MRI-Based Brain Age Estimation Using Supervised Contrastive Learning

Simon Joseph Clément Crête

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Computer Science (Computer Science) at

Concordia University

Montréal, Québec, Canada

July 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Simon Joseph Clément Crête**

Entitled: **MRI-Based Brain Age Estimation Using Supervised Contrastive Learning**

and submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science (Computer Science)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Tristan Glatard

_____ External Examiner
Dr. Habib Benali

_____ Examiner
Dr. Tristan Glatard

_____ Co-supervisor
Dr. Yiming Xiao

_____ Co-supervisor
Dr. Marta Kersten-Oertel

Approved by _____
Dr. Joey Paquet, Chair
Department of Computer Science and Software Engineering

_____ 2024

_____
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

MRI-Based Brain Age Estimation Using Supervised Contrastive Learning

Simon Joseph Clément Crête

Brain age estimation models aim to accurately assess a subject's biological brain age based on neuroanatomical features. Various factors, including neurodegenerative diseases, can cause accelerated brain aging and measuring this phenomena could serve as a biomarker for clinical applications. While promising results have been achieved in previous works, there is no consensus on an optimal model for accurate prediction or clinical utility. This thesis proposes using supervised contrastive learning with Rank-N-Contrast (RNC) loss and Grad-RAM for explainability utilizing structural T1w MRI data. Results indicate that the supervised contrastive strategy significantly outperformed ResNet models, achieving a mean absolute error of 4.28 years and an R^2 of 0.93 with a limited dataset of aging subjects. Benchmark comparisons with state-of-the-art models demonstrated that the supervised contrastive approach achieved comparable performance in our test sample, particularly among older age groups. The Grad-RAM analysis revealed anatomically relevant regions associated with aging, with the more nuanced capabilities exhibited by the supervised contrastive learning approach. Analyses of disease populations revealed significantly higher brain age gaps in Alzheimer's Disease (AD) patients, correlating strongly with ADAS-cog scores ($r = 0.37$, $p = 0.0098$), suggesting its potential as a biomarker for assessing AD severity. However, no significant correlation was found between brain age gap and UPDRSIII scores in Parkinson's disease (PD) patients. The Grad-RAM focuses on regions known to be linked to AD and PD, but with little difference to the healthy population. Our study demonstrates the potential of supervised contrastive learning with an RNC loss in brain age prediction, highlighting its ability to outperform other models in smaller datasets.

Acknowledgments

First and foremost, I would like to thank both my co-supervisors Dr. Marta Kersten-Oertel and Yiming Xiao. You have created an environment of learning and collaboration that allows all your students to grow and achieve their potential. I find myself fortunate to have been under the supervision of some of the hardest working and most caring professors I know. Xiao, thank you for your patient and thoughtful guidance that allowed me to learn the technical and research skills to complete this thesis. Marta, thank you for your supportive words, positive mindset, and crucial advice on writing and presenting.

My friends, the ones from childhood and the ones I made along the way, thank you for always being there for me through my best and worst times.

To my parents, who have emotionally and financially supported me throughout this long academic journey, without you none of this would have been possible. I will never be able to thank you enough for your patience and unconditional support.

These have been the most enjoyable and transformative years of my life. Moving to Montreal to complete a masters was a leap of faith, and I have faced many more academic and personal challenges than I had foreseen. But through the guidance and support all of you offered, I managed to persevere and grow as a person.

Contents

List of Figures viii

List of Tables xi

1 Introduction 1

1.1 Motivation . . . 2

1.2 Contributions . . . 2

1.3 Organization . . . 3

2 Background 4

2.1 Human Neuroanatomy . . . 4

2.2 Brain Aging . . . 6

2.3 Neurodegenerative Disorders of Interest . . . 8

2.3.1 Alzheimer's Disease . . . 8

2.3.2 Parkinson's Disease . . . 9

2.4 Magnetic Resonance Imaging . . . 10

2.5 MRI Preprocessing . . . 11

2.5.1 Denoising . . . 11

2.5.2 Bias Field Correction . . . 12

2.5.3 Skull Stripping . . . 12

2.5.4 Registration . . . 13

2.5.5 Intensity Normalization . . . 13

2.6 Deep Learning for Medical Image Analysis . . . 14
2.6.1 Convolutional Neural Networks . . . 15
2.6.2 Global-Local Learning . . . 17
2.6.3 Transformers . . . 18
2.6.4 Contrastive Learning . . . 18
2.7 Data Augmentation . . . 20
2.8 Brain Age Estimation . . . 21
2.9 Explainability for Deep Learning Models . . . 23
2.9.1 Gradient-weighted Class Activation Mapping . . . 24
2.9.2 Gradient-weighted Regression Activation Mapping . . . 25
2.9.3 Explainable DL Applications in Brain Age Estimation . . . 27
2.10 Summary . . . 27

3 Brain Age Estimation With Data-Efficient Supervised Contrastive Learning and Grad-RAM-based Visual Explanation 28

3.1 Introduction . . . 28
3.2 Related Work . . . 29
3.2.1 Aging-Related Anatomical Changes . . . 29
3.2.2 Brain Age Estimation Methods . . . 30
3.2.3 Applications to Neurodegenerative Diseases . . . 31
3.2.4 Explainability Methods . . . 32
3.3 Methods and Materials . . . 33
3.3.1 Dataset . . . 33
3.3.2 Preprocessing . . . 34
3.3.3 Deep Learning Models and Experimental Setup . . . 35
3.3.4 Evaluation metrics . . . 39
3.3.5 Grad-RAM for Aging Bias Analysis . . . 39
3.3.6 Analysis of Brain Age for Neurodegenerative Conditions . . . 40
3.4 Results . . . 41

3.4.1 Evaluation of Deep Learning Models 41
3.4.2 Grad-RAM Analysis of Healthy Population 43
3.4.3 Brain Age Estimation for Alzheimer's and Parkinson's Disease Patients . . 46
3.4.4 Grad-RAM Analysis for Alzheimer's and Parkinson's Disease Patients . . 47
3.5 Discussion 49
3.5.1 Deep Learning Model Performance 49
3.5.2 Analysis of Grad-RAM results 50
3.5.3 Alzheimer's and Parkinson's Disease 51
3.6 Conclusion 53

4 Conclusion **55**
4.0.1 Future Works 56
4.1 Summary 56

Appendix A **58**

Bibliography **64**

List of Figures

Figure 2.1 Diagram of brain lobes (left) and brain divisions (right). 5
Figure 2.2 Diagram of the inner brain, including the limbic system and basal ganglia [78]. 6
Figure 2.3 Axial slice of raw T1w (a) and T2w (b) full head MRI of the same subject. . 11
Figure 2.4 (a) Axial slice of raw T1w MRI and (b) preprocessed T1w MRI with bias field correction, skull stripping, intensity normalization, registration to ICBM 152 brain template of (a). 14
Figure 2.5 Diagram of a 3D CNN architecture for binary classification. [35] 15
Figure 2.6 Comparison of a classic CNN block (left) and a ResNet block with a skip connection (right). As seen on the skip connection, parts of the input can flow through the model and not be affected by the convolutional layer 16
Figure 2.7 Detailed layer information for various 2D ResNet variants with different network depths [32] 17
Figure 2.8 Diagram comparing contrastive losses: Supervised contrastive loss (left), contrasts all samples from the same class as positives against the negatives from the batch, augmentation of the anchor is also possible. Self-supervised loss (right) considers an anchor and various augmentations of the same sample and a set of negatives consisting of the entire remainder of the batch without class information [45] 19
Figure 2.9 Learned representations of L1 (MAE) loss, SupCon loss [45] and RNC loss on a temperature regression task [91] 20

Figure 2.10 Examples of various data augmentation techniques applied to a T2w brain MRI [66] 21
Figure 2.11 Visual explanation of an image classified as "West Highland white terrier" by a VGG16 CNN model trained on the ImageNet dataset. 24
Figure 3.1 Visualization of relative age distribution of cognitively normal patients per dataset (n = 1618) 34
Figure 3.2 Axial slice of raw, preprocessed, and resized data for comparison 35
Figure 3.3 Overview of the 3D ResNet-50 architecture used for brain age estimation. . 36
Figure 3.4 Overview of the training process for the RNC model supervised contrastive learning model, with the first step (left) training the ResNet backbone with the RNC loss (contrastive), and the second step (right) training a FC regression layer with MAE/L1 loss. 38
Figure 3.5 (a) Predictions of the ResNet 50 w. Data Aug and the RNC model brain ages vs chronological ages (b) Batch size vs. MAE for RNC models trained on 1x1x1 resolution data and 2x2x2 mm^3 resolution. 42
Figure 3.6 Averaged Grad-RAM heatmaps produced from the ResNet 50-highRez model based on the healthy control test set for the age groups of 20-40 yo (left) and 80+ yo (right). 44
Figure 3.7 Averaged Grad-RAM heatmaps produced from the RNC-highRez model based on the healthy control test set for the age groups of 20-40 yo (top left), 40-60 yo (bottom left), 60-80 yo (top right), and 80+ yo (bottom right). 45
Figure 3.8 Group-averaged Grad-RAM intensities produced from the RNC-highRez model of male (left) and female healthy subjects (right) of the entire test dataset. . . 46
Figure 3.9 Violin plot of differences between the estimated brain age and the ground truths for the ResNet50-highRez and RNC-highRez models on healthy and diseased populations. 47
Figure 3.10 Averaged Grad-RAM heatmaps produced from the RNC-highRez model for the AD (left) and PD (right) patients. 48

Figure A.1 Normalized average Grad-RAM heatmap intensities per AAL atlas parcellation for each sub-population for the ResNet50-highRez model. 61
Figure A.2 Normalized average Grad-RAM heatmap intensities per AAL atlas parcellation for each sub-population for the contrastive model. 62
Figure A.3 Normalized average Grad-RAM heatmap intensities per AAL atlas parcellation for a sub-populations and model, and difference of parcellation intensities for AD vs CN (60-100), PD vs CN (60 to 100), and male vs. female. 63

List of Tables

Table 3.1 Statistics of data sets used for cognitively normal, AD, and PD. 33

Table 3.2 Accuracy assessments of MRI-based brain age estimation of ResNet and RNC models with different image resolutions and data augmentation conditions. (* indictates $p < 0.01$) 42

Table 3.3 Results of the best performing ResNet, proposed RNC and benchmark models, with addition of male (M) and female (F) subgroup performance. Our methods are highlighted in gray. (* indictates $p < 0.01$). 43

Table 3.4 Metrics of interest for a disease population for each model, including the correlation between brain age (BA) and disease severity metrics: ADAS-cog 11 for AD, UPDRSIII for PD. 47

Table A.1 AAL atlas brain regions and corresponding IDs 60

Chapter 1

Introduction

Aging has complex mechanisms and multifaceted impacts on humans, making it a focal point of scientific research. Despite significant advancements, the mechanisms and impacts of aging remain only partially understood. Understanding these mechanisms extends beyond academic inquiry, bearing important implications for public health, particularly in the context of our currently aging population.

Age-related diseases, such as cognitive decline and dementia, are on the rise, necessitating more sophisticated tools and approaches to manage these conditions. Due to large individual variability and the importance of personalized healthcare, the concept of biological age has gained interest in research. Biological age is the distinction between chronological age and physiological features which reflects an individual's state. It can provide a more nuanced understanding of an individual's health status, potentially serving as an important biomarker for various age related conditions. This is particularly relevant for the brain due to neurodegenerative diseases like Alzheimer's and Parkinson's that are often characterized by changes resembling advanced brain aging. Therefore, measuring biological brain age could offer valuable insights into overall brain health and treatment of diseases.

Deep learning (DL) has shown great promise in medical applications, providing new tools for diagnosis, treatment planning, and analysis of complex biological processes. DL models can analyze neuroimaging data to accurately estimate brain age, but few studies have shown its use as a biomarker to detect or measure neurodegenerative diseases. These methods not only enhance our

ability to predict brain age with high precision but also enable the identification of subtle biomarkers that may signify early signs of neurological disorders.

1.1 Motivation

Understanding brain aging is increasingly important as our population ages and neurodegenerative diseases become more prevalent. Accurately assessing brain age can provide insights into cognitive decline, disease progression, and overall brain health. Traditional methods often fall short in capturing the complexity of brain aging, prompting the exploration of new techniques. DL, and in particular Convolutional Neural Networks (CNN), have emerged as a powerful a tool in this domain, offering sophisticated approaches to analyze complex patterns in neuroimaging data. Despite significant progress in this domain, there remains a lack of robust testing of these models as they have been applied across different datasets and populations without thorough validation. Moreover, certain cutting edge DL architectures have yet to be applied in the domain of brain aging. Thus, research is still needed to find optimal methods to measure brain age and understand its utility as a biomarker. In this thesis, we aim to introducing a novel application of supervised contrastive learning for brain age estimation with the goal of further advancing the growing body of research.

1.2 Contributions

The primary contributions of this thesis are threefold, focusing on novel application, the integration of explainability techniques, and analysis of neurodegenerative diseases in brain age estimation:

- We introduce an application of supervised contrastive learning with the Rank-N-Contrast (RNC) loss for brain age estimation. To the best of our knowledge, this approach has not previously been applied to brain age prediction using T1-weighted (T1w) structural MRIs.
- When trained on a limited dataset of aging healthy population, the RNC architecture significantly outperformed a ResNet trained on the same data and demonstrated better or comparable results to certain state-of-the-art (SOTA) models on our test set sample.

- To provide insights into the neuroanatomical regions associated with aging and disease we integrated Gradient-weighted Regression Activation Mapping (Grad-RAM) into our methods.
- We explored correlations between estimated brain age gap and severity metrics for Alzheimer's and Parkinson's diseases, with further Grad-RAM analyses providing insights into the potential anatomical regions affected by these diseases.

1.3 Organization

The remainder of this thesis is organized as follows. Chapter 2 offers the foundational knowledge necessary for comprehending the thesis, covering human neuroanatomy and its aging, neurodegenerative diseases of interest, structural magnetic resonance imaging (MRI), brain MRI preprocessing, DL, and approaches to brain age estimation. In Chapter 3, we present our main body of research, consisting of a brain age prediction model using supervised contrastive learning with RNC loss and Grad-RAM for explainability. We assess the method's performance on healthy subjects, Alzeihmer's, and Parkinson's patients. We compare results against our own ResNet and other SOTA techniques in the literature and use Grad-RAM to understand our DL model's task-specific image features. Finally, Chapter 4 concludes the findings of the thesis and discusses future avenues of research.

Chapter 2

Background

2.1 Human Neuroanatomy

The brain is considered the most complex organ system in the human body. It is composed of various tissue types and can be divided into several parts, based on the specific anatomical tissue structure, or on its function. The following neuroanatomy descriptions are based on the Atlas of Functional Neuroanatomy [34]. The brain is divided into four main lobes which have distinct functions and characteristics: the frontal, temporal, parietal and occipital lobes. Furthermore, the cortex, the outer layer of the cerebrum, is divided into several regions or cortical areas that are involved in various high-level brain function. For example, the frontal lobe is responsible for reasoning and is the part of the brain which takes longest to fully develop, it houses the motor cortex which handles voluntary movement. The temporal lobe is responsible for auditory processing and houses the somatosensory cortex, which is a sensory receiver for taste, temperature, touch and other environmental factors. The parietal lobes handle taste, smell, arithmetic and reading tasks. The occipital lobe processes vision and memories of visual medium. The illustration of the general human brain anatomy is shown in Figure 2.1.

The inner brain, illustrated in Figure 2.2, contains regions that directly connect to the central nervous system and contribute to regulation of emotions, perceptions, and, responses. The limbic system is a complex set of structures in the brain that deal with emotions and memory. The amygdala is involved in emotion processing, particularly fear and pleasure responses. The hypothalamus is

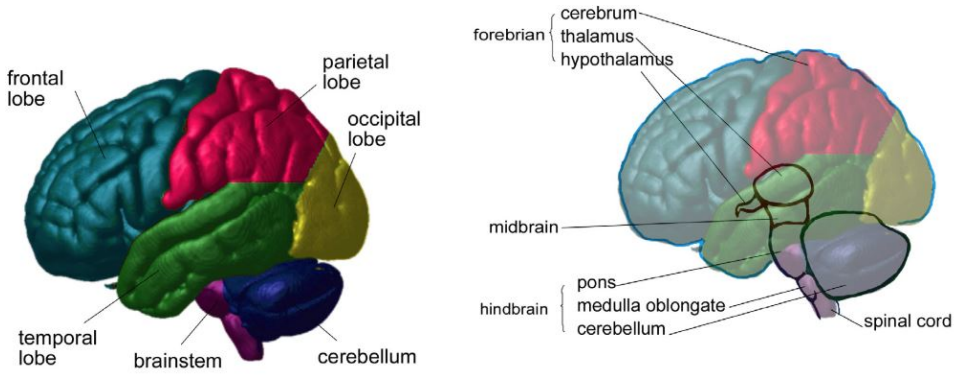


Figure 2.1: Diagram of brain lobes (left) and brain divisions (right).

responsible for the control of hormones and chemicals to appropriately respond to any stimulus, and attempt to keep the body in homeostasis. It helps manage the body temperature, hunger and thirst, mood, sex drive, blood pressure and sleep. The thalamus processes all sensory and motor information except smell before it reaches the cerebral cortex. The hippocampus is a embedded deep into the temporal lobe, it is responsible for memory formation and retrieval. The thalamus acts as a relay station for sensory information directing it to cerebral cortex. The cingulate gyrus plays a role in processing emotions and regulating behavior. And finally the parahippocampal gyrus surrounds the hippocampus and is also also responsible for memory retrieval and encoding.

The basal ganglia are a group of structures near the center of the brain that filters and selects voluntary motor, cognitive and emotional responses. The caudate is a C-shaped structure that arches over the thalamus; it is involved in various higher neurological functions, including learning, memory, and voluntary movement control. The putamen is a round structure located at the base of the forebrain primarily involved in regulating movements and influencing various types of learning. The globus pallidus that lies medially to the putamen plays a role in the regulation of voluntary movement. The substantia nigra is a darkly pigmented structure and contains dopaminergic neurons that are critical for movement control and reward processing. The subthalamic nucleus is a small lentil-shaped structure located below the thalamus. It is part of the indirect pathway of the basal ganglia circuitry and plays a vital role in regulating motor output.

The brain consists of three main tissue types, grey matter, white matter, and cerebrospinal fluid

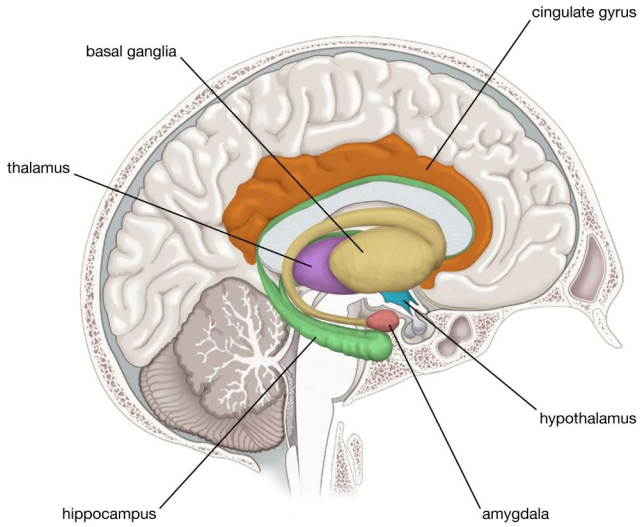


Figure 2.2: Diagram of the inner brain, including the limbic system and basal ganglia [78].

(CSF). Grey matter houses neural cell bodies, axon terminals, nerve synapses and dendrites. This tissue allows for processing and release of information. In contrast, white matter is composed of groups of axons that process and conduct nerve signals. Cerebrospinal fluid is a watery liquid that is produced and circulates through the brains ventricles (i.e., hollow cavities), and around the outer surface of the brain. This liquid acts as a cushion, regulates pressure inside the skull, and compensates in volume when the brain shrinks.

2.2 Brain Aging

Aging, the time-related deterioration of physiological functions, is an irreversible and not fully understood phenomena. Functional decline, such as reduced physical fitness and reduced cognitive abilities, typically accompany aging. Understanding the biological, cognitive and psychological changes that occur during aging is of great interest for many researchers due to its important impact on human health and its complex mechanisms. Brain aging has been linked to physiological changes that lead to functional decline, increased susceptibility to certain diseases, and increased mortality. Understanding brain aging can enhance healthy aging, early detection of neurodegenerative diseases, disease prevention, and quality of life.

Infancy and early childhood are marked by rapid brain growth and development. In an infant's

first year, the brain doubles in volume. By age 5 to 6, the brain is roughly ninety percent of adult size [36]. During this time, there is significant development in areas responsible for basic sensory processing, cognitive ability, and motor skills. Gray matter has a higher proportion of volume during this period. During pre-adolescence the brain continues to develop its cognitive and emotional capacities, with an increase in myelination of neurons, leading to a steady increase in white matter. Adolescence brings a notable increase in the dendritic pruning process, where the brain eliminates weaker synaptic connections and strengthens the more frequently used ones [1]. This is particularly evident in the frontal lobes, which are responsible for decision making, planning, and social behavior. The pruning causes a decrease in gray matter volume around puberty and a steady increase in white matter. The brain doesn't reach full maturity and volume until the mid-20s, particularly in the frontal areas. Typically the consensus for the end of brain development is around 25 years old [1].

