navia, Xavier Pennec, Eric C. Porges, Cecilia M. Shikuma, April D. Thames, Victor Valcour, Matteo Vassallo, Adam J. Woods, Paul M. Thompson, Ronald A. Cohen, Robert Paul, Dan J. Stein, Neda Jahanshad, and for the ENIGMA-HIV Working Group. Smaller limbic structures are associated with greater immunosuppression in over 1000 HIV-infected adults across five continents: Findings from the ENIGMA-HIV Working Group, August 2019. URL https://www.biorxiv.org/content/10.1101/724583v1. Pages: 724583 Section: New Results. 38, 49

[146] Debapriya Sengupta, Phalguni Gupta, and Arindam Biswas. A survey on mutual information based medical image registration algorithms. *Neurocomputing*, 486: 174–188, 2021. URL https://api.semanticscholar.org/CorpusID:244121171. 43

[147] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Neural Information Processing Systems*, 2019. 46, 47, 49

[148] Lukas Biewald. Experiment tracking with weights and biases, 2020. URL https://www.wandb.com/. Software available from wandb.com. 48

[149] A. H. Zhu. neuroharmonize. GitHub repository, 2023. URL https://github.com/ahzhu/neuroHarmonize. 49

[150] R Li, D Duffee, and M F Gbadamosi-Akindele. Cd4 count. In: StatPearls [Internet], 2023. URL https://www.ncbi.nlm.nih.gov/books/NBK470231/. 51

[151] Centers for Disease Control and Prevention. 1993 revised classification system for hiv infection and expanded surveillance case definition for aids among adolescents and adults. https://www.cdc.gov/mmwr/preview/mmwrhtml/00018871.htm, 1992. Accessed: 2024-05-25. 51

[152] Vaughn E. Bryant, Destin D. Shortell, Jason S DeFelice, Ashley Huxhold, Robert L. Cook, Eric C. Porges, and Ronald A Cohen. The association of depression and anxiety symptoms to three different types of activities of daily living in persons with and without hiv. *AIDS care*, pages 1–8, 2023. URL https://api.semanticscholar.org/CorpusID:261335816. 51

[153] David Jakabek, Caroline D. Rae, Bruce J. Brew, and Lucette A. Cysique. Brain aging and cardiovascular factors in HIV: a longitudinal volume and shape MRI study. *AIDS*, 36(6):785, May 2022. ISSN 0269-9370. doi: 10.1097/QAD.0000000000003165. URL https://journals.lww.com/aidsonline/fulltext/2022/05010/brain_aging_and_cardiovascular_factors_in_hiv__a.4.aspx. 51

[154] Talia M. Nir, Jean-Paul Fouche, Jintanat Ananworanich, Beau M. Ances, Jasmina Boban, Bruce J. Brew, Linda Chang, Joga Chaganti, Christopher R. K. Ching, Lucette A. Cysique, Thomas Ernst, Joshua Faskowitz, Vikash Gupta, Jaroslaw Harezlak, Jodi M. Heaps-Woodruff, Charles H. Hinkin, Jacqueline Hoare, John A. Joska, Kalpana J Kallianpur, Taylor P Kuhn, Hei Y. Lam, Meng Law, Christine Lebrun-Frénay, Andrew J. Levine, Lydiane Mondot, Beau K Nakamoto, Bradford Navia, Xavier Pennec, Eric C. Porges, Cecilia M. Shikuma, April Denise Thames, Victor G. Valcour, Matteo Vassallo, Adam J. Woods, Paul M. Thompson, Ronald Cohen, Robert H. Paul, Dan J Stein, and Neda Jahanshad. Smaller limbic structures are associated with greater immunosuppression in over 1000 hiv-infected adults across five continents: Findings from the enigma-hiv working group. *bioRxiv*, 2019. URL https://api.semanticscholar.org/CorpusID:201200162. 63

[155] Sunniva M. K. Bøstrand, Kadi Vaher, Laura de Nooij, Matthew A. Harris, James H. Cole, Simon R. Cox, Riccardo E. Marioni, Daniel L. McCartney, Rosie M. Walker, Andrew M. McIntosh, Kathryn L. Evans, Heather C. Whalley, Robyn E. Wootton, and Toni-Kim Clarke. Associations between alcohol use and accelerated biological ageing. *Addiction biology*, 27:e13100 – e13100, 2020. URL https://api.semanticscholar.org/CorpusID:227153943. 84

[156] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando C Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79:151–175, 2010. URL https://api.semanticscholar.org/CorpusID:8577357.

