UNIVERSIDADE D
COIMBRA

Erica Maria Vicente Prioste

AGE ESTIMATION OF THE CENTRAL NERVOUS SYSTEM FROM RETINAL IMAGES

Dissertation for the degree of Master in Biomedical Engineering with specialization in Clinical Informatics and Bioinformatics supervised by Professor Doctor Rui Manuel Dias Cortesão dos Santos Bernardes and Master Ana Teresa Casimiro Nunes and presented to the Faculty of Sciences and Technology of the University of Coimbra.

October of 2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

Erica Maria Vicente Prioste

Age Estimation of the Central Nervous System from Retinal Images

Thesis submitted to the Faculty of Sciences and Technology
of the University of Coimbra for the degree of
Master in Biomedical Engineering with specialization
in Clinical Informatics and Bioinformatics

Supervisors:
Rui Bernardes, PhD
Ana Nunes, MSc

Coimbra, 2020

Abstract

This project aimed to predict the age value for the Central Nervous System (CNS), as close as possible to the true chronological age, for healthy individuals, based on texture features computed from optical coherence tomography (OCT) data. OCT volumetric scans from 100 subjects were used to compute mean value fundus images from the six layers of the neuroretina. Texture features (local and global) were then extracted from these images. Neural Networks were applied to estimate the CNS's age, after identifying the best neural network architecture, and the most discriminative features for each retinal layer. Depending on the neuroretinal layer, the neural network architecture and the best retinal features were different. The ganglion cell layer and the outer nuclear layer were found to be the most discriminative retinal layers, and their information was concatenated into one single vector. Discrepancies between the right and left eyes were discovered for both male and female groups. For each subject, the mean of both eye's predicted age was computed, and a root mean square error (in the training set) was obtained for the female and male groups. In conclusion, the results suggest that the computed texture features contain information on the CNS's age. These findings represent an encouraging discovery that might simplify the estimation age of the CNS since the imaging method used is not as time-consuming as MRI or other methods. Further work is required for a more in-depth exploration, as factors such as the dataset size and the COVID-19 pandemic imposed restrictions on the experiments performed.

Keywords: Optical Coherence Tomography, Ageing, Retina, Central Nervous System, Neural Networks.

Resumo

Este projeto teve como objetivo a obtenção do valor da idade estimada do Sistema Nervoso Central (SNC), tão próximo quanto possível à verdadeira idade cronológica, em indivíduos saudáveis, através da utilização de parâmetros de textura obtidos a partir de dados de tomografia de coerência ótica (OCT). *Scans* volumétricos de OCT obtidos de 100 indivíduos foram utilizados para obter imagens de fundo ("mean value fundus images") das seis camadas da neuroretina, a partir das quais foram determinados os respetivos parâmetros de textura (locais e globais). A partir destes dados, utilizaram-se redes neuronais para estimar a idade do SNC, tendo para tal sido identificada a arquitetura mais favorável e os parâmetros mais discriminativos de cada camada da neuroretina. Dependendo das camadas da neuroretina utlizada, diferentes redes neuronais e diferentes parâmetros foram considerados melhores. A camada de células ganglionares e a camada nuclear externa foram identificadas como as camadas da retina mais discriminativas e a informação destas camadas foi concatenada num único vetor. Foram encontradas discrepâncias entre os olhos direito e esquerdo na idade estimada. Para cada indivíduo, a média da idade estimada dos dois olhos foi calculada e, a raiz quadrada do erro quadrático médio para os grupos feminino e masculino foi a obtida. Em conclusão, os resultados sugerem que os parâmetros de textura obtidos contêm informação sobre a idade do SNC. Estas conclusões representam uma descoberta que poderá simplificar a estimação da idade do SNC, uma vez que os métodos de imagiologia utilizados não consomem tanto tempo como o MRI ou outros métodos. No entanto, ainda existe um longo percurso na análise de dados e desenho de redes neuronais, o que não foi possível devido às restrições impostas nas experiências realizadas pelo tamanho do conjunto de dados e pela pandemia de COVID-19.

Palavras-chave: Tomografia de Coerência Ótica, Envelhecimento, Retina, Sistema Nervoso Central, Redes Neuronais.

Acknowledgements

I want to start thanking my tutor and professor Rui Bernardes and my co-tutor Ana Nunes. Without their guidance, patience and help, none of this would have been possible. They were amazing and super supportive.

My family had a major role. To mami, papi and Betos and to *mis abuelitos* - for believing in me since the first day (and still at the airport), and supporting me in every decision I made (and not gonna lie, having always a little weight on my decisions). They made me fight in every single moment of my life. They spent countless hours on phone calls when I needed the most, and I am sure they will never fail me. Never did.

To the ones that made me fight this fight through all these five years and have the strength to keep going, and going, and going. They made me change in many aspects, and I am thankful for that. Life is lived learning, and all the people involved are important. To Matilde and Maria, my crazy ones - can't advance in details; To Juliana, Cátia and Mónica for the nights spent at home, eating, laughing and just existing; To Joana, Marli, João, Rute - no need to say anything. To the ones I didn't mention but were fully supportive, you know who you are.

And more importantly: to me. For keeping up on my feet even when life gave me lemons. For fighting against all the unforeseen events on the way, and for don't letting my mind go to waste when the strongest and evil thought came across my mind. It was a rough journey, but I can't wait to rock the next one.

Contents

List of Figures

List of Tables

List of Abbreviations

AD Alzheimer's disease. 1, 8, 23, 59

CA Chronological Age. 3, 4, 8–16, 39, 44, 52, 57

CNN Convolutional Neural Network. 11, 13, 18, 57, 59

CNS Central Nervous System. 1–4, 55, 59

DM *Diabetes Mellitus*. 9

DNN Deep Neural Networks. 16

DTCWT Dual-Tree Complex Wavelet Transform. 27, 28

DWT Discrete Wavelet Transform. 27

EEG Electroencephalogram. 1, 4, 7, 12, 55

GLCM Grey-Level Co-occurrence Matrix. 3, 25–28

MAE Mean Absolute Error. 8, 10–16, 18, 32, 56

ML Machine Learning. 3

MRI Magnetic Resonance Imaging. 2–4, 7–16, 55, 57, 59

MVF Mean Value Fundus. ix, 2, 24, 25, 33, 34

NAF Neighbourhood Approximation Forest. 9

NN Neural Network. 3, 10, 30–32, 34–36, 39, 42–44, 55–58

OCT Optical Coherence Tomography. ix, 2–4, 7, 16–18, 21–23, 25, 33, 34, 58, 59

PA Predicted Age. 4, 10–13, 15, 35, 39, 51, 52, 55, 57

PCA Principal Component Analysis. 7, 8, 10, 12, 15, 17

PD Parkinson's disease. 1, 23, 59

RF Random Forest. 10, 12

RMSE Root Mean Squared Error. 8–10, 12, 32, 38, 39, 42–44, 52, 56, 57, 59

RVM Relevance Vector Machine. 8, 9

RVoxM Relevance Voxel Machine. 8, 9

RVR Relevance Vector Regression. 9

SVM Support Vector Machine. 10, 11, 17

SVR Support Vector Regression. 10, 12, 16

Introduction

Ageing is a biological phenomenon, which may be accelerated by lifestyle or disease [1–4]. In the course of this project, two terms ought to be considered: chronological age (CA) and biological age. While the former can be defined as the age since the day of birth, the latter considers ageing as regulated by intrinsic mechanisms (at cellular level), and as influenced by extrinsic factor (such as environment and lifestyle) [5]. Chronological and biological ages do not always correspond. Human age prediction is applicable in different fields and can involve various sources such as DNA methylation, dental and face images and even electroencephalogram signals (EEG) [6].

The CNS is known as the body's central control centre. It is composed of the brain and the spinal cord, responsible for the body function such as movement, memory, sensation, and others [7]. Similarly to the rest of the body, CNS ageing may not progress at the same rate as chronological ageing, and it may differ between individuals of the same population. One intriguing characteristic of the CNS is that, as opposed to the Peripheral Nervous System, its capacity to regenerate its own cells is significantly reduced, leading to the degeneration of the brain tissue [8] over time. CNS ageing is influenced by several external and internal factors such as neurodegenerative diseases, dementia, and changes at structural, functional, or chemical levels. The project herein focuses on the structural changes that can be analysed through imaging the retina. Nowadays, the incidence of neurological pathologies such as Alzheimer's disease (AD), Parkinson's disease (PD), Diabetes Mellitus (DM) and other pathologies, is elevated in a worrying level. One thing that those pathologies have in common is that all of them affect the CNS in a way that ages it at a faster pace than would be verified in normal circumstances. That being said, the early identification of accelerated ageing is beneficial in the diagnosis of brain alterations that are typically identified only at later, advanced stages of development. This project is centred on findings from healthy controls, which will establish the foundations to further comparison to unhealthy individuals and, potentially, between distinct neurodegenerative disorders.

Collecting images of the CNS (*in vivo* and *in situ*) gives a more in-depth understanding about its structure [9]. The easy access to human retinal images allows us to get creative about how to use images of an organ that cannot be accessed in other non-invasive ways. In the literature, one can find that many researchers tried to estimate the brain age using different methods such as MRI (T1-weighted, sMRI, fMRI, etc.). The problem here is that those methods, although non-invasive, are slow, prone to movement artefacts, difficult to achieve, with limited access and expensive.

The eye and the brain share the same embrionary origin, and the retina is composed of several layers, some of which are similar to the ones in the brain [10]. As such, there is a portion of the retina (its six innermost layers, known as the neuroretina) whose structure mirrors that of the brain. This makes the eye, which is a directly observable organ accessible through optical imaging techniques, a "window" to brain, and therefore a good candidate to estimate the CNS age [11].

This project is centred on features extracted from retinal images, acquired from healthy controls via optical coherence tomography (OCT) imaging. OCT is a non-invasive imaging technique that provides detailed information on the retina, namely well-defined variations in content and structural organization. In addition, the acquisition time of 2.4 seconds per volume, allows an easy and fast data collection, with limited movement artefacts [12]. OCT uses low-coherence infra-red light, allowing for a high accuracy. Typically, in OCT imaging, transversal sections of the retina (A-scans), forming several B-scans (2-dimensional images) are obtained as a representation of the retinal structure. These generated (i.e. not real) images can be used to assess the retina's structure integrity, like an *in vivo* histological section.

In this project, a different approach was taken to analyse OCT data. Mean value fundus (MVF) images were generated foreach of the six neuroretinal layers. These MVF images were used as fundus references for OCT data [9]. Texture analysis was performed on these images to obtain features highlighting subtle changes in the retina's structure that would otherwise remain unnoticed order to obtain features from those images.

While two major types of texture can be defined, pattern and random texture, the one present in the images under study (MVF images of the layers of the neuroretina) is of the random type, where visually no repetitive pattern can be seen; however, it has a meaning when analysed. Texture analysis is a useful way of finding information embedded in images, allowing to obtain a much more considerable amount of information, in comparison to what is visible to the naked eye.

In texture analysis, statistical methods use the distribution of pixel values to obtain mathematical parameters (texture features), that can later be used to characterize or quantify differences in the image. Examples of statistical methods are image histograms, the run-length matrix, the grey-level co-occurrence matrix (GLCM), and absolute gradients. The GLCM is a second-order statistical methods, which allows for the extraction of statistical information regarding the distribution of pairs of pixels in the image. From the GLCM, metrics like image contrast, homogeneity, correlation and entropy can be obtained. Through all those metrics, we aim to build the model for healthy controls and further compare it with the non-healthy ones.

As previously demonstrated by the MSc candidate's research group [13–16], the texture features under study carry information on the brain status. Unfortunately, the unexpected constraints due to the COVID-19 pandemic led to limited access to the required computing power, which severely limited the possibility of using a significant number of therefore severely limiting the information content available for the aim of the work herein presented. Furthermore, each of the original texture features already present a significant compression of information regarding the amount of information gathered by the imaging technique.

By using Neural Networks (NN), fed by retinal data obtained through OCT images of healthy controls, the goal of this work was to estimate the CNS age and identify possible deviations from the CA and, hence, potentially identify unhealthy cases. The reason for the use of NNs is that it is a machine learning (ML) approach that is able to identify complex structures embedded in data and seems the most promising for the aimed tasks. Indeed, it is the ML approach most commonly used for these types of problems, and also a common choice for similar approaches in the identification of the CNS age, based on MRI [6, 17–19].

Similarly to the brain, which contains neurons as its basic unit, NNs uses perceptrons, which are the building blocks of a NN layer [20]. They incorporate several components, such as the input layer, weights, bias, net sum and activation function. Their weighted inputs are summed-up, added to the respective bias, and the result is applied to the activation function that will standardise the value as an output, which will serve as the next layer's input [21]. NNs end up classifying or predicting results based on inputs with outstanding performances. Individual perceptrons may be assembled into an extensive network to solve complex problems as is nowadays standard in our daily life (speech or object recognition) [20, 21].

NNs allow a computer to learn through an algorithm just by incorporating new data along with the expected output. Also, deep learning (DL) might be used through deep neural networks

by identifying patterns and models that explain the relationships between variables [22]. These algorithms are capable of recognize hidden features and learn complex representations of highly multidimensional data [4].

It is necessary to understand the healthy ageing in order to distinguish it from the ageing caused by neurodegenerative diseases. In addition, with the early identification of accelerated ageing, future health issues might be anticipated. A model was created to estimate the age of the CNS based on textures from images computed from OCT volumetric scans. The main goal was to obtain a predicted age (PA) value as close as possible to the true CA. A secondary aim of this project was to determine if all six layers of the neuroretina are equally significant for the CNS age estimation, or if a specific subset of these layers provides a superior classification performance. Moreover, following the same line of reasoning, do all texture features have an equal contribution? A third question that later emerged was whether or not there are any differences in the PAs between females and males. Finally, are there any significant differences between the right and left eye of the same individual?

1.1 Dissertation's Structure

This dissertation is organized into six core chapters:

- *Introduction* - where the aims of this project, their relevance and the used approach are first introduced (see Chapter 1);
- *State of the Art* - where the current state of knowledge about this topic is analysed. Some of the different approaches, theories, models and conclusions will be presented are used for further comparison and development of this study (see Chapter 2). The techniques used to acquire the data (MRI, EEG, OCT and others), details on the datasets under study, the methodologies used for age estimation, and the respective results and conclusions are presented. A summary about the information to retain from each paper from the state of the art can be seen in Table 2.0.1.
- *Methods* - where it is explained how the study was carried, which methods were used and why did we choose to follow this path, justifying all our choices. Moreover, all steps beginning with the data acquisition and pre-analysis, how the were features extracted, which metrics and heuristics were used and described (see Chapter 3).
- *Results* - in this chapter it is presented what resulted from the methodology used (see

Chapter 4);

- *Discussion* - all the data interpretation and analysis will be presented in this chapter. The meaning of the results and its relation with the research questions, the methodology and its relations with the literature (see Chapter 5);
- *Conclusion* - this chapter is dedicated to answering our main research question (and secondary ones), explain the implications of our findings and the limitations of the implemented research (see Chapter 6);

State of the Art

A considerable amount of studies were performed since the 20th century, encouraging the development and the deepening of knowledge on the subject of brain age. Predicting peoples brain's age, and the different forms to attain it, has been attracting the curiosity of several researchers. An appreciable number of techniques were used, from MRI and EEG to OCT. Since MRI is most used the imaging technique, it will be the first one analysed in the state of the art, and in a chronological order.

Franke *et al.* (2010) [23] estimated the age of healthy subjects from T1-weighted MRI scans using kernel methods, such as smoothing Full Widht at Half Maximum (FWHM), non-linear and polynomial kernels. Their motivation was the possible early identification of brain anatomy deviating from normal patterns of growth and atrophy in neurodegenerative pathologies such as the Alzheimer's disease [23]. They used the T1-weighted MRI from 550 healthy subjects from the IXI database [24] within the age range 19-86 years-old, that was divided into train and test sets (respectively, with 410 – age range 20-86 years-old – and 136 subjects – age range 19-83 years old). They also had additional samples of their own, 108 subjects (age range 20-59 years old), that were used for testing.

Principal Component Analysis (PCA) was used for data reduction, resulting in 410 principal components per subject. Support Vector Regression (SVR) was used for regression to estimate the subjects' age. Relevance Vector Regression (RVR) was also used as a regression method, in order to explore the influence of the regression method. The accuracy of these processes was established based on the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) (see Equations (2.0.1) and (2.0.2)). They explored the influences of the ability to generalize across scanners, data reduction method (PCA *versus* no-PCA), the preprocessing method used (the different kernels used and different spatial resolutions), and the influence of the size of the training dataset. The ADNI database [25] (people with early Alzheimer's Disease (AD) and cognitively normal elderly control subjects) was used in order to test the potential

of the model on age estimation. The age of both test samples was estimated and successfully proved efficient with a correlation of 0.92 and a mean absolute error of 5 years between the real and the estimated age. Their results suggests that the subjects' sex and the use of different scanners did not influence results. However, the accuracy of age estimation depended mostly on the number of subjects on the training set, on the image preprocessing, and the dimensionality reduction. For the AD group from the ADNI database, the estimated age exceeded the CA by ten years. The authors conclude, nevertheless, that their age estimation framework can help the identification of brain atrophy before the manifestation of clinical symptoms, leading to an early diagnosis of neurodegenerative diseases and their early treatment [23].

$$MAE = \frac{1}{n} * \sum_{i} |BA_i - CA_i| \tag{2.0.1}$$

$$RMSE = [\frac{1}{n} * \sum_{i} (BA_i - CA_i)^2]^{\frac{1}{2}} \tag{2.0.2}$$

In 2011, Sabuncu and Van Leemput [26] used the Relevance Voxel Machine (RVoxM), a Bayesian multivariate pattern analysis method (MVPA) designed for prediction purposes from image data, to test age prediction from structural brain MRI. The predicted outcome is a linear combination of a small number of spatially clustered batch of voxels. A total of 336 T1-weighted MRI images were used from healthy subjects within the age range 18-93 years-old from the OASIS dataset [27]. The authors also applied two additional methods as a benchmark: the Relevance Vector Machine (RVM) and the Relevance Voxel Machine with no spatial regulation (RVoxM-NoReg). The former method uses PCA to reduce the dimensionality of the images after the application of the linear RVM, and the latter has the same implementation as RVoxM but without the spatial regularisation. The RVoxM algorithm achieved the best RMSE (9.5 years) of the two.

In 2012 [28], the same authors went forward and did two different experiments again based on T1-weighted structural MRI scans. The first aimed to predict a subject's age from a volumetric grey matter segmentation (regression), and second to discriminate healthy controls and AD patients using surface-based cortical thickness measurements (classification). Their approach is an upgraded version of the RVoxM algorithm presented in [26],where now it also performed classification and presented additional details on theoretical derivations and extensive experimental results. They used the OASIS dataset [27], and 100 of the patients over 60 years-old were di-

agnosed with very mild to moderate AD. For the healthy controls, an RMSE of 7.9 years was achieved from the RVoxM method. RVM and RVoxM-NoReg were again used, and RVoxM still maintained the best RMSE value.

Franke *et al.* (2013) [1] also studied the brain age effects of type 2 DM in adults, as well as the relationships between brain ageing, risk factors, and functional measures. As referenced before, diabetes may accelerate the brain ageing, and this study used the BrainAGE approach ("Brain Age Gap Estimation" - the difference between CA and estimated brain age), determining the intricate, multidimensional ageing pattern. Kernel regression methods were applied (the same as used in other works from the same author [23]) to T1-weighted MRI anatomical brain images with the purpose of distinguishing brain atrophy that deviates from normal brain ageing. The authors used 561 healthy subjects from the IXI database [24] with an age range 20-86 years old [23] to train the age estimation framework. The BrainAGE analysis for this study used existing records of 185 subjects where 98 of them were diagnosed with type 2 DM. The same preprocessing and data reduction from [23] was executed, and an Relevance Vector Regression (RVR) model was used. Results suggest that lifestyle factors (such as alcohol consumption and smoking) has a significant role in the variance between CA and estimated brain age. BrainAGE scores for males were higher when compared with females.