While the brain has reached its peak in size by early adulthood, it continues to change and adapt. As individuals age, there is a gradual decrease in brain volume and weight, particularly in the later decades of life. This shrinking with age occurs at a rate of roughly five percent per decade, accelerating around the age of 70 [25]. Decreases of 0.3% per year for individuals in their 40s and of 0.5% in their 80s were found. There is also an average annual increase in ventricle volume of 1.8%, with a range of 1.0% per year among individuals in their 40s to 2.5% for individuals in their 80s. Rates of volume decline are not constant as inter-individual and region specific differences are present [25]. At the same time, as vasculature ages and blood pressure rises, the risk of stroke, ischaemia, and white matter lesions all increase. The functional capabilities of the brain decline progressively during aging, which manifests as decrements in learning and memory, attention, decision-making speed, sensory perception (vision, hearing, touch, smell, and taste), and motor coordination. Older adults often have difficulty understanding rapid speech, with reduced comprehension of complex sentences and difficult word retrieval ability. The time course of the age-related decline in brain performance roughly parallels the reduced performance of other organ systems with notable acceleration to the brain beyond 50 years of age [73].

It has been shown, however, that biological aging of the brain is not strictly tied to chronological aging and constant across subjects [72]. Genetics, usage of the brain for complex tasks, healthy diet, low alcohol intake, and regular exercise may slow aging by a significant factor [58, 56]. This

individual-specific aging shows the need for personalized aging biomarkers that surpass the generalization of chronological age.

Aging and Sex

Sex refers to a set of biological attributes in humans and animals. It is primarily associated with physical and physiological features, including chromosomes, gene expression, hormone levels and function, and reproductive/sexual anatomy. Sex is usually categorized in the binary of female or male, but there are variations of the biological attributes that comprise sex and how these attributes are expressed. Studies that have found sex differences in brain aging suggest that male and female brains age in different patterns, though no thorough consensus has been reached [3, 90]. Gender, the multidimensional construct that encompasses gender identity and expression, as well as social and cultural expectations associated with certain sex traits, is not extensively explored in brain aging literature.

2.3 Neurodegenerative Disorders of Interest

2.3.1 Alzheimer's Disease

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia among older adults [8]. It is currently impacting around 350 thousand Canadians and is expected to increase to over a million by 2050 [8]. AD is characterized by a gradual decline in cognitive functions, such as memory, language skills, and the ability to carry out simple tasks. The exact cause of the disease remains largely unknown, although it is believed to result from a combination of genetic, lifestyle, and environmental factors that affect the brain over time. AD is identified by the presence of amyloid-beta plaques and tau protein tangles within the brain, which disrupt neural function and lead to neuronal death [76]. AD brings about significant anatomical changes in the brain, most notably atrophy of brain tissue. This atrophy initially affects the hippocampus and entorhinal cortex before spreading to other areas leading to a reduction of cortical thickness, particularly in the temporal and parietal lobes [81]. The ventricles therefore enlarge as a result of this loss of brain volume.

The severity of Alzheimer's Disease can be assessed using cognitive and functional measures. The Alzheimer's Disease Assessment Scale (ADAS) is a comprehensive test battery that evaluates memory, attention, reasoning, language, orientation, and praxis [79]. The cognitive part (ADAS-Cog) is the most commonly used section of the scale to assess the severity and progression of cognitive impairment in AD. The Mini-Mental State Examination (MMSE) is a brief 30-point questionnaire used extensively in clinical and research settings to measure cognitive impairment. It assesses various cognitive functions, including arithmetic, memory, orientation, and language abilities. The MMSE score can help in diagnosing AD and monitoring its progression over time [2].

Recent studies have explored the relationship between brain age estimation and AD. These models have identified that individuals with AD often have a higher brain age compared to their chronological age, suggesting accelerated brain aging [60, 26]. Some studies have measured the accelerated aging in brain structures, and used it as a potential biomarker to differentiate healthy and AD subjects [67].

2.3.2 Parkinson's Disease

Parkinson's Disease (PD) is a neurodegenerative disease affecting over 100 thousand Canadians [75]. It primarily impacts motor control in the central nervous system. The principal symptoms of PD include tremors at rest, rigidity, bradykinesia (slowness of movement), and postural instability. The disease further causes a wide range of non-motor symptoms, including cognitive impairment, mood disorders, depression, and autonomic dysfunction. The most significant pathological feature of PD is the loss of dopaminergic neurons in the substantia nigra, which is crucial for motor control. Neuroimaging studies have shown various brain structural changes associated with PD such as noticeable depigmentation in the substantia nigra, atrophy and shape changes of parts of the basal ganglia [77], hippocampus, and cortical thinning in the frontal and temporal lobes [74].

The severity and progression of Parkinson's Disease can be evaluated using various scales. The Unified Parkinson's Disease Rating Scale (UPDRS) is a comprehensive tool used to quantify the severity of PD, encompassing mental, behavioral, and mood assessments, self-evaluation of the activities of daily life, and a clinical evaluation of motor function [29]. In particular, UPDRS-III which focuses on the motor disability evaluation is linked with anatomical brain changes [48].

The Hoehn and Yahr Scale (HY) classifies patients into five stages based on mobility and balance impairments. It is straightforward and is commonly used to provide a clinical description of PD progression.

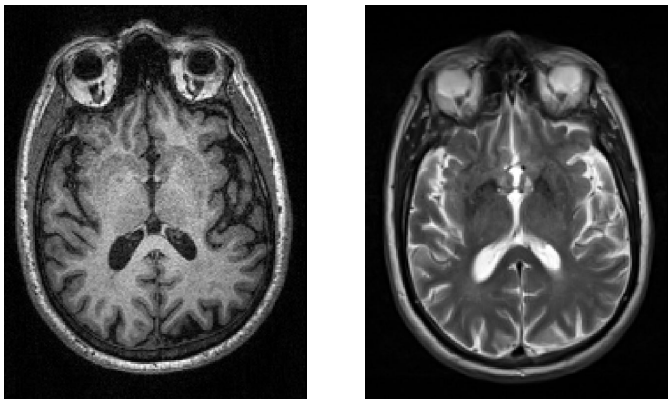
Several studies have investigated whether PD is associated with accelerated brain aging. Studies looking at more severe PD cases have found a higher brain age discrepancy between healthy and PD patients [17]. However, other studies have failed to disclose significant morphological changes between PD and healthy controls [93, 69], potentially due to poor contrast for anatomical structures related to PD (e.g., the Basal Ganglia) in more commonly seen T1w MRI [67].

2.4 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive widely used imaging technique that can capture 3D images of internal organs allowing complex analysis of structures and functions of the human body. With MRI, a magnetic field is applied with specific radio frequency (RF) pulses to excite protons in hydrogen nuclei of water molecules, causing these protons to change alignment. Turning off the RF pulse returns the atoms to their prior state and releases energy in the form of radio waves. These radio waves are captured and spatially encoded to form an image. The time taken for protons to realign with the magnetic field and the amount of energy released is dependent on the environment and the chemical nature of the molecule. This results in different contrasts in the image, making it possible to distinguish various tissue types in the images.

T1 and T2 weighted (T1w, T2w) imaging are often the most common modalities to examine soft tissue in the body. Different MRI contrasts between different tissue types can be obtained by varying the MR imaging acquisition parameters, including Repetition time (TR), Echo time (TE), and flip angle (FA). The repetition time is the amount of time between successive pulse sequences applied to the same slice. The echo time is the time between the delivery of the RF pulse and the receipt of the echo signal, and the flip angle is the amount of angle the net magnetization is tipped away during application of a radiofrequency pulse. T1w MRIs typically use shorter TR and TE, while T2w ones use longer times. This results in tissues with shorter relaxation times to appear brighter in T1w MRI scans, and ones with longer times to appear brighter in T2w images. T1w

MRIs are typically preferred for assessing the general anatomy, specifically fatty tissues, and T2w MRIs are preferred for identifying pathological changes involving fluids, like edema, inflammation, or infection. Further, due to longer acquisition time (e.g., long TR), T2w scans are often less commonly found in medical imaging datasets, further contributing to T1w images being the staple contrast for neuroanatomical analysis. A visual comparison is presented in Fig. 2.3.



(a) T1w head MRI (b) T2w head MRI

Figure 2.3: Axial slice of raw T1w (a) and T2w (b) full head MRI of the same subject.

2.5 MRI Preprocessing

High image quality is desirable for brain images analysis, but limitations in imaging systems can introduce image noise and artifacts, which often interfere with data analyses. Denoising and artifact removal methods should be devised based on the specific data characteristics of medical imaging systems (e.g., hardware, reconstruction algorithms, etc.), and the preprocessing pipelines should be consistent across all samples to allow correct and reproducible studies. An example of a raw and preprocessed T1w MRI is presented in Fig. 2.4.

2.5.1 Denoising

MR image noise primarily comes from thermal noise and physiological noise. The noise can reduce the quality of the image making it harder to clearly differentiate refined anatomical features. While the statistical characteristics of MRI noise can be complex, often, it is approximated with

a Gaussian distribution. To best preserve anatomical details, a number of denoising algorithms have been proposed, including the non-local means algorithm [54], principal component analysis denoising [55], and DL-based approaches [53]. While MRI denoising is crucial in field applications, for DL-based analysis, this preprocessing is not often incorporated due to the inherent ability in focusing on key features from the networks.

2.5.2 Bias Field Correction

Image artifacts are image features that don't exist in the original object, and can mislead diagnosis and analysis. These can be caused by factors, such as patient motion (e.g., motion artifacts), image reconstruction algorithms (e.g., aliasing artifacts), and limitation of the scanner (e.g., image distortion). One typical MRI artifact is the presence of a low-frequency intensity non-uniformity in the image data, also known as bias or inhomogeneity. The artifact is induced by various factors, such as the non-uniform sensitivity of RF coils and/or the influence of external magnetic fields. The inhomogeneous brightness of the same tissue class can cause difficulty in robust image recognition of anatomical structures. N4 Bias Field Correction [87] is an effective method to mitigate this issue. This algorithm employs an iterative process between deconvolving the image histogram by a Gaussian function, estimating the "corrected" intensities and spatially smoothing the resulting bias field estimation using a B-spline model.

2.5.3 Skull Stripping

Brain extraction or skull stripping, which involves removing non-brain tissues including skin, scalp and dura, is an important step in many neuroimaging analyses such as registration, tissue classification, and segmentation. While it is not inherently required for DL analysis, it allows the model to focus solely on the physiology of the brain, and not the rest of the head, and has become the standard procedure. A variety of algorithms with high accuracy have been developed to attain this goal [70], one commonly used method is 'BEaST', which is based on the non-local means segmentation within a multi-resolution framework for computational efficiency [18].

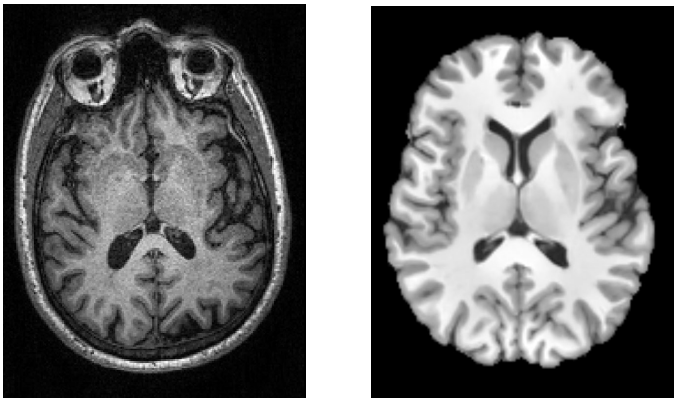
2.5.4 Registration

Registration in MRI medical image analysis involves aligning or matching two or more images. In many cases, all images used in a study are registered to a common template which enables more accurate comparison of the information contained in these images. Rigid (linear) registration aligns images using rotation and translation, thus preserving distances and angles. With additional parameters of scaling and shear, affine registration allows more flexibility in spatial transformation. In contrast, non-rigid (non-linear) registration allows local deformations to align images for more refined anatomical matching. Automatic medical image registration requires a well-defined loss function to optimize. Typically for inter-contrast MRI registration, cross-correlation (CC) is a popular choice while inter-modal/contrast registration often relies on the optimization of mutual information between the image pairs. A number of automatic medical image registration algorithm and software packages (e.g., ANTs, MINC Toolkit, and FSL) have been proposed to facilitate neuroimage analysis.

2.5.5 Intensity Normalization

Methods are required to standardize the image contrast and intensity range so that the variation due to different scanners and imaging protocols is mitigated. One way to solve this contrast difference is through histogram matching. Using a population-averaged MRI atlas as the histogram template, individual MRI's histogram can be transformed to follow the template's distribution. As the histogram profile can be influenced by the proportion of tissue types, it is advisable to use MRI atlases that match the characteristics of the target population. Min-Max and Z-Score normalization are commonly used and simple solutions. Min-Max scales intensities between 0 and 1 inclusively, Z-score scales intensities such that the mean is 0 and the standard deviation 1.

$$\text{Min-Max: } M(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad \text{Z-Score: } Z(x) = \frac{x - \text{mean}(x)}{\text{stddev}(x)}$$



(a) Raw T1w head MRI

(b) Preprocessed T1w brain MRI

Figure 2.4: (a) Axial slice of raw T1w MRI and (b) preprocessed T1w MRI with bias field correction, skull stripping, intensity normalization, registration to ICBM 152 brain template of (a).

2.6 Deep Learning for Medical Image Analysis

DL has proven a powerful method to analyze real-world images in classical applications such as image classification, handwriting recognition, and image segmentation. Automated methods to interpret medical images could enable a shift in clinical application by reducing costs, time, and human error, and provide analysis beyond human capabilities by compiling huge amounts of data and perceiving patterns not easily recognizable to humans.

Several major challenges unique to the domain have resulted in very few applications of such tools to advance clinical adoption. DL models require large amounts of high-quality and varied data to be effective. Moreover, three-dimensional imaging data is memory intensive and requires massive computational power to process. Further, acquiring such datasets is challenging due to immense costs, privacy concerns, governance laws, poor data infrastructure, and variety in acquisition methods and quality. Moreover, medical datasets for DL algorithm development can unknowingly suffer from imbalances in training data sampling, with bias towards certain conditions or demographics if care is not taken. This can be particularly problematic for rare diseases and under-represented populations. These models would require precise and robust explainability and uncertainty features,

allowing clinicians and patients to confidently use information provided. Furthermore, the deployment of such models must navigate complex regulatory landscapes while ensuring patient privacy, data security, and ethics.

2.6.1 Convolutional Neural Networks

Convolutional Neural Networks (CNN), a class of feed-forward neural networks which learns to extract features through kernel optimization, was one of the first model architectures used on medical images and are extensively used to this day. In 1993, Lo et al. [50] introduced CNNs for the analysis of medical images by developing a model for lung nodule detection in chest radiographs. Various advances to the CNN architecture have enabled the creation of extremely effective models for medical imaging classification, segmentation, or regression.

CNNs are composed of three types of layers: convolutional layer, pooling layer, and fully connected layer. The convolution layer uses filters that perform convolution operations with a specified kernel size and stride creating a feature map. The pooling layer performs a downsampling operation. Max or average pooling are typically used. The fully connected layer, the last layer of the CNN, performs class scores or regression. It is used after flattening the feature space. A visual illustration of a 3D binary classification CNN is shown in Fig. 2.5.

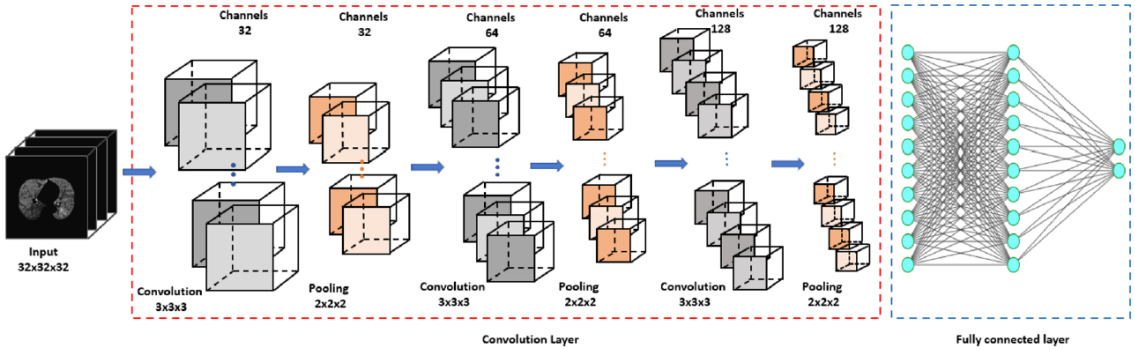


Figure 2.5: Diagram of a 3D CNN architecture for binary classification. [35]

As the required tasks become more difficult and nuanced, deeper networks with more trainable parameters will be required. Deeper networks can capture more complex features as the layers learn hierarchical features, with early layers capturing low-level features, such as edges and textures, and

deeper layers capturing high-level features such as object parts and shapes. Issues arise from using deeper models, as error gradients can accumulate during an update and result in exploding gradients. These in turn create large updates to the network weights, resulting in an unstable network that diverges instead of converging. Vanishing gradients can make models harder to train as the large number of layers with many activation functions can bring the loss function gradient to zero, making ineffective learning.

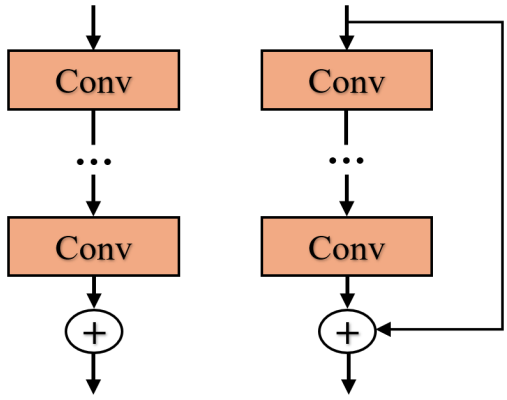


Figure 2.6: Comparison of a classic CNN block (left) and a ResNet block with a skip connection (right). As seen on the skip connection, parts of the input can flow through the model and not be affected by the convolutional layer

Several advancements have been made to mitigate these issues. One such advance is skip connections (see Fig. 2.6), which are used in residual CNN networks, such as ResNets [32]. Skip connections allow gradients to flow past certain layers to train the network more easily thus enabling the creation of much deeper CNNs without suffering from exploding gradients, but they remain prone to overfitting with increased depth. Popular ResNet architectures include ResNets-18, 50 and 101, with the number representing its depth and layer configuration, see Fig. 2.7. The residual blocks in a ResNet-18 consist of two convolutional layers each, with batch normalization and ReLU activations. The skip connections add the input of each block directly to the output of the block, helping preserve the gradient flow during backpropagation. The overall architecture consists of some initial convolutional and maxpooling layers, followed by the residual blocks, and a final average pooling and fully connected layer. The ResNet-50 and its larger variants use a bottleneck

architecture for the residual blocks to improve computational efficiency and to better handle network depth. They consist of three convolutional layers instead of two: a 1x1 convolution layer to reduce the dimensionality, a 3x3 convolution layer to process the data, and another 1x1 convolution layer to restore the dimensionality. Likewise, each of these convolutional layers is followed by Batch Normalization and ReLU activation. The skip connection then adds the input of the block to the output of the third convolutional layer. These models are adaptable to 3D inputs by simply using the 3D variant of the operations of each layer.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 2.7: Detailed layer information for various 2D ResNet variants with different network depths [32]

2.6.2 Global-Local Learning

Due to inductive bias in CNN, some CNN-based DL methods have proposed to leverage both global and local context by using one model for the entire image input and another for its local patches, with their final outputs combined to improve prediction tasks. Specifically, the global models take the entire image as input, providing the general spatial relationship of image features across the field-of-view. The local models take image patches as input to only consider local context for a better understanding of fine-grained details.

2.6.3 Transformers

To better model the spatial context of image features across the image, Vision Transformers (ViTs) are a class of DL models that have gained prominence in medical image analysis due to impressive results. ViTs are inspired by transformers conceptualized for natural language processing (NLP). Unlike CNNs, ViTs divide images into patches and treat them as sequences of tokens, similar to words in a sentence in natural language processing. They utilize the self-attention mechanism, which learns the importance between all pairs of tokens to understand the global context of an image. ViTs excel in capturing the global context of images, which is often crucial in medical imaging for accurate diagnosis.

2.6.4 Contrastive Learning

Contrastive learning (CL) has also been gaining popularity in medical imaging and achieved promising results in disease classification [37] and segmentation [9]. In its classic form, contrastive learning was designed as an unsupervised or self-supervised learning method. This creates a data efficiency advantage in areas, where data with high-quality annotation is sparse, as is the case for medical image segmentation. Typically, two samples are compared, an anchor and a "positive" sample are pulled closer in the embedding space, and a negative sample is pushed from the anchor. For cases with no labels, positive samples consist of an augmentation of the anchor, and negative samples are other random cases from the same dataset.