[157] USC INI. URL https://www.ini.usc.edu/.

[158] Sebastian Ruder. An overview of gradient descent optimization algorithms. *ArXiv*, abs/1609.04747, 2016.

[159] Leslie N. Smith and Nicholay Topin. Super-convergence: very fast training of neural networks using large learning rates. In *Defense + Commercial Sensing*, 2019.

[160] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[161] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

[162] Robert K Heaton, Donald R. Franklin, Reena Deutsch, Scott L. Letendre, Ronald J. Ellis, Kaitlin B Casaletto, Maria Marquine, Steven Paul Woods, Florin Vaida, Joseph Hampton Atkinson, Thomas D Marcotte, John Allen McCutchan, Ann C Collier, Christina M. Marra, David B. Clifford, Benjamin B. Gelman, Ned Sacktor, Susan Morgello, David M. Simpson, Ian Abramson, Anthony Collins Gamst, Christine Fennema-Notestine, David M. Smith, and Igor Grant. Neurocognitive change in the era of hiv combination antiretroviral therapy: the longitudinal charter study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 60 3:473–80, 2015.

[163] Trang T. Le, Rayus Kuplicki, Brett A. McKinney, Hung wen Yeh, Wesley Kurt Thompson, and Martin P. Paulus. A nonlinear simulation framework supports adjusting for age when analyzing brainage. *Frontiers in Aging Neuroscience*, 10, 2018.

[164] James H. Cole and Katja Franke. Predicting age using neuroimaging: Innovative brain ageing biomarkers. *Trends in Neurosciences*, 40:681–690, 2017.

[165] Stephen M. Smith, Lloyd T. Elliott, Fidel Alfaro-Almagro, Paul McCarthy, Thomas E. Nichols, Gwenaëlle Douaud, and Karla L. Miller. Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *eLife*, 9, 2020.

[166] Xinyuan Zhang, Yanqiu Feng, Wufan Chen, Xin Li, Andreia V. Faria, Qianjin Feng, and Susumu Mori. Linear registration of brain mri using knowledge-based multiple intermediator libraries. *Frontiers in Neuroscience*, 13:909, 2019. ISSN 1662-453X. doi: 10.3389/fnins.2019.00909. URL `https://www.frontiersin.org/article/10.3389/fnins.2019.00909`.

[167] Ari Kolbeinsson, Jean Kossaifi, Yannis Panagakis, Adrian Bulat, Anima Anandkumar, Ioanna Tzoulaki, and Paul Matthews. Tensor dropout for robust learning. *IEEE Journal of Selected Topics in Signal Processing*, 15:630–640, 2021.

[168] Jean Kossaifi, Zachary Chase Lipton, Aran Khanna, Tommaso Furlanello, and Anima Anandkumar. Tensor regression networks. *ArXiv*, abs/1707.08308, 2020.

[169] Kaida Ning, Lu Zhao, William Matloff, Fengzhu Sun, and Arthur W. Toga. Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants. *Scientific Reports*, 10, 2020.

[170] Arinbjörn Kolbeinsson, Jean Kossaifi, Yannis Panagakis, Adrian Bulat, Anima Anandkumar, Ioanna Tzoulaki, and Paul Matthews. Robust deep networks with randomized tensor regression layers. *arXiv: Learning*, 2019.

[171] Hyunho Hwang, Hafiz Zia Ur Rehman, and Sungon Lee. 3d u-net for skull stripping in brain mri. *Applied Sciences*, 2019.

[172] Ellyn R. Butler, Andrew A. Chen, Rabie Abd El-Tawab Ramadan, Trang T. Le, Kosha Ruparel, Tyler Maxwell Moore, Theodore Daniel Satterthwaite, Fengqing Zhang, Haochang Shou, Ruben C. Gur, Thomas E. Nichols, and Russell T. Shinohara. Pitfalls in brain age analyses. *Human Brain Mapping*, 42:4092 – 4101, 2021.

[173] Jessica A. Bernard and Rachael D. Seidler. Moving forward: Age effects on the cerebellum underlie cognitive and motor declines. *Neuroscience & Biobehavioral Reviews*, 42:193–207, 2014.