Konukoglu *et al.* (2013) [29] used random forests for neighbourhood approximation. They introduced a supervised learning algorithm, Neighbourhood Approximation Forest (NAF), which can, from an image training database and a user defined distance metric between images, learn to use appearance-based features to cluster images and to approximate the neighbourhood structured induced by the distance (k-nearest neighbour clustering method (kNNc)). The algorithm is efficient in inferring nearest neighbours of an out-of-range image. The authors used it for age prediction from 414 T1-weighted brain MRI images from the OASIS database [27] (and other applications that do not concern our subject). The authors claim to have similar results to the studies of Sabuncu *et al.* [26] and Franke *et al.* [23] since they have applied their method to a subset of the same database. With the NAF algorithm, the obtained RMSEs for the regression between within 8.6-10.9 years, depending on the k value in use (number of neighbours), and using kNNc for unsupervised clustering of features RMSEs between 11.9-21.0 years, also depending on the k values in use. kNNc defines the neighbourhoods by the L2 distance between feature vectors. NAF has a better performance than kNNc because it learns which features to approximate and defines a neighbourhood for the specific task of age regression, in contrast with the kNNc that

assumes that feature distances are correlated to the differences between ages and does not learn.

NNs were used by Lin *et al.* (2015) [19] to predict healthy older adults brains' age based on structural connectivity networks. The age range of the 112 participants was 50-79 years old. All participants were screened for any dementia, mild cognitive impairment, and depression. T1-weighted MRI images were acquired from all the subjects. The authors made use PCA and of backpropagation artificial neural networks (BPANN) to predict brains' age. Also, to improve the performance and reduce training time, they used a hybrid algorithm that integrates the genetic algorithm (GA) and the Levenberg-Marquardt algorithm (LM). Their best RMSE, of 5.11 years, was obtained through the use of the BPANN with GA and LM models and was achieved within the lowest computing time per trial (1.3 hours). For individuals younger than 55 years old, an overestimation of the biological age was detected. For individuals older than 70 years old the opposite happened.

Liem *et al.* (2017) [30] studied how to improve the prediction of the brains' age from multi-modal imaging data from MRI. Instead of only using structural or functional brain data, they used cortical anatomy and whole-brain functional connectivity. The participants in this study were randomly selected, ending up with 2354 individuals with ages between 19-82 years-old. These kinds of data allow improving brains' age prediction with an MAE of 4.29 years, using SVR and RF as machine learning methods. They state that the discrepancy between the CA and PA captures cognitive impairment. Their results were robust to confounding effects like movement artefacts. Also, they state that, given their findings, machine learning methods are promising for establishing brain-based biomarkers to help to diagnose neurocognitive disorders.

Valizadeh *et al.* (2017) [31] studied the possibility to predict the brain's age based on brain anatomical measures such as cortical volume, thickness, and area, among other, of healthy subjects from sMRI and fMRI data. The authors made use of three different datasets, and only subjects who had no reported history of neurological diseases were included in the study; the authors ended-up with 3144 subjects with an age range between 7-96 years old. Six statistical techniques, multiple linear regression, ridge regression, NN, k-nearest neighbourhood (kNN), SVM, and random forest (RF), were used allowing obtaining high prediction accuracies (84% using NN and SVM, and 73% with multiple linear regression), even with a small but age-representative dataset. Their results suggest that CAs can be well predicted based on anatomical measures only.

Cole *et al.* (2017) [17] applied a CNN as a deep-learning based-predictive model on T1-weighted MRI data (raw data). The objective was to determine if age-associated functional measures and mortality were related. The authors made use of healthy adults (N=2001; mean±SD age= 36.95±18.12 years; age range = 18–90 years; # males = 1016; # females = 985) to demonstrate the model's accuracy. They tried to demonstrate the heritability of brain-PA with samples of monozygotic and dizygotic female twins (N = 62). Then they examined the reliability of the brain-PA using two samples. The resulting PAs were compared, on all datasets, by a Gaussian Process Regression (GPR) approach. The authors managed to accurately predict the brains' age using the grey-matter (GM) and white matter (WM). For the GM, a correlation between ages of 0.96, and an MAE of 4.16 years were obtained. For all models and all input data, BPA was $h^2 \geqslant 0.5$, suggesting that it is a heritable phenotype. All tests were made on Lothian Birth Cohort 1936 (LBC1936) study participants, in which 669 of 1091 them had MRI (male = 352 and female = 317).

From the authors' words, the "brain-PA represents an accurate, highly reliable and genetically-influenced phenotype, that has potential to be used as a biomarker of brain ageing" [17] and raw data generated good results, which reduces the computational resources required, making the process suitable to be used in the clinical context. To prove that all healthy participants were truly healthy, a screening was made to check if they had either any neurological or psychological disorder, head trauma, or major medical health issues. For the first time, 3D-CNN network (3-Dimensional Convolutional Neural Networks) was shown to estimate the brain age from MRI data accurately by Cole *et al.* [6, 17]. They concluded that the neuroimaging could help in the creation of a feasible ageing biomarker and contribute with prognostic information.

Lancaster *et al.* (2018) [32] also used T1-weighted MRI images for brain age classification and prediction. Bayesian optimisation for neuroimaging preprocessing was used instead of normalisation and re-sampling to a common voxel size, often with arbitrarily selected parameters, followed by spatial smoothing. In this work, re-sampling parameters were optimised using a Bayesian optimisation to identify the optimal voxel size and smooth kernel sizes for each task. The authors used data of 2003 healthy individuals with an age range of 16-90 years old, and trained the SVMs to distinguish between old (>50 years old) and young (<22 years old) brains and to predict their CA. For the age prediction, they managed to achieve an MAE of 5.08 years. The authors state that the use of *a priori* unbiased parameter optimisation could be highly beneficial in machine learning contexts, where researchers may not have expertise on every relevant experimental parameter, stating that brain ageing could benefit from applying those kinds of

optimisation approaches to improve the study sensitivity and reduce the bias. Moreover, they state that large datasets are necessary for the optimization of the work.

Al Zoubi *et al.* (2018) [33] went further and predicted age from brain (EEG) signals through a machine learning approach. Their objective was to move away from MRI studies on BrainAGE and use a combination of EEG signals and machine learning approaches, not commonly used then, investigating if brain EEG signals were affected by age-related changes and if the CA could be predicted. Regression algorithms such as Elastic Net (ENET), SVR, RF, eXtreme gradient boosting Tree (XgbTree), and Gaussian Process with Polynomial Kernel (gaussprPoly), were performed. Their methods included the Nested-Cross-Validation (NCV) to provide an unbiased prediction for age and stack-ensemble learning from EEG features, obtaining a correlation between CA and PAs of 0.6. The population used was in the age range of 18-57 years old and the best results were achieved with the SVR with the RBF kernel (m/SD): $R^2 = 0.34(0.056)$, MAE = 7.01(0.68) years and RMSE = 8.7(0.63) years. Also, the stack-ensemble improved the performance with $R^2 = 0.37(0.064)$, MAE = 6.87(0.69) years and RMSE = 8.46 (0.59) years. No differences between females and males were identified in the study. The authors concluded that their machine learning framework and their EEG signals' features allowed an accurate estimation of CA and BrainAGE.

Spatial and temporal features of the human brain functional networks were also used to predict brain age by Zhai and Li (2019) [34]. Two different datasets were used, the Enhanced Nathan Kline Institute—Rockland Sample (NKI-RS-E) [35] where the authors selected 496 individuals with the age range of 6-85-years-old, and the Nathan Kline Institute—Rockland Sample (NKI-RS) [35] where they selected 207 subjects between the ages of 4 and 85-year-old. fMRI images were used to capture brain activity and measure the organisation of human brain networks. The authors performed PCA, and used three different regression models: ordinary linear regression (OLS), SVR and Lasso were trained through the extracted features. The fractional amplitude of low-frequency fluctuation (fALFF) method provides the temporal features. The authors computed edge features through a new feature reduction method and, depending on the regression model, different MAE values were obtained (OLS with MAE = 14.2 years, SVR with MAE = 11.4 years and Least Absolute Shrinkage and Selection Operator (LASSO) with MAE = 9.2 years). Temporal features based on fALFF used as predictor variables, and the three different regression models were learned to make the brain age prediction.

Sajedi and Pardakhti (2019) [6] did a survey on age prediction based on brain MRI images and the used approaches, including the preprocessing and algorithms. The following Brain Age Estimation (BAE) methods were categorised based on two factors: MRI image processing, and the generation of machine learning algorithms. The deep learning and shallow classification methods are the ones that simplify the computation and provide the most accurate results.

Following the previous study of Sajedi and Pardakhti [6], Gialluisi *et al.* (2019) [4] refer to machine learning approaches for the estimation of biological ages. The biological age is sometimes computed via biomarkers (circulating – such as blood – or non-circulating) that are fundamental for monitoring healthy ageing and prevent clinical events. This study [4] focused on the potential use of blood biomarkers as a biological age estimation on the grounds of the speed and ease estimation. Some machine learning approaches are reviewed in this work, such as CNN and Gaussian Process Regressions (GPR), the former providing the highest correlation between the estimated and CAs. The authors also discuss genetic and environmental, mortality, and others, in order to assess the potential of a biomarker of ageing and hypothesise new approaches at the single organ and whole organism level for biological age estimation.

These authors studied two tissues: blood and brain. For each, a machine learning algorithm was trained and tested with different populations obtaining different MAE values, between 5.59 and 6.36 years for blood and deep neural networks, and between 4.16 and 11.81 years for the brain (based on structural MRI images) using GPR and CNN. For these tissues, the authors proved that environmental influences have a positive (meditation, education, music playing) and negative (tobacco smoking, malnutrition, and others) impact in the biological age such as health conditions increase biological age (HIV, Down Syndrome, and others).

A DL approach was used by Jonsson *et al.* (2019) [18] to predict the brain's age using T1-weighted MRI images of 1264 healthy individuals(18-75 years old) from an Iceland population. These images were used to train the model later tested on two different datasets: the IXI Dataset [24], with 544 individuals (20-86 years old), and UK Biobank [36], with 12395 individuals(46-79 years old). The difference between CA and PA holds promise for the study of normal brain ageing and diseases, and genetic discovery via genome-wide association studies [18]. The authors made use of transfer learning to improve the accuracy of the age predicting method and discovered two sequence variants. The sequence variants were associated with the brain structure and the brain ageing. CNN was used and Exponential Linear Unit was used as activation function. Their best MAE value (3.58 years) was achieved when combining CNN predictions. Their deep learning

methods proved to be useful when studying different diseases.

Hu *et al.* (2019) [37] proposed a two-stage age prediction method for infants, through hierarchical rough-to-fine model (HRtoF), based on cortical features that capture the rapid and heterogeneous changes during the brain development, achieving high accuracy on infant age prediction. For the task at hand, the authors had a dataset of T1-weighted MRI images from 50 infants aged between 1-24 months old. Due to the rapid brain development, as an infant, conventional regression models are not accurate enough hence the proposed two-stage prediction model: first, there is a rough prediction of the age, followed by a fine prediction based on a learned, age-group-specific, regression model. Then these are combined with another prediction method (HRtoF), splitting those rapid changes into several time ranges. By using this last model, the correlation coefficient between predicted and CA reached 0.963, which shows that the cortical maturation is strongly correlated with the CA. Also, the MAE for this model varies between 11.0-60.7 days, depending on the infants' age.

In [38], Sone *et al.* (2019) used T1-weighted MRI images and machine learning methods to estimate an individual's brain age in diverse forms of epilepsy. Healthy subjects (1196 subjects with an age range between 20-89 years old) and unhealthy subjects (318 epilepsy' patients over 20 years old) were used. Support vector regression was used on healthy controls to build a brain's age prediction model for which an MAE = 5.28 years was attained. According to the developed model, increased brain's age was found in most of the subjects with epilepsy, suggesting that brain's age prediction might provide insights into changes unfolding in the epilepsy brain.

Franke and Gaser (2019) [3] wrote a review article about the insights gained during the last ten years about the BrainAGE method as a neuroimaging biomarker of brain ageing, which summarised all studies on this matter and evaluated the effects of the interaction of genes, environment, life-burden diseases, and lifetime on individual neuroanatomical ageing. Supervised, linear or non-linear, pattern recognition techniques are used to quantify the accelerated or decelerated brain ageing through structural MRI. Once again, there is a reminder for the importance of linking the processes of brain ageing to the mechanisms of neurodegenerative diseases to find new approaches to justify the increasing numbers of neurodegenerative diseases [3] brought by the extended lifespan and growth of the population.

Since ageing is a complex time-dependent functional decline, and unique in its way, it is an important topic of study. This is an approach that establishes reference curves for healthy brain's

ageing from adulthood until a state of agedness [3] and also the curves for brain maturation from childhood into young adulthood. The brain's ageing deviation from the healthy brain ageing reference curve was examined in neurodegenerative diseases, also for predicting the worsening of cognitive functions and their conversion to AD [3]. Effects of lifestyle and health factors were analysed and adapted to experimental studies with rodents and non-human primates. An automated pre-selection of the MRI data reduces the computational resources needed, avoids the overfitting usually present in pattern recognition works. A Relevance Vector Regression (RVR) is applied to establish ageing patterns in the brain. The authors then applied PCA to reduce data dimensionality.

Cross-validation was used to generate and validate the model of brain's ageing [3]. The F statistics are also used to analyse the fit between BA and CA in regression models. To test the model usability in new datasets, and the strength of the model as a diagnostic/prediction tool, the model was validated on independent test samples composed of healthy and patient subjects.

Further analyses on the same work demonstrated that the number of subjects in the reference sample has the principal role in the brain age prediction accuracy. The preprocessing and model-training algorithm are also important parameters that influence the performance of the model. Results of BrainAGE were correlated with smoking, alcohol consumption, and depression, among others. When gender-specific effects were studied, the health parameters explained 39% of the observed variance with Body Mass Index (BMI), uric acids, and others, for male subjects. For females subjects, 32% of the variance observed in the difference between the CA and PA was explained by health parameters.

In summary, premature brain ageing can originate from AD, HIV, traumatic brain injury, chronic pain, low physical and mental fitness, and others. The combination of structural and functional biomarkers of brain's age allows increasing the sensitivity and sensibility of error detection in BA in comparison with the CA in different neurodegenerative diseases.

Ning *et al.* (2020) [2] studied the relative brain's age in association with smoking tobacco, the consumption of alcohol, and some genetic variants, using structural T1-weighted MRI scan data from 17308 subjects of the UK Biobank [36]. Subjects' age ranged between 45.2 years and 80.7 years old and were all cognitively normal. The authors built a model for the brain's age prediction of linear regression with LASSO regularisation. They used 403 quantitative brain measures as predictors with no pre-selection (due to the large training set in comparison with the number of predictors). Between the predicted and CAs in the training set, the MAE was

3.5 years. The older subjects tend to have a smaller PBA in comparison with the CA. The opposite effect was found for the younger subpopulation. Results suggest that brain ageing can be associated with smoking, alcohol consumption, and chromosome 17 (2Mb locus) [2].

Niu *et al.* (2020) [39] explored an improved prediction for the brain's age through multimodal neuroimaging data, recognising that machine learning has a remarkable ability to help on the understanding of cognition and mental disorders. They used 36 combinations of machine learning models (which includes deep learning) and imaging features and evaluated their corresponding prediction performance since they felt the need for systematic comparison. The authors carried a follow-up study to investigate how to combine data from different sources such as MRI, DTI (diffusion tensor imaging), and rs-fMRI, to improve the prediction accuracy of brain's age, based on a dataset with 839 subjects with various health conditions, in which they used single and multimodal brain imaging.

They also found that the traditional approach to predicting the brain's age gap had a systematic bias (underestimates the age of older subjects and overestimates it for younger ones) and proposed a method to correct it by taking into account different features such as gender,CA, and their interactions. The behavioural performance across subjects was also one of the targets to study. Depending on the regression model used, MAE values also varied (Ridge with MAE between 1.40-1.96 years, SVR with MAE between 1.43-2.0 years, GPR with MAE between 1.38-1.97 years, and DNN with MAE between 1.38-2.39 years). The authors found a gender difference on the brain development trajectory. The authors concluded that their findings could be advantageous to the optimisation of the existent methodologies in brain's age prediction [39].

As shown above, most of the state-of-the-art focused on MRI imaging and machine learning approaches. However, only a few of them address the use of OCT and the retina, despite the much simpler data acquisition process, the much broader availability of the imaging technique, and the more inexpensive instrumentation and operation costs.

In 2010, Liu *et al.* [40] made use of OCT images of the retina (centred at the fovea) to diagnose multiple macular pathologies based on an automatic method by only giving a manually-selected x-z 2D slice (B-scan). The proposed method can discriminate between the normal macula and maculas presenting macular pathologies such as macular hole (MH), macular oedema (ME), and age-related macular degeneration(AMD). The dataset is made of 326 OCT B-scans from 136 patients. A multi-scale spatial pyramid (MSSP) and the dimension-reduced Local Binary

Pattern (LBP) histogram based on PCA were used to construct a global descriptor for the aligned image to represent the retina at different spatial scales and granularities. The MSSP was used to obtain the geometry of the aligned retina at multiple scales and spatial resolutions since it preserves spatial organisation in conjunction with multi-scale modelling. The LBP was used to obtain the micro-patterns at local block sites, as it summarises the local structure and is highly discriminative. An histogram of the LBP was computed for each block of pixels in the MSSP to encode the statistical distribution of different micro-patterns such as spots, edges, corners, and flat areas [40]. PCA was used to reduce the dimension of LBP histograms since the concatenation of those to achieve a global feature vector results in a considerable dimension.

A 2-class non-linear Support Vector Machine (SVM) with radial basis function (RBF) kernel was trained using a labelled training set for each pathology. Their results validate the proposed approach and demonstrate its effectiveness. The authors went forward in 2011 [12] by building a classifier to differentiate cases within the macular hole category and demonstrating that their approach is also accurate for other cross-sections slices (B-scans) as long as the labelled slices from that same location were also collected for the training of the classifier. Here, a binary non-linear SVM was used. However, their approach presents some limitations such as only analysing a single slice and not the full OCT volume, being that one slice manually selected by ophthalmologists, henceforth requiring one expert to be operated. On the other hand, they have the advantages of being able to extend the approach to identify additional pathologies, used in other retinal locations besides the foveal one [12].

The OCT data of the human retina contains more than what we expect as suggested by Bernardes *et al.* (2012) [41], such as the co-existence of the blood-retinal barrier status information in the healthy status and information on the ageing of the retina regarding the disease of diabetes from the healthy status. An SVM classifier was used with an accuracy of 64% in correctly classification of eyes in their proper group. They demonstrated the ability of the system to distinguish between the eyes of healthy volunteers' from the ones of the diabetic patients even with matching ages, opening the possibility of using the human eye to check on changes in the brain in a non-invasive way.

In 2019, Liu *et al.* [10] explored the potential of using retinal fundus images to estimate the biological age as a classification problem. Two datasets were used: the Yangxi Eye Study [42], and the Shenzhen Eye Study. The authors did not specify which equipment was used to collect the fundus images, and refer to it in general terms as "multiple types of fundus cameras". The

training was made using only the healthy individuals from the Yangxi Dataset, while the validation used healthy individuals from the Yangxi Dataset and healthy and unhealthy individuals from the Shenzhen Dataset.

The subjects age range was 50-94 years old, for the healthy individuals from the Yangxi Dataset, 50-73 and 50-93 years old for the healthy and unhealthy individuals from the Shenzhen Dataset, respectively. A total of 12000 images were used for training purposes and 8081 images for validation purposes. To the best of our knowledge, this is the study with the biggest number of subjects studied so far. The authors used the VGG-19 CNN architecture with batch size = 64, learning rate = 0.001, weight decay = 0.008, and momentum = 0.9. Input images were cropped and underwent a thorough detail enhancement (method based on domain transform filtering) and later resized to 299x299 pixels. Their best validation score was for the Yangxi dataset, for the method in which the authors combined the network with the detail enhancement and a Gaussian label distribution loss, achieving an MAE = 3.67 years. For validation, the Shenzhen Dataset was used with the CNN plus detail enhancement and joint loss function to achieve an MAE of 3.39 and 4.21 years, respectively for the healthy and unhealthy groups. The authors concluded that, with their method, the prediction of the biological age based on retinal imaging could be used as an effective biomarker of ageing.