The main difference between self-supervised and supervised contrastive learning is the ability of the latter to use labels on all samples to learn. As seen in Fig. 2.8 with the black and white dog (bottom), taking class labels into account results is a more accurate embedding space, where samples from the same class are pulled closer together. Supervised contrastive learning has an information advantage over its counterpart when the data is labeled in its entirety. This is useful in areas with larger quantities of well annotated data. This method has achieved top-1 image retrieval accuracy of 81.4% on the ImageNet dataset, outperforming ResNets trained with cross entropy loss and unsupervised representation learning [45].

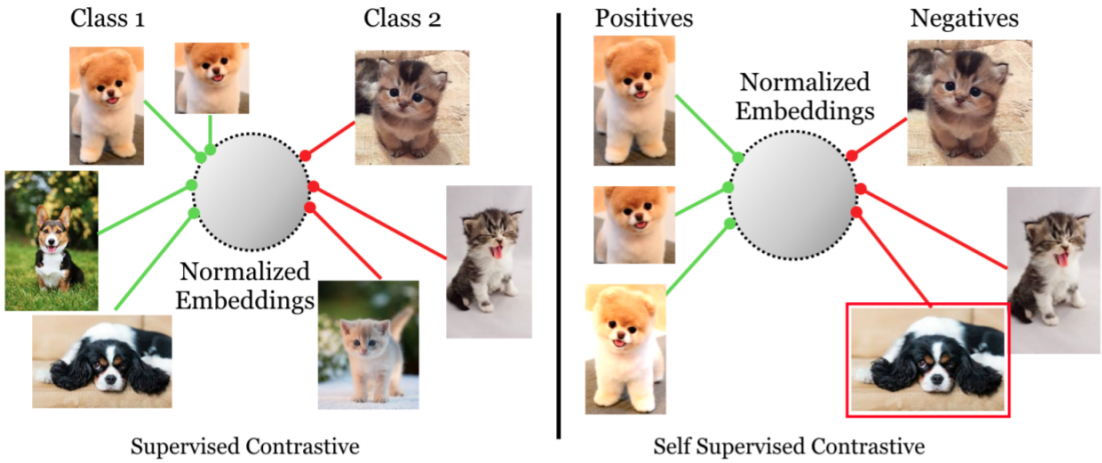


Figure 2.8: Diagram comparing contrastive losses: Supervised contrastive loss (left), contrasts all samples from the same class as positives against the negatives from the batch, augmentation of the anchor is also possible. Self-supervised loss (right) considers an anchor and various augmentations of the same sample and a set of negatives consisting of the entire remainder of the batch without class information [45] .

Contrastive learning has often been applied in classification tasks. However, these concepts can be employed to regression tasks as well. A novel method was presented in "Rank-N-Contrast: Learning Continuous Representations for Regression" by Zha et al. [91], where continuous valued labels are handled by introducing a ranking-based contrastive loss. Specifically, a ranking mechanism is used that orders samples, to which a contrastive loss that respects this ordering is applied. This allows the model to learn representations, such that the distance in the embedding space reflects the relative differences in the target values. Figure 2.9 shows the learned representations of different loss methods, showcasing the continuous representation of RNC. This is achieved by contrasting pairs of samples according to their rank order, encouraging the model to place samples closer together or further apart based on the similarity of their continuous target values. This method allows the model to capture the underlying continuous structure of the data, making it particularly suitable for tasks like brain age estimation, where the goal is to estimate a continuous age value rather than to categorize into discrete classes.

The Rank-N-Contrast (RNC) has shown SOTA performance in other image processing domains such as face age estimation with the AgeDB [62] and IMBD-WIKI datasets [80], containing 16,488 images of celebrities, and temperature prediction with the SkyFinder dataset [59], which contains

35,417 images captured by 44 outdoor webcam cameras. Training on sub-sampled datasets showed that RNC was more robust against reduced training data size compared to L1 loss[91]. This can be particularly useful for medical image DL, where acquiring large datasets can be challenging. One drawback is the large batch size required to obtain optimal results [91], as it allows a greater number of pairs per anchor to be contrasted. This can cause memory complexity issues with large 3D medical images.

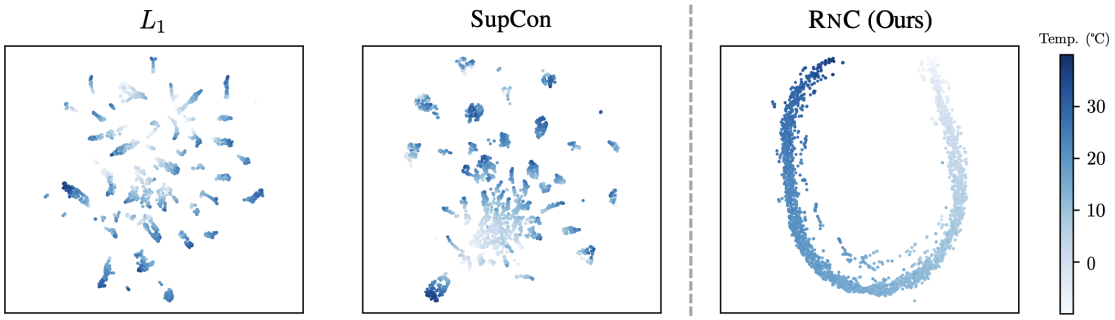


Figure 2.9: Learned representations of L1 (MAE) loss, SupCon loss [45] and RNC loss on a temperature regression task [91]

2.7 Data Augmentation

Data augmentation can reduce overfitting to training data and help increase the performance of a model without changing the architecture or adding more data. This widely used technique consists of generating new training examples by applying various transformations to the existing data. The goal is to increase the diversity in the data, incorporate random noise, or create synthetic data points, allowing the model to become less biased towards the training sample. A synthetic dataset can be created and incorporated with real data, which is particularly useful for small datasets. Another option is to randomly apply data augmentation to a portion of the training set during each training epoch to create variety.

A wide range of augmentation techniques can be used, as illustrated in Fig. 2.10. Geometric transformations include rotating images by random angles, translation, horizontal and vertical flipping, random cropping, and resizing. These transforms help the model generalize to various poses and scales of the image, which is particularly useful for datasets registered to a template space (e.g.,

brain MRI in the MNI-ICBM152 space). Intensity transformations aim to make the model invariant to different contrast conditions and noise. The intensity of the pixels or voxels can be randomly adjusted altogether with brightness and contrast transforms, or Gaussian noise can be injected to simulate noisy scans. In natural image datasets, color jittering has a similar effect.

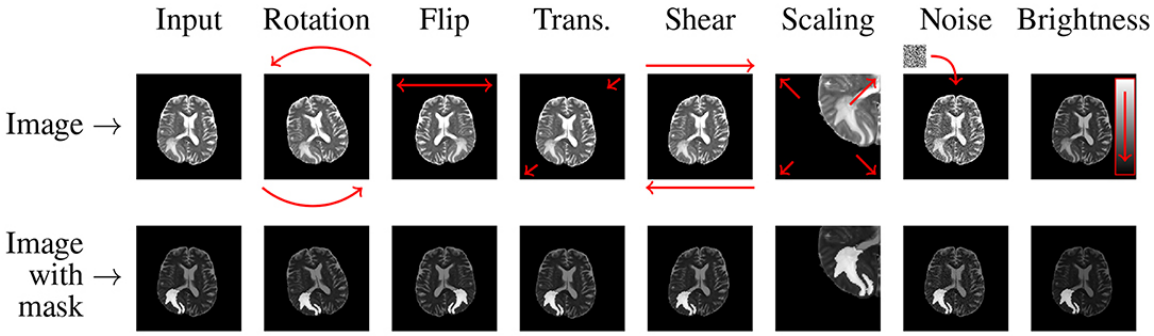


Figure 2.10: Examples of various data augmentation techniques applied to a T2w brain MRI [66]

Some more advanced methods exist to create synthetic data. For example, Generative Adversarial Networks (GANs) can be used to generate images that are nearly indistinguishable from the real ones. Mix-up involves creating new training samples by taking linear combinations of pairs of images and their labels. However, these methods can be difficult to apply in radiological settings, as creating completely synthetic variations to mitigate insufficient data could render the dataset non-descriptive of medical reality.

Data augmentation has been shown to contribute to significant improvements in various medical imaging tasks, such as tumor detection [66], disease classification and detection [14], and brain age prediction [67]. By artificially adding variability to the existing training dataset, data augmentation helps mitigate the pressing challenges posed by the limited availability of annotated medical images and possible biases towards the training population.

2.8 Brain Age Estimation

The concept of biological age is an estimation of individual physiological age, based on various biomarkers in the entirety of the body, or specific organs [46, 65]. Brain age is a biological biomarker that estimates the age of the brain based on anatomical or functional features observed

through neuroimaging techniques. These are traditionally morphological features such as cortical thickness, brain volume and the integrity of various brain regions. The brain age gap can be defined as the difference of an estimated brain age and the chronological age of a patient. A positive gap suggests a less healthy, more aged brain than what is expected for a person of a given chronological age while a negative one shows a healthier and less aged brain. Several studies have developed accurate models to estimate brain age due to natural aging and/or neurodegenerative diseases by examining the brain's anatomical structures [84]. The computed brain age and brain age gap can serve as potential biomarkers to assess an individual's brain development and/or neurodegeneration state, and thus facilitate personalized treatment.

The vast majority of brain age estimation (also referred to as brain age prediction in the literature) models are trained on a healthy population to predict their chronological age with high accuracy [84, 86, 67]. These models can then be tested on other healthy individuals to assess their performance. Downstream applications include applying the model to classify patients with neurodegenerative diseases [67, 88] or to find correlation between brain age and the severity of their condition [5, 90]. Ideally, the brain age gap should be positively correlated with the severity and progression of the disease. Training the model on healthy populations without controlling for patient demographics and genetics (if applicable) can introduce bias in the model and result in sub-optimal accuracy in real deployment. In addition, uncertainty can also arise from deep learning models with limited training samples and variety, and thus such factors should be considered in the interpretation of performance metrics reported. T1w MRIs are commonly used for structural neuroimaging analysis, with multiple widely used datasets, such as the Human Connectome Project, UK Biobank, ABIDE, OASIS, and IXI. Metrics to evaluate MRI-based brain age estimation DL models usually include the mean absolute error (MAE), which measures the average magnitude of errors in the predictions, and the coefficient of determination (R^2), which measures the proportion of the variance in the observed data that is explained by the independent variables. DL models have shown success in this target task, with SOTA MAEs ranging from 2.3 [71] to 4 years and R^2 values of 0.90 to 0.97 for aging populations [86], though this is highly dependent on the datasets used and tested.

Deep Learning Brain Age Estimation Models

DL approaches to brain age estimation generally use full brain T1w MRIs with minimal preprocessing. As opposed to ML which used hand picked features, DL approaches allow the model to extract features through training. Using the whole image, DL models may aim to measure the morphological properties that match trends of healthy chronological aging considering all potential factors, with the goal of identifying key common patterns that help achieve an accurate prediction.

CNNs have been foundational for applications of DL in the domain. A pioneering study by Cole et al. [11] demonstrated that 3D-CNNs could predict brain age of aging healthy subjects with an MAE of around 4 years and R^2 of 0.94, outperforming most traditional machine learning (ML) methods. Optimized CNNs have drastically improved the current SOTA by reaching MAEs as low as 3.90 [47] and 2.14 years[71] on healthy subjects.

Several techniques have used CNN models as backbones for feature extraction, with more advanced methodologies used to analyze the features. An example of this approach is the Global-Local Transformer for Brain Age Estimation by He et al. [33], which processes brain images through two parallel pathways: one focusing on the entire brain and the other zooming into specific regions. The information extracted is then processed with a transformer for brain age estimation. The authors achieved a mean absolute error (MAE) of 2.7 years on a dataset population spanning from infants to elderly (0-97 years old). While the application of ViTs in medical imaging has shown promise, no dedicated studies exist specifically for brain age prediction using ViTs.

One study used a supervised contrastive learning model with a custom Y-aware regression loss (different from the RNC loss) [4]. This model achieved an MAE of 2.55 years on a dataset biased towards younger individuals (mostly around 20 years old), using gray matter volumes from voxel-based morphometry MRIs.

2.9 Explainability for Deep Learning Models

DL explainability refers to the ability to understand, interpret, and trust the decisions made by DL models. Given the usual 'black box' nature of these models, explainability is crucial in ensuring that model predictions and decisions are justifiable in risk-sensitive deployment (e.g., clinical

diagnosis). In addition, the ability to explain the automatic decisions from different angles will allow further improvement of model architecture and explore novel insights into the target task. For example, localizing key brain structures in MRI-based brain age estimation. Typically, explanation for DL models can be achieved through techniques like feature importance, layer-wise relevance propagation, and visualization of neural network activations.

2.9.1 Gradient-weighted Class Activation Mapping

Gradient-weighted Class Activation Mapping (Grad-CAM) [82] is a technique used to provide visual explanations for CNN decisions. Grad-CAM uses gradients of a target concept flowing into the final convolutional layer to generate a localization map, highlighting important regions in the image for prediction. This method offers insights into the decision-making process of the model by indicating which parts of the input image are influential in terms of the model prediction. Grad-CAM is tailored to each class and subject, showcasing different activation patterns for different concepts. It is commonly implemented in classification tasks, using backpropagation through the correct class output dimension to visualize relevant activations. Figure 2.11 shows the Grad-CAM activation for the class "West Highland white terrier" of a 2D CNN (VGG-16) trained on the ImageNet dataset, with the heatmap correctly focusing on the dog itself rather than the background objects.

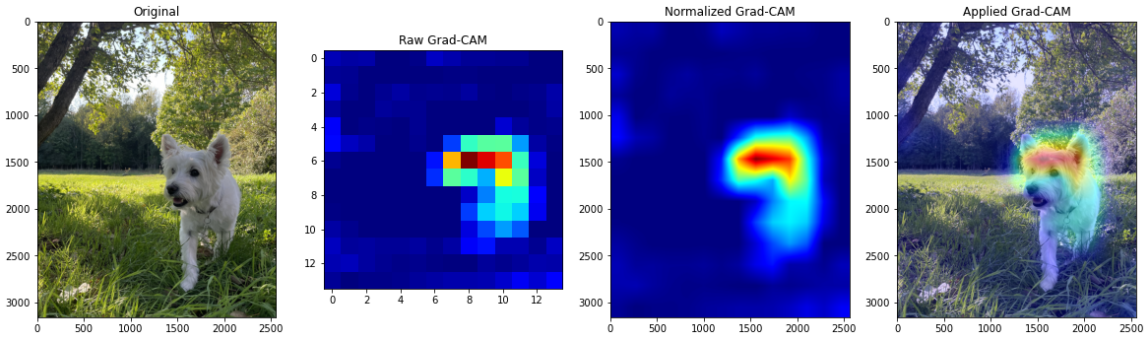


Figure 2.11: Visual explanation of an image classified as "West Highland white terrier" by a VGG16 CNN model trained on the ImageNet dataset.

In natural image processing, challenges arise in creating averaged Grad-CAM heatmaps across diverse scenes to probe any consensus in image patterns of a specified image class. However,

in medical scans such as brain MRIs, where registration towards a common anatomical atlas can be achieved, we can perform group-wise averaging to reveal common patterns of the GradCAM insights derived from DL models. This may offer unique insights for medical image analysis besides the more classic voxel-based and deformation-based morphometric analyses.

The following steps are followed to achieve these results:

1. **Forward Pass:** Compute the score for a target class y^c before the softmax layer by propagating an input image through the CNN..

2. **Gradients of the Target Class:** Calculate the gradients of the score for class c with respect to the feature maps A^k of a chosen convolutional layer (typically the last one):

$$\frac{\partial y^c}{\partial A^k}$$

3. **Global Average Pooling of Gradients:** Perform global average pooling on the gradients to obtain the weights α_k^c:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

where Z is the number of pixels in the feature map A^k, and i and j are the spatial dimensions.

4. **Weighted Combination of Feature Maps:** Compute the weighted combination of the feature maps using the gradients as weights:

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right)$$

The ReLU activation applied ensures that only positive influences on the class score are considered, effectively filtering out non-contributory features.

2.9.2 Gradient-weighted Regression Activation Mapping

Gradient-weighted Regression Activation Mapping (Grad-RAM) is an adaptation of Grad-CAM for regression tasks. While Grad-CAM is traditionally used for classification, Grad-RAM extends

this approach to handle continuous outputs, adapting it to brain age estimation tasks. By computing the gradients of the output with respect to the final convolutional layer, Grad-RAM generates a similar heatmap that highlights the areas of the input that most influence the regression output. This technique enables researchers to visualize which brain regions are contributing to the predicted age, offering valuable insights into the biological processes associated with aging.

1. **Forward Pass:** Compute the regression output y by performing a forward pass through the network.

2. **Gradients of the Regression Output:** Calculate the gradients of the regression output with respect to the feature maps A^k:

$$\frac{\partial y}{\partial A^k}$$

3. **Global Average Pooling of Gradients:** Perform global average pooling on the gradients to obtain the weights α_k:

$$\alpha_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y}{\partial A_{ij}^k}$$

where Z is the number of pixels in the feature map A^k, and i and j are the spatial dimensions.

4. **Weighted Combination of Feature Maps:** Compute the weighted combination of the feature maps using the gradients as weights:

$$L_{\text{Grad-RAM}} = \text{ReLU}\left(\sum_k \alpha_k A^k\right)$$

or:

$$L_{\text{Grad-RAM}} = \sum_k \alpha_k A^k$$

The method with the ReLU operation will limit the visualization of features with a negative influence on the regression output. The appropriate method will therefore depend on the application and the purpose of the visualization.

2.9.3 Explainable DL Applications in Brain Age Estimation

Grad-CAM and Grad-RAM have been implemented in brain age models to visualize the regions most related to aging, with findings fairly consistent with classic morphometric analysis [92, 43, 90]. Other explainability methods have been utilized with similar success. U-Net-based approaches have demonstrated comparable outcomes [67, 27]. Patch-based models can compute a brain age estimate for each local region, identifying those that show greater or lesser aging effects and employing patch weighting to determine their significance in overall age prediction [33].

2.10 Summary

Understanding the intricate connections between neuroanatomy and aging is important for advancing our knowledge of the brain and neurodegenerative diseases. Alzheimer's disease (AD) and Parkinson's disease (PD) pose unique challenges in terms of diagnosis and treatment. Personalized anatomical brain biomarkers offer an opportunity to enhance current diagnostic methods, severity metrics, and treatment plans. Brain age estimation can also shed light on how factors such as sex, genetics, and environmental conditions influence brain aging.

Deep learning, particularly CNNs, consistently achieve SOTA results in estimating brain age using MRI data. However, many advanced DL techniques widely applied in other domains have yet to be fully explored in this field. Further studies are needed across diverse datasets to determine the optimal model types, preprocessing steps, data augmentation strategies, and training methods.

Explainability of DL models is also important for interpreting model decisions and identifying anatomical features critical for understanding aging processes in healthy and diseased populations.

In summary, the intersection of neuroanatomy, neurodegenerative disease research, and advanced DL techniques holds great promise for enhancing our understanding and clinical management of brain aging and related diseases. In the following chapter, we present using supervised contrastive learning with Rank-N-Contrast (RNC) loss and Grad-RAM for explainability utilizing structural T1w MRI data to further explore the use of DL on brain age prediction and its association with AD and PD.

Chapter 3

Brain Age Estimation With Data-Efficient Supervised Contrastive Learning and Grad-RAM-based Visual Explanation

3.1 Introduction

Chronological age is commonly used to describe patients or define health patterns in populations, however, it does not take into account individual deviations from the normal aging progression that could be due to disease, genetics, or external factors [21, 22]. In contrast, measuring biological aging offers a more precise metric for clinical applications [46]. This approach is particularly useful for studying the human brain, which undergoes significant anatomical and functional changes due to aging [73, 25]. Further, most neurodegenerative diseases are characterized by anatomical changes similar to advanced or accelerated brain aging [15, 41, 12]. Thus, the difference between an estimated brain age and chronological age (brain age gap) can provide a valuable biomarker and enable better management and understanding of neurological diseases [11].

A growing body of research has shown that deep learning (DL) models can accurately predict

the age of healthy patients using structural MRIs [86, 71, 47], with many methods employing CNN-style architectures [44, 68, 90]. However, several challenges and open questions remain. First, there is a lack of comparative studies between various DL models due to the diversity of datasets used. Additionally, the inherent black-box nature of CNNs hinders straightforward interpretability, resulting in no consensus on brain region saliency concerning aging. Moreover, there is no agreement on the utility of brain age as a biomarker for neurodegenerative diseases [21, 42], nor on whether DL models can accurately associate aging and age-related decline with relevant neuroanatomical regions.