[174] Abien Fred Agarap. Deep learning using rectified linear units (relu). *ArXiv*, abs/1803.08375, 2018.

[175] Joaquim Raduà, Eduard Vieta, Russell T. Shinohara, Peter V. Kochunov, Yann Quidé, Melissa J. Green, Cynthia Shannon Weickert, Thomas W. Weickert, Jason Bruggemann, Tilo T. J. Kircher, Igor Nenadić, Murray J. Cairns, Marc L. Seal, Ulrich Schall, Frans A. Henskens, Janice M. Fullerton, Bryan J. Mowry, Christos Pantelis, and Julian Pineda-Zapata. Increased power by harmonizing structural mri site differences with the combat batch adjustment method in enigma. *NeuroImage*, 218:116956 – 116956, 2020. URL `https://api.semanticscholar.org/CorpusID:219105763`.

[176] Mio Sakai, Masahiro Higashi, Takuya Fujiwara, Tomoko Uehira, Takuma Shirasaka, Katsuyuki Nakanishi, Nobuo Kashiwagi, Hisashi Tanaka, Hitoshi Terada, and Noriyuki Tomiyama. MRI imaging features of HIV-related central nervous system diseases: diagnosis by pattern recognition in daily practice. *Japanese Journal of Radiology*, 39(11):1023–1038, 2021. ISSN 1867-1071. doi: 10.1007/s11604-021-01150-4. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8202053/`.

[177] Ravi Bansal, Xuejun Hao, Feng Liu, Dongrong Xu, Jun Liu, and Bradley S. Peterson. The effects of changing water content, relaxation times, and tissue contrast on tissue segmentation and measures of cortical anatomy in mr images. *Magnetic resonance imaging*, 31 10:1709–30, 2013. URL `https://api.semanticscholar.org/CorpusID:39142575`.

[178] Lucette A. Cysique, Kirsten Moffat, Danielle M. Moore, Tammy A. Lane, Nicholas W. S. Davies, Andrew Carr, Bruce J. Brew, and Caroline Rae. HIV, vascular and aging injuries in the brain of clinically stable HIV-infected adults: a (1)H MRS study. *PloS One*, 8(4):e61738, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0061738.

[179] Shammi More, Georgios Antonopoulos, Felix Hoffstaedter, Julian Caspers, Simon B. Eickhoff, and Kaustubh R. Patil. Brain-age prediction: A systematic comparison of machine learning workflows. *NeuroImage*, 270:119947, April 2023. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2023.119947. URL `https://www.sciencedirect.com/science/article/pii/S1053811923000940`.

[180] Juhyuk Han, Seo Yeong Kim, Junhyeok Lee, and Won Hee Lee. Brain Age Prediction: A Comparison between Machine Learning Models Using Brain Morphometric Data. *Sensors*, 22(20):8077, January 2022. ISSN 1424-8220. doi: 10.3390/s22208077. URL `https://www.mdpi.com/1424-8220/22/20/8077`. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[181] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks : the official journal of the International Neural Network Society*, 61:85–117, 2014. URL `https://api.semanticscholar.org/CorpusID:11715509`.

[182] Corinna Cortes and Vladimir Naumovich Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995. URL https://api.semanticscholar.org/CorpusID:52874011.

[183] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. URL https://api.semanticscholar.org/CorpusID:89141.

[184] J. Ross Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, 1986. URL https://api.semanticscholar.org/CorpusID:13252401.

[185] Yann LeCun, Bernhard E. Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne E. Hubbard, and Lawrence D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989. URL https://api.semanticscholar.org/CorpusID:41312633.

[186] Robert H. Paul, Ronald Cohen, Bradford Navia, and Karen T. Tashima. Relationships between cognition and structural neuroimaging findings in adults with human immunodeficiency virus type-1. *Neuroscience & Biobehavioral Reviews*, 26: 353–359, 2002. URL https://api.semanticscholar.org/CorpusID:26609750.

[187] ImageNet. 2015 imagenet large scale visual recognition challenge (ilsvrc). https://image-net.org/challenges/LSVRC/2015/, 2015. Accessed: 2023-11-16.

[188] Jean-Philippe Fortin and Will Foran. Combatharmonization: Harmonization of multi-site imaging data with combat. https://github.com/Jfortin1/ComBatHarmonization, 2024.