The bibliography is far from being fully covered. There is a lot to work with on the brain's age topic. Machine learning, neural networks and deep learning are the most referenced methods thus far on predicting brain's age. The use of OCT would be a considerable advance in assessing the central nervous system's age due to the fast and ease of operation, the non-invasive nature and the ease of access to a much larger population. Table 2.0.1 summarises the essential information to retain from all the methods aforementioned and will later help when comparing the state-of-the-art with achieved results.

Table 2.0.1: Summary of important information to retain from the state of the art.

Summary of the State of the Art						
Authors	**Year**	**Method**	**N. of Study Subjects**	**Age Range (years)**	**Ratio Train/Test(%)**	**RMSE/MAE**
Franke et al. [23]	2010	T1-Weighted MRI	654	19-86 (healthy)	63/27	MAE = 5 years (Controls) MAE = +10 years (AD patients)
Sabuncu and Van Leemput [28]	2012	T1-Weighted MRI	336 (healthy) 100 (AD patients)	18-93	80/20	RMSE = 7.9 years
Konukoglu et al. [29]	2013	T1-Weighted MRI	414	18-93	Variable	RMSE = 8.57-10.91 years (NAF) RMSE = 11.88-21.01 years (kNNc for Unsupervised Clustering of Features)
Lin et al. [19]	2015	T1-Weighted MRI	112(healthy)	50-79	Variable	MAE = 4.31-6.14 years RMSE = 5.11-6.77 years
Liem et al. [30]	2017	MRI	2354	19-82	50/50	MAE = 4.29-7.29 years
Valizadeh et al. [31]	2017	sMRI and fMRI	3144 (healthy)	7-96	50/50	MAE = 1.23-5.23 years
Cole et al. [17]	2017	T1-Weighted MRI	2001 (healthy)	18-90	80/20	MAE = 4.16-11.81 years RMSE = 5.31-15.10 years
Lancaster et al. [32]	2018	T1-Weighted MRI	2003 (healthy)	16-90	90/10	MAE = 5.08 years
Al Zoubi et al. [33]	2018	EEG Signals	468 (healthy and patients)	18-57	Variable	MAE = 6.87(0.69) years RMSE = 8.46 (0.59) years
Zhai et al. [34]	2019	fMRI	703	4-85	Not explicit	MAE (OLS) = 6.7-9.8 years (network and edge-based) and 14.2 years (fALFF) MAE (SVR) = 7.1-9.9 years (network and edge-based) and 11.4 years (fALFF) MAE (LASSO) = 6.5-9.2 years (network and edge-based) and 9.2 years (fALFF)
Gialluisi et al. [4]	2019	Structural MRI (brain) and	24325	35-99	Not explicit	MAE(blood) = 5.59-6.36 years MAE(brain) = 4.16-11.82 years
Jonsson et al. [18]	2019	haemochrome markers (blood)	14203 (healthy)	18-75 (train) 20-86 and 46-79 (test)	8.9/91.1	MAE = 3.388-5.149 years (test set)
Hu et al. [37]	2019	T1-weighted MRI	50 (healthy)	1-24 months	Not explicit	MAE (HRtoF) = 11.0-60.7 days
Sone et al. [38]	2019	T1-weighted MRI	1514 (1196 healthy and 318 patients)	20-89 (healthy) >20 (patients)	79/21	MAE = 5.28 years (healthy) MAE = 10.6 years (patients)
Ning et al. [2]	2020	T1-weighted MRI	17308	45.2-80.7	30/70	MAE = 3.8 years (training set)
Niu et al. [39]	2020	T1-weighted MRI	839 (healthy and patients)	8-21	90/10	MAE(Ridge) = 1.396-1.959 years MAE(SVR) = 1.426-2.003 years MAE(GPR) = 1.384-1.965 years MAE(DNN) = 1.381-2.392 years
Liu et al. [10]	2019	Fundus images and CNN	14016 (healthy)	50-94	85.6/14.4	MAE = 3.67-5.02 years

Methods

MATLAB R2020a (The MathWorks Inc., Natick, MA, USA) was used to perform all the needed processing.

3.1 Participants

Our dataset is composed of 100 healthy controls, of which 50 are females, and 50 are males. Participants were selected to safeguard the age-matched between groups and a sex-balanced dataset. Their age ranges between 19 and 74 years old; however, the dataset is skewed towards the younger population. OCT data collection was performed under a data collection protocol, approved by the Ethics Committee of the Faculty of Medicine of the University of Coimbra, and performed in compliance with the Declaration of Helsinki [43]. Both eyes of the participants were analyzed. However, we have a total of 99 acquisitions for each sex, making a total of 198 eyes for study since, respectively, one right eye and one left eye from the female and male groups, were excluded due to the low quality of the scan [13]. The age distribution across the dataset is presented in Table 3.1.1.

Table 3.1.1: Demographic data of the dataset.

Demographic Data					
Group	**N**	**Age(years) Mean±STD**	**Age(years) Min-Max(Median)**	**Eyes Right(Left)**	**Total Acquisitions**
All Females	50	40 ± 17.1	19-74(40)	49(50)	99
All Males	50	40 ± 17.1	19-74(39)	50(49)	99

3.2 OCT Imaging

Data was gathered by the Cirrus SD-OCT 5000 (Carl Zeiss Meditec, Dublin, CA, USA) using the 512x128 Macular Cube Protocol (centred on the macula).

Figure 3.2.1: OCT image acquisition (source: https://www.glaucoma.org/treatment/imaging-of-the-optic-nerve-what-is-it-and-why-is-it-needed.php)

Optical Coherence Tomography (OCT) is a non-invasive, contactless imaging technique which performs optical sectioning, therefore providing cross-sectional tomograms of the retinal tissue at microscopic resolution. Its working principle is analogous to that of the ultrasound [12,40,41,44] allowing *in vivo* and *in situ* imaging of the human retina [9]. The system has a 5 μm depth resolution, a 20 μm transversal resolution and a scanning speed of 27,000 A-scans per second, hence allowing for the fundus image in 2.4 seconds for a volume of 128 B-scans, each composed of 512 A-scans, for a total of 65536 A-scans [41]. The imaging modality is based in the interferometer principle [9].

It uses a near-infrared light beam from a superluminescent diode that passes through the retina. The images are generated from the reflections and backscattering originated from the refractive index changes along the light-path, which depend on the structural organization of the eye [41]. That being said, it provides layer-wise information of the retina [45] such as retinal thickness, being useful in the evaluation of changes in the retina, possibly associated with neurodegenerative diseases [14].

This is done with A-scans and B-scans: A-scans convey information on the refractive index change throughout the beam light-path (anterior-to-posterior direction), while B-scans are a 2D array of A-scans coded in grey-scale to produce histology-like images of the retina to convey

its structure and organization (Figure 3.2.2). The 3D OCT data is organized as a sequence of slices (B-scans, see Figure 3.2.3) corresponding to a sequence of x-z scans [12]. The generated images not real but they are an *in vivo* histological section of the retina. Any possible color is pseudo-color, only used for representation purposes.

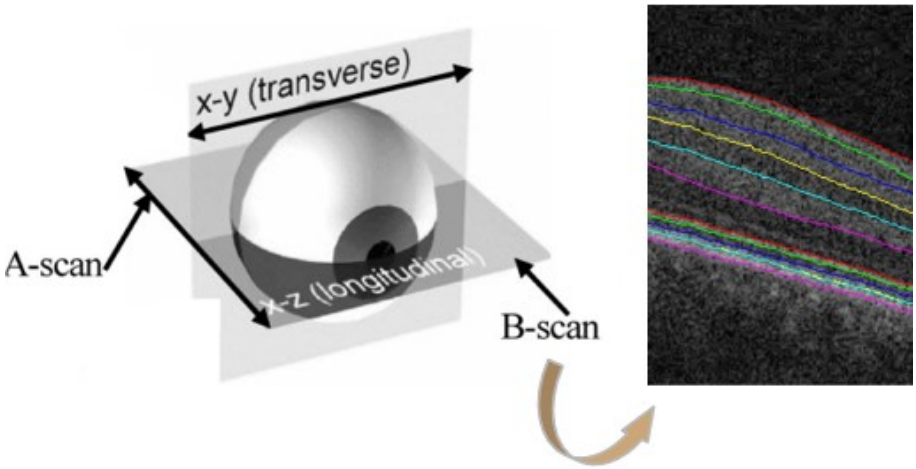


Figure 3.2.2: Representation of the OCT image acquisition process.

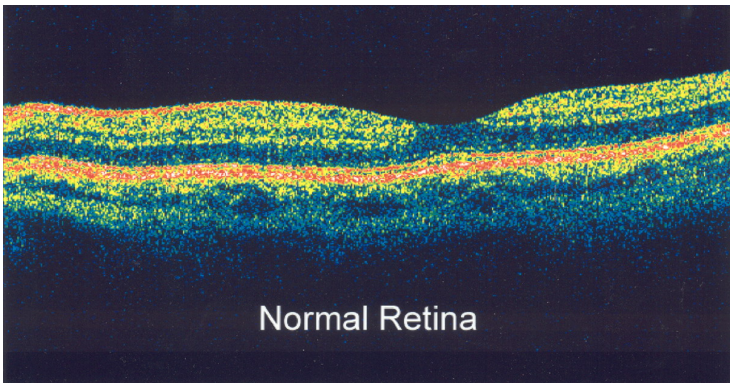


Figure 3.2.3: Generated image of the retina (source: `http://www.njvision.net/our-services/optical-coherence-tomography/`) .

Data obtained through OCT has been several times suggested as a new biomarker of ocular and neurodegenerative diseases (like AD, PD , multiple sclerosis, and others) [14, 15, 46–49]. Using texture analysis of the OCT data itself (in humans and animals) can also be promising [14, 15, 48].

3.3 Image Processing

The segmentation of the six anterior layers of the retina (the neuroretina) is necessary to distinguish/separate the layers. The OCT Explorer software (Retinal Image Analysis Lab, Iowa

Institute for Biomedical Imaging, Iowa City, IA, USA) [50–52] was used to identify:

- the Retinal Nerve Fibre Layer (RNFL),
- the Ganglion Cell Layer (GCL),
- the Inner Plexiform Payer (IPL),
- the Inner Nuclear Layer (INL),
- the Outer Plexiform Layer (OPL), and
- the Outer Nuclear Layer (ONL).

From the segmentation of the retinal layers and for each one of them, ocular fundus images are generated. The colours are only for representation purposes (pseudo-color).

The Mean Value Fundus (MVF) images were used as a retinal fundus reference for OCT data. Each pixel of the MVF images corresponds to the average of the A-scan values comprised within the respective layer. All remaining A-scan values are ignored [13]. When the light gets to the region of interest, occurs an hyper-reflectivity of the region, generating a clear distinction to the regions where the light was absorbed by hemoglobin present in the retina vessels (see Figure 3.3.1) [9].

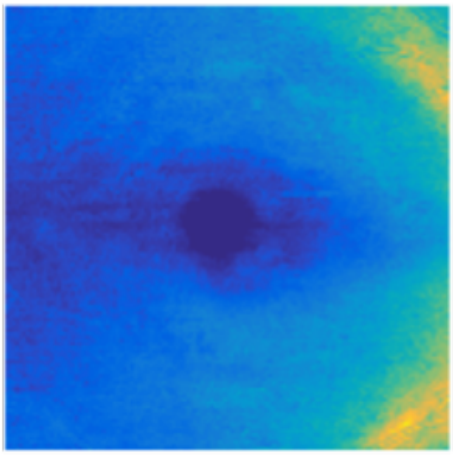


Figure 3.3.1: Colour-coded Mean Value Fundus (MVF) image from the Retinal Nerve Fibre Layer.

Since both eyes of all participants were considered, MVF images of all left-eyes were horizontally flipped to match the right ones. This procedure allows to account for the asymmetry of the human macula and ensures that local texture metrics can be compared within the entire dataset [13].

3.4 Texture Analysis

Texture is a function of spatial variation in terms of pixel intensity. It may be associated with the local variation in the intensity, or colour, of pixels of an image [53]. Because OCT does not convey any colour information, only the pixels' intensity will be considered onwards. If we look to an image, we can easily identify or perceive different textures (either random or presenting a pattern), and because texture allows telling two images apart, it may be regarded as a potential discriminator between sets of images [53]. As such, a number of metrics (statistical, structural, and others) can be computed based on texture to characterise different images and, potentially, leading to the identification of differences of interest, like the ones discriminating between healthy control groups and patients with a particular disease [15]. Some imaging modalities and imaged organs will generate random texture images, which the human visual system cannot categorise. However, texture analysis methods like the grey-level co-occurrence matrices (GLCM), allow for the computation of informative image features like coarseness, contrast, roughness, and others [54–56].

For each one of the MVF images, two different texture analysis was made, one based on the Grey-Level Co-occurrence Matrix (GLCM) (Section 3.4.1) and one based on the Dual-Tree Complex Wavelet Transform (DTCWT) (Section 3.4.2), which is a transform-based method of texture analysis. The former allows analysing pixel pairs for local intensity (grey-level) variations while the latter allows appraising the macular region at a larger scale.

3.4.1 Grey-Level Co-occurrence Matrix (GLCM)

Each image is composed by pixels and each one of them has a certain intensity. GLCM method is based on a tabulation of how many occurrences of transition between two pixel-intensities verify in a particular direction and at a certain distance. GLCMs are characterised by being square in which the number of rows/columns equals the number of grey-levels of the image, that is, the quantization level of the image. Furthermore, considering as the same transitions in a particular orientation, irrespectively of the direction (angles) or the distances between two pixels, renders the matrix diagonally symmetrical. While conveying all the required information, these matrices cannot be used directly since it only apprehends the properties of the image texture, hence being used through metrics derived from the distribution of its values [57].

Each one of the images were down-sampled to 128x128 pixels in order to make the samples isotropic in the vertical and horizontal directions (isotropic sampling). The matrix size was

reduced by converting the image to a 16 grey-level. To obtain local GLCM features, the images were divided into 7x7 blocks to make the central block to overlap with the fovea. Each block was independently analysed, except blocks on the 4^{th} row and the 4^{th} column (the central ones, as represented by the shaded blocks in Figure 3.4.1), as they encompass the foveal region where the neuroretina is considerably reduced or even nonexistent. In Figure 3.4.1 each one of the 3x3 white blocks was aggregated to form quadrants, and, consequently, reduce the number of features derived per image.

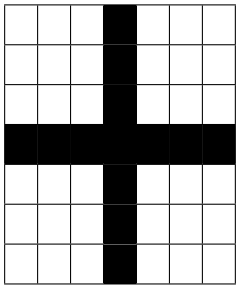


Figure 3.4.1: Representation of a computed fundus image from the volumetric macular cube scan.

Four pixel pairs orientations were used (0°, 45°, 90° and 135°) with a distance of one pixel ($d = 1$) to compute four GLCMs per block. Opposite angles (180° apart) were considered the same. For each block and every four directions, 20 features were computed. The maximum value across all directions of a particular feature (metric) was chosen as the feature value for the block in order to integrate existent differences in orientation (for example for nerve fibres). Quadrants were made from aggregating 3x3 blocks: the temporal superior (Q1), the nasal superior (Q2), the temporal inferior (Q3), and the nasal inferior (Q4) [13]. Finally, the ultimate feature value for each quadrant was the average of each one of the features across the 3x3 blocks [13].

The extracted features were Autocorrelation [55], Cluster Prominence [58], Cluster Shade [58], Correlation [57], Difference Entropy [57], Difference Variance [57], Dissimilarity [56], Entropy [57], Homogeneity (also known as Inverse Difference Moment [57]), Inertia (also known as Contrast [57]), Information Measure Correlation 1 - IMC1 [57], Information Measure Correlation 2 - IMC2 [57], Inverse Difference Moment Normalized - IDN [54], Inverse Difference Normalized - INN [54], Maximum Probability [55], Sum Average [57], Sum Entropy [57], Sum of Squares [57], Sum Variance [57], and, finally, Uniformity (also known as Angular Second Moment, or Energy [57]). Henceforth, a total of 80 GLCM-based features per layer were computed amounting to a total of 480 features per eye.

3.4.2 Dual-Tree Complex Wavelet Transform (DTCWT)

The Discrete Wavelet Transform is used as an image processing technique yielding a multi-scale representation of an image (or signal). However, it lacks directionality, which complicates modelling and processing ridges and edges (geometric image features), and is vulnerable to any shifts (shift variance). When using the same filter (either high or low-pass filter) in both directions (vertical and horizontal), it mixes the -45 and +45 degrees turning the direction check difficult [59,60].

The Dual-Tree Complex Wavelet Transform is implemented to improve the directional selectivity jeopardized by the DWT. When using a 2-dimensional DTCWT, six directionally selective sub-bands ($\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$) are produced for both real and imaginary parts per scale, together forming the DTCWT. The magnitude of the DTCWT coefficients is given by the magnitude of the complex-values [60].

The variance feature for each decomposition level and direction can be extracted from the coefficients of DTCWT. It measures the spread of grey-level distribution of the subbands, reflecting the contrast of the texture. With a larger variance, comes a wider variation of the grey-levels of the image [60]. This is a complement to the use of the GLCM method. While the GLCM was used to compute local features, the DTCWT was used to compute global features, i.e. across the whole image area (without any block division). Six different features were computed, based on the DTCWT coefficients at the six directionally selective subbands mentioned above: Variance-15°, Variance+15°, Variance-45°, Variance+45°, Variance-75° and Variance+75°.

Combining the DTCWT and GLCM features results in 86 features per layer and a total of 516 features per eye.

Because of the confinement imposed by the COVID-19 pandemic, it was not possible to access the computing infrastructure of the research group. For this reason, the number of features was severely reduced to a number possible to be dealt with a personal laptop computer. As such, the local features (the ones computed block-wise and averaged into four quadrants, using the GLCM method) were averaged for the whole imaged region by averaging those of the four retinal quadrants. The analysis is now made as a whole and not by quadrants. In consequence, an unprecedented level of compression was achieved from the individual A-scan values to an image, and from the 36 image blocks to a single value for the entire imaged region. Expected results were, therefore, harshly compromised and will expectedly lead to implication in the system's

performance. This resulted on a meagre pool of 26 features per layer (20 from the GLCM and 6 from DTCWT), reflecting in 156 features per eye to choose.

3.5 Feature Pre-Selection: Correlation

While traditionally these type of applications make use of a significant number of features compared to ours, this figure (26 per retinal layer), is still appreciable because of the number of cases at hand to train and test the intended network, in addition to the testing capacity limitation imposed by the context in which this work was developed.

Because of the limitations disclosed, feature pre-selection is indispensable. On the one hand, it will allow decreasing the complexity of the neural network, and, therefore, the number of parameters/weights used, and, on the other hand, it will allow keeping the features that better correlate with the targeted output.

Since the most advantageous scenario would be to find a feature highly linearly correlated with the targeted output, the subject's age, the natural choice for the correlation is the Pearson's correlation Equation (3.5.1). It would not matter if the found correlation is positive or negative, as the network would easily adjust its weights with the appropriate sign, but only the strength of the correlation, i.e. its absolute value.

$$\text{Correlation Coefficient} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{3.5.1}$$

In Equation (3.5.1) the x and the y represent the values of feature x and y respectively, and the $\overline{x}$ and $\overline{y}$ represents the means of x and y, respectively.