Motivated by these knowledge gaps, we aim to advance the field by applying recent advancements in CNN models from other domains to brain age estimation. Supervised contrastive learning, for example, has been shown to be more accurate than traditional cross-entropy in many applications [45, 91, 37, 9]. In this paper, we investigate the application of supervised contrastive learning with Rank-N-Contrast (RNC) loss [91] for brain age estimation, using a dataset of 1,618 T1w MRIs. We compare this approach to a commonly used ResNet model and assess the impact of various parameters on performance, using cohorts with neurodegenerative conditions to evaluate correlation with disease severity and employing Grad-RAM for anatomical feature analysis.

3.2 Related Work

3.2.1 Aging-Related Anatomical Changes

Studies have shown statistical correlations between certain regions and natural aging, but differences exist across studies and the methods used. For example, Fjell et al. [20] used cortical grey matter (GM) thickness measures to reveal that the superior and inferior frontal gyri and the superior parts of the temporal lobe are subject to aging-related cortical thinning. Later, further evidence was found for atrophy in subcortical regions with a similar method [19]. Mu et al. [63] found that the hippocampal formation, the amygdala, and the temporal horn all had significant volume decreases with older age. The hippocampus loss in volume was further reinforced by Convit et al. [13]. A review by Gunning-Dixon et al. showed that prefrontal cortical white matter (WM) thinned with age [30]. A more recent study by Fujita et al. on a 653-patient longitudinal dataset confirmed that

each brain structure showed characteristic age-dependent volume and atrophy changes. In particular, cortical GM showed consistent volume loss, WM showed an accelerated atrophy rate with age, and volume increased in the lateral ventricle [25]. Another recent study by Dima et al. [16] on a dataset with 18,605 subjects found significant association between aging and the volumes of the basal ganglia, thalamus, amygdala and the hippocampus for subjects over 60 years old (yo), while the lateral ventricles showed continuous enlargement throughout the lifespan.

3.2.2 Brain Age Estimation Methods

Over the last decade, significant work has been produced to create accurate brain age models using neuroimaging data [21]. Structural T1w MRI data are the main modality and contrast used [61], with machine learning (ML) and deep learning (DL) approaches able to produce accurate predictors of brain ages [84]. Typically, a brain age estimation model will be trained on healthy subjects and aim to be as accurate as possible at predicting the subjects chronological ages. Initially, more traditional ML regression algorithms used engineered features extracted from MRIs, such as cortical thickness and volumetric measurements. For instance, Franke et al [23] used relevance vector regression on segmented GM data and achieved a mean absolute error (MAE) of 5 years. Liem et al. [49] used linear support vector regression models on multimodal data, including cortical thickness, cortical surface area, and subcortical volume measurements to achieve an MAE of 4.29 years. A more recent approach by Han et al. [31] used regularized linear regression on extracted morphometric features and achieved an MAE of 3.7 years.

The introduction of DL has enabled the use of entire structural T1w MRIs, leading to significant performance improvements in models within this domain [84]. A model developed in 2016 by Cole et al. [11] demonstrated that 3D-convolutional neural networks (CNNs) trained on brain structural MRIs could predict brain age with an MAE of around 4 years with an R^2 of 0.94, outperforming most traditional ML methods at the time. Advancements to CNN architectures have also drastically improved the current state-of-the-art (SOTA) accuracy; Yin et al. [90] achieved an MAE of 2.3 years with a customized 3D CNN, Leonardsen et al.[47] and Peng et al. [71] achieved 3.90 years and 2.14 years respectively with SFCN architectures (VGG-like). MAEs of 3.39 [44] to 2.7 [68] have also been achieved with residual CNNs (ResNet). Techniques using CNN backbones as feature

extractors have proven to be more or equally accurate. He et al. [33] presented a global-local approach with ResNet backbones, coupled to a transformer to reach an MAE of 2.7 years for a population with an age range of 0-97 yo. A U-Net style approach proposed by Nguyen et al. [67] achieved an MAE of 3.83 years on an aging cohort and 1.91 years on a young one. Lastly, one study that used a supervised contrastive learning method with a custom Y-aware regression loss achieved an MAE of 2.55 years on a mostly young dataset using gray matter volumes from voxel-based morphometry (VBM) [4]. To the best of our knowledge, no supervised contrastive approach has been used directly on T1w structural MRIs.

3.2.3 Applications to Neurodegenerative Diseases

Several papers[61] have suggested that brain age estimation techniques could be used to detect or quantify the severity of certain neurological diseases. One observation is that the mean brain age gap between the biological and chronological ages of a disease population is typically higher compared to a healthy one. Sendi et al. [83] found a mean age gap of +3.3 years in their AD sample using support vector regression. Similarly, Eickhoff et al. [17] found a mean age gap of +2.8 years in their PD sample using an ensemble of linear support vector regression models. Nguyen et al. [67] used a U-net and showed an increased brain age gap for neurodegenerative patients with means of +3.3 years for AD and a more modest +1.3 years for PD. They then computed the brain age gap of specific brain structures to classify patients with neurodegenerative diseases. ML approaches using morphological data found that brain age could help predict mild cognitive impairment (MCI), conversion to AD with an accuracy of 86% [26] and classify healthy and AD patients with an accuracy of around 90% [88].

Investigations into correlations between brain age and neurodegenerative disease cognitive assessments have yielded varied results. Studies using ML on processed MRIs have found correlations between brain age gap and ADAS-cog 11 scores of 0.25 to 0.56 by Lowe et al. [51] and 0.26 by Beheshti et al. [5]. Yin et al. [90] found that their CNN produced brain age gaps for their MCI and AD sample that correlated with neurocognitive measures such as ADAS, CDRSB, or FAQ. For PD, [17] found a weak correlation (0.14-0.17) between a predicted brain age gap and UPDRSIII, using a model based on gray and white matter features [17]. Others have found no correlation between

PD and UPDRSIII [10]. While these findings are promising, more studies are required to truly understand the potential clinical use of brain age as a biomarker, especially with DL approaches.

3.2.4 Explainability Methods

Due to their black-box nature, explainability techniques must be applied to CNN-based models to understand and map the regions with the strongest relation to aging. Grad-CAM/Grad-RAM has been implemented on DL CNN models with findings mostly consistent with anatomical regions known to alter with age [43]. Joo et al [43] found that the corpus callosum, internal capsule, and brain regions adjacent to the lateral ventricle were most important in predicting age for all healthy subjects. Yin et al. [90] used a similar saliency map method and found the ventricles, frontal, temporal, and hippocampal cortices to be most salient in CN aging, with a shift to cerebral WM for cognitively impaired patients. He et al. [33] estimated patch saliency by weighing the model's patches based on accuracy and found a shift in brain region over age.

Previous methods produced relatively blurry heatmaps, but recent advancements have attempted to address this issue by estimating saliency at the voxel level. Occlusion-map based techniques by Bintsi et al. [6] had findings consistent with the existing literature, highlighting that the hippocampus and the ventricles are the most relevant regions. Certain methods (U-Net style) perform brain age estimation at the voxel level which directly creates the saliency maps, this is particularly useful to see which regions contribute to an older or younger brain age. Using this approach, Gianchandani et al. [28] found higher brain age gaps in the caudate nuclei which resonates with lower caudate nuclei volumes in AD compared to healthy control subjects [52]. Nguyen et al. [67] found that in their AD group regions surrounding the hippocampus show the most accelerated aging, a region well-known to be related to AD [38, 24, 64]. For their PD cohort, all regions seem to be similar to the healthy populations. Explainability features are essential for model confidence, especially in future clinical applications, to provide insights and to understand areas associated with aging for neurodegenerative patients. However, no consensus exists on which method is most accurate or brain regions are most salient.

3.3 Methods and Materials

3.3.1 Dataset

We collected a total of 1,618 healthy subjects' 3T T1w brain MRIs with the corresponding chronological age and sex (age=52.8±18.5 yo, range=20-100yo, 872 females) for the development and testing of the proposed method from five publicly available data sets, including the Human Connectome Project Aging (HCP) [7], Alzheimer's Disease Neuroimaging Initiative (ADNI) [40], IXI Dataset3.1, OpenNeuro (Neurocognitive aging data release with behavioral structural and multi-echo functional MRI measures [85]), and Parkinson's Progressive Markers Initiative (PPMI) [57]. The details for each dataset and totals are provided in Table 3.1.

We used a train/validation/test split with 80%/10%/10% of the full data respectively. More specifically, the test dataset (149 subjects with 79 females, age=51.26±19.32 yo) have 41, 46, 47, and 15 subjects in the 20-40 yo, 40-60 yo, and 60-80 yo, and 80+yo age groups, respectively. To further explore the application of the proposed method in gauging progression of neurodegenerative conditions, we have also collected 61 Alzheimer's disease patients (age=78.1±7.3yo, 22 female) with ADAS-cog 11 score of 25.56±9.46 from the ADNI database, and 80 Parkinson's disease patients (age=66.8±9.1yo, 24 female) with the UPDRSIII and H&Y scores of 27.91±13.88 and 2.28±0.58, respectively. Specifically, for the AD patients, we selected those that were scanned 24 months after their baseline visits, and PD patients's data 48 months after baseline were selected. For both patient cohorts, we ensure a variety of mild and severe disease cases, and a variety of age ranges.

Table 3.1: Statistics of data sets used for cognitively normal, AD, and PD.

Dataset	$N_{samples}$	Mean Age ±Std.Dev.	Age range	Sex(F/M)
◊ IXI	562	48.7 ±16.5	20-86.32	311/251
◊ HCP	635	59.8 ±15.5	20-100	355/280
◊ OpenNeuro	223	41.0 ±23.3	20-89	129/94
◊ ADNI NC	136	78.0 ±6.5	69-93	55/81
◊ PPMI NC	62	61.4 ±9.8	40.2-81.1	22/40
CN Totals	1618	52.8 ±18.5	20-100	872/746
△ ADNI AD	61	78.1 ±7.3	59-91	22/39
△ PPMI PD	80	66.8 ±9.1	43-86	24/56

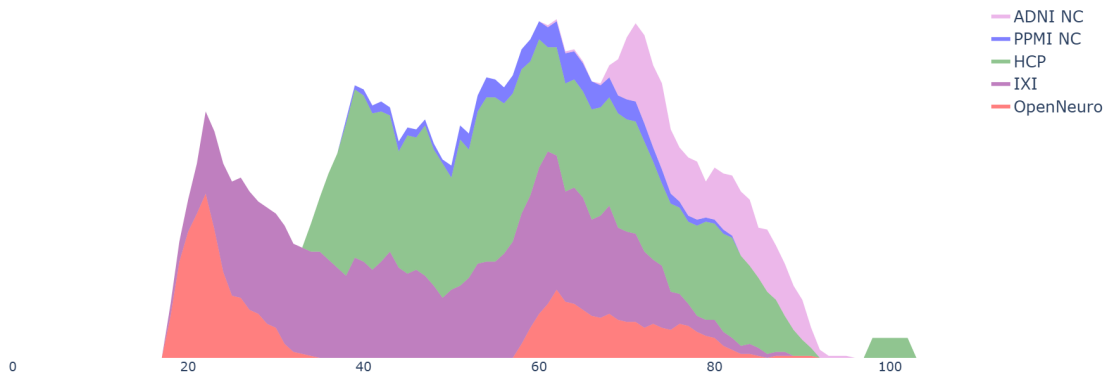
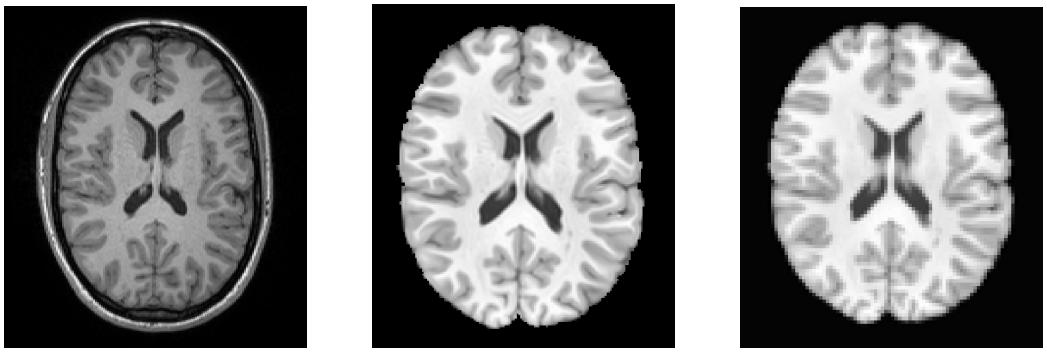


Figure 3.1: Visualization of relative age distribution of cognitively normal patients per dataset (n = 1618)

3.3.2 Preprocessing

All collected T1w MRIs were preprocessed with the MINC Toolkit (https://bic-mni.github.io/) to ensure data quality and facilitate image analysis. Specifically, we obtained the brain mask with the BEaST algorithm [18] and performed non-uniformity correction within the constraints of the brain mask using the N4 inhomogeneity correction algorithm [87] for each scan. Afterwards, the skull-stripped MRIs were registered and resampled to the MNI ICBM space with a 12-parameter affine transformation using normalized cross-correlation as the cost function. The final images are of dimension 193x229x193 voxels with 1x1x1 mm^3 resolution. To help with data efficiency, we also generated a downsampled version of the same dataset at 2x2x2 mm^3 resolution and size of 96x114x96 voxels. A visual quality control is performed on each sample. An example of the images can be seen in Fig. 3.2



(a) Raw T1w (b) Preprocessed 1x1x1 mm^3 T1w (c) Preprocessed 2x2x2 mm^3 T1w

Figure 3.2: Axial slice of raw, preprocessed, and resized data for comparison

Data Augmentation

To improve learning efficiency and mitigate overfitting, we adopted four data augmentation techniques for the training dataset, each with a 50% probability of being applied, including 1) random affine translation of ±10 voxels in each direction, 2) random rotation transformation of 0.1 to 0.5 radians in any direction, 3) random Gaussian noise addition with 0 mean and 0.025 standard deviation, and 4) random image cropping with the size no smaller than 70% of the original image (and is resized to the original image dimensions).

3.3.3 Deep Learning Models and Experimental Setup

Overview of experimental setup

We propose to use the ResNet models [32] as the backbone to build our proposed DL algorithm for MRI-based brain age estimation, with the RNC contrastive learning technique. To optimize the target method, we inspected the performance of ResNet depth, data augmentation, and image resolution for different variants of the ResNet architecture, which are compared against our contrastive learning approach, together with two other SOTA techniques. In addition, we implemented the Grad-RAM technique to provide visual saliency maps for the DL-based age estimation results to reveal insights regarding brain aging. Finally, an AD and PD cohort was used to assess the model's abilities on subjects with neurodegenerative conditions, where correlations between brain age gaps

and disease severity were evaluated. All models were trained on the AllianceCan Narval cluster with 4 Nvidia A100-40gb GPUs.

Backbone Model

We evaluated ResNets [32] of varying depths, specifically 18, 50, and 101, to determine the optimal backbone architecture for our intended contrastive learning approach. Figure 3.3 shows the architecture for the basic ResNet-50 model used in the study. Each version of the ResNet architecture underwent training for 300 epochs with early stopping to regress age estimates based on preprocessed T1-weighted MRI scans. MAE) was used as the loss function with Adam optimizer and a learning rate of 0.001. For images with resolutions of 2x2x2 mm^3, a batch size of 8 was used, while for 1x1x1 mm^3 resolution images, a batch size of 4 was employed. We conducted experiments with both augmented and unaugmented datasets to assess the impact on accuracy improvement.

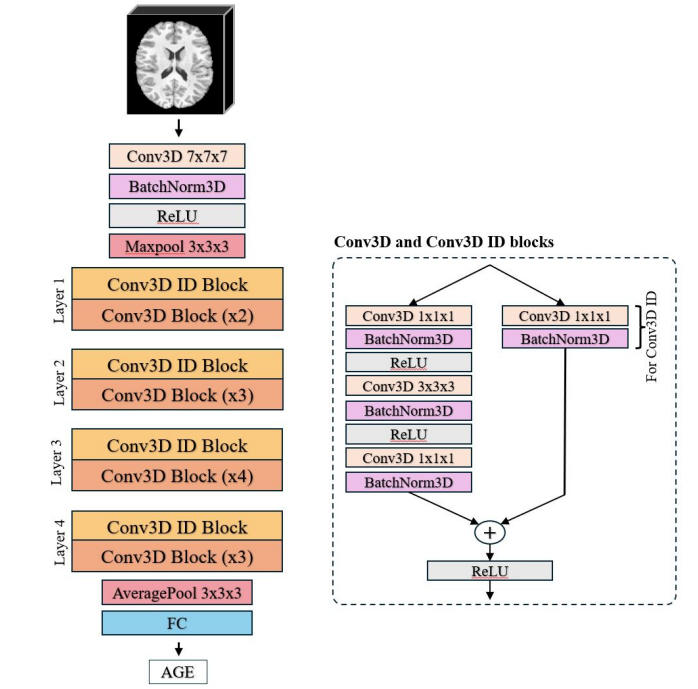


Figure 3.3: Overview of the 3D ResNet-50 architecture used for brain age estimation.

Contrastive Learning with Rank-N-Contrast

Contrastive learning has recently emerged as a powerful technique for learning representations, especially in image classification tasks, but its application in regression tasks has yet to be fully explored. This approach learns data feature embeddings by comparing the similarity and dissimilarity. A recent variation, Rank-N-Contrast, proposed by Zha et al. [91] is designed to handle continuous labels (e.g., brain age) by allowing the image encoder to learn representations, such that the distance in the embedding space reflects the relative differences in the target values. So far, it has provided excellent results in various regression tasks with natural images. We leverage this technique for MRI-based brain age estimation for the first time.

Specifically, for our application, the original Rank-N-Contrast [91] model architecture and loss function were modified to process 3D brain imaging data. The implementation of this method uses the same data, same split of 80%/10%/10%, and the same augmentation techniques described previously. To obtain the optimal backbone encoder, multiple architectures and hyperparameters were tested. Due to the memory intensive 3D data used, batch sizes, a crucial factor in contrastive learning had to be maximized to increase performance while not creating GPU memory issues. Thus, for the 2x2x2 mm^3 resolution data, a ResNet-50 backbone with batchsize of 48 was used and we name the resulting model RNC-lowRez while for 1x1x1 mm^3 resolution data, a ResNet-18 backbone with a batch size of 24 was employed and we name the model RNC-highRez. Note that ResNet-101 was omitted due to computational and memory complexity requirements. Model training hyperparameters included a learning rate of 0.5, decay rate of 0.1, and momentum of 0.9 with stochastic gradient descent (SGD) optimizer. The backbone encoders were trained for 1,000 epochs with the RNC loss. Afterwards, a regression head is added to the last layer of the backbone model, consisting of a fully connected (FC) to estimate the age from the encoded image features. In the last step, the backbone encoder weights were frozen and an MAE (L1) loss was used to train the FC layer, with the same training data for 100 epochs using the SGD optimizer (learning rate = 0.05, decay rate = 0.2, and momentum = 0.9). Figure 3.4 shows an overview of the training steps.

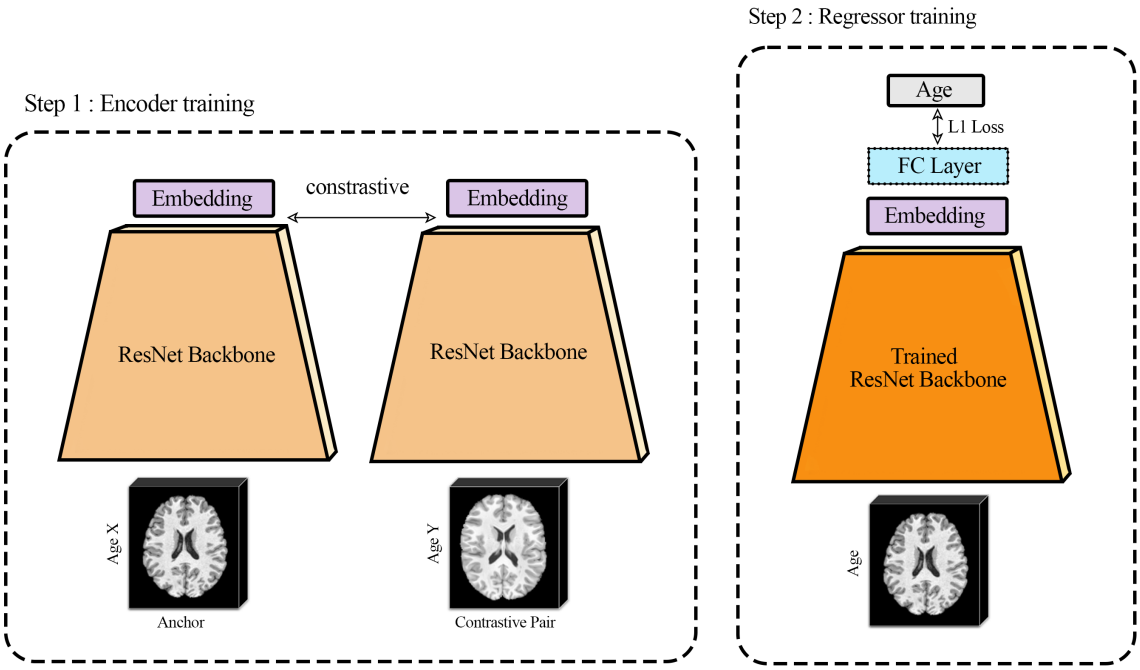


Figure 3.4: Overview of the training process for the RNC model supervised contrastive learning model, with the first step (left) training the ResNet backbone with the RNC loss (contrastive), and the second step (right) training a FC regression layer with MAE/L1 loss.

Comparison to the State-of-the-Art Methods

We compared our contrastive learning technique for MRI-based brain age estimation against two popular pre-trained public DL models, including the Simple Fully Convolutional Network with regression output (SFCN-reg) [47] and BrainStructureAges (BSA) [67], on the same test dataset. Note that to best reflect the accuracy of these models, we decided not to train these models from scratch based on our curated dataset, but directly used their pretrained weights. Furthermore, to mitigate the impact of domain shift that can degrade performance, we reprocessed our test data as instructed by each algorithm.

The SFCN-reg model [47] was trained on a large and diverse neuroimaging dataset consisting of 42,829 T1-weighted MRI scans from 21 public datasets, covering ages from 3 to 95. HCP is the only dataset used in both our method and SFCN-reg. In short, the model's architecture consists of a VGG-like structure with five repeated convolutional blocks, each including a three-dimensional convolutional layer, batch normalization, ReLU activation, and max-pooling followed by a global

average pooling layer and a single output predicting age. It achieved an MAE of 2.47 years on an in-domain test set and 3.90 years on an external dataset, making it one of the best-performing models in the literature. Our raw test dataset was preprocessed using the preprocessing script published by the authors on GitHub [47], and the images were subsequently ran on their trained model through their publicly available docker.

The BSA model [67] predicts a 3D aging map at the voxel level and subsequently computes specific brain structure ages using a 3D segmentation mask. The brain age of the structures is then used as a biomarker for various downstream applications such as disease classification. The model was trained on 2,887 T1w brain MRIs from various sources. ADNI is the only overlapping dataset with our study. The BSA model achieved an MAE of 1.88 years on a young population (14.8±9.3 yo) and 3.83 years on an older population (64.2±7.9 yo), achieving great accuracy. The trained model is hosted on the volBrain website, our raw test dataset was processed as per the instructions.

3.3.4 Evaluation metrics

To evaluate the accuracy for MRI-based brain age estimation, we used the mean absolute error (MAE), coefficient of determination, and Spearman correlation. When comparing the absolute errors between DL models based on the test set, two-sided paired-sample t-tests were conducted, and a difference with p-value below 0.05 is considered statistically significant.

3.3.5 Grad-RAM for Aging Bias Analysis

Grad-RAM visualization

Gradient-weighted Regression Activation Mapping (Grad-RAM) is an adaptation of Grad-CAM for regression tasks. While Grad-CAM is traditionally used for classification, Grad-RAM extends this approach to handle continuous outputs, adapting it to brain age estimation tasks. Grad-RAM generates a heatmap that highlights the areas of the input that most influences the regression output. In this implementation, a ReLU operation on the gradient activations allows the heatmap to solely represent areas that contribute to increasing the age values. The heatmaps are dependent on the input and output and vary per individuals; averaging the heatmaps of specific populations can reveal

the most important regions to their aging.

To reveal the insights provided by the DL models, Grad-RAM is performed on all test dataset subjects to obtain relevant saliency maps (with the normalized value range of [0,1] for each MRI). Averaged saliency maps were computed for different age groups to visualize the relevant anatomical regions that are closely related to aging at different stages of the lifespan. To facilitate the analysis, The Automated Anatomical Labeling (AAL)-16 atlas was used to help compute the mean Grad-RAM heatmap scores within each of the brain parcellations. Averaged Grad-RAM heatmaps and the associated parcellation-based quantitative values were obtained for the following four subgroups of our testing data: 20-40 yo, 40-60 yo, 60-80 yo, and above 80 yo. To further inspect the potential sex-based differences in brain aging, we performed a similar analysis by regrouping and averaging the Grad-RAM heatmaps with respect to sex. Note that the Grad-RAM results of both the RNC and best performing ResNet model are produced for comparison as well.

3.3.6 Analysis of Brain Age for Neurodegenerative Conditions

In neurodegenerative disorders, the patient's brain age gap between biological and chronological aging and disease severity could be correlated [17, 51, 5], and may serve as an indicator for disease progression. To explore the application of our proposed method in this domain, the T1w MRIs of the selected AD and PD patients from the respective ADNI and PPMI datasets were used in two different experiments. Note that for both experiments, we compared the proposed contrastive learning method and the best ResNet variant that were trained with the 1x1x1 mm^3 images.

In the first experiment, we estimated the biological brain age using the two designated DL models, and then we computed the correlation between the brain age gap, representing the difference between biological and chronological aging and the associated disease severity metrics. Specifically, for AD, we used the ADAS-cog 11 [79] score, which evaluates memory, language, praxis, and orientation, and is widely regarded as a gold standard for assessing cognitive function in Alzheimer's disease patients. On the other hand, for PD, we used the UPDRSIII [29] score that focuses specifically on motor symptoms, the primary clinical manifestations of Parkinson's disease, and is a more objective metric compared to other sub-evaluations of the UPDRS assessment. In the second experiment, population-averaged Grad-RAM heatmaps based on DL-based age estimation were obtained

to reveal the relevant anatomical patterns that are associated with aging under the influence of neurodegenerative conditions in comparison to healthy aging. The Grad-RAM heatmaps of the ADNI AD and PPMI PD patients were evaluated as a whole.

3.4 Results

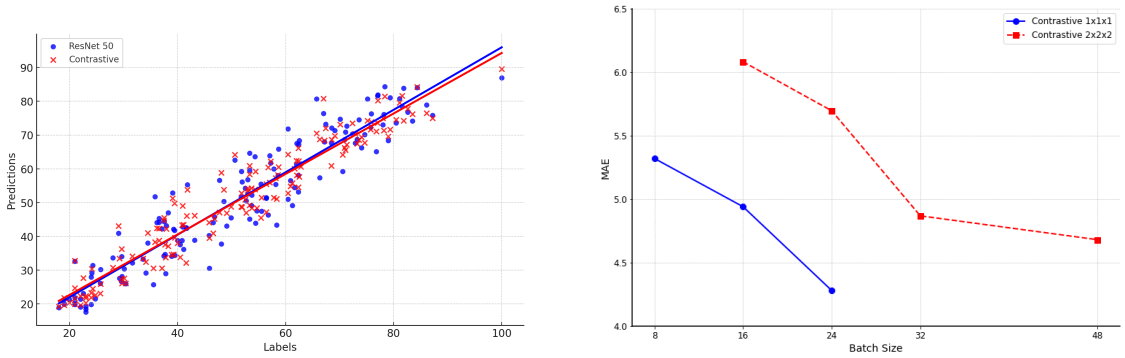
3.4.1 Evaluation of Deep Learning Models

Brain Age Estimation Accuracy

The results of the proposed contrastive learning method and the variants of the ResNets models for MRI-based age estimation are shown in Table 3.2, where results from both the high-resolution (1x1x1 mm^3) and lower-resolution (2x2x2 mm^3) are compared. In addition, we also compared the performance of the ResNet50 model with and without data augmentation during training. First, when examining the difference between the ResNet50 variants with and without data augmentation, we have observed an evident performance boost from data augmentation, particularly at the 1x1x1 mm^3 resolution. Secondly, with the incorporation of data augmentation, the depth of 50 for the ResNets has shown better performance compared to both the deeper 101-layer and shallower 18-layer versions. Notably, increasing the data resolution from 2x2x2 mm^3 to 1x1x1 mm^3 significantly enhanced model performance across all ResNet models ($p < 0.05$). Lastly, the RNC models exhibited the best performance across both resolutions, with the 1x1x1 mm^3 version outperforming all other models across all metrics with an MAE of 4.28 years, Std. Dev of 3.34, and R^2 of 0.93. As the sizes of the backbone ResNets for RNC models at different resolutions differ, the MAE and R^2 improvements are limited at the higher resolution. In addition, as shown in Fig. 3.5, the batch size had an important impact on the RNC model. The results with the largest possible batch size without a computational memory issue are presented in Table 3.2, with 48 for 2x2x2 mm^3 and 24 for 1x1x1 mm^3. Based on this investigation, we used ResNet50-highRez and RNC-highRez for the rest of our experiments and analysis.

Table 3.2: Accuracy assessments of MRI-based brain age estimation of ResNet and RNC models with different image resolutions and data augmentation conditions. (* indictates $p < 0.01$)

Model	MAE ± Std. Dev.	R^2	Correlation
1x1x1 mm^3			
ResNet50 w/o Data Augmentation	6.2 ±5.25	0.74	0.86*
ResNet18-highRez	5.10 ±4.19	0.87	0.93*
ResNet50-highRez	4.93 ±3.90	0.90	0.94*
ResNet101-highRez	5.92 ±5.14	0.84	0.92*
RNC-highRez	4.28 ±3.34	0.93	0.96*
2x2x2 mm^3			
ResNet50 w/o Data Augmentation	6.4 ±6.11	0.79	0.84*
ResNet18-lowRez	6.03 ±5.09	0.79	0.88*
ResNet50-lowRez	5.45 ±4.94	0.82	0.89*
ResNet101-lowRez	6.62 ±5.82	0.74	0.87*
RNC-lowRez	4.68 ±3.88	0.90	0.92*



(a) (b)

Figure 3.5: (a) Predictions of the ResNet 50 w. Data Aug and the RNC model brain ages vs chronological ages (b) Batch size vs. MAE for RNC models trained on 1x1x1 resolution data and 2x2x2 mm^3 resolution.

Comparison with State-of-the-art methods

Table 3.3 compares the performance of the best performing ResNet50-highRez and RNC-highRez against the two benchmarking models. The SFCN-reg model had a significantly higher MAE of 5.26±3.89 years ($p = 0.002$) but a lower R^2 of 0.85 than the proposed RNC-highRez, despite being trained on a much larger dataset. On the other hand, the BSA model [67] obtained a significantly better MAE at 4.08±3.23 years ($p < 0.01$), but had a lower R^2 and correlation than RNC-highRez. However, when inspecting the estimation accuracy for the older population group

over 60 yo (N=62), RNC-highRez offered an MAE of 4.30±3.23 years and R^2 of 0.77 while BSA resulted in slightly inferior MAE of 4.36±3.44 years and R^2 of 0.74.

Table 3.3: Results of the best performing ResNet, proposed RNC and benchmark models, with addition of male (M) and female (F) subgroup performance. Our methods are highlighted in gray. (* indictates $p < 0.01$).

Model	Train Data	MAE ± Std. Dev.	M/F MAE	Mean	M/F Mean	R^2	Correlation
ResNet50-highRez	1294	4.93 ±3.90	4.97/4.89	-0.36	-0.86/+0.11	0.90	0.94*
RNC-highRez	1294	4.28 ±3.34	4.38/4.19	-0.60	-1.0/-0.17	0.93	0.96*
BSA	2887	4.08 ±3.23	4.06/4.09	-0.61	-0.40/-0.88	0.89	0.94*
SFCN-reg	42 829	5.46 ±3.89	5.13/5.76	-2.50	-2.47/-2.52	0.85	0.92*

3.4.2 Grad-RAM Analysis of Healthy Population

As an exploratory study, we computed the mean values from the population-averaged Grad-RAM heatmaps for 116 brain parcellations based on the AAL116 atlas. A visualization of the numerical results for different DL models' Grad-RAM heatmaps across different age groups and sexes is presented in Appendix A.

Upon inspecting the population-averaged Grad-RAM results across different age groups, the ResNet50-highRez model presented fairly consistent areas of interest across the age groups, with virtually no differences across sexes. Specifically, for all age groups, the main areas the model focuses on are in the left hemisphere (in order of decreasing magnitude): pallidum, amygdala, and putamen. The younger population (20-40 yo) has a high concentration on the left hippocampus and left pallidum. As the population ages, several other areas become more prominent, especially for the population above 80 yo, and these include the right pallidum, thalamus, caudate, left parahippocampal gyrus, left hippocampus, and posterior cingulum, vermis, left frontal orbital gyrus. Based on the observation, as we age, the ResNet50-highRez model focuses on wider and more diverse regions of interest while the primary focus is still on subcortical regions. A visual inspection of the averaged Grad-RAMs (see Fig. 3.6) further confirms our results, showing the younger population (20-40 yo) having a more concentrated activation area, and the older population expanding the activation area in the subcortical regions, ventricles, and the cortex.

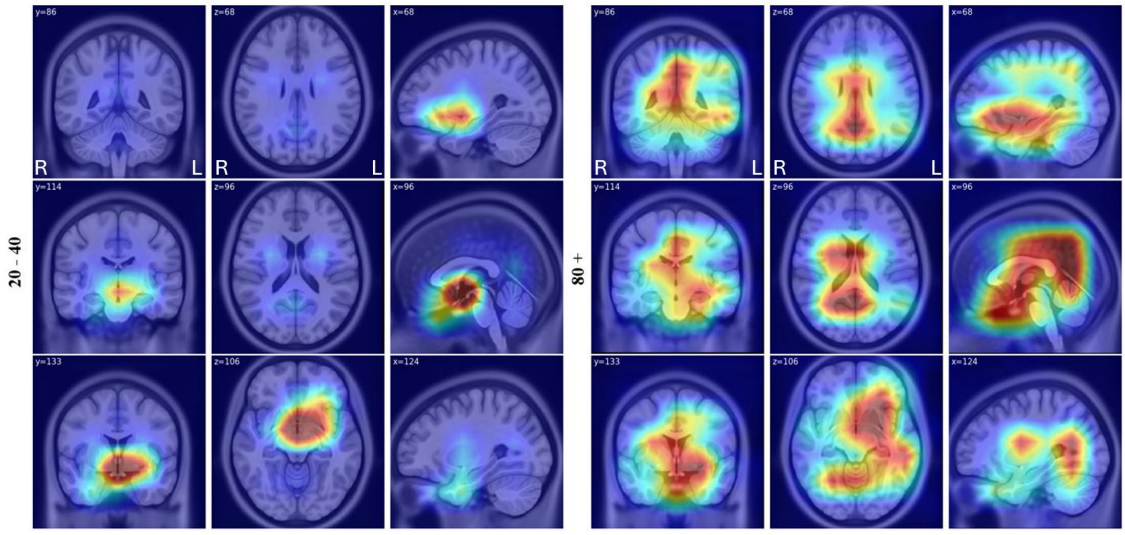


Figure 3.6: Averaged Grad-RAM heatmaps produced from the ResNet 50-highRez model based on the healthy control test set for the age groups of 20-40 yo (left) and 80+ yo (right).

In contrast to ResNet50-highRez, the RNC-highRez model showed more distinct Grad-RAM activation patterns across different age groups. For the 20-40 yo group, the activated areas from the DL model are on the right parietal and lateral frontal lobe, including (in the order of decreasing magnitude) inferior parietal, right postcentral, precentral, supramarginal, and angular gyrus, and right paracentral lobule. For the 40-60 yo group, the model shifts the attention with a focus towards the left lateral frontal lobe, particularly the frontal inferior operculum, frontal inferior triangularis, rolandic operculum, and with a broader and weaker focus on areas around it, such as the left insula. Finally, the 60-80 yo group and the 80+ yo group have nearly identical activation patterns, focusing on subcortical regions and to a lesser extent, the left lateral frontal lobe. The main regions of interest remain on the left hemisphere, including (in the order of decreasing magnitude) the thalamus, pallidum, putamen, hippocampus, amygdala, caudate, and vermis. A visual inspection (see Fig. 3.7) confirms these findings from the parcellation-wise mean Grad-RAM values and suggests that the younger population (20-40 yo) has more activation in the lateral frontal and parietal lobes, shifting towards the left hemisphere's subcortical regions and lateral ventricles along with the aging process.

Similar to the case of ResNet50-highRez, the RNC-highRez model also did not demonstrate drastic Grad-RAM pattern differences between male and female aging (see Fig. 3.8) although some nuanced distinctions in the strengths of activation were present for the overall population and within

the same age groups. The more pronounced differences are for the youngest group of 20-40 yo. On average, the right parietal inferior lobe, right angular gyrus, right paracentral lobule, right parietal superior lobe, and left frontal superior lobe, are about 40% more activated for males, with the contribution coming mainly from the 20-40 group.

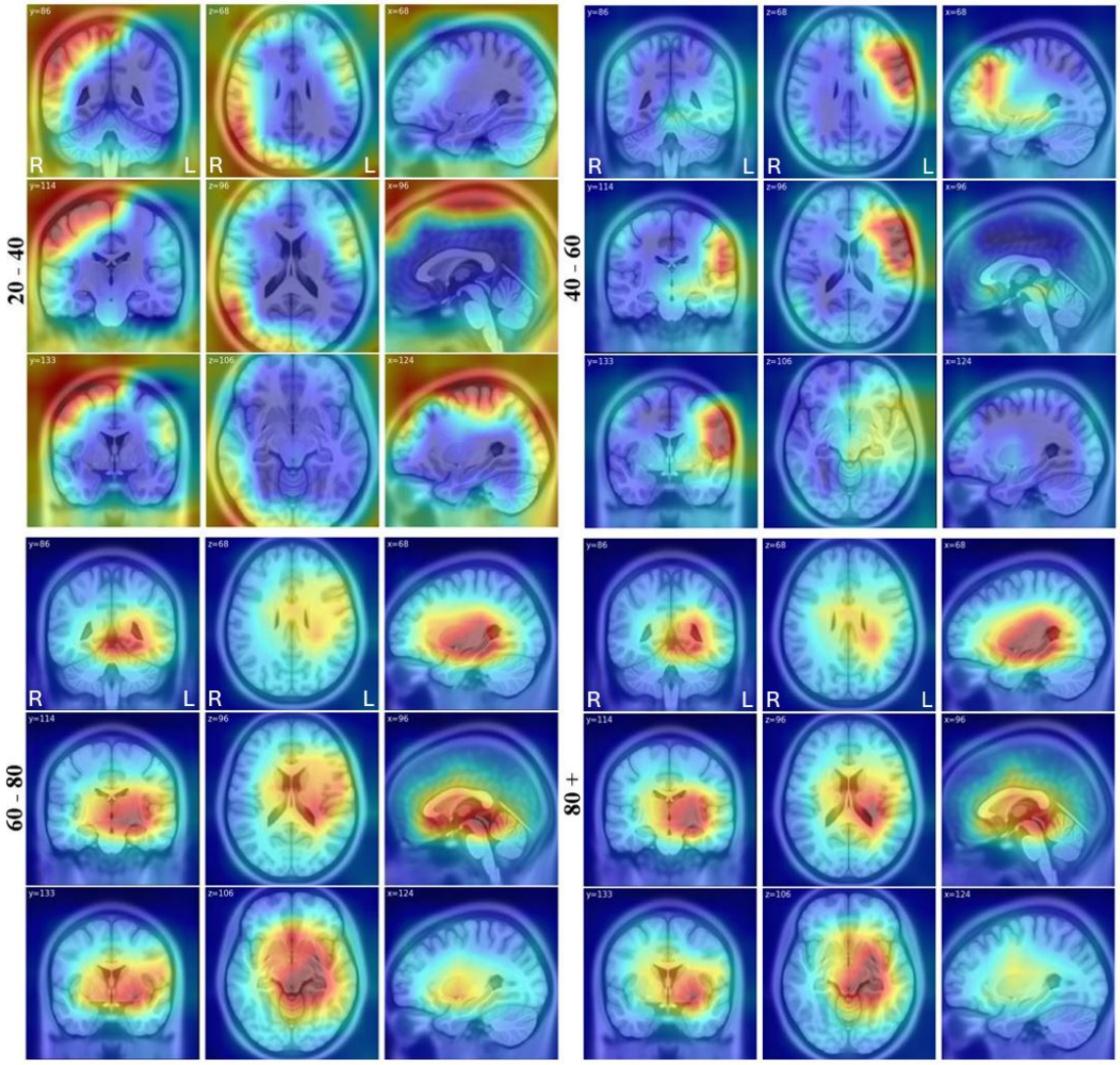


Figure 3.7: Averaged Grad-RAM heatmaps produced from the RNC-highRez model based on the healthy control test set for the age groups of 20-40 yo (top left), 40-60 yo (bottom left), 60-80 yo (top right), and 80+ yo (bottom right).