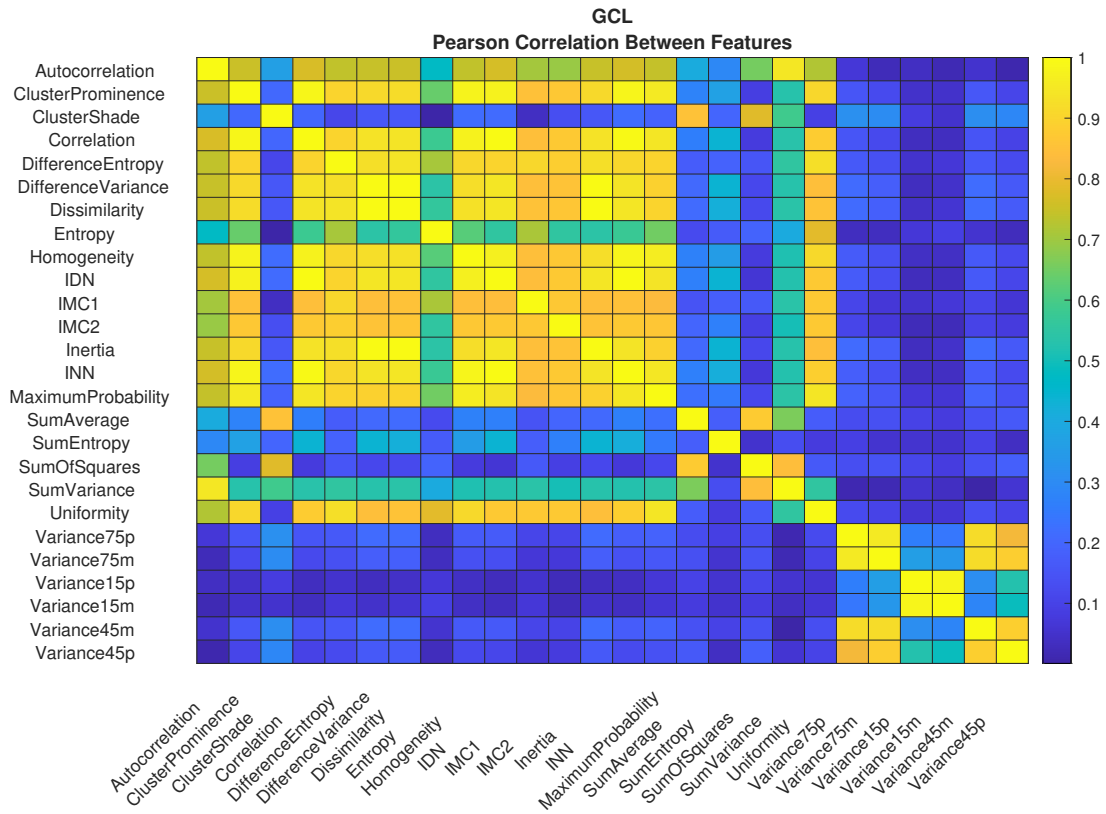
In our case, since we have a large number of features, a heatmap was used to help the visualization of the results (see Figures 3.5.1a, 3.5.1b and I.1.0a to I.1.0j).

Heatmaps were computed using the Pearson Correlation between features (Figures 3.5.1a, I.1.0a, I.1.0c, I.1.0e, I.1.0g and I.1.0i) and between target and features (Figures 3.5.1b, I.1.0b, I.1.0d, I.1.0f, I.1.0h and I.1.0j) in order to better understand the relationships between them (only the GCL heatmaps are shown here, heatmaps for the remaining retinal layers can be found in Appendix I.1).

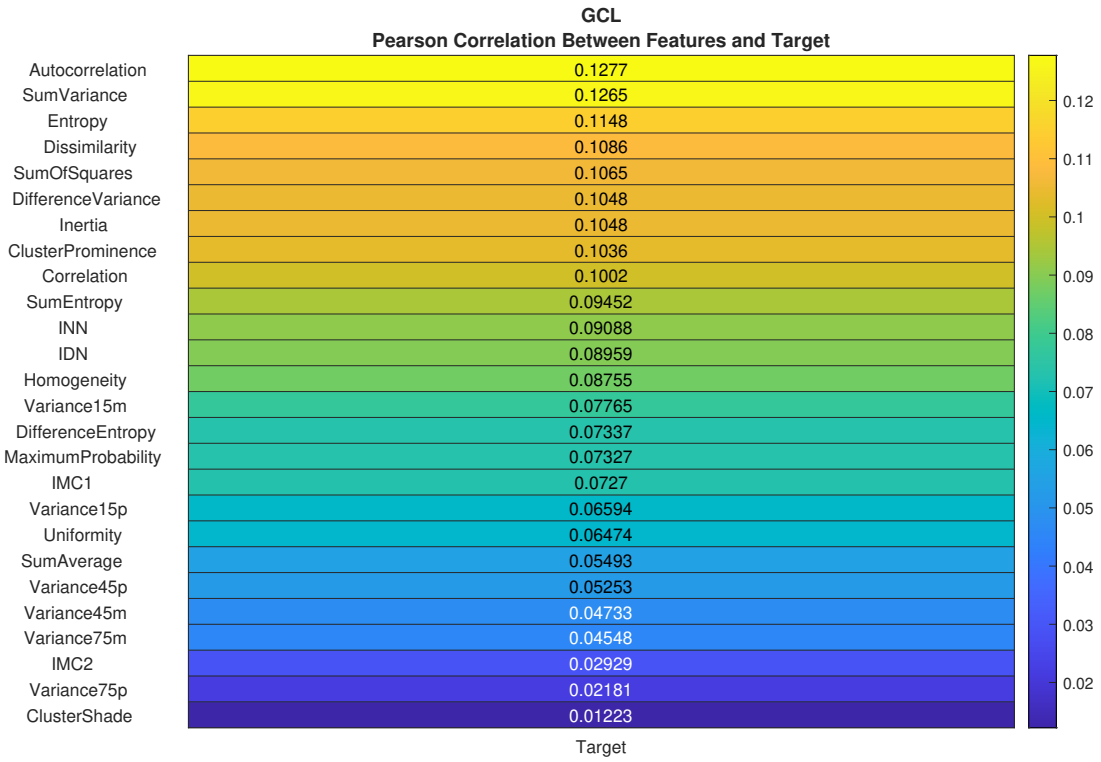
The chosen path was first to check the correlation between each feature and the subject's age (the targeted output) for each retinal layer to shed light on the layer possibly contributing the

most for the solution. Because features are highly correlated with one another, only the six ones most highly correlated to the subject's age were kept for each layer to built a set of features representative of the entire scan.

This approach was chosen over a computationally exhaustive method due to limitations already mentioned. It is an obvious approach because we are trying to find the age-discriminant features, i.e. the ones with a higher correlation to the subjects' age. Starting from the 6 features already chosen, the ones that are highly correlated with them should be excluded since they do not add any useful information besides the one provided by the chosen features. That being said, each one of the layers will have different selected features. These results are merely a pre-selection of the features and will be compared with the real feature selection in Table 3.10.1 (chosen through the neural networks).



(a) GCL - Correlation between features.



(b) GCL - Correlation features-target.

Figure 3.5.1: Pearson Correlation between features and between features and target, for GCL retinal layer.

3.6 Neural Networks

Neural Networks are systems inspired by animals' brains, however, they are not identical since it is not possible to recreate all the connections present in an animal's brain [61]. Nonetheless, the principle is that these systems, just like an animal's brain, learn to execute tasks only by experience, e.g. observation of examples, identifying characteristics and patterns present on those (either images or any arrangement of data). The advantage of learning from examples is that neural networks do not require prior information about the subject at hand to deliver a prediction on newly presented input data.

In resemblance to the animal brain, neurons are the NN basic unit. NNs are organized in a sequence of connected layers in which two are mandatorily present: the input and output layers. The former is the one who receives the stimuli, while the latter will provide the output based on the inputs and the previously learnt model. Between these two layers is where the hidden layers reside. Hidden layers are responsible for processing the information received at the input layer [21].

These hidden layers are connected to the input and output layers by weighted (either positive or negative) connections. Each layer receives information from the previous layer, multiplied by the respective weight (determined during the learning/training process) [21].

Each unit (neuron) present in the layer can receive multiple inputs from the previous layers, and those inputs are all summed up and passed to an activation function that will feed the next layer.

A feedback mechanism is responsible for the learning process by adjusting current weights based on the difference between the network's output and the expected result. This renders neural networks an incredible adaptive capacity [21].

There are several types of NNs, and we can use the one which best suits to our dataset and objective needs. The most commonly used are: the Perceptron, the simplest of the networks, composed of a single neuron, but able to perform classification tasks for decision in between two classes [21]; the Feed-forward Neural Network, where the data moves in only one direction, and does not allow loops [21]; the Recurrent Neural Network Long Short-Term Memory, where the output from a particular layer is fed back into the input, acting like a memory node, which is typically used for vocabulary speech recognition [62]; the Radial Basis Function Neural Network, which makes use of the distance of a point to the center (or the origin) and it is used for signal processing, system modeling, control and fault diagnosis [63]; Convolutional Neural Networks which are commonly used to compute features from input images, and whose applications include computer vision and object recognition because of their ability to make correct assumptions about the nature of images [64].

In line with the aim of every other model, the NN is also trying to make the right prediction, and for that, the predictions must be as close as possible to the target values. Hidden layers are responsible for that prediction or classification. Every neuron has its bias term, and the connections in between neurons have a weight that will transform the input given to the neuron. Furthermore, each layer has an activation function that will influence the prediction/classification, and the particular choice for the activation function should suit the particular objectives of the problem being tackled [21]. The connection and the neuron, together form the fundamental components of the neural network.

Increasing the complexity of the network (including more layers, more neurons and more connections) will allow the NN to learn further complex relations concealed in data. For any neural network, the intermediate output of each hidden layers is given by Equation (3.6.1), where [W]

is the matrix of weights, [X] is the matrix/vector of the original inputs, or the output from the activation functions of the previous hidden-layer, and [Bias] is the matrix/vector of neuron biases (see Figure 3.6.1).

$$[W] \times [X] + [Bias] = [Z] \tag{3.6.1}$$

Afterwards, the activation function is applied to each element of [Z], to produce the neuron's output for the current layer, and the process repeats from the first to the last layers. This is known as forward propagation [21].

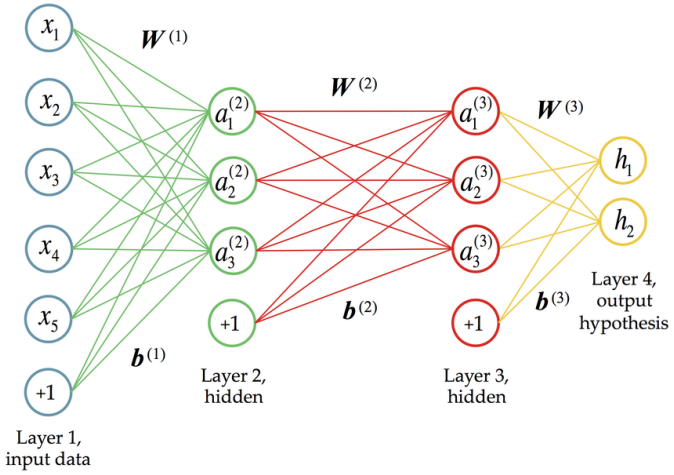


Figure 3.6.1: Neural network representation (source: https://www.researchgate.net/figure/An-example-of-a-deep-neural-network-with-two-hidden-layers-The-first-layer-is-the-input_fig6_299474560).

In the training step, the objective is to minimize the error between the predicted value and the actual target values, and this error can be measured resorting to distinct metrics. This cost function can make use of distinct metrics such as the RMSE (Equation (2.0.2)) and the MAE (Equation (2.0.1)), providing a measure for the performance of the NN for the task at hand. The aim is to minimise the cost function by adjusting the weights and bias for all the neurons of the NN.

One way of spreading the error towards adjusting weights and bias is through gradient descendant [21]. The process is repeated until a minimum of the cost function is achieved, or a stop condition is met. Backpropagation is used to move the error computed at the output backwards via the same weights and connections that were used to propagate forward the signal.

The error in each neuron is important; its weight and bias can change the whole prediction, and the backpropagation allows calculating the correction for each neuron.

3.7 Metrics and Heuristics

Different strategies were used to analyse the suitability of the neural network developed towards the prediction of healthy subjects' age from texture metrics of computed OCT MVF images.

Because initial results were relatively poor, the network was converted from the regression to the classification type. As such, subjects were split into three age groups, with balanced numbers in each, and the possibility to establish a NN to classify subjects into the appropriate group was assessed. In consequence, for this test, accuracy and F1-scores were used as per the results ahead from the establish confusion matrix.

To define those terms its required to understand the values represented in a confusion matrix. Precision, Recall, Accuracy, and F1-Score are defined in Equations (3.7.1) to (3.7.4). A Confusion Matrix is a performance measurement for classification, and tabulates the number of cases a particular class distributes by all classes, allowing the computations of all the above metrics [65].

In a multi-class problem, as opposed to the traditional binary classification, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) cases need to be defined for each of the classes. An example is provided by the confusion matrix shown in Figure 3.7.1, for classes A, B, and C. When referring to class A, there are seven TP, one FP, 22 TN, and three FN.

		True Class		
		A	**B**	**C**
Predicted Class	**A**	7	1	0
	B	1	11	0
	C	2	3	8

Figure 3.7.1: Example for the Confusion Matrix.

Because no classifier is perfect, errors are expected. Hence, metrics are mandatory to assess the performance of the proposed method/solution. Which particular metric should be used depends on the specific application and impact of FP and FN cases [65]. For an unbalanced number of cases per class, metrics such as the F1-score are typically used [66]. The F1-Score Equation (3.7.4) allows a balance between the Precision and the Recall metrics, presenting an advantage compared to the accuracy metric [66]. F1-scores range from 0 to +1, with the latter meaning a perfect classification mode.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{True\ Positive}{Total\ Predicted\ Positive} \tag{3.7.1}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{True\ Positive}{Total\ Actual\ Positive} \tag{3.7.2}$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \tag{3.7.3}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.7.4}$$

3.8 Classification for Data Analysis

As aforementioned, initial results were relatively poor, casting the doubt about the possibility of texture from computed OCT MVF images of the human retina to convey information on the CNS's age, all despite the many efforts.

A cascade forward network was used because it is similar to the feed-forward network with the advantage of encompassing connections between the input and every other layer (including the output layer). Furthermore, each layer connects to every one of the following ones. This type of NN has the advantage of learning any relationship between the input and the output, maintaining the non-linear relationship and also not eliminating the previously existing linear relationship [67].

That being said, the NN had three hidden layers with 15, 50 and 150 neurons, and the *tansig* as the non-linear transfer function. For the output layer, we opted for the use of a linear transfer function (*purelin*). The training function was limited at that time to the *trainscg* - Scaled Conjugate Gradient backpropagation instead of *trainlm* - Levenberg-Marquardt because the GPU available in the laptop does not support the Jacobian training (limitations imposed by the impossibility to access computing resources at UC premises).

Because results were not satisfactory, the target was modified to an age group where the subject's eye should be classified though a "*patternnet*" network. That means that our NN went from a linear regression one into a classifier. This approach would give us some insights about the

information present in the data and if it was useful or not (for our research question).

Also, regression NNs were tested within each group, as these were more homogeneous groups (age-wise). The deviations, in each group, between the actual age and the PA could, potentially, provide information about the non-generalisation power because of the range of ages presented in the dataset, covering all the adulthood, from 19 years-old to 74 years-old.

The division of the dataset in the three age groups was made keeping the original balance in data, notably the sex-balanced and age-matching. The first group corresponds to the younger population of 68 individuals, with ages ranging from 19 and 31 years-old. The second group corresponds to the middle age with 71 individuals in between 32 and 48 years-old. Finally, the third group corresponds to the older population, with 59 individuals in between 49 and 74 years-old (Table 3.8.1). This split also sought for balanced groups concerning the number of cases. Ideally, the system would be able to identify each age as an age group but due to the scarce number of cases available, it will not be possible.

Table 3.8.1: Division of the dataset in age groups with the minimum and maximum age and the number of respective cases.

Division of the dataset in Age Groups		
Age Group	**Ages(years) Min(Max)**	**Number of Cases**
Young (1)	19(31)	68
Middle-aged (2)	32(48)	71
Old (3)	49(74)	59

All data was z-score normalized since data values had significant discrepancies between features. This way it is guaranteed that they have the same scale, allowing the comparison of different feature values with different ranges, without distorting differences between them.

3.9 Neural Network Architecture

Each and every case are unique in its own way, and what was the perfect architecture for other authors mentioned in Chapter 2 may not be the optimal solution for our dataset. As such, a *script* was made to search for the best NN's architecture based on the F1-score applied to the test dataset.

Only networks with one layer and 1 to 500 neurons, two layers and 1 to 100 neurons in each layer, and three layers and 1 to 50 neurons in each layers were tested because of the required

computing time, and the constraints mentioned above. The quest for the best architecture was independently run for each layer twice: one using the default parameters of maximum of 1000 epochs, minimum gradient of 1^{-7}, performance function 'mse' and another with handcrafted specifications. All best NNs' architectures were attained with two hidden layers and maximum of 1000 epochs, minimum gradient of 10^{-9}, learning rate of 0.01, performance function 'msereg' (mean squared error with regularisation performance function), performance goal of 10^{-8} and maximum fail of 100. The best NNs' architectures are summarized in Table 3.9.1 for all retinal layers.

Initially the data was divided in 68.7% (136 cases) for training and 31.3% (62 cases) for a hold-out set, that will be used as a test set for a first analysis of the NN. This hold-out set does not have in account if both eyes of an individual are present in the training set or in the test set. The automatic split of the training data into the training, validation and test sets was possible and readily available in MATLAB. Data were split randomly with the following division: 70% for training, 15% for validation and 15% for testing. For each one of the 100 runs, the splits were naturally different from the others, because of its randomness.

Table 3.9.1: Neural Networks chosen architectures. The numbers inside the brackets are the number of neurons for each hidden layer.

Neural Network Best Architectures						
	Layers					
	GCL	**RNFL**	**INL**	**IPL**	**ONL**	**OPL**
Number in Hidden Layer	[47 21]	[46 29]	[23 9]	[15 33]	[40 13]	[22 44]
F1-Score hold-out (%)	0.5253	0.5831	0.5502	0.5566	0.5211	0.5421

3.10 Feature Selection per Retinal Layer

With the NN's architecture established, it was decided to proceed towards the identification of the features contributing the most for the intended regression network.

Therefore, based on the classification network established, and resorting to the F1-score applied to the hold-out data, a backward elimination approach was applied to successively eliminate features with a negative contribution to the correct classification of cases in the respective age group. Because the split of data for train and test is randomly established, and this split impacts the measured performance, the average over 20 runs was established as a suitable metric. Although the number of runs was handpicked, it took into account the time constraints and the

runtime per series on the available hardware. The particular features selected per retinal layer are listed in 3.10.1.

Table 3.10.1: Final features for each retinal layer.

Feature Selection for each Retinal Layer						
	Retinal Layers					
Features	RNFL	GCL	IPL	INL	OPL	ONL
Autocorrelation	✓	✓	✓		✓	✓
Cluster Prominence		✓		✓		✓
Cluster Shade	✓				✓	
Correlation		✓	✓	✓		✓
Difference Entropy	✓	✓			✓	
Difference Variance						✓
Dissimilarity				✓	✓	
Entropy	✓		✓			✓
Homogeneity				✓		
IDN	✓			✓	✓	✓
IMC1	✓	✓	✓			✓
IMC2	✓				✓	✓
Inertia		✓	✓			
INN	✓			✓		✓
Maximum Probability		✓	✓		✓	✓
Sum Average	✓				✓	
Sum Entropy	✓		✓	✓		✓
Sum Of Squares		✓		✓		
Sum Variance	✓	✓	✓		✓	
Uniformity						✓
Variance+75º		✓			✓	✓
Variance-75º	✓	✓				✓
Variance+15º			✓		✓	
Variance-15º	✓			✓		✓
Variance-45º		✓			✓	
Variance+45º	✓	✓		✓		
Total Number of Features	14	13	9	10	12	15

Taking a look at the first correlations between target and features (see heatmaps in Section 3.5) it is observable that for different retinal layers, different features were shown to be more correlated to the target. For GCL, for example, the most correlated features were Autocorrelation, Sum Variance, Entropy, Dissimilarity, Sum of Squares and Difference Variance. However, in Table 3.10.1, those features are not present in the feature selection of that layer in particular. That happened with all retinal layers, which might be due to the low number of times the net was ran (20 times) or also because F1-Score might not be adequate for this kind of procedure. The incompatibility between features (also seen in the correlation between features heatmaps) might also be the reason why the final feature selection was different than the obtained feature

pre-selection with Pearson correlation. This might result in some differences in the final results.