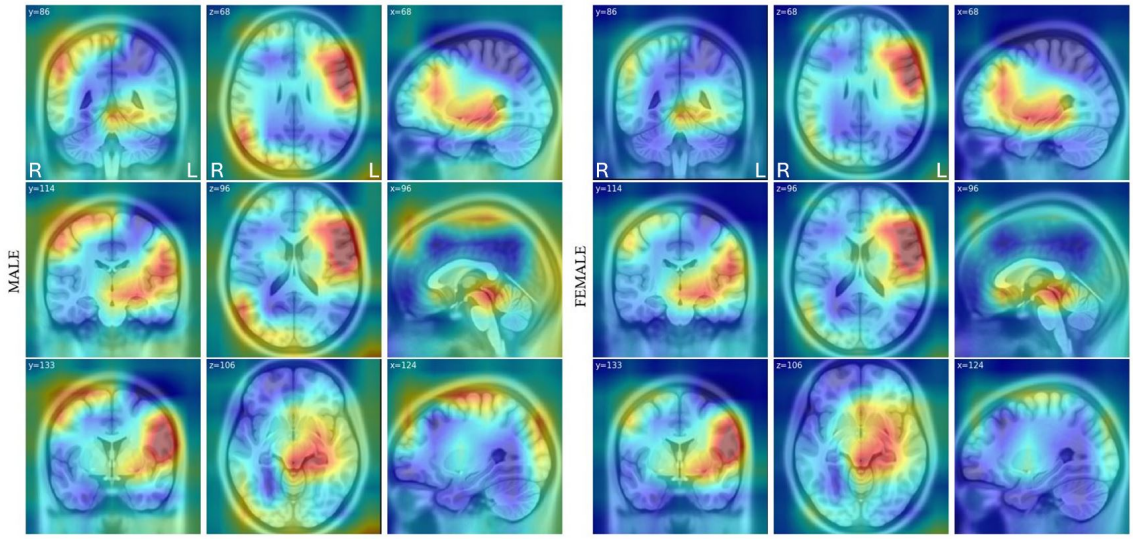


Figure 3.8: Group-averaged Grad-RAM intensities produced from the RNC-highRez model of male (left) and female healthy subjects (right) of the entire test dataset.

3.4.3 Brain Age Estimation for Alzheimer's and Parkinson's Disease Patients

In Table 3.4, we list the age estimation errors in terms of the MAE and mean age gap ("prediction - ground truth") and "age gap vs. disease severity" correlations for AD and PD patient groups. In both disease groups and across all models, the MAE and standard deviation exhibit an increase, as anticipated in diseased patients. Notably, the mean error between the healthy and diseased populations also increases, as illustrated in Fig. 3.9. As expected, the healthy control group exhibits a mean error closest to 0, with the ResNet50-highRez showing an error of -0.36 years and the RNC-highRez model of -0.60 years. In contrast, the AD group shows a substantial increase in mean error to +0.70 years for the ResNet50-highRez ($p < 0.01$) and +2.10 years for the RNC-highRez model, which is significantly higher than that of the ResNet50-highRez ($p < 0.01$). These errors are consistent with those reported by Nguyen et al. [67], who found a discrepancy of +3.4 years, and Sendi et al. [83], who reported a difference of +3.3 years.

We observed that the brain age gaps between the prediction and ground truths are correlated to ADAS-cog 11 scores, with the RNC-highRez model achieving the highest correlation of 0.37 ($p = 0.0098$). In addition, the RNC-highRez model tends to predict older ages than the ResNet50-highRez counterpart, and shows a stronger correlation between brain age gap and disease severity.

In the PD group the ResNet50-highRez model exhibits a modest increase in mean error without a statistically significant difference, whereas the RNC-highRez model shows a significantly higher mean error of +0.47 ($p < 0.01$). These findings are consistent with Nguyen et al. [67], who reported a brain age gap of +1.0 years for PD, but differ from Eickhoff et al. [17], who found a larger gap of +2.8 years. In our tested PD cohort, no correlation was found between the brain age gap and UPDRSIII scores.

Table 3.4: Metrics of interest for a disease population for each model, including the correlation between brain age (BA) and disease severity metrics: ADAS-cog 11 for AD, UPDRSIII for PD.

Population	Model	MAE ± Std. Dev.	Mean Error ± Std. Dev.	Corr. BA vs. Disease
AD (n=80)	ResNet50-highRez	5.49 ±4.42	0.70 ±7.05	0.25 ($p = 0.032$)
	RNC-highRez	5.61 ±4.40	2.10 ±6.70	0.37 ($p = 0.0098$)
PD (n=61)	ResNet50-highRez	6.01 ±5.35	0.24 ±8.07	0.03 ($p = 0.80$)
	RNC-highRez	5.85 ±5.21	0.47 ±7.84	0.043 ($p = 0.63$)

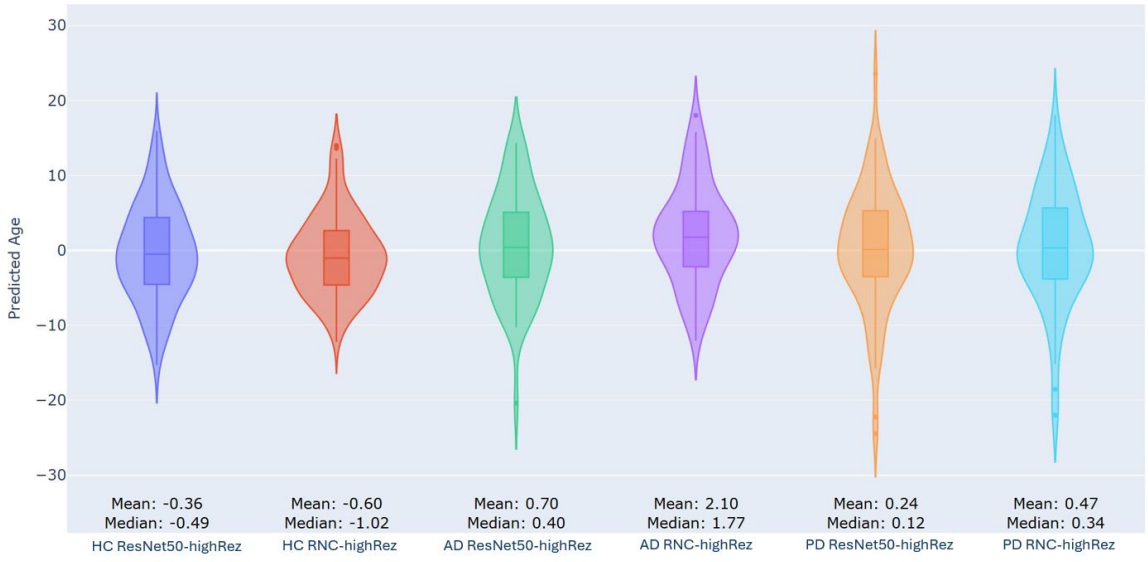


Figure 3.9: Violin plot of differences between the estimated brain age and the ground truths for the ResNet50-highRez and RNC-highRez models on healthy and diseased populations.

3.4.4 Grad-RAM Analysis for Alzheimer's and Parkinson's Disease Patients

Similar to the Grad-RAM analysis for healthy aging, a visualization of the numerical results for different DL models' Grad-RAM heatmaps for the AD and PD groups is presented in Appendix A Fig.A.3. The ResNet50-highRez model produced almost no difference compared to the healthy

cohort of the same age range. The RNC-highRez model presented some changes, the Grad-RAM outputs are shown in Fig. 3.10.

For AD patients, the RNC-highRez revealed similar activation patterns to the healthy controls of the same age range, which are concentrated on the left hemisphere (in the order of decreasing mean grad-RAM values): thalamus, pallidum, putamen, hippocampus, amygdala, and caudate. Compared to HC activations with the same model there are slight increases in intensity to the left temporal and lateral frontal lobe, caudate and putamen. Noticeably, the model intensities are consistent with the RNC-highRez model on the left side of the brain, but intensities slightly increased towards the longitudinal fissure and right hemisphere of the brain. Visually, the hippocampus, left lateral ventricle and midbrain are more activated.

For PD patients, with the RNC-highRez model, the main activation patterns remain similar to those of the HC group for the same age range, once again on the left hemisphere (in the order of decreasing mean grad-RAM values): thalamus, pallidum, putamen, hippocampus, amygdala, and caudate. There are some increases in intensity to the left temporal and lateral frontal lobe compared to the HC cohort. Visually, the left lateral ventricle, basal ganglia, and midbrain are more activated.

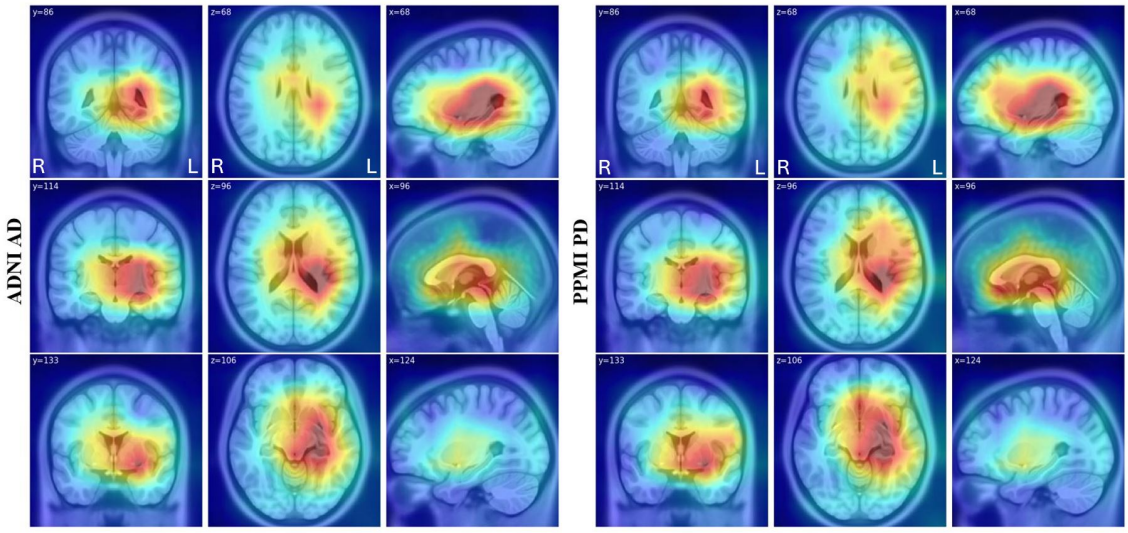


Figure 3.10: Averaged Grad-RAM heatmaps produced from the RNC-highRez model for the AD (left) and PD (right) patients.

3.5 Discussion

3.5.1 Deep Learning Model Performance

The absence of a standardized dataset for brain age estimation introduces significant variability in accuracy results across studies, as outcomes can depend on dataset size, origin (e.g., scanner types and imaging protocols), and age distribution. Studies with older cohorts (over 25yo) reported higher MAEs compared to those with younger populations (0-18yo) [67, 86]. Consequently, it is important to consider how training or testing datasets influence the accuracy of models.

When applying models trained on significantly larger datasets to our testing set, we observe an underperformance compared to their expected out-of-domain accuracy using the same data preprocessing procedures. BSA [67] trained on roughly 2,887 subjects achieved an MAE of 4.08 years compared to their out-of-domain results of 3.83 years, although they provided an improvement in R^2 of 0.89 compared to 0.62. Similarly, SFCN-reg [47] trained on 42,829 T1w MRIs achieved an MAE of 5.46 years compared to their test set's MAE of 3.90 years, and R^2 of 0.89 instead of 0.94. When applying the same ResNet model to a much larger dataset, performance can improve significantly. For instance, Peng et al. [71] utilized a ResNet-50 trained on 12,949 subjects (44-80 yo) using similar data augmentation methods to ours, and achieved an MAE of 2.32 years on their test set. Therefore, the comparatively smaller training dataset and demographic differences may contribute to the lower MAEs observed with our proposed method compared to the best SOTA DL model. The RNC-highRez model significantly outperformed all our ResNet variants and the SFCN-reg model, and had comparable performance to the BSA model. This performance relative to our ResNet variants and two SOTA benchmarking techniques trained on much more data show potential for the RNC approach to achieve superior performance if trained on a larger dataset. This observation is supported by the findings from other brain age estimation studies, which have demonstrated that increasing dataset size substantially enhances model performance [86, 67, 71]. This was also highlighted by the authors of RNC [91]. Our results also confirmed the benefit of contrastive learning in data efficient DL training, which is particularly significant in medical image analysis. Additionally, we observed a positive impact of finer brain MRI resolution on the age estimation, suggesting that structural changes related to aging can be more nuanced.

The main limitation of our model development and accuracy evaluation is the lack of an independent out-of-domain test set for evaluating accuracy to provide more rigorous comparisons between our models and the benchmark models. Another approach would be to train the benchmark models from scratch with our dataset for a more fair comparison. However, this is often challenging due to many models not being publicly available and the difficulty to exactly replicate the original studies. On the other hand, curation and management of large brain MRI datasets can be highly resource-demanding.

Several other parameters were found to significantly impact performance. The combination of random affine translation and rotation, random Gaussian noise, and random crop resize as data augmentation techniques improved the performance of our ResNet models. Additionally, increasing the batch size for the RNC model improved performance, consistent with the original supervised contrastive loss study [45]. Finally, a higher image resolution resulted in increased performance across all models, suggesting that reducing resolution to save computational and memory resources may severely reduce performance.

3.5.2 Analysis of Grad-RAM results

Our Grad-RAM analyses provided insights into the brain regions that are most associated with healthy aging by the selected DL models across different age groups and sexes. The saliency map results revealed similarity to existing literature, as well as notable differences between the two models: ResNet50-highRez and RNC-highRez.

The ResNet50-highRez model presented consistent areas of interest in the Grad-RAM result across different age groups, focusing primarily on the limbic system, basal ganglia, cortices and ventricles. For MRI-based age estimation, the model mainly focuses on subcortical regions, which are known to have significant volume and morphological changes through aging [19, 25, 13, 39]. The saliency of the orbital cortex also resonates with known aging-related cortical thinning and ventricular enlargement [73, 20, 25]. On the other hand, the RNC-highRez model's saliency maps showed a more dynamic pattern with regions of interest varying significantly with age. For younger populations (20-40 yo), the model focused on the right parietal and lateral frontal lobes, shifting towards the left lateral frontal lobe in the 40-60 yo age group. For older groups (60+ yo), the focus

transitioned to subcortical regions similar to those identified by the ResNet50-highRez counterpart, showing once again influence on the limbic system and basal ganglia from aging, but with the notable addition of the cerebellum. These observations echo with existing literature, which reported volume reduction to the limbic system and basal ganglia [63, 39] and ventricle enlargement [25, 73].

Other studies using CNN based explainability methods [90, 43] also find similar activation patterns in the subcortical regions, but reported greater activations on the cortices compared to our study. Aging results in substantial thinning of cortical structures [30, 73] which is often the focus of most existing studies using the popular VBM [31, 49]. In contrast, our DL-based analysis revealed heavier weights on the sub-cortical regions in aging. The RNC-highRez model also showed slight sex-based differences, particularly in younger populations, similar to Yin et al [90]. However, the impact of sex is not strong in our analysis and may be a result of data sampling.

The Grad-RAMs gradient activations are a 7x8x7 voxel heatmap, which is interpolated back to the 193x229x193 voxel MRI. This results in a blurry activation heatmap reducing the precision of our explainability analysis. Consequently, regions unrelated to aging located near ones of true anatomical interest may be false positives and small structures of interest in the brain may be difficult to precisely visualize. Lastly, Grad-RAM only highlights regions of saliency for age estimation and does not provide regional differences in age, unlike some voxel-based methods [67, 28]. This can limit the specificity and precision in identifying relevant regions in aging.

3.5.3 Alzheimer's and Parkinson's Disease

Brain Age Estimation for Alzheimer's and Parkinson's Disease Patients

Our findings show an increased MAE and mean age gap in both disease groups compared to healthy controls, which aligns with the expected impacts of neurodegenerative conditions on the brain aging processes.

The AD cohort has a mean gap of +0.70 years for the ResNet50-highRez and a larger +2.10 years for the RNC-highRez model. These results are similar to other significant deviations reported by Nguyen et al. of +3.4 years [67] and Sendi et al. of +3.3 years [83]. The significant differences ($p < 0.01$) between the ResNet50-highRez and RNC-highRez models suggest that the RNC-highRez

model may be more sensitive to nuanced anatomical changes due to the disease. For the PD cohort, the ResNet50-highRez model shows a modest, non-significant increase in mean error, whereas the RNC-highRez model displays a significant mean error of +0.47 years ($p < 0.01$). These findings are similar to Nguyen et al. of +1.0 years [67], but are weaker than the larger gap reported by Eickhoff et al. of +2.8 years [17]. A similar significant difference between the two models was also found, confirming the robustness of the employed contrastive learning strategy.

When associating the brain age gaps between the estimated and ground truth values with the ADAS-cog 11 scores in AD, we found a weak-to-moderate correlation, with the RNC-highRez model obtaining the highest correlation of 0.37 ($p = 0.0098$). Our result is comparable to that of Lowe et Al [51] ranging between 0.25 to 0.56, and slightly higher than the correlation of 0.26 obtained by Behesheti et al. [5]. Conversely, there is no significant correlation between the brain age gap and UPDRSIII scores in our tested PD cohort. This is similar to the conclusion of Cheng et al. [10], but Eickhoff et al. found a significant yet weak correlation of 0.14 to 0.17 [17]. This potentially implies that anatomical alterations due to PD can be masked by natural aging [67, 60]. It is also possible that 3 Tesla T1w MRIs do not provide adequate details of small subcortical regions and that 7 Tesla MRIs are required [89]. However, 7T scans have drawbacks such as high cost, less accessibility, and can tend to have large distortions in some brain regions. Alternatively, it is also possible that T1w MRIs are not the optimal contrasts for the task due to weaker visualization of the subcortical structures.

Grad-RAM Analysis for Alzheimer's and Parkinson's Disease Patients

The Grad-RAM result of the RNC-highRez model on the AD cohort showed similar regions of interest in aging as those on healthy controls in the same age groups, focusing on the left hemisphere's limbic system, lateral ventricle, and basal ganglia. However, higher saliency values in AD than in HC are seen in the left temporal and lateral frontal lobe, the hippocampus, the caudate and putamen. The hippocampus and surrounding areas present themselves as the most relevant region for aging. This mirrors the common concept of significant hippocampal atrophy during AD progression [81] [24, 64]. Additionally, Yin et al. [90], who also used saliency maps from a CNN found similar dependence on the limbic system for MCI and AD cohorts. In the voxel-based brain age

estimation by Nguyen et al. [67], they found more advanced brain aging in the hippocampus and amygdala than the rest of the brain while Gianchandani et al.[28] reported the caudate, insula and putamen to be the most prominent in brain aging. These findings further reinforce the importance of the limbic system, basal ganglia, and specifically the hippocampus on potential disease progression of AD that can introduce accelerated biological aging.

In terms of the PD cohort, the population-averaged saliency map from Grad-RAM demonstrated similar patterns with respect to age estimation to the healthy population in the same age groups, highlighting the left hemispheres' thalamus, pallidum, putamen, hippocampus, amygdala, and caudate. Although the basal ganglia and limbic system are known to undergo atrophy and morphological changes in PD [77, 10, 74], the prominence of frontal and temporal cortices also noted in previous studies [10, 74] was not observed in our saliency maps. Conversely, some studies also reported no significant areas associated with PD in terms of accelerated aging patterns [67].

In general, the Grad-RAM results from the RNC-highRez model demonstrated greater saliency values in areas impacted by AD and PD compared to those of the ResNet50-highRez counterpart. This observation suggests that RNC-highRez may be more adapted at identifying disease-specific features within our sample cohort. However, strong similarities are still present in Grad-RAM saliency patterns between AD/PD and HC, and the distinct aging-related and disease-induced changes may be entangled in the saliency patterns that were produced. This warrants further investigation to determine whether Grad-RAM can reliably identify regions of accelerated aging for neurodegenerative disease samples with high specificity.

3.6 Conclusion

We have proposed the RNC-HighRez model, which adapts the Rank-N-Contrast loss function in a contrastive learning strategy for brain age estimation based on 3D MRI for the first time. The proposed method has significantly outperformed multiple ResNet variants in brain age estimation across multiple metrics, and has shown better or similar performance against existing DL models trained with much larger datasets. By using Grad-RAM on the proposed technique, which visually explains the decision-making process of the algorithm, we highlighted anatomical regions that are

known to be associated with brain aging, and the proposed RNC-highRez model did so with more nuance than the ResNet counterpart. For neurodegenerative diseases, we demonstrated that our DL techniques can detect accelerated aging in diseased brains that is associated with disease severity, with the model trained using contrastive learning offering more robust detection.

Chapter 4

Conclusion

In this thesis, we have implemented a new approach to brain age estimation using supervised contrastive learning with Rank-And-Contrast loss on a limited aging dataset with the addition of Grad-RAM for explainability. Our approach aimed to accurately estimate brain age while providing insights into the neuroanatomical regions associated with aging and neurodegenerative diseases.

Our results demonstrate that a supervised contrastive learning strategy can significantly outperform traditional ResNet training in brain age estimation across multiple metrics. This performance on a small training sample (1,618 subjects) shows its benefit on data efficient deep learning (DL) training and potential for achieving state-of-the-art results on larger, more diverse datasets. We additionally found that data augmentation, batch size, and image resolution all had a significant impact on model performance.