3.11 Regression Neural Network

Having defined the best architecture and set of features for the classification problem, these were adopted as the starting point for the regression problem. The number of hidden layers and number of neurons per hidden layer was maintained. Since the type of network in used was changed (from *patternnet* to *cascadeforwardnet*), a cycle was created in order to decide the best training function. Available training options, the *trainlm* (Levenberg-Marquardt), the *trainscg* (Scaled Conjugate Gradient), and the *trainbr* (Bayesian Regularization), were tested to find the *trainlm* (Levenberg-Marquardt) the best one. This testing was now possible because the access to UC premises was granted by the end of June, allowing access to better-suited hardware for the task (where the GPU supported Jacobian Training).

Since the hold-out set used to test the NN was composed by a selection of individuals eye's and not having in account if both eyes of an individual were in the training set or in the test set, a hold-out group only with pairs of eyes belonging to the same individuals was created. In other words, a hold-out group where none of the eyes were present on the training set in order to have a better validation of the model. This hold-out group had 60 cases, equivalent to 30 individuals (two eyes per individual) and it was sex-matched (15 females and 15 males).

Results from a regression neural network are dependent on the complexity of the network, on the optimisation process, and the limited dataset. All these factors would require a proper selection of features and network architecture. Different ways to maximize the performance were explored by isolating and combining the layers who seemed to have a more significant contribution, meaning the two retinal layers with the best RMSE values. The idea was to concatenate them in a single vector, and go over the same process because now, even though each retinal layer was already simplified in terms of feature selection, the junction of those two may end up with features that can be highly correlated with each other, jeopardizing the results. Explore the best NN architecture for that new vector, and perform feature selection one more time, and run the regression NN to obtain the RMSE. This was made for the test set. However, there was the necessity to validate the results with an external dataset, never seen by the NN: the hold-out set. Other validations were performed such as the distribution of the PA along the age range, the analysis by biological sex and per eyes, and the analysis of the mean of the PA *versus* the CA, for both genders.

Results

The output from the implemented NN is the estimation of the CNS age based on the input and target given. In this case, a graph comparing the CA *versus* the PA is more informative when compared to one displaying the regression coefficient since slight deviations between the CA and PA would jeopardize the coefficient itself.

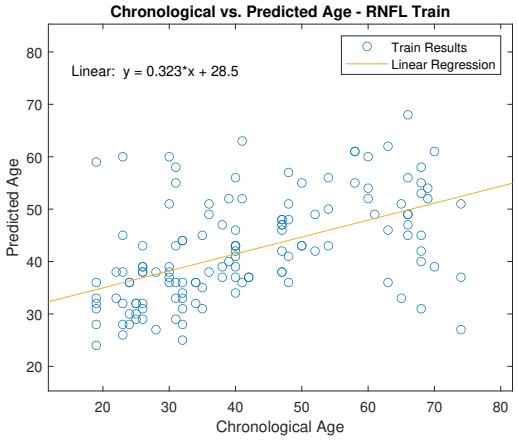
4.1 CNS Age Estimation

Because specific training and validation sets lead to a distinct set of parameters, even when maintaining the same network's architecture, all results herein presented correspond to the average performance of 100 runs. Also, because the aim is the correct estimation of the CNS's age, the RMSE was chosen as the performance metric. Table 4.1.1 presents the performance on the CNS's age estimation based on each of the individual neuroretina layers. On the one hand, it shows the individual ability of specific layers in conveying the ageing process. On the other hand, it shows how close the estimation is to the subject's real age.

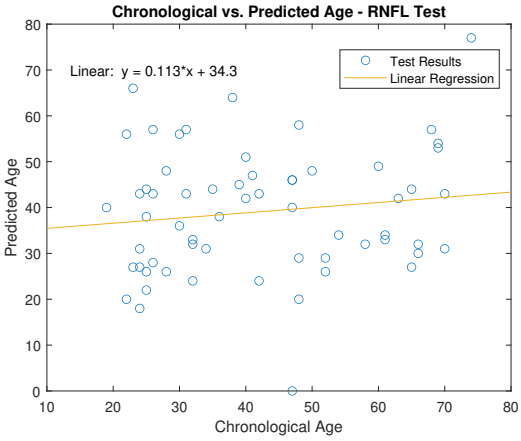
Table 4.1.1: Performance for Regression Neural Network relative to the test set.

Performance for Regression Neural Network						
Layers	**RNFL**	**GCL**	**IPL**	**INL**	**OPL**	**ONL**
RMSE (years)	22.64	26.24	**18.28**	**17.53**	23.16	19.22

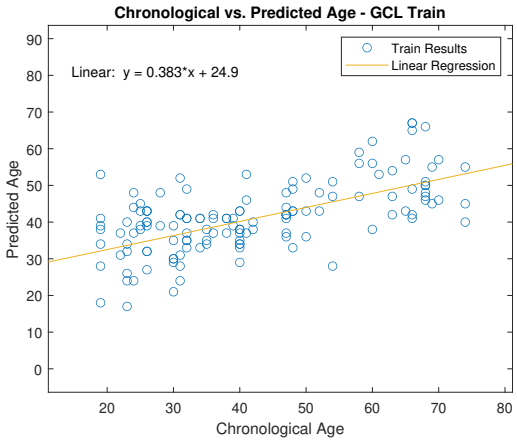
These results, far from optimal, represent a significant step from our initial ones demonstrating the advantage of the careful selection of features from the global pool, and the need for the proper selection of the network's architecture. Because of the low accuracy of the estimates from the test set (Table 4.1.1), it became clear that the established NNs are unable to learn from individual neuroretina layers. In Figure 4.1.1 the obtained graphs for the CA *versus* the PA are displayed for the train and test sets, and for each retinal layer. It is also shown the inability to learn, suggested in Table 4.1.1.



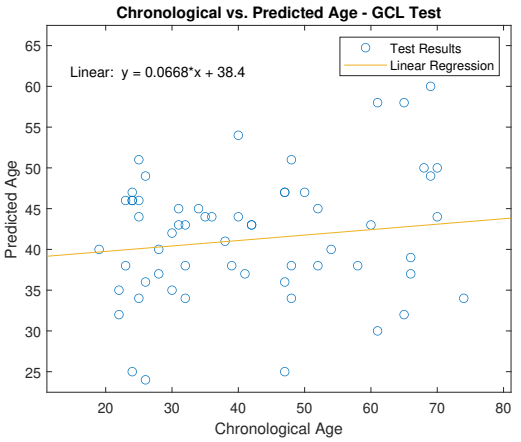
(a) RNFL - Train Results.



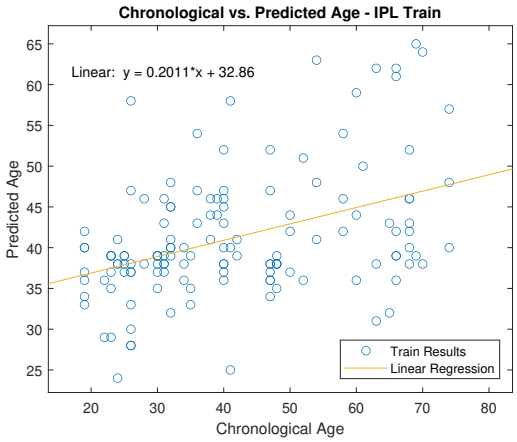
(b) RNFL - Test Results.



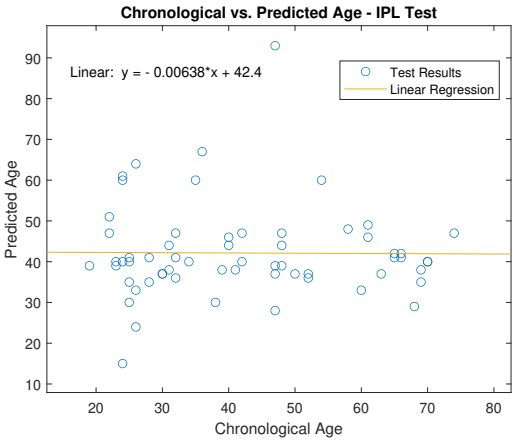
(c) GCL - Train Results.



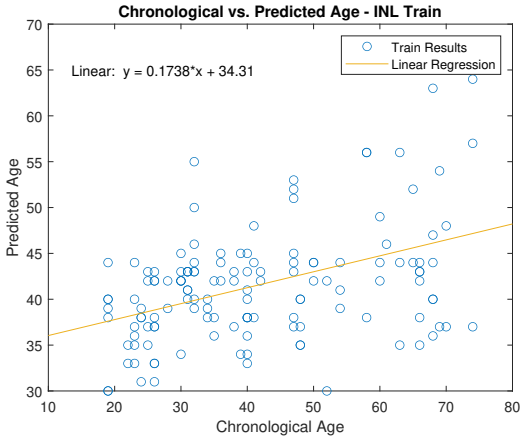
(d) GCL - Test Results.



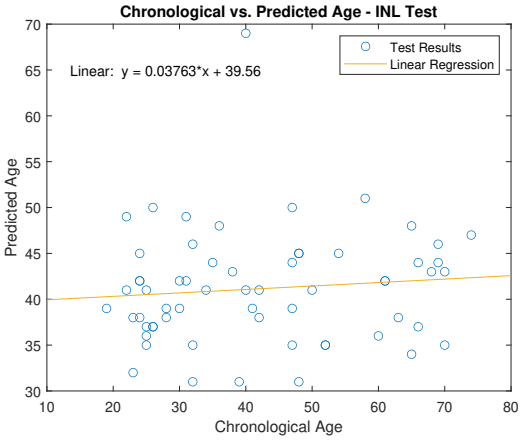
(e) IPL - Train Results.



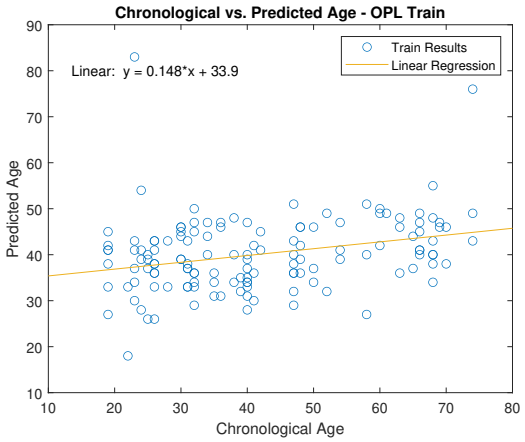
(f) IPL - Test Results.



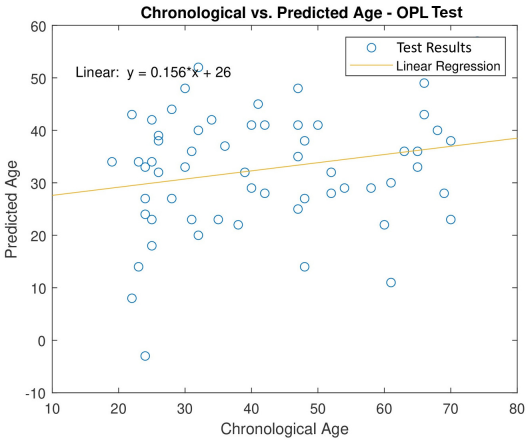
(g) INL - Train Results.



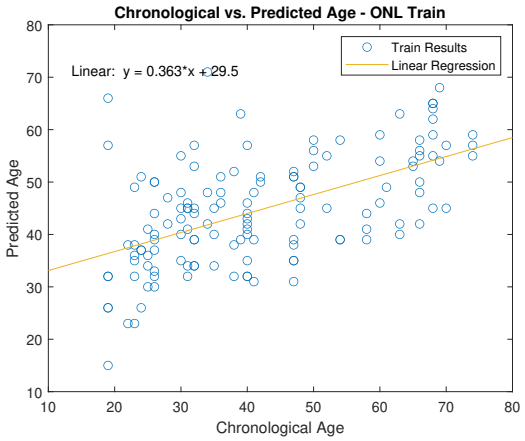
(h) INL - Test Results.



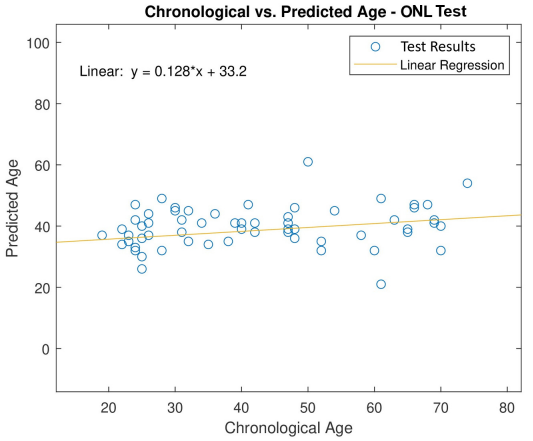
(i) OPL - Train Results.



(j) OPL - Test Results.



(k) ONL - Train Results.



(l) ONL - Test Results.

Figure 4.1.1: Chronological *versus* Predicted Age graphs for each one of the retinal layers, for train and test sets[1].

[1] Test set with 62 cases (first hold-out set, see Sections 3.9 and 3.11).

As we can see, in each figures from the training set (Figures 4.1.1a, 4.1.1c, 4.1.1e, 4.1.1g, 4.1.1i and 4.1.1k), there is a tendency (represented by the yellow regression line) meaning that the information exists. Obviously, the perfect regression line would be the one going through the graph origin (slope = 1), but with the size of our dataset and all the simplifications previously made to the data, a slope in the range of 0.148-0.383 (worst and best case, respectively) is reasonable.

It would be expected to obtain better results with a larger dataset. Obviously, for the test set (Figures 4.1.1b, 4.1.1d, 4.1.1f, 4.1.1h, 4.1.1j and 4.1.1l) it is always expectable to have a worse performance than for the training set. It is observable that the best slope is 0.156 which is not excellent, but understandable regarding the conditions.

The two retinal layers with the best RMSE values were the INL and IPL. They were concatenated in a single vector, resulting in a vector with 19 features (10 from INL and 9 from IPL). The vector went over the process of finding the best NN architecture because the junction of those two initial vector may end up with features that can be highly correlated with each other, influencing the results.

With the INL NN architecture (two hidden layers with 23 and 9 neurons, respectively) an RMSE of 18.62 years was obtained. For the IPL NN architecture (two hidden layers with 15 and 33 neurons, respectively) an RMSE = 20.99 years. The best architecture for this new vector was a NN with three hidden layers with 18, 10 and 15 neurons, respectively, obtaining a F1-Score of 0.5639 for the test set, better than the one for the training set (0.5431). An RMSE of 18.90 years was obtained.

Feature selection was made in order to reduce the computational complexity, following the same procedure as the one mentioned in Section 3.10. Recalling we had a vector with 19 features (see Table 3.10.1), and ended up with only 8 features: *Correlation, Homogeneity, IDN, Sum Entropy, Sum of Squares, Variance -15°* from INL; and from IPL only *IMC1* and *Sum Variance*.

Finally, the regression was made using the 3 different architectures: for the last one computed an RMSE of 19.59 years was obtained, for the INL architecture RMSE = 19.44 years, and for IPL architecture RMSE = 19.38 years. It is observable that with or without feature reduction, the RMSE value converges to 19.5 years.

ONL was added to the INL and IPL layers since it was the third best RMSE value. However, the results were not satisfactory (RMSE = 22.64 years).

As mentioned before, for the hold-out set, the cases were chosen in order to be as representative as possible of the age range present in the dataset since we chose data points throughout the age range in order to cover all ages. The train/hold-out ratio was 69.7/30.3%. It is important to remark that making the data of the hold-out set completely independent from the data of the training set, makes the process even more complex and challenging for the neural network to classify. With the test set, there was the possibility of having the two eyes of the same individual distributed in the training and test sets, creating a dependence, which eased the process.

Two approaches were made: the first one was to ran the regression NN for every retinal layer with the features reduced (available on Table 3.10.1) and obtained the RMSE in Table 4.1.2.

Table 4.1.2: Performance for Regression Neural Network for the hold-out set and first feature selection.

Performance for Regression Neural Network						
Layers	**RNFL**	**GCL**	**IPL**	**INL**	**OPL**	**ONL**
RMSE (years)	26.02	20.66	**22.38**	**22.19**	27.49	23.58

The second approach was made maintaining all original features and repeating the processes mentioned before, such as finding the best architecture, feature reduction and regression NN (see Table 4.1.3).

Table 4.1.3: Performance for Regression Neural Network for the hold-out set and all original features.

Performance for Regression Neural Network						
Layers	**RNFL**	**GCL**	**IPL**	**INL**	**OPL**	**ONL**
RMSE (years)	25.25	25.85	**24.83**	**22.22**	27.30	28.06

And again, the best two retinal layers (INL and IPL) were concatenated in one single vector and the best NN architecture was determined (three hidden layers with 20, 17 and 12 neurons in each layer, respectively) with an F1-Score = 62.60% for the hold-out set. Feature reduction was performed in the same terms as mentioned in Section 3.10.

Due to the COVID-19 pandemic, an unpredictable change in the hardware forced the change of the training function from *trainlm* to *trainscg* (Levenberg-Marquardt to Scaled Conjugate Gradient) since the CPU does not support Jacobian training. It resulted in expectable differences in the results, which is normal since we are dealing with parameterization problems. In this

regard, the new two best layers were GCL (train RMSE = 13.03 years and hold-out RMSE = 28.78 years) and ONL (train RMSE = 13.52 years and hold-out RMSE = 30.57 years). The results are also consistent with the results presented in Figure 4.1.1. The NN architecture used was a three hidden layer network with 1, 1 and 8 neurons in each layers, respectively. After feature selection, only 13 features (6 from GCL and 7 from ONL) from the original 29 (14 from GCL and 15 from ONL) were used for further analysis.

Figures 4.1.2 to 4.1.7 present the differences between estimated ages to the actual ones for the hold-out set for individual layers. There are a few outliers, particularly in the RNFL and IPL. The younger the individual, the larger the distribution of differences between the PA and CA. However, for younger individuals, the median is closer to the CA. The GCL and the ONL were the layers with a smaller distribution, which might explain why the results were better for those layers.

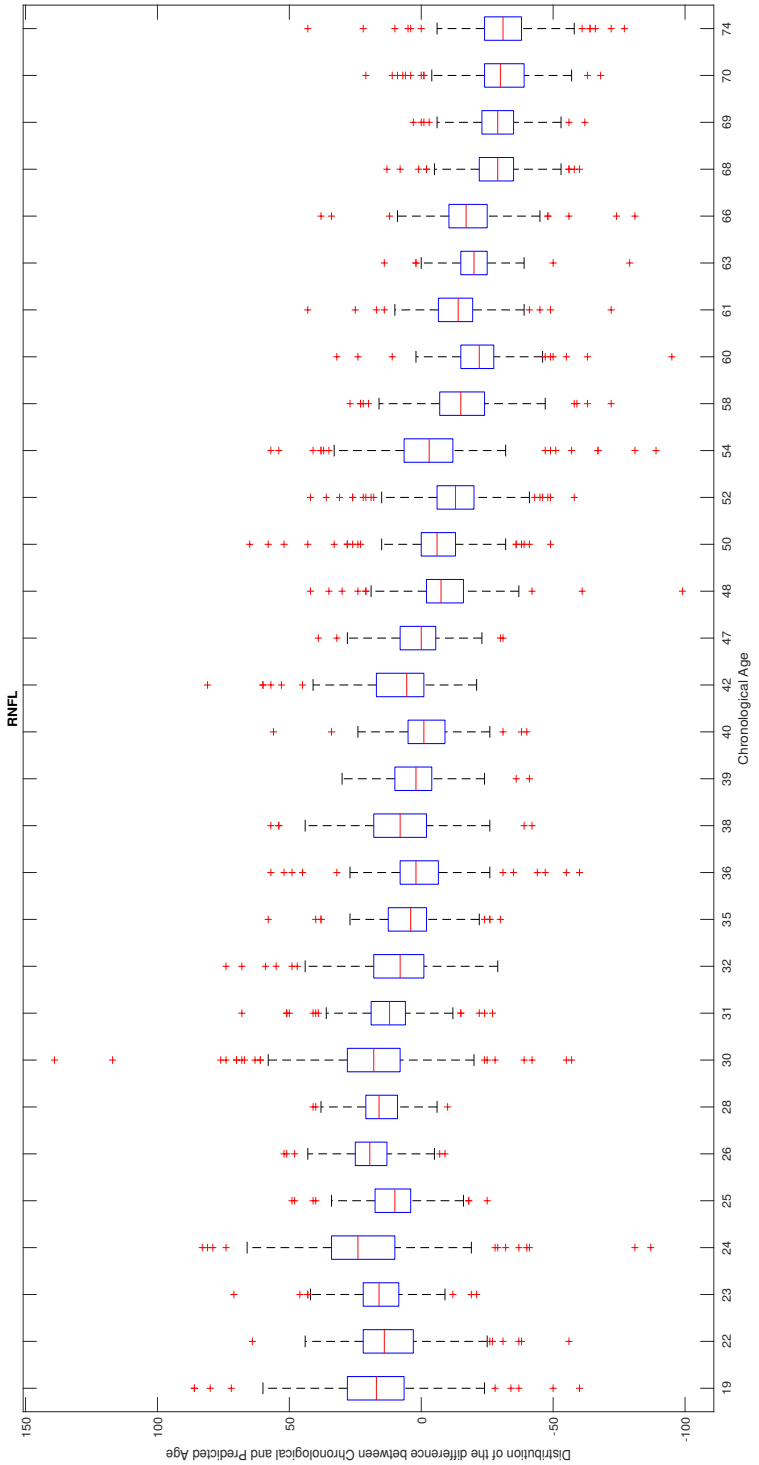


Figure 4.1.2: Distribution of the difference between the chronological and predicted age (RNFL: Hold-out set).

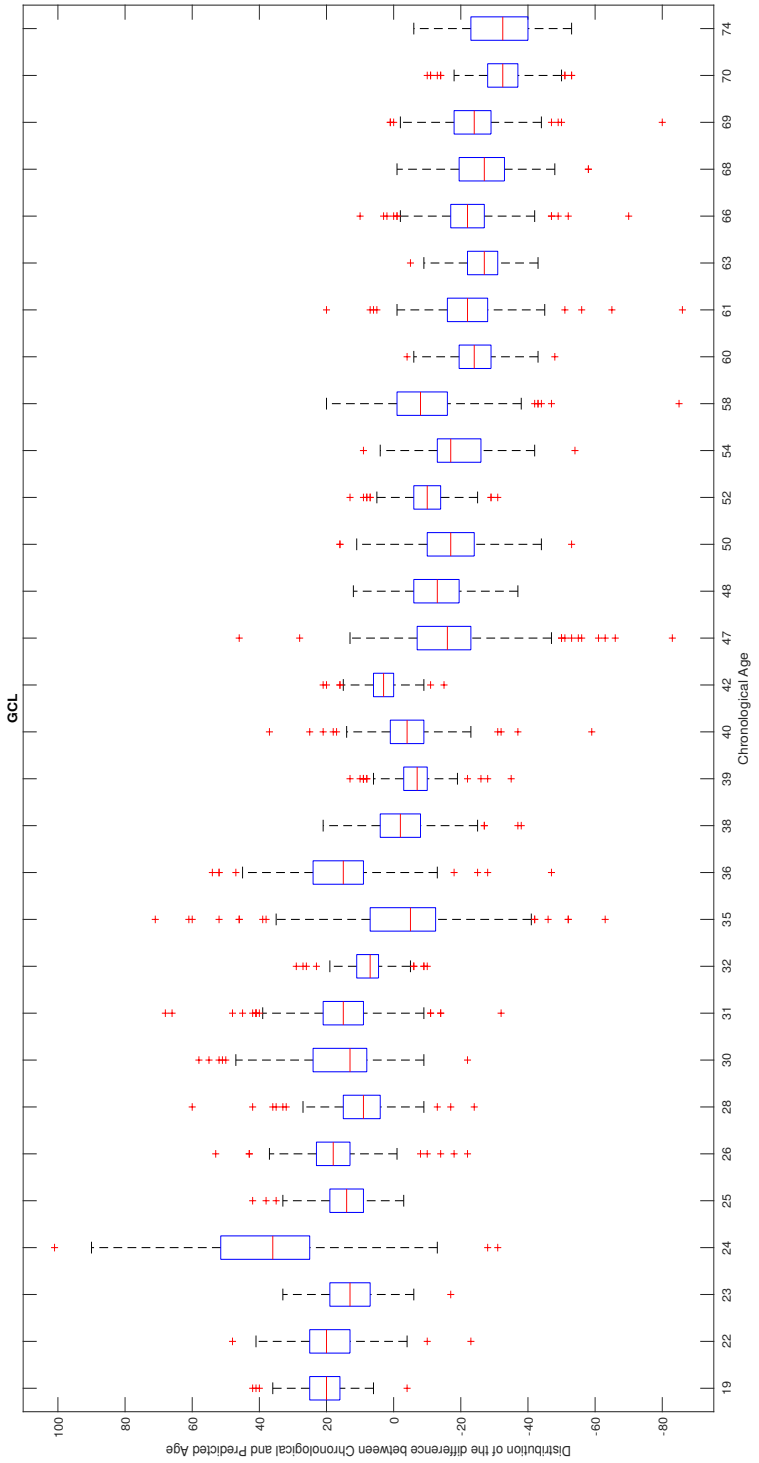


Figure 4.1.3: Distribution of the difference between the chronological and predicted age (GCL: Hold-out set).

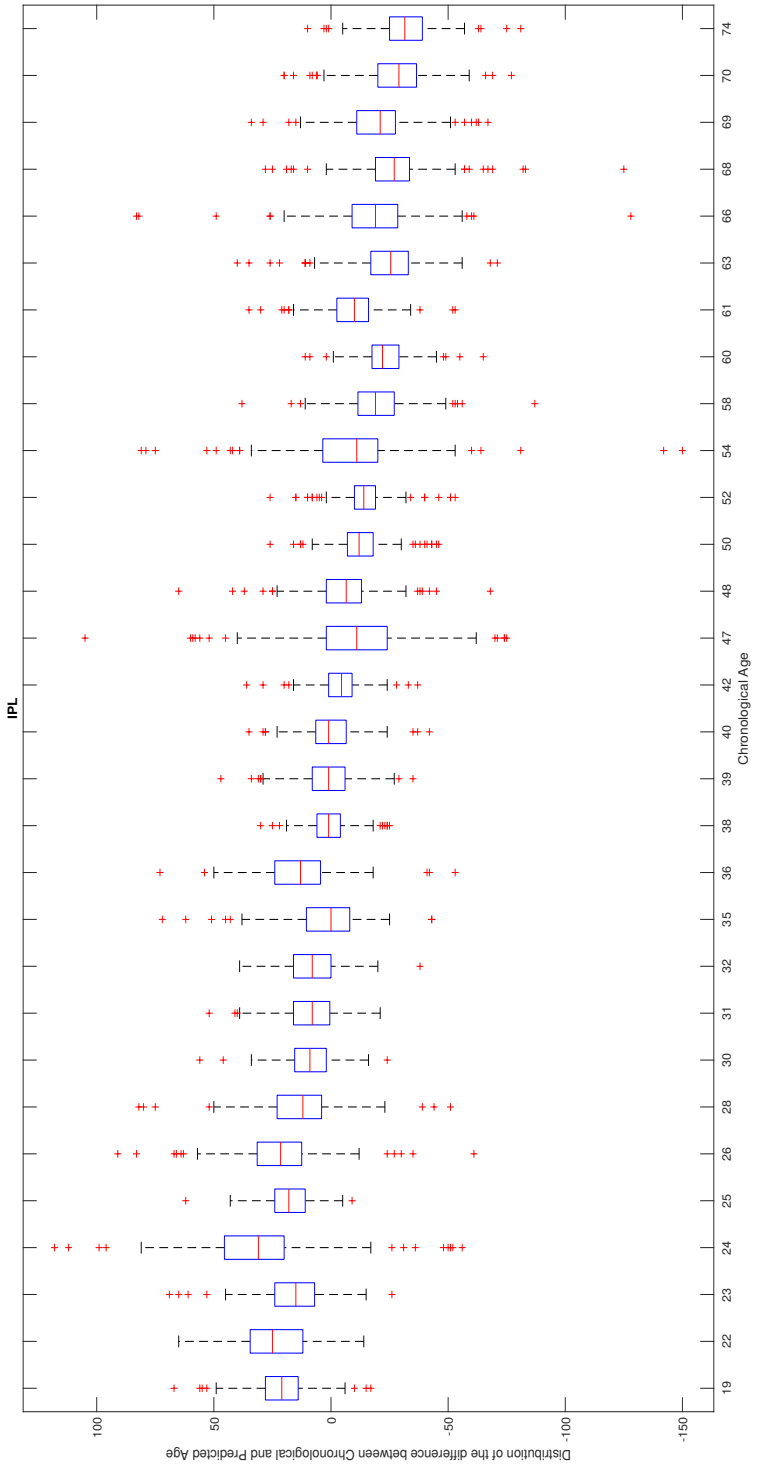


Figure 4.1.4: Distribution of the difference between the chronological and predicted age (IPL: Hold-out set).

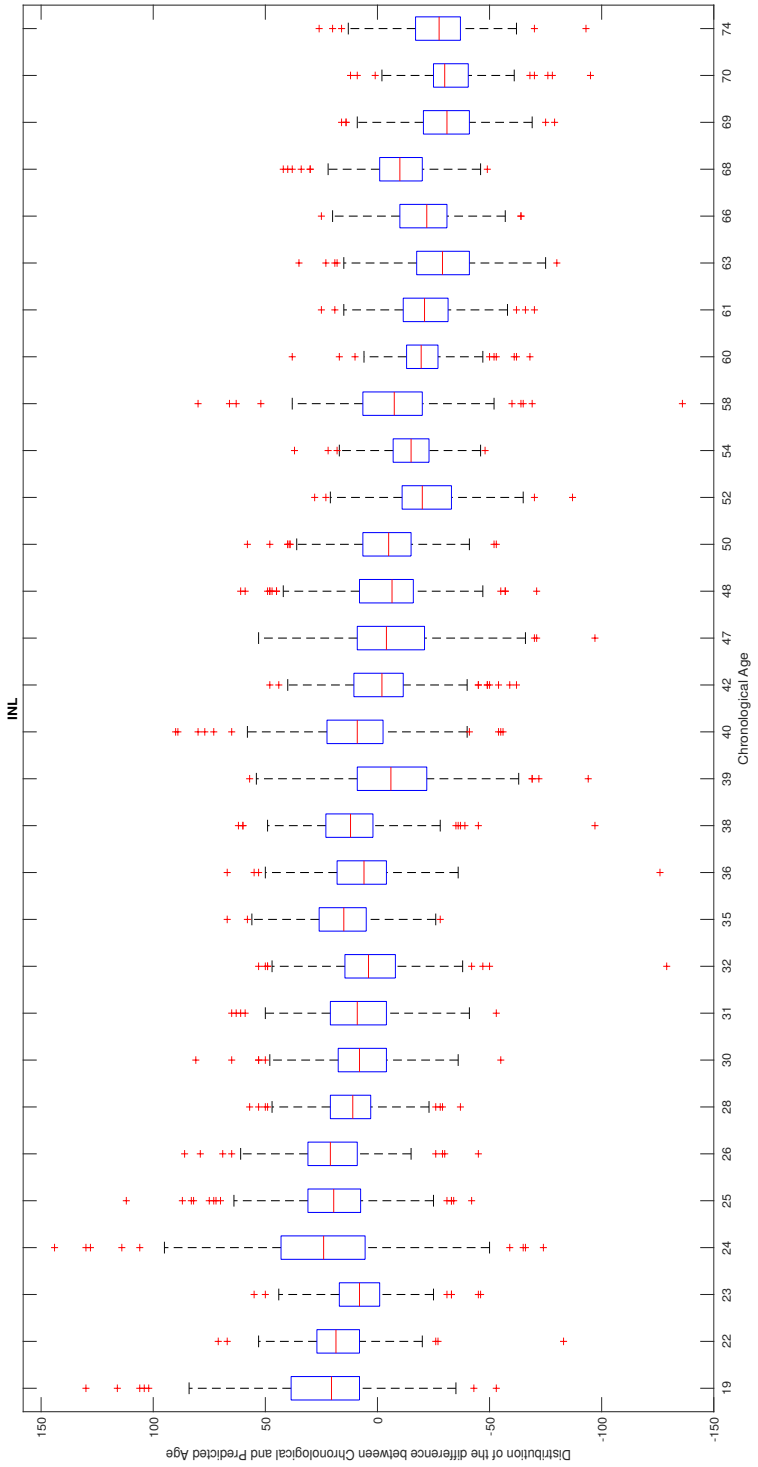


Figure 4.1.5: Distribution of the difference between the chronological and predicted age (INL: Hold-out set).

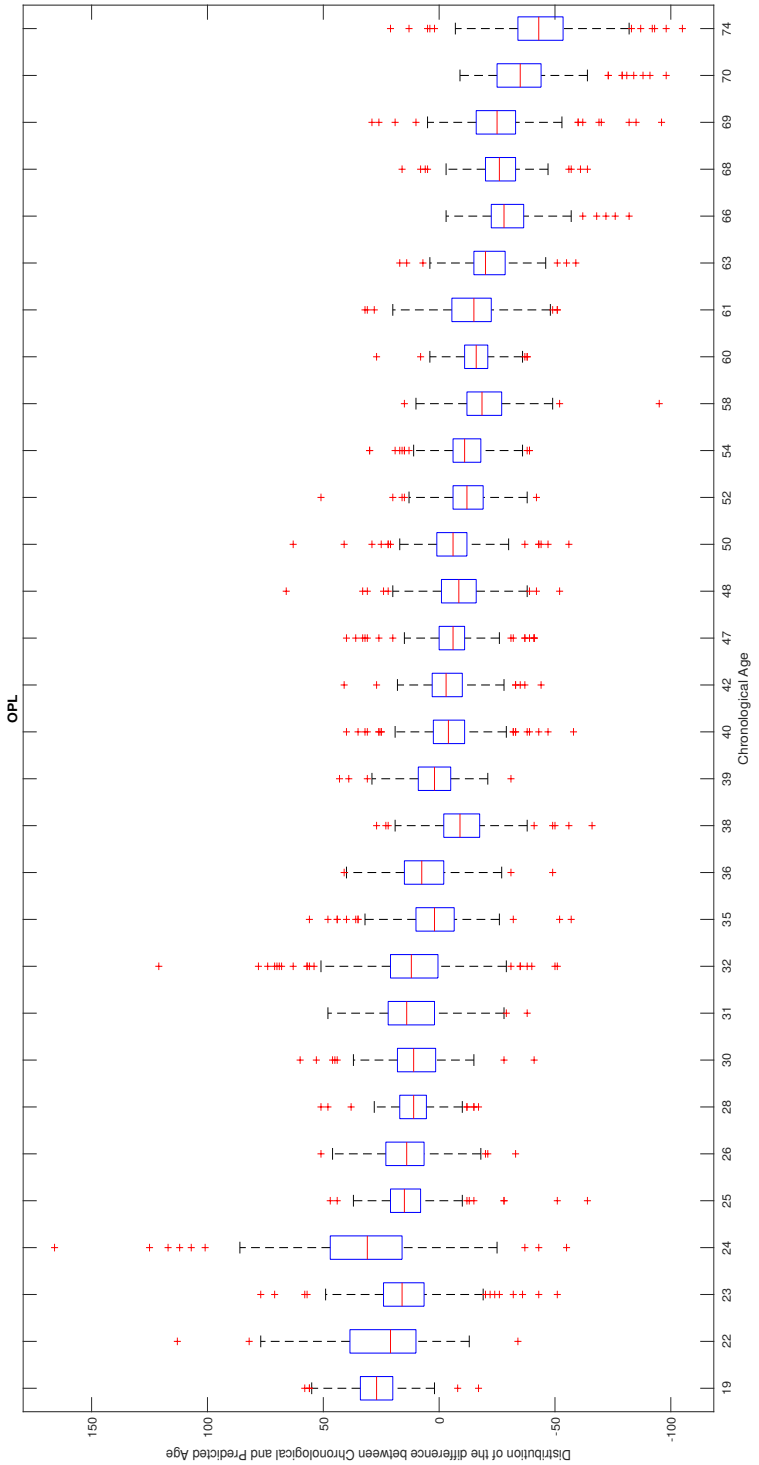


Figure 4.1.6: Distribution of the difference between the chronological and predicted age (OPL: Hold-out set).

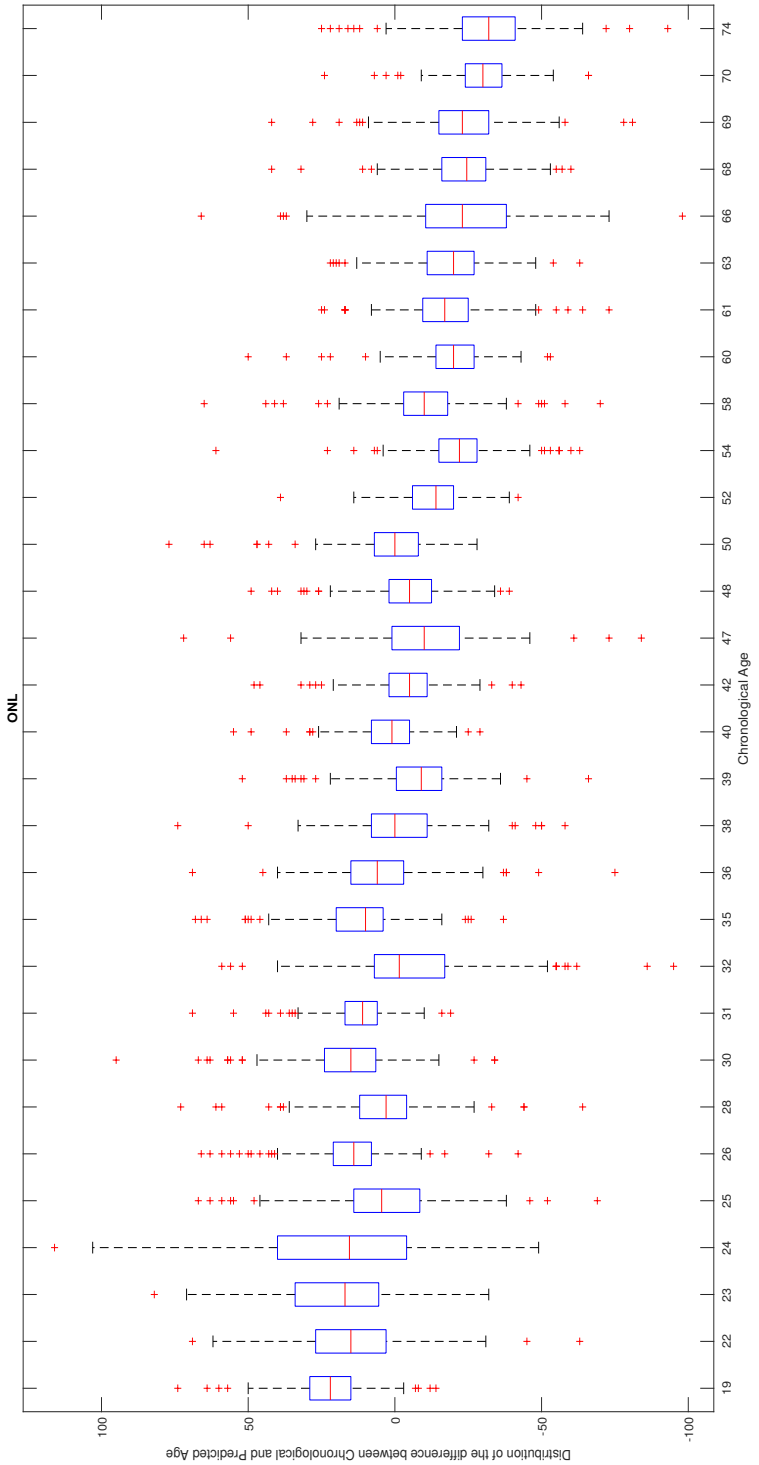
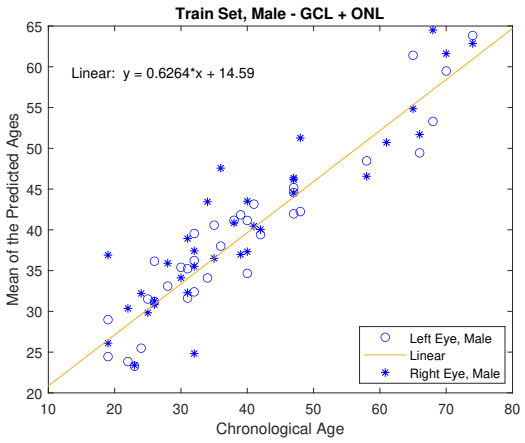
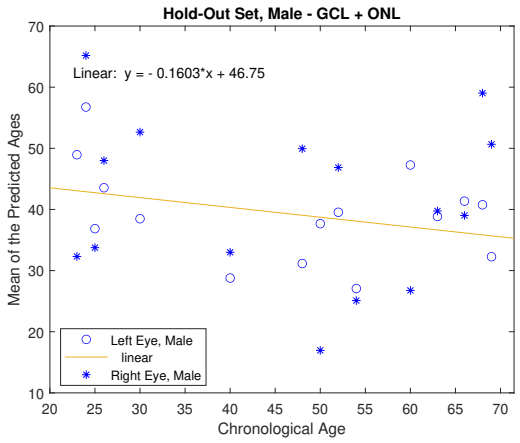


Figure 4.1.7: Distribution of the difference between the chronological and predicted age (ONL: Hold-out set).