Grad-RAM effectively highlighted neuroanatomical regions known to be associated with brain aging, with the supervised contrastive learning approach demonstrating greater nuance across different ages and diseases compared to a ResNet approach. Our models identified subcortical structures as having greater significance to healthy aging compared to more classic voxel-based morphology and similar DL studies. This suggests that these regions play a more critical role in the aging process than previously estimated, particularly in older age groups. The Grad-RAM for both AD and PD focused on areas known to undergo disease-related anatomical changes. However, as there were only a few variations from the healthy cohort, it is challenging to assess its ability to differentiate areas of accelerated aging in AD or PD patients.

Disease cohort analyses revealed that AD brain age gaps are significantly higher compared to healthy controls and showed significant correlation to ADAS-cog scores. In contrast, PD patients exhibit a significant but less pronounced brain age gap and no significant correlation between brain age gap and UPDRSIII scores was found. The supervised contrastive learning approach produced higher brain age gaps for AD and PD and stronger correlation between AD brain age gaps and severity scores than the ResNet approach. This suggests that a supervised contrastive learning approach has superior performance in discrimination of healthy and disease patients. The observed increase in brain age gaps and their correlations to disease severity metrics, particularly in AD, highlight the potential biomarker and clinical applications of these models. However, the variability in prediction errors across different diseases and models requires further investigation and refinement to enhance clinical relevance.

4.0.1 Future Works

Despite the promising results, our study has limitations and areas for potential improvement. The relatively small dataset should be expanded to match the scale of larger studies for a more accurate assessment of this model's performance, including the use of a larger and out-of-domain test set. Benchmarking models should be trained from scratch to truly compare the models. Additionally, using a finer-grained saliency map method could offer more precise anatomical demarcation and certainty. This avenue of future work could serve as an additional biomarker by using the saliency output to identify regions with higher brain age and using it to discriminate neurodegenerative diseases. Additionally, we could quantify divergences in aging trajectories between healthy and diseased populations using longitudinal data furthering our understanding of brain age as a biomarker throughout disease progression. Finally, a sex-based exploration of the AD and PD cohorts should be pursued as the diseases affect each sex differently.

4.1 Summary

In summary, this work contributes to the field by offering a new approach to brain age estimation that leverages advanced DL techniques and emphasizes the importance of explainability in

model predictions. We demonstrate the effectiveness of deep learning models, particularly ResNet architectures and supervised contrastive learning strategies, in MRI-based brain age estimation. The integration of explainability tools, such as Grad-RAM, provides additional insights into the specific neuroanatomical changes associated with aging and neurodegenerative diseases. Finally, by identifying relationships between brain age gaps and neurodegenerative diseases, our approach offers insights for future clinical applications. We aimed to not only enhance the understanding of brain aging but also underscore the utility of deep learning in medical imaging for precise assessments.

Appendix A

Parcellation ID Legend

Region	Left	ID	Right	ID
Lateral frontal	Precentral_L	2001	Precentral_R	2002
Lateral frontal	Frontal_Sup_L	2101	Frontal_Sup_R	2102
Lateral frontal	Frontal_Sup_Orb_L	2111	Frontal_Sup_Orb_R	2112
Lateral frontal	Frontal_Mid_L	2201	Frontal_Mid_R	2202
Lateral frontal	Frontal_Mid_Orb_L	2211	Frontal_Mid_Orb_R	2212
Lateral frontal	Frontal_Inf_Oper_L	2301	Frontal_Inf_Oper_R	2302
Lateral frontal	Frontal_Inf_Tri_L	2311	Frontal_Inf_Tri_R	2312
Lateral frontal	Frontal_Inf_Orb_L	2321	Frontal_Inf_Orb_R	2322
Lateral frontal	Rolandic_Oper_L	2331	Rolandic_Oper_R	2332
Lateral frontal	Supp_Motor_Area_L	2401	Supp_Motor_Area_R	2402
Medial frontal	Olfactory_L	2501	Olfactory_R	2502
Medial frontal	Frontal_Sup_Medial_L	2601	Frontal_Sup_Medial_R	2602
Medial frontal	Frontal_Med_Orb_L	2611	Frontal_Med_Orb_R	2612
Medial frontal	Rectus_L	2701	Rectus_R	2702
Insula	Insula_L	3001	Insula_R	3002
Cingulum_Ant + Mid	Cingulum_Ant_L	4001	Cingulum_Ant_R	4002
Cingulum_Ant + Mid	Cingulum_Mid_L	4011	Cingulum_Mid_R	4012

Cingulum_Post	Cingulum_Post_L	4021	Cingulum_Post_R	4022
Hippocampus + amyg.	Hippocampus_L	4101	Hippocampus_R	4102
Hippocampus + Amyg.	ParaHippocampal_L	4111	ParaHippocampal_R	4112
Hippocampus + Amyg.	Amygdala_L	4201	Amygdala_R	4202
Calcarine	Calcarine_L	5001	Calcarine_R	5002
Occipital	Cuneus_L	5011	Cuneus_R	5012
Occipital	Lingual_L	5021	Lingual_R	5022
Occipital	Occipital_Sup_L	5101	Occipital_Sup_R	5102
Occipital	Occipital_Mid_L	5201	Occipital_Mid_R	5202
Occipital	Occipital_Inf_L	5301	Occipital_Inf_R	5302
Parietal	Fusiform_L	5401	Fusiform_R	5402
Parietal	Postcentral_L	6001	Postcentral_R	6002
Parietal	Parietal_Sup_L	6101	Parietal_Sup_R	6102
Parietal	Parietal_Inf_L	6201	Parietal_Inf_R	6202
Parietal	SupraMarginal_L	6211	SupraMarginal_R	6212
Parietal	Angular_L	6221	Angular_R	6222
Parietal	Paracentral_Lobule_L	6401	Paracentral_Lobule_R	6402
Precuneus	Precuneus_L	6301	Precuneus_R	6302
Caudate + Putamen	Caudate_L	7001	Caudate_R	7002
Caudate + Putamen	Putamen_L	7011	Putamen_R	7012
Globus Pallidus	Pallidum_L	7021	Pallidum_R	7022
Thalamus	Thalamus_L	7101	Thalamus_R	7102
Temporal	Heschl_L	8101	Heschl_R	8102
Temporal	Temporal_Sup_L	8111	Temporal_Sup_R	8112
Temporal	Temporal_Pole_Sup_L	8121	Temporal_Pole_Sup_R	8122
Temporal	Temporal_Mid_L	8201	Temporal_Mid_R	8202
Temporal	Temporal_Pole_Mid_L	8211	Temporal_Pole_Mid_R	8212

Temporal	Temporal_Inf_L	8301	Temporal_Inf_R	8302
Cerebellum_Superior	Cerebellum_Crus1_L	9001	Cerebellum_Crus1_R	9002
Cerebellum_Superior	Cerebellum_3_L	9021	Cerebellum_3_R	9022
Cerebellum_Superior	Cerebellum_4_5_L	9031	Cerebellum_4_5_R	9032
Cerebellum_Superior	Cerebellum_6_L	9041	Cerebellum_6_R	9042
Cerebellum_Inferior	Cerebelum_Crus2_L	9011	Cerebellum_Crus2_R	9012
Cerebellum_Inferior	Cerebellum_7b_L	9051	Cerebellum_7b_R	9052
Cerebellum_Inferior	Cerebellum_8_L	9061	Cerebellum_8_R	9062
Cerebellum_Inferior	Cblm_Tonsil_9_L	9071	Cblm_Tonsil_9_R	9072
Cerebellum_Inferior	Cerebelum_10_L	9081	Cerebelum_10_R	9082
Vermis	Vermis_1	9100		
Vermis	Vermis_2	9110		
Vermis	Vermis_3	9120		

Table A.1: AAL atlas brain regions and corresponding IDs

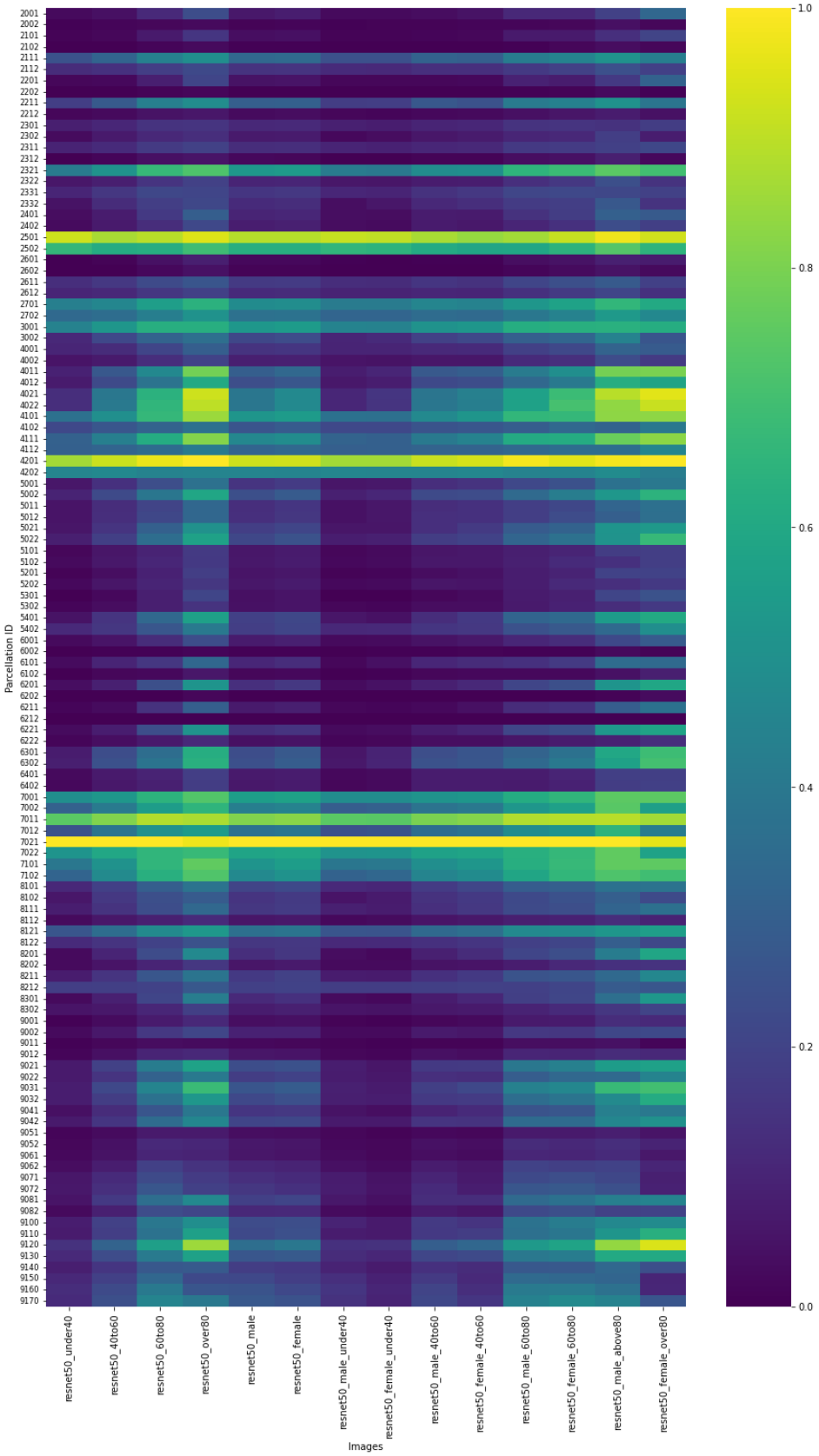


Figure A.1: Normalized average Grad-RAM heatmap intensities per AAL atlas parcellation for each sub-population for the ResNet50-highRez model.

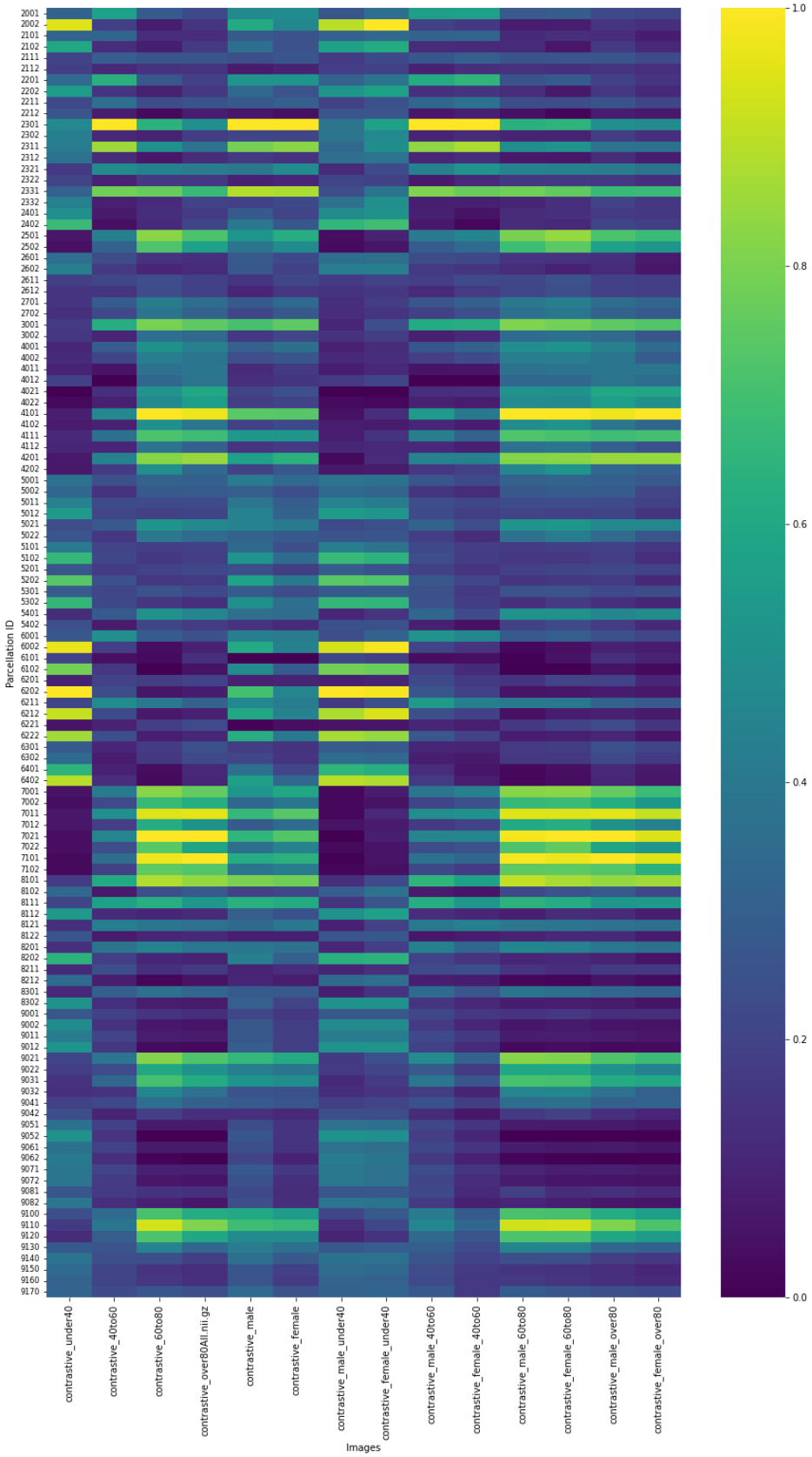


Figure A.2: Normalized average Grad-RAM heatmap intensities per AAL atlas parcellation for each sub-population for the contrastive model.

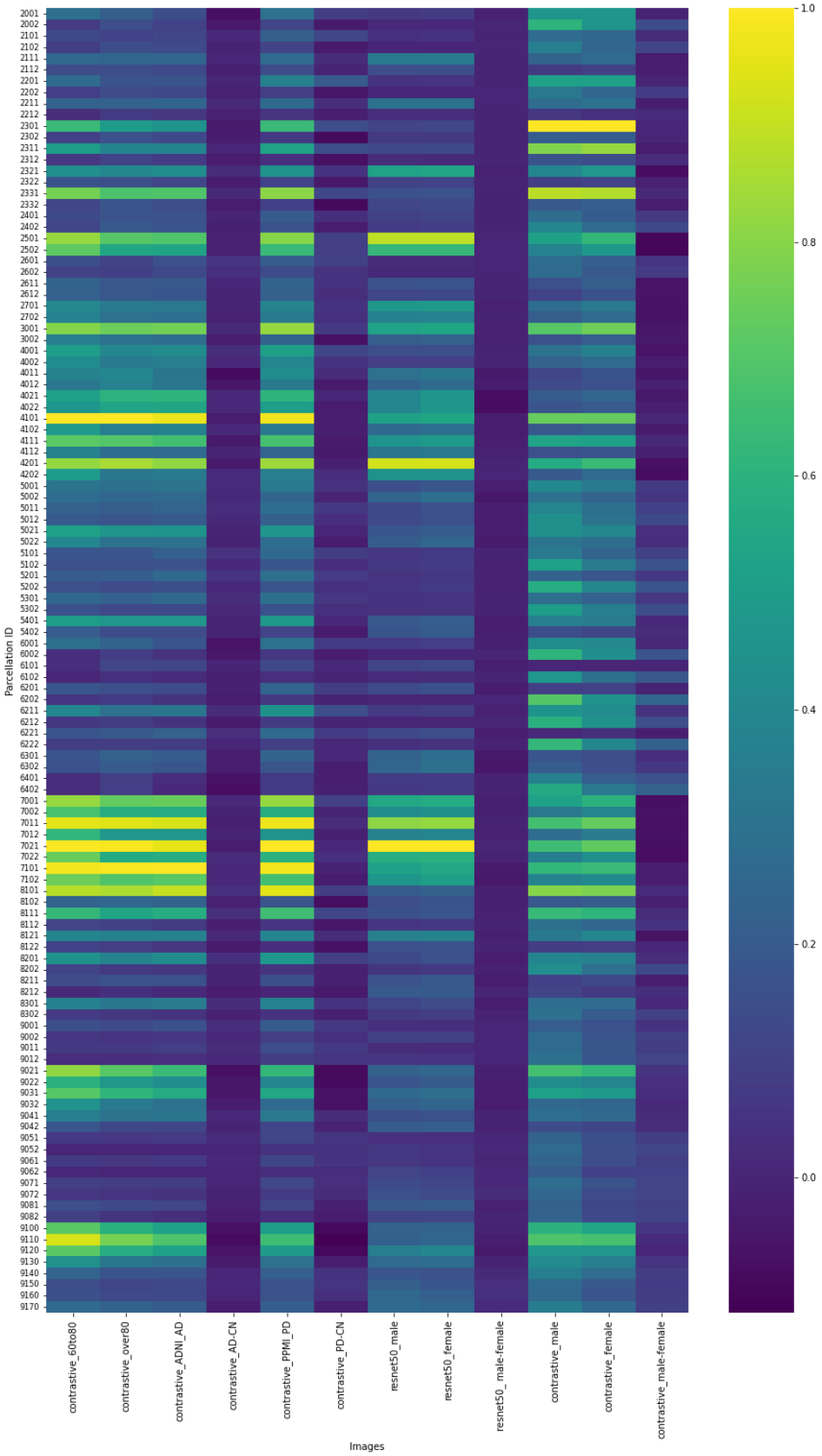


Figure A.3: Normalized average Grad-RAM heatmap intensities per AAL atlas parcellation for a sub-populations and model, and difference of parcellation intensities for AD vs CN (60-100), PD vs CN (60 to 100), and male vs. female.

Bibliography

[1] Mariam Arain, Maliha Haque, Lina Johal, Puja Mathur, Wynand Nel, Afsha Rais, Ranbir Sandhu, and Sushil Sharma. Maturation of the adolescent brain. *Neuropsychiatric disease and treatment*, 9:449–61, 04 2013.

[2] Ingrid Arevalo-Rodriguez, Nadja Smailagic, Marta Roqué-Figuls, Agustín Ciapponi, Erick Sanchez-Perez, Andry Giannakou, Olga Pedraza, Xavier Bonfill, and Sarah Cullum. Mini-mental state examination (mmse) for the early detection of dementia in people with mild cognitive impairment (mci). *Cochrane Database of Systematic Reviews*, 2021, 07 2021.

[3] Nicole Armstrong, Yang An, Lori Beason-Held, Jimit Doshi, Guray Erus, Luigi Ferrucci, Christos Davatzikos, and Susan Resnick. Sex differences in brain aging and predictors of neurodegeneration in cognitively healthy older adults. *Neurobiology of Aging*, 81, 06 2019.

[4] Carlo Alberto Barbano, Benoit Dufumier, Edouard Duchesnay, Marco Grangetto, and Pietro Gori. Contrastive learning for regression in multi-site brain age prediction. 11 2022.

[5] Iman Beheshti, Norihide Maikusa, and Hiroshi Matsuda. The association between "brain-age score" (bas) and traditional neuropsychological screening tools in alzheimer's disease. *Brain and Behavior*, 8:e01020, 06 2018.

[6] Kyriaki-Margarita Bintsi, Vasileios Baltatzis, Alexander Hammers, and Daniel Rueckert. *Voxel-Level Importance Maps for Interpretable Brain Age Estimation*, pages 65–74. 09 2021.