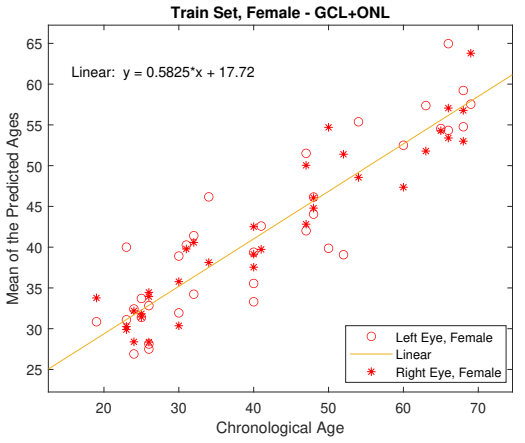
Because all the above data were presented based on individual eyes and irrespectively of the subject's sex, separated analyses were carried to assess the differences in estimated ages by identifying left- and right-eyes and subjects' sex to find that no differences existed on the learning ability and prediction accuracy Figure 4.1.8.



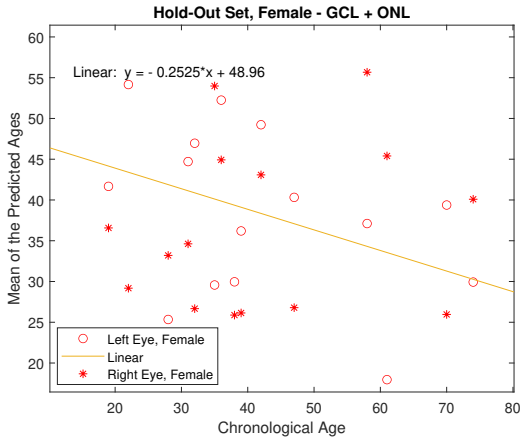
(a) Train Set - Male.



(b) Hold-Out Set - Male.



(c) Train Set - Female.

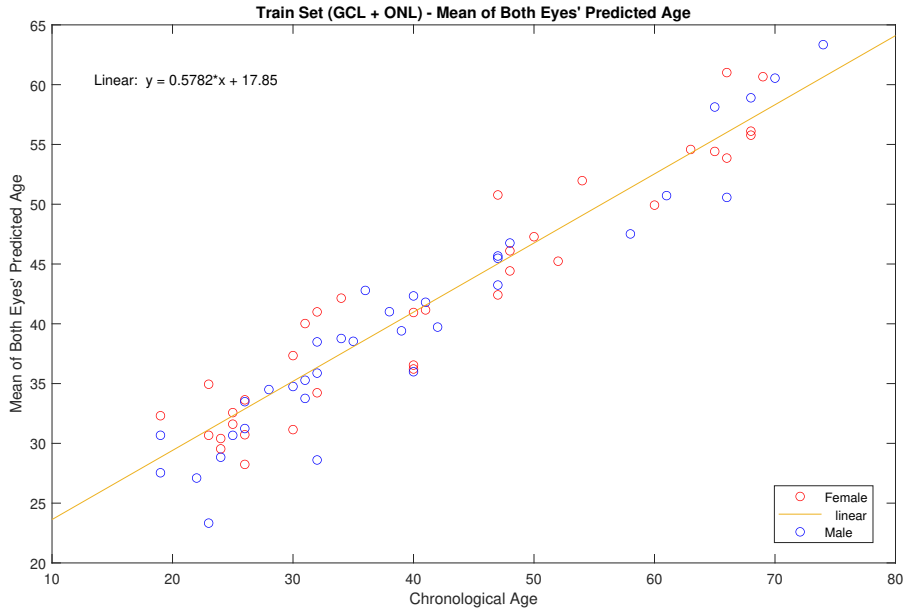


(d) Hold-Out Set - Female.

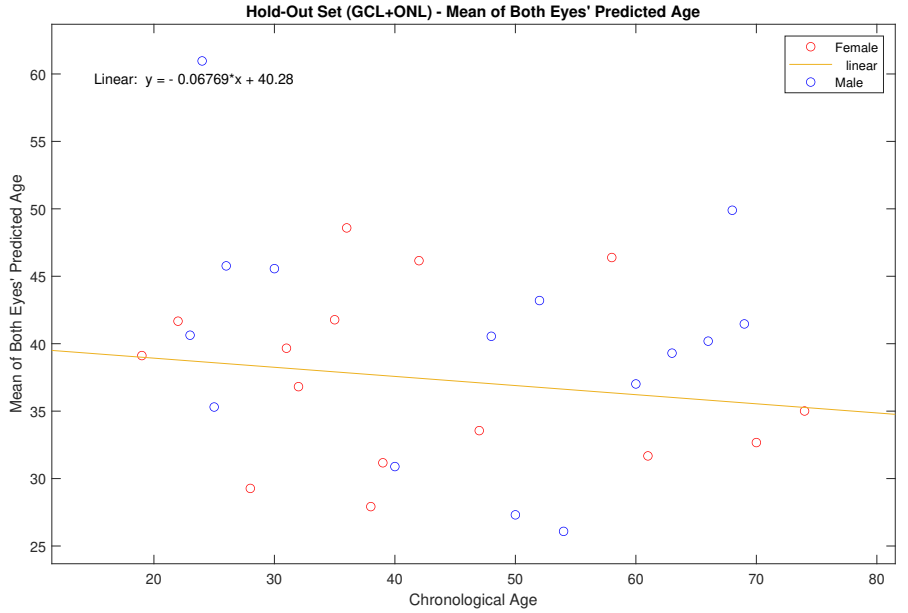
Figure 4.1.8: Mean of the predicted ages *versus* the chronological age for train and hold-out sets, separated by sex.

As can be observed, for the train set, the regression coefficient is obviously better than for the hold-out set, once again showing that the model could not extrapolate the train information for the hold-out set, which can be related with the low number of cases. However, for the train set, either for females or males, the PA from the left- and right-eyes are not significantly discrepant between them, which was expected and once again, suggests that the information about the CNS age is present in the data. In general, for younger ages, the error is smaller, validating the results given by the boxplots.

The next logical approach was to do a double validation were the mean of both eyes' PAs of an individual are combined into one and compared with the CA. The results are represented in Figure 4.1.9. For the train set, an RMSE of 7.33 and 6.45 years was obtained for the female and male groups, respectively.



(a) Train Set.



(b) Hold-Out Set.

Figure 4.1.9: Mean of both eyes predicted age for each individual *versus* the chronological age.

The results shown in Figure 4.1.8 demonstrate that the number of features may be insufficient, these do not convey enough information on the ageing process, or both, even when averaging the individual age estimates for the left- and right-eyes (Figure 4.1.9).

Discussion

At the beginning of this study, the expectation was to develop a model that would match, as closely as possible, the CNS PA to the chronological one on healthy individuals.

In the literature, the majority of the studies makes use of MRI as their primary imaging method [2, 4, 18, 19, 23, 28–32, 34, 37–39, 68]. Other studies made use of EEG signals [33] or haemochrome markers (blood) [4]. Liu *et al.* [10] used retinal (fundus) images obtained from a digital non-mydriatic fundus camera to diagnose macular pathologies using different types of image analysis (texture analysis being one of them), which proves that these kind of images and analysis can provide information on the ageing of the CNS.

One of the dilemmas that arose at the beginning of this study was if the features carried the information about the CNS age or not, since the initial NN was not producing any meaningful results. As a testbed, age groups were created to find the possibility to correctly classify eyes in the appropriate age group based on the same information, as mentioned and explained in Section 3.8. The results suggested that the information about the CNS age was present in the data.

While adjusting the parameters of the NN, the question about which weights and bias should the net be initialized with arose. There are a lot of different methods but none of them is a standard and universally accepted. That is a problem where the transfer learning method would really be useful. For the reader to be familiarized with this method, it is also a machine learning method; however the model is developed for a specific task and always reused as the starting point for another task [69]. It is trained with a significant amount of samples (running into millions) so that the weights and bias are well defined. A critical aspect of this method is that the net is usually trained with resource of the human, meaning that the classification and the validation of the images used are visually made by humans. In virtue of these facts, we can clearly understand why we cannot use transfer learning in this specific case: first, there is no

pre-existent neural network trained with similar data, and second, the used data cannot, in any way, be validated by a human.

A summary of the information about the literature may be found in Table 2.0.1 where the MAE and RMSE values, age range used and train/test ratio are represented. For the majority of the cases there was no information if the MAE/RMSE values were from the train or the test sets, so we cannot make a direct comparison with the obtained train and test values. The RMSE in the literature values vary between 5.1 and 21.01 years.

The values for the hold-out set (RMSE of 28.78 and 30.57 years for the GCL and ONL, respectively) are not within the RMSE range mentioned in the literature. However, the results cannot be considered inadequate due to the framework, since the used dataset is composed by 198 cases and an age range between 19 and 74 years old. The worst RMSE value (21.01 years) in the literature [29] also had a small dataset (414 cases) and an age range of 18-93 years old. The results in the literature tend to present a lower RMSE when the dataset size increases. There are studies with 50 cases and other with almost 25000. The reduced size of the used dataset in this study is due to the impossibility of having sufficient computational resources due to the restrictions imposed by the COVID-19 pandemic. If the dataset was not compressed, the information of the quadrants would still be available, which would result in a more complex dataset, and the results could have a different outcome - a lower RMSE or even information about which quadrants for the chosen retinal layers would have been the best. Even though these results significantly differ from the best ones achieved through more complex imaging systems (such as MRI) and many cases to train more complex networks, the figures seem, nevertheless, promising.

Another influence in the results is the train/test ratio used. We used the standard 70/30 ratio (approximately) but the literature is not coherent about it. Some use 50/50, others 80/20, 90/10, 10/90 and even variable ratios. A hold-out set was created due to the need for an external validation. The train/hold-out ratio was 69.7/30.3%. The difference between the test set and the hold-out set is that the former was made in order to not have repeated ages but to be distributed along the age range and the latter one is the same but without the possibility of having one of the individuals eyes in each set. This means that for the test set the RMSE was expected to be better than for the hold-out set, since it is expectable that if the NN has an eye from the same individual present in the training set, it will be perform better when the other eye is presented in the hold-out set. And as can be observed in Tables 4.1.1 to 4.1.3, the results confirmed that expectation. Considering the small dataset size, the percentage of cases used for the hold-out

set (30.3%) may be regarded as excessively high. However, given the wide age range (19-74), the cases sampled for the hold-out set should be evenly (and somewhat closely) spaced across the full dataset, for an accurate representation of its age distribution. This would not be achieved by any other method.

Also, the number of cases was limited leading to the reduced number of learning cases for the NN. This contributed to the non-extrapolation of the learning for the hold-out cases, and consequently to the increase in the RMSE values. Lancaster *et al.* [32] stated that larger datasets are necessary for the optimization of the work. Franke *et al.* [23] also mentioned that the size of the dataset influences the results. Cole *et al.* [17] used T1- weighted MRI data with a dataset of 2001 healthy subjects with age ranging from 18 to 90 years old, and obtained an RMSE of 5.31-6.54 years from a CNN and 5.43-15.10 years from a GPR method. Lin *et al.* [19] used MRI images from 112 healthy subjects with age ranging from 50 to 79 years old and obtained an RMSE between 4.31 and 6.14 years, depending on the artificial neural network model used. A much simpler NN was used in this project. However the hold-out results were not close, the training ones are near, which suggest promising results for future work.

Figure 4.1.1 depicts the CA *versus* PA for the train and test sets, for each of the neuroretina layers. For the training set, GCL, IPL, INL and ONL were the layers with the best performance, while for the test set the RNFL, GCL, OPL and ONL stood out. Nonetheless, having in account the low regression coefficient, it was questioned whether or not the train set size would be sufficient, since the information it contained could not be extrapolated to the hold-out set. However, as shown by the results (Figures 4.1.1a, 4.1.1c, 4.1.1e, 4.1.1g, 4.1.1i and 4.1.1k), it is suggested that the information exists in the data.

For the train set, an RMSE of 7.33 and 6.45 years for the female and male groups, respectively, was achieved. It is within the RMSE range proposed by the literature, however, it was already expectable having in account the anterior results. Some other comparisons with the literature can be made such as the dependency on biological sex proposed by Niu *et al.* [39] - this dependency can be observed in the results (Figures 4.1.8 and 4.1.9), where the male group has a smaller RMSE than the female group. Franke *et al.* [1] obtained a larger difference between the CA and PA for males than for females - however, our results did not suggest that, on the contrary. Also, the fact that older individuals present a smaller PA in comparison with other age groups [2, 19] - which can be observed in the results from the boxplots (Figures 4.1.2 to 4.1.7).

The approach developed in this work is promising, as it is based on an imaging method (OCT) that entiles a short acquisition time relatively to the methods used in the literature, leading to fewer artefact errors and a more expedite data collection process. Also, it is important to remember that for the most of the literature, the data was acquired with the MRI method, so there is no direct point of comparison with our data.

The lack in distribution present in the age range (19-74 years old) can be considered a restraint, since it influences the learning and incentives the error. In addition, while some of the subjects under study were selected and tested as healthy controls (in the scope of specific projects running in the institute), some of the individuals were not subjected to any testing before the OCT acquisition, as they were volunteers or companions of patients visiting the institute to realize other exams. For these sporadic cases, the subjects' word on their healthy status was the only thing taken in consideration to classify them as healthy individuals. No control about caffeine intake, tobacco consuming, or other substances was obtained. These factors also influence the age prediction as mentioned in [1, 2, 4], which might have lead to errors or bias in the results. Also, the compression and simplification of the dataset made it clearly difficult for the NN to learn and could have hidden some important information. Due to time limitation and lack of computational resources as a result of the COVID-19 pandemic, some of the simplifications and decisions were made in order to simplify the procedure and without them, different results could have been generated.

Conclusion

From the work developed herein, it can be concluded that information on the CNS age is present in OCT data, even though a high level of data compression and simplification was applied, providing good perspectives for future work. It was also discovered that not all retinal layers are necessary to estimate the CNS age. Similarly, not all the features under study are required, as some of them provide the exact same information and can be discarded (depending on the retinal layer). The layers of the neuroretina contributing the most to CNS age estimation were found to be the GCL and the ONL. Some differences were found between female and male healthy controls: the male group had a smaller RMSE for the estimated age than the female group. For the same individual, the difference between eyes is not severely significant, which can be regarded as a validation of the CNS age information being present in the data. These findings represent an encouraging discovery that might simplify the estimation age of the CNS since the imaging method used is not as time-consuming as MRI or other methods, and presents a decreased probability of movement artefacts due to its reduced acquisition time. Once the developed model is optimized to estimate the CNS age with insignificant RMSE and high accuracy in healthy controls, the extrapolation of the information to patients diagnosed with diabetes or neurodegenerative diseases like AD or PD should be straightforward. The integration of OCT scanning in primary care settings would allow diagnosing those diseases years before their manifestation and without submitting the patients to MRI scans. A simple OCT scan should do the job and improve the life quality of the patient. For future work it is clear that the limitations mentioned in Chapter 5 ought to be overcome. In addition, the methodology should be improved in further studies. The introduction of patients diagnosed with AD, PD and other neurodegenerative disorders would be particularly intriguing. Finally, the use of CNNs to directly analyse the images obtained from OCT data would be interesting. Those images would represent a step further from the extensively compressed dataset used in this study (composed of features that summarize the information contained in OCT data, which were further simplified by quadrant averaging and feature selection/reduction).

Bibliography

[1] K. Franke, C. Gaser, B. Manor, and V. Novak, "Advanced BrainAGE in older adults with type 2 diabetes mellitus," *Frontiers in Aging Neuroscience*, vol. 5, no. DEC, pp. 1–9, 2013. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnagi.2013.00090/abstract

[2] K. Ning, L. Zhao, W. Matloff, F. Sun, and A. W. Toga, "Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants," *Scientific Reports*, vol. 10, no. 1, p. 10, 12 2020. [Online]. Available: http://www.nature.com/articles/s41598-019-56089-4

[3] K. Franke and C. Gaser, "Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained?" *Frontiers in Neurology*, vol. 10, no. JUL, 8 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneur.2019.00789/full

[4] A. Gialluisi, A. Di Castelnuovo, M. B. Donati, G. de Gaetano, L. Iacoviello, and Moli-sani Study Investigators, "Machine Learning Approaches for the Estimation of Biological Aging: The Road Ahead for Population Studies." *Frontiers in medicine*, vol. 6, p. 146, 7 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmed.2019.00146/full

[5] S. Rodríguez-Rodero, J. L. Fernández-Morera, E. Menéndez-Torre, V. Calvanese, A. F. Fernández, and M. F. Fraga, "Aging genetics and aging." *Aging and disease*, vol. 2, no. 3, pp. 186–95, 6 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22396873

[6] H. Sajedi and N. Pardakhti, "Age Prediction Based on Brain MRI Image: A Survey," *Journal of Medical Systems*, vol. 43, no. 8, p. 279, 2019. [Online]. Available: http://link.springer.com/10.1007/s10916-019-1401-7

[7] L. Thau, V. Reddy, and P. Singh, *Anatomy, Central Nervous System*, 2020. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/31194336

[8] F. M. Mar, A. Bonni, and M. M. Sousa, "Cell intrinsic control of axon regeneration," *EMBO reports*, vol. 15, no. 3, pp. 254–263, 3 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/embr.201337723

[9] P. Guimarães, P. Rodrigues, C. Lobo, S. Leal, J. Figueira, P. Serranho, and R. Bernardes, "Ocular fundus reference images from optical coherence tomography," *Computerized Medical Imaging and Graphics*, vol. 38, no. 5, pp. 381–389, 7 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0895611114000329

[10] C. Liu, W. Wang, Z. Li, Y. Jiang, X. Han, J. Ha, W. Meng, and M. He, "Biological Age Estimated from Retinal Imaging: A Novel Biomarker of Aging," in *Miccai*, ser. Lecture Notes in Computer Science, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, vol. 11764, no. October, pp. 138–146. [Online]. Available: http://link.springer.com/10.1007/978-3-030-32239-7_16

[11] R. Bernardes, G. Silva, S. Chiquita, P. Serranho, and A. F. Ambrósio, "Retinal Biomarkers of Alzheimer's Disease: Insights from Transgenic Mouse Models," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10317 LNCS, pp. 541–550. [Online]. Available: http://link.springer.com/10.1007/978-3-319-59876-5_60

[12] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and J. M. Rehg, "Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding," *Medical Image Analysis*, vol. 15, no. 5, pp. 748–759, 10 2011. [Online]. Available: http://dx.doi.org/10.1016/j.media.2011.06.005

[13] A. Nunes, P. Serranho, H. Quental, A. F. Ambrósio, M. Castelo-Branco, and R. Bernardes, "Sexual dimorphism of the adult human retina assessed by optical coherence tomography," *Health and Technology*, vol. 10, no. 4, pp. 913–924, 7 2020. [Online]. Available: http://link.springer.com/10.1007/s12553-020-00428-3

[14] A. Nunes, G. Silva, C. Alves, S. Batista, L. Sousa, M. Castelo-Branco, and R. Bernardes, "Textural information from the retinal nerve fibre layer in multiple sclerosis *," in *2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG)*. IEEE, 2 2019, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/8692454/

[15] A. Nunes, G. Silva, C. Duque, C. Januário, I. Santana, A. F. Ambrósio, M. Castelo-Branco, and R. Bernardes, "Retinal texture biomarkers may help to discriminate between Alzheimer's, Parkinson's, and healthy controls," *PLOS ONE*, vol. 14, no. 6, p. e0218826, 6 2019. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0218826

[16] A. Nunes, P. Serranho, H. Quental, A. F. Ambrósio, M. Castelo-Branco, and R. Bernardes, "Sexual Dimorphism of the Adult Human Retina Assessed by Optical Coherence Tomography," in *IFMBE Proceedings*, ser. IFMBE Proceedings, J. Henriques, N. Neves, and P. de Carvalho, Eds. Cham: Springer International Publishing, 2020, vol. 76, pp. 1830–1834. [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-030-31635-8_222

[17] J. H. Cole, R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector, and G. Montana, "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker," *NeuroImage*, vol. 163, no. March, pp. 115–124, 12 2017. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2017.07.059

[18] B. A. Jonsson, G. Bjornsdottir, T. E. Thorgeirsson, L. M. Ellingsen, G. B. Walters, D. F. Gudbjartsson, H. Stefansson, K. Stefansson, and M. O. Ulfarsson, "Brain age prediction using deep learning uncovers associated sequence variants," *Nature Communications*, vol. 10, no. 1, p. 5409, 12 2019. [Online]. Available: http://www.nature.com/articles/s41467-019-13163-9

[19] L. Lin, C. Jin, Z. Fu, B. Zhang, G. Bin, and S. Wu, "Predicting healthy older adult's brain age based on structural connectivity networks using artificial neural networks," *Computer Methods and Programs in Biomedicine*, vol. 125, pp. 8–17, 3 2016. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2015.11.012

[20] B. Widrow and M. Lehr, "30 years of adaptive neural networks: perceptron, Madaline, and backpropagation," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415–1442, 1990. [Online]. Available: http://ieeexplore.ieee.org/document/58323/

[21] M. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015. [Online]. Available: https://books.google.pt/books?id=STDBswEACAAJhttp://neuralnetworksanddeeplearning.com/index.html

[22] G. Di Franco and M. Santurro, "Machine learning, artificial neural networks and social research," *Quality & Quantity*, no. 0123456789, 9 2020. [Online]. Available: https://doi.org/10.1007/s11135-020-01037-y

[23] K. Franke, G. Ziegler, S. Klöppel, and C. Gaser, "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *NeuroImage*, vol. 50, no. 3, pp. 883–892, 4 2010. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2010.01.005

[24] "Ixi dataset - brain development," http://brain-development.org/ixi-dataset/.

[25] "Adni - alzheimer's disease neuroimaging initiative," http://adni.loni.usc.edu/.

[26] M. R. Sabuncu and K. Van Leemput, "The Relevance Voxel Machine (RVoxM): A Bayesian Method for Image-Based Prediction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6893 LNCS, no. PART 3, pp. 99–106. [Online]. Available: http://link.springer.com/10.1007/978-3-642-23626-6_13

[27] "Oasis brains - open access series of imaging studies," https://www.oasis-brains.org/.

[28] M. R. Sabuncu and K. Van Leemput, "The Relevance Voxel Machine (RVoxM): A Self-Tuning Bayesian Model for Informative Image-Based Prediction," *IEEE Transactions on Medical Imaging*, vol. 31, no. 12, pp. 2290–2306, 12 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6307878/

[29] E. Konukoglu, B. Glocker, D. Zikic, and A. Criminisi, "Neighbourhood approximation using randomized forests," *Medical Image Analysis*, vol. 17, no. 7, pp. 790–804, 10 2013. [Online]. Available: http://dx.doi.org/10.1016/j.media.2013.04.013

[30] F. Liem, G. Varoquaux, J. Kynast, F. Beyer, S. Kharabian Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock, S. Riedel-Heller, T. Luck, M. Loeffler, M. L. Schroeter, A. V. Witte, A. Villringer, and D. S. Margulies, "Predicting brain-age from multimodal imaging data captures cognitive impairment." *NeuroImage*, vol. 148, no. November 2016, pp. 179–188, 3 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811916306103

[31] S. Valizadeh, J. Hänggi, S. Mérillat, and L. Jäncke, "Age prediction on the basis of brain anatomical measures," *Human Brain Mapping*, vol. 38, no. 2, pp. 997–1008, 2 2017. [Online]. Available: http://doi.wiley.com/10.1002/hbm.23434

[32] J. Lancaster, R. Lorenz, R. Leech, and J. H. Cole, "Bayesian Optimization for Neuroimaging Pre-processing in Brain Age Classification and Prediction," *Frontiers in Aging Neuroscience*, vol. 10, p. 28, 2 2018. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnagi.2018.00028/full

[33] O. Al Zoubi, C. Ki Wong, R. T. Kuplicki, H.-w. Yeh, A. Mayeli, H. Refai, M. Paulus, and J. Bodurka, "Predicting Age From Brain EEG Signals—A Machine Learning Approach," *Frontiers in Aging Neuroscience*, vol. 10, no. JUL, pp. 1–12, 7 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnagi.2018.00184/full

[34] J. Zhai and K. Li, "Predicting Brain Age Based on Spatial and Temporal Features of Human Brain Functional Networks," *Frontiers in Human Neuroscience*, vol. 13, no. February, pp. 1–17, 2 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnhum.2019.00062/full

[35] "Enhanced nathan kline institute - rockland sample," http://fcon_1000.projects.nitrc.org/indi/enhanced/.

[36] "Uk biobank," https://www.ukbiobank.ac.uk/.

[37] D. Hu, Z. Wu, W. Lin, G. Li, and D. Shen, "Hierarchical Rough-to-Fine Model for Infant Age Prediction based on Cortical Features," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8632694/

[38] D. Sone, I. Beheshti, N. Maikusa, M. Ota, Y. Kimura, N. Sato, M. Koepp, and H. Matsuda, "Neuroimaging-based brain-age prediction in diverse forms of epilepsy: a signature of psychosis and beyond," *Molecular Psychiatry*, pp. 1–10, 6 2019. [Online]. Available: http://www.nature.com/articles/s41380-019-0446-9

[39] X. Niu, F. Zhang, J. Kounios, and H. Liang, "Improved prediction of brain age using multimodal neuroimaging data," *Human Brain Mapping*, vol. 41, no. 6, pp. 1626–1643, 4 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24899

[40] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and J. M. Rehg, "Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid with local binary patterns." *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 13, no. Pt 1, pp. 1–9, 2010. [Online]. Available: http://link.springer.com/10.1007/978-3-642-15705-9_1

[41] R. Bernardes, P. Serranho, T. Santos, V. Gonçalves, and J. C. Vaz, "Optical Coherence Tomography - Automatic Retina Classification Through Support Vector Machines," *European Ophthalmic Review*, vol. 06, no. 04, p. 200, 2012. [Online]. Available: https://repositorioaberto.uab.pt/handle/10400.2/2765

[42] G. Jin, X. Ding, W. Xiao, X. Xu, L. Wang, X. Han, O. Xiao, R. Liu, W. Wang, W. Yan, L. An, J. Zhao, and M. He, "Prevalence of age-related macular degeneration in rural southern China: the Yangxi Eye Study," *British Journal of Ophthalmology*, vol. 102, no. 5, pp. 625–630, 5 2018. [Online]. Available: https://bjo.bmj.com/lookup/doi/10.1136/bjophthalmol-2017-310368

[43] World Medical Association (WMA), "World Medical Association. World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects," *Journal International de Bioéthique*, vol. 15, no. 1, p. 124, 1 2004. [Online]. Available: https://www.degruyter.com/doi/10.1515/9783110208856.233

[44] G. Virgili, F. Menchini, G. Casazza, R. Hogg, R. R. Das, X. Wang, and M. Michelessi, "Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy," *Cochrane Database of Systematic Reviews*, vol. 2015, no. 1, 1 2015. [Online]. Available: http://doi.wiley.com/10.1002/14651858.CD008081.pub3

[45] S. Athira, R. M. Roy, and R. Aneesh, "Computerized Detection of Macular Edema Using OCT Images Based on Fractal Texture Analysis," in *2018 International CET Conference on Control, Communication, and Computing (IC4)*, vol. 00. IEEE, 7 2018, pp. 326–330. [Online]. Available: https://ieeexplore.ieee.org/document/8530952/

[46] N. Mekhasingharak, P. Laowanapiban, S. Siritho, C. Satukijchai, N. Prayoonwiwat, J. Jitprapaikulsan, and N. Chirapapaisan, "Optical coherence tomography in central nervous system demyelinating diseases related optic neuritis," *International Journal of Ophthalmology*, vol. 11, no. 10, pp. 1649–1656, 10 2018. [Online]. Available: http://www.ijo.cn/gjyken/ch/reader/view_abstract.aspx?file_no=20181012&flag=1

[47] J. P. Cunha, R. Proença, A. Dias-Santos, D. Melancia, R. Almeida, H. Águas, B. O. Santos, M. Alves, J. Ferreira, A. L. Papoila, C. Louro, and A. Castanheira-Dinis, "Choroidal thinning: Alzheimer's disease and aging," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 8, pp. 11–17, 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352872917300192

[48] H. Ferreira, J. Martins, A. Nunes, P. I. Moreira, M. Castelo-Branco, A. F. Ambrósio, P. Serranho, and R. Bernardes, "Characterization of the retinal changes of the 3Tg-AD mouse model of Alzheimer's disease," *Health and Technology*, vol. 10, no. 4, pp. 875–883, 7 2020. [Online]. Available: http://link.springer.com/10.1007/s12553-020-00413-w

[49] A. Nunes, A. F. Ambrosio, M. Castelo-Branco, and R. Bernardes, "[Regular Paper] Texture Biomarkers of Alzheimer's Disease and Disease Progression in the Mouse Retina," in *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 10 2018, pp. 41–46. [Online]. Available: https://ieeexplore.ieee.org/document/8567455/

[50] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis." *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5660089/http://www.ncbi.nlm.nih.gov/pubmed/22275207http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3131209

[51] M. Garvin, M. Abramoff, Xiaodong Wu, S. Russell, T. Burns, and M. Sonka, "Automated 3-D Intraretinal Layer Segmentation of Macular Spectral-Domain Optical Coherence Tomography Images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1436–1447, 9 2009. [Online]. Available: http://ieeexplore.ieee.org/document/4799172/

[52] Kang Li, Xiaodong Wu, D. Chen, and M. Sonka, "Optimal Surface Segmentation in Volumetric Images-A Graph-Theoretic Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 119–134, 1 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1542036/

[53] L. Armi and S. Fekri-Ershad, "Texture image analysis and texture classification methods - A review," no. January, 4 2019. [Online]. Available: http://arxiv.org/abs/1904.06554

[54] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 1 2002. [Online]. Available: http://www.tandfonline.com/doi/abs/10.5589/m02-004

[55] R. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979. [Online]. Available: http://ieeexplore.ieee.org/document/1455597/

[56] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 780–795, 3 1999. [Online]. Available: http://ieeexplore.ieee.org/document/752194/

[57] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 11 1973. [Online]. Available: http://library.seg.org/doi/10.1190/segam2015-5927230.1

[58] R. W. Conners, M. M. Trivedi, and C. A. Harlow, "Segmentation of a high-resolution urban scene using texture operators," *Computer Vision, Graphics, and Image Processing*, vol. 25, no. 3, pp. 273–310, 3 1984. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0734189X8490197X

[59] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 11 2005. [Online]. Available: http://ieeexplore.ieee.org/document/1550194/

[60] S. Wang, S. Lu, Z. Dong, J. Yang, M. Yang, and Y. Zhang, "Dual-Tree Complex Wavelet Transform and Twin Support Vector Machine for Pathological Brain Detection," *Applied Sciences*, vol. 6, no. 6, p. 169, 6 2016. [Online]. Available: http://www.mdpi.com/2076-3417/6/6/169

[61] D. S. Badde, A. Gupta, and V. K. Patki, "Cascade and Feed Forward Back propagation Artificial Neural Network Models for Prediction of Compressive Strength of Ready Mix Concrete," *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, no. 2278-1684, pp. 1–6, 2009. [Online]. Available: http://www.iosrjournals.org/iosr-jmce/papers/sicete(civil)-volume3/26.pdf

[62] X. Li and X. Wu, "Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2015-Janua, pp. 3219–3223, 10 2014. [Online]. Available: http://arxiv.org/abs/1410.4281

[63] J. Gomm and D. Yu, "Selecting radial basis function network centers with recursive orthogonal least squares training," *IEEE Transactions on Neural Networks*, vol. 11, no. 2, pp. 306–314, 3 2000. [Online]. Available: http://ieeexplore.ieee.org/document/839002/

[64] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 5 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/3065386

[65] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 7 2019. [Online]. Available: https://doi.org/10.1016/j.patcog.2019.02.023

[66] G. Hripcsak and A. S. Rothschild, "Agreement, the F-measure, and reliability in information retrieval," *Journal of the American Medical Informatics Association*, vol. 12, no. 3, pp. 296–298, 2005.

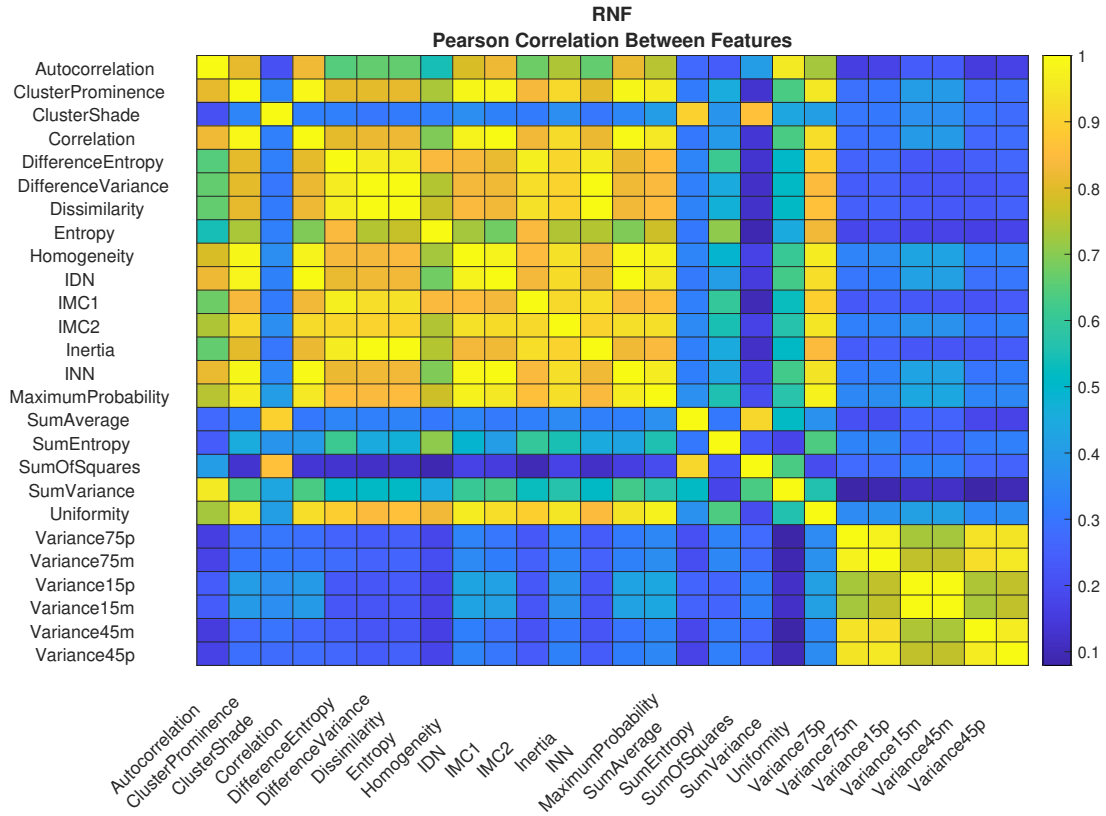
[67] B. Warsito, R. Santoso, Suparti, and H. Yasin, "Cascade Forward Neural Network for Time Series Prediction," *Journal of Physics: Conference Series*, vol. 1025, no. 1, p. 012097, 5 2018. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1025/1/012097

[68] J. H. Cole, S. J. Ritchie, M. E. Bastin, M. C. Valdés Hernández, S. Muñoz Maniega, N. Royle, J. Corley, A. Pattie, S. E. Harris, Q. Zhang, N. R. Wray, P. Redmond, R. E. Marioni, J. M. Starr, S. R. Cox, J. M. Wardlaw, D. J. Sharp, and I. J. Deary, "Brain age predicts mortality." *Molecular psychiatry*, vol. 23, no. 5, pp. 1385–1392, 5 2018. [Online]. Available: http://dx.doi.org/10.1038/mp.2017.62

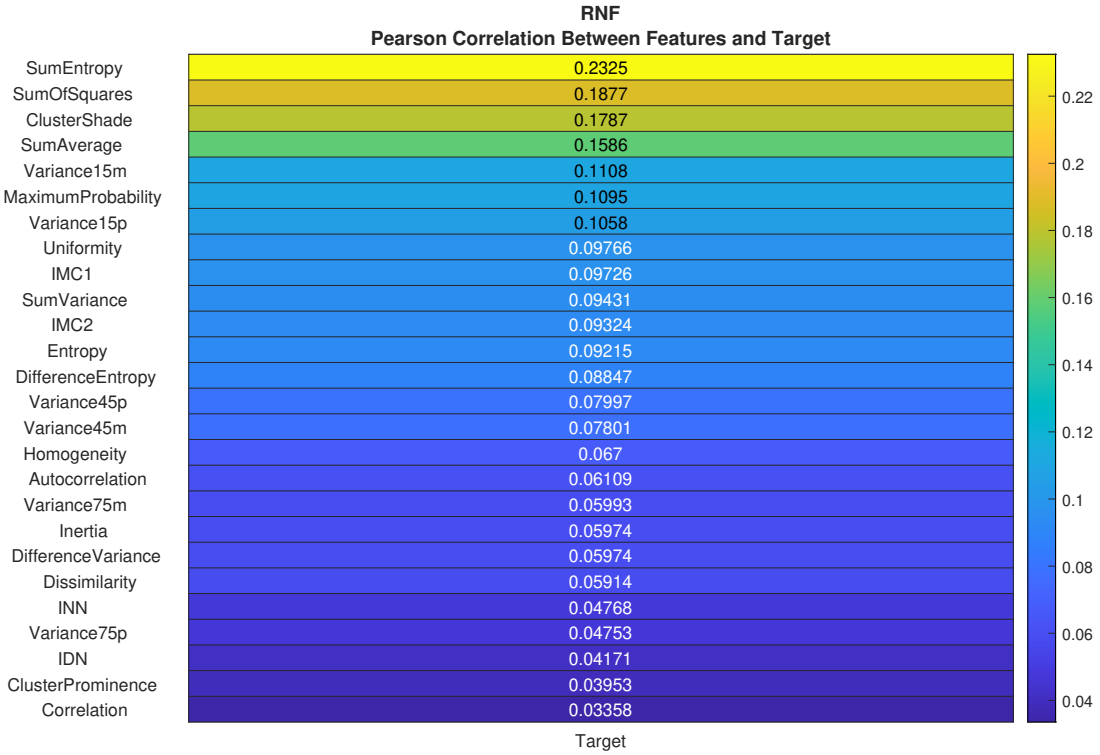
[69] L. Torrey and J. Shavlik