[7] Susan Bookheimer, David Salat, Melissa Terpstra, Beau Ances, Deanna Barch, Randy Buckner, Gregory Burgess, Sandra Curtiss, Mirella Diaz-Santos, Jennifer Elam, Bruce Fischl, Douglas Greve, Hannah Hagy, Michael Harms, Olivia Hatch, Trey Hedden, Cynthia Hodge, Kevin

Japardi, Taylor Kuhn, and Essa Yacoub. The lifespan human connectome project in aging: An overview. *NeuroImage*, 185, 10 2018.

[8] Alzheimer Society Canada. Navigating the path forward for dementia in canada. 2022.

[9] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. 06 2020.

[10] Yueh-Sheng Chen, Chen-Yuan Kuo, Cheng-Hsien Lu, Yuan-Wei Wang, Kun-Hsien Chou, and Wei-Che Lin. Multiscale brain age prediction reveals region-specific accelerated brain aging in parkinson's disease. *Neurobiology of Aging*, 140, 05 2024.

[11] James Cole, Rudra Poudel, Dimosthenis Tsagkrasoulis, Matthan Caan, Claire Steves, Tim Spector, and Giovanni Montana. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163, 12 2016.

[12] James H. Cole and Katja Franke. Predicting age using neuroimaging: Innovative brain ageing biomarkers. *Trends in Neurosciences*, 40(12):681–690, 2017.

[13] A. Convit, M J de Leon, M J Hoptman, C Tarshish, S De Santi, and H Rusinek. Age-related changes in brain: I. magnetic resonance imaging measures of temporal lobe volumes in normal subjects. *Psychiatric Quarterly*, 1995.

[14] Manuel Cossio. Augmenting medical imaging: A comprehensive catalogue of 65 techniques for enhanced data analysis. 03 2023.

[15] Pierrick Coupé, Jose Manjon, Enrique Lanuza, and Gwenaelle Catheline. Lifespan changes of the human brain in alzheimer's disease. 12 2019.

[16] Danai Dima, Amirhossein Modabbernia, Efstathios (Steven) Papachristou, Gaelle Doucet, Ingrid Agartz, Moji Aghajani, Theophilus Akudjedu, Anton Albajes-Eizagirre, Dag Alnæs, Kathryn Alpert, Micael Andersson, Nancy Andreasen, Ole Andreasssen, Philip Asherson, Tobias Banaschewski, Nuria Bargalló, Sarah Baumeister, Ramona Baur-Streubel, Alessandro Bertolino, and Sophia Frangou. Subcortical volumes across the lifespan: Data from 18,605 healthy individuals aged 3-90 years. *Human brain mapping*, 43, 02 2021.

[17] Claudia Eickhoff, Felix Hoffstaedter, Julian Caspers, Kathrin Reetz, Christian Mathys, Imis Dogan, Katrin Amunts, Alfons Schnitzler, and Simon Eickhoff. Advanced brain aging in parkinson's disease is related to disease duration and individual impairment. *Brain Communications*, 3, 08 2021.

[18] Simon F. Eskildsen, Pierrick Coupé, Vladimir Fonov, José V. Manjón, Kelvin K. Leung, Nicolas Guizard, Shafik N. Wassef, Lasse Riis Østergaard, and D. Louis Collins. Beast: Brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373, 2012.

[19] Anders M Fjell and Kristine B. Walhovd. Structural brain changes in aging: Courses, causes and cognitive consequences. *Reviews in the Neurosciences*, 21(3):187–222, 2010.

[20] Anders M. Fjell, Lars T. Westlye, Inge Amlien, Thomas Espeseth, Ivar Reinvang, Naftali Raz, Ingrid Agartz, David H. Salat, Doug N. Greve, Bruce Fischl, Anders M Dale, and Kristine B. Walhovd. High Consistency of Regional Cortical Thinning in Aging across Multiple Samples. *Cerebral Cortex*, 19(9):2001–2012, 01 2009.

[21] Katja Franke and Christian Gaser. Ten years of brainage as a neuroimaging biomarker of brain aging: What insights have we gained? *Frontiers in Neurology*, 10, 2019.

[22] Katja Franke, Christian Gaser, Brad Manor, and Vera Novak. Advanced brainage in older adults with type 2 diabetes mellitus. *Frontiers in aging neuroscience*, 5:90, 12 2013.

[23] Katja Franke, Gabriel Ziegler, Stefan Klöppel, and Christian Gaser. Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, 50:883–92, 04 2010.

[24] Giovanni Frisoni, Nick Fox, Clifford Jack, Ph Scheltens, and Paul Thompson. The clinical use of structural mri in alzheimer's disease. *Nature reviews. Neurology*, 6:67–77, 02 2010.

[25] Shohei Fujita, Susumu Mori, Kengo Onda, Shouhei Hanaoka, Yukihiro Nomura, Takahiro Nakao, Takeharu Yoshikawa, Hidemasa Takao, Naoto Hayashi, and Osamu Abe. Characterization of brain volume changes in aging individuals with normal cognition using serial magnetic resonance imaging. *JAMA network open*, 6:e2318153, 06 2023.

[26] Christian Gaser, Katja Franke, Stefan Klöppel, Nikolaos Koutsouleris, Heinrich Sauer, and ADNI. Brainage in mild cognitive impaired patients: Predicting the conversion to alzheimer's disease. *PLoS ONE*, 8:e67346, 06 2013.

[27] Neha Gianchandani, Mahsa Dibaji, Mariana Bento, Ethan MacDonald, and Roberto Souza. Reframing the brain age prediction problem to a more interpretable and quantitative approach. 08 2023.

[28] Neha Gianchandani, Mahsa Dibaji, Johanna Ospel, Fernando Vega, Mariana Bento, M. Ethan MacDonald, and Roberto Souza. A voxel-level approach to brain age prediction: A method to assess regional brain aging. *Machine Learning for Biomedical Imaging*, 2:761–795, 04 2024.

[29] Christopher G. Goetz, Barbara C. Tilley, Stephanie R. Shaftman, Glenn T. Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B. Stern, Richard Dodel, Bruno Dubois, Robert Holloway, Joseph Jankovic, Jaime Kulisevsky, Anthony E. Lang, Andrew Lees, Sue Leurgans, Peter A. LeWitt, David Nyenhuis, C. Warren Olanow, Olivier Rascol, Anette Schrag, Jeanne A. Teresi, Jacobus J. van Hilten, and Nancy LaPelle. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results. *Movement Disorders*, 23(15):2129–2170, 2008.

[30] Faith M. Gunning-Dixon, Adam M. Brickman, Janice C. Cheng, and George S. Alexopoulos. Aging of cerebral white matter: a review of mri findings. *International Journal of Geriatric Psychiatry*, 24(2):109–117, 2009.

[31] 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), 2022.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[33] Sheng He, P. Ellen Grant, and Yangming Ou. Global-local transformer for brain age estimation. *IEEE Transactions on Medical Imaging*, 41(1):213–224, 2022.

[34] W Hendelman. Atlas of functional neuroanatomy. *CRC Press*, 2000.

[35] Thao Ho, Taewoo Kim, Woo Jin Kim, Chang Hyun Lee, Kum Chae, So Bak, Sung Kwon, Gong Jin, Eun-Kee Park, and Sanghun Choi. A 3d-cnn model with ct-based parametric response mapping for classifying copd subjects. *Scientific Reports*, 11:34, 01 2021.

[36] Suzanne Houston and Megan Herting. The neurobiology of childhood structural brain development: Conception through adulthood. *Current topics in behavioral neurosciences*, 16, 12 2013.

[37] Shih-Cheng Huang, Anuj Pareek, Malte Jensen, Matthew Lungren, Serena Yeung, and Akshay Chaudhari. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *npj Digital Medicine*, 6, 04 2023.

[38] B. T. Hyman, G. W. Van Hoesen, A. R. Damasio, and C. L. Barnes. Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science*, 1984.

[39] Clifford R. Jack, Ronald C. Petersen, Peter C. O'Brien, and Eric G. Tangalos. Mr‐based hippocampal volumetry in the diagnosis of alzheimer's disease. *Neurology*, 42(1):183–183, 1992.

[40] Clifford R. Jack Jr., Matt A. Bernstein, Nick C. Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J. Britson, Jennifer L. Whitwell, Chadwick Ward, Anders M. Dale, Joel P. Felmlee, Jeffrey L. Gunter, Derek L.G. Hill, Ron Killiany, Norbert Schuff, Sabrina Fox-Bosetti, Chen Lin, Colin Studholme, Charles S. DeCarli, Gunnar Krueger, Heidi A. Ward, Gregory J. Metzger, Katherine T. Scott, Richard Mallozzi, Daniel Blezek, Joshua Levy, Josef P. Debbins, Adam S. Fleisher, Marilyn Albert, Robert Green, George Bartzokis, Gary Glover, John Mugler, and Michael W. Weiner. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

[41] Xiuqin Jia, Peipeng Liang, Y Li, L Shi, and D Wang. Longitudinal study of gray matter changes in parkinson disease. *AJNR. American journal of neuroradiology*, 36, 09 2015.

[42] Robert J. Jirsaraie, Aaron J. Gorelik, Martins M. Gatavins, Denis A. Engemann, Ryan Bogdan, Deanna M. Barch, and Aristeidis Sotiras. A systematic review of multimodal brain age studies: Uncovering a divergence between model accuracy and utility. *Patterns*, 4(4):100712, 2023.

[43] Yoonji Joo, Eun Namgung, Hyeonseok Jeong, Ilhyang Kang, Jinsol Kim, Sohyun Oh, In Lyoo, Sujung Yoon, and Jaeuk Hwang. Brain age prediction using combined deep convolutional neural network and multi-layer perceptron algorithms. *Scientific Reports*, 13, 12 2023.

[44] Benedikt Jónsson, Gyda Bjornsdottir, Thorgeir Thorgeirsson, Lotta Ellingsen, Gudmundur Walters, Daníel Guobjartsson, Hreinn Stefansson, Kari Stefansson, and Magnus Ulfarsson. Deep learning based brain age prediction uncovers associated sequence variants. 04 2019.

[45] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning, 04 2020.

[46] Ksenia Kudryashova, Kseniia Burka, Anton Kulaga, Nataliya Vorobyeva, and Brian Kennedy. Aging biomarkers: From functional tests to multi-omics approaches. *PROTEOMICS*, 20:1900408, 02 2020.

[47] Esten H. Leonardsen, Han Peng, Tobias Kaufmann, Ingrid Agartz, Ole A. Andreassen, Elisabeth Gulowsen Celius, Thomas Espeseth, Hanne F. Harbo, Einar A. Høgestøl, Ann-Marie de Lange, Andre F. Marquand, Didac Vidal-Piñeiro, James M. Roe, Geir Selbæk, Øystein Sørensen, Stephen M. Smith, Lars T. Westlye, Thomas Wolfers, and Yunpeng Wang. Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage*, 256:119210, 2022.

[48] Xingfeng Li, Yue Xing, Antonio Martin-Bastida, Paola Piccini, and Dorothee Auer. Patterns of grey matter loss associated with motor subscores in early parkinson's disease. *NeuroImage: Clinical*, 17, 11 2017.

[49] Franziskus Liem, Gael Varoquaux, Jana Kynast, Frauke Beyer, Shahrzad Kharabian Masouleh, Julia Huntenburg, Leonie Lampe, Mehdi Rahim, Alexandre Abraham, Cameron Craddock, Steffi Riedel-Heller, Tobias Luck, Markus Loeffler, Matthias Schroeter, Veronica Witte, Arno Villringer, and Daniel Margulies. Predicting brain-age from multimodal imaging data captures cognitive impairment. 11 2016.

[50] Shih-Chung Lo, Jyh-Shyan Lin, Matthew Freedman, and Seong Mun. Computer-assisted diagnosis of lung nodule detection using artificial convolution neural network. *Proceedings of SPIE - The International Society for Optical Engineering*, 1898:859–869, 02 1993.

[51] L. C. Löwe, Christian Gaser, and Katja Franke. The effect of the apoe genotype on individual brainage in normal aging, mild cognitive impairment and alzheimers disease. *PLOS ONE*, 2016.

[52] S.K. Madsen, A.J. Ho, X. Hua, P.S. Saharan, A.W. Toga, C.R. Jack, M.W. Weiner, and P.M. Thompson. 3d maps localize caudate nucleus atrophy in 400 alzheimer's disease, mild cognitive impairment, and healthy elderly subjects. *Neurobiology of Aging*, 31(8):1312–1325, 2010. Alzheimer's Disease Neuroimaging Initiative (ADNI) Studies.

[53] José Manjón and Pierrick Coupé. Mri denoising using deep learning. 01 2018.

[54] José V. Manjón, José Carbonell-Caballero, Juan J. Lull, Gracián García-Martí, Luís Martí-Bonmatí, and Montserrat Robles. Mri denoising using non-local means. *Medical Image Analysis*, 12(4):514–523, 2008.

[55] José V. Manjón, Pierrick Coupé, and Antonio Buades. Mri noise estimation and denoising using non-local pca. *Medical Image Analysis*, 22(1):35–47, 2015.

[56] Jordi Manuello, Joosung Min, Paul McCarthy, Fidel Alfaro-Almagro, Soojin Lee, Stephen Smith, Lloyd Elliott, Anderson Winkler, and Gwenaëlle Douaud. The effects of genetic and modifiable risk factors on brain regions vulnerable to ageing and disease. *Nature Communications*, 15, 03 2024.

[57] Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Christopher Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, Werner Poewe, Brit Mollenhauer, Paracelsus-Elena Klinik, Todd Sherer, Mark Frasier, Claire Meunier, Alice Rudolph, Cindy Casaceli, John Seibyl, and Peggy Taylor. The parkinson progression marker initiative (ppmi). *Progress in Neurobiology*, 95:629–635, 12 2011.

[58] Thayza Martins Melzer, Luana Meller Manosso, Suk-yu Yau, Joana Gil-Mohapel, and Patricia S. Brocardo. In pursuit of healthy aging: Effects of nutrition on brain function. *International Journal of Molecular Sciences*, 22(9), 2021.

[59] Radu Paul Mihail, Scott Workman, Zach Bessinger, and Nathan Jacobs. Sky segmentation in the wild: An empirical study. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–6, 2016. Acceptance rate: 42.3%.

[60] Shiwangi Mishra. T1-weighted mri-driven brain age estimation in alzheimer's disease and parkinson's disease. *Aging and Disease*, 11:3, 06 2019.

[61] Shiwangi Mishra, Iman Beheshti, and Pritee Khanna. A review of neuroimaging-driven brain age estimation for identification of brain disorders and health conditions. *IEEE Reviews in Biomedical Engineering*, PP:1–1, 08 2021.

[62] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: The first manually collected, in-the-wild age database. pages 1997–2005, 07 2017.

[63] Qiwen Mu, Jingxia Xie, Zongyao Wen, Yaqin Weng, and Zhang Shuyun. A quantitative mr study of the hippocampal formation, the amygdala, and the temporal horn of the lateral ventricle in healthy subjects 40 to 90 years of age. *American Journal of Neuroradiology*, 20(2):207–211, 1999.

[64] Yangling Mu and Fred Gage. Mu y, gage fh. adult hippocampal neurogenesis and its role in alzheimer's disease. mol neurodegener 6: 85. *Molecular neurodegeneration*, 6:85, 12 2011.

[65] Eitaro Nakamura and Kenji Miyao. A Method for Identifying Biomarkers of Aging and Constructing an Index of Biological Age in Humans. *The Journals of Gerontology: Series A*, 62(10):1096–1105, 10 2007.

[66] Jakub Nalepa, Michał Marcinkiewicz, and Michal Kawulok. Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 13, 12 2019.

[67] Huy-Dung Nguyen, Michaël Clément, Boris Mansencal, and Pierrick Coupé. Brain structure ages—a new biomarker for multi-disease classification. *Human Brain Mapping*, 45, 01 2024.

[68] Kaida Ning, Ben Duffy, Meredith Franklin, Will Matloff, Lu Zhao, Nibal Arzouni, Fengzhu Sun, and Arthur Toga. Improving brain age estimates with deep learning leads to identification of novel genetic factors associated with brain aging. *Neurobiology of Aging*, 105, 04 2021.

[69] Charles Nyatega, Li Qiang, Mohammed Adamu, and Halima Bello Kawuwa. Gray matter, white matter and cerebrospinal fluid abnormalities in parkinson's disease: A voxel-based morphometry study. *Frontiers in Psychiatry*, 13, 10 2022.

[70] Kalavathi Palanisamy and Surya Prasath. Methods on skull stripping of mri head scan images—a review. *Journal of Digital Imaging*, 29, 12 2015.

[71] 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.

[72] Jarrad Perron and Ji Hyun Ko. Brain age estimation: Premise, promise and problems. 03 2024.

[73] Ruth Peters. Ageing and the brain. *Postgraduate medical journal*, 82:84–8, 03 2006.

[74] Peter Pieperhoff, Martin Südmeyer, Lars Dinkelbach, Christian Hartmann, Stefano Ferrea, Alexia Moldovan, Martina Minnerop, Sandra Diaz, Alfons Schnitzler, and Katrin Amunts. Regional changes of brain structure during progression of idiopathic parkinson's disease – a longitudinal study using deformation based morphometry. *Cortex*, 151, 03 2022.

[75] Ronald Postuma and Julius Anang. *Parkinson's Disease: An Introductory Guide*. McGill University Health Centre Patient Education Office and Parkinson Canada, 2017.

[76] Anirudh Prabhu. Alzheimer's disease: A comprehensive review of its causes, diagnosis, and treatment. *Journal of Student Research*, 12, 11 2023.

[77] K. G. Prakash, B. M. Bannur, M. D. Chavan, K. Saniya, K. S. Sailesh, and A Rajagopalan. Neuroanatomical changes in parkinson's disease in relation to cognition: An update. *Journal of advanced pharmaceutical technology and research*, 2016.

[78] S. Pai Raikar. Primary components of the limbic system. 2024.

[79] W.G. Rosen, Richard Mohs, and K.L. Davis. A new rating scale for alzheimer's disease am j psychiatry 141: 1356-1364. *The American journal of psychiatry*, 141:1356–64, 12 1984.

[80] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. 12 2015.

[81] Rachael I. Scahill, Jonathan M. Schott, John M. Stevens, Martin N. Rossor, and Nick C. Fox. Mapping the evolution of regional atrophy in alzheimer's disease: Unbiased analysis of fluid-registered serial mri. *Proceedings of the National Academy of Sciences*, 99(7):4703–4707, 2002.

[82] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[83] Mohammad S.E. Sendi, David H Salat, and Vince D Calhoun. Brain age acceleration as biomarker of alzheimer's disease progression: Functional network connectivity analysis. *Alzheimer's & Dementia*, 17(S1):e050562, 2021.

[84] L.K. Soumya Kumari and R. Sundarrajan. A review on brain age prediction models. *Brain Research*, 1823:148668, 2024.

[85] R. Nathan Spreng, Roni Setton, Udi Alter, Benjamin Cassidy, Bri Darboh, Elizabeth DuPre, Karin Kantarovich, Amber Lockrow, Laetitia Mwilambwe-Tshilobo, Wen-Ming Luh, Prantik Kundu, and Gary Turner. Neurocognitive aging data release with behavioral, structural and multi-echo functional mri measures. *Scientific Data*, 9:119, 03 2022.

[86] 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, 2023.

[87] Nicholas Tustison, Brian Avants, Philip Cook, Yuanjie Zheng, Alexander Egan, Paul Yushkevich, and James Gee. N4itk: improved n3 bias correction. *Medical Imaging, IEEE Transactions on*, 29:1310 – 1320, 07 2010.

[88] Ali Varzandian, Miguel Razo, Michael Sanders, Akhila Atmakuru, and Giuseppe Di Fatta. Classification-biased apparent brain age for the prediction of alzheimer's disease. *Frontiers in Neuroscience*, 15:673120, 05 2021.

[89] Thomas Welton, Septian Hartono, Yao-Chia Shih, Stefan Schwarz, Yue Xing, Eng-King Tan, Dorothee Auer, Noam Harel, and Ling Chan. Ultra-high-field 7t mri in parkinson's disease: ready for clinical use?-a narrative review. *Quantitative Imaging in Medicine and Surgery*, 13:7607–7620, 11 2023.

[90] 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, 2023.

[91] Kaiwen Zha, Peng Cao, Jeany Son, Yuzhe Yang, and Dina Katabi. Rank-n-contrast: Learning continuous representations for regression. 2023.

[92] Yu Zhang, Rui Xie, Iman Beheshti, Xia Liu, Guowei Zheng, Yin Wang, Zhenwen Zhang, Weihao Zheng, Zhijun Yao, and Bin Hu. Improving brain age prediction with anatomical feature attention-enhanced 3d-cnn. *Computers in Biology and Medicine*, 169:107873, 2024.

[93] Dan Zheng, Chuang Chen, WenChun Song, ZhongQuan Yi, PanWen Zhao, Jianguo Zhong, ZhenYu Dai, HaiCun Shi, and Pinglei Pan. Regional gray matter reductions associated with mild cognitive impairment in parkinson's disease: A meta-analysis of voxel-based morphometry studies. *Behavioural Brain Research*, 371:111973, 05 2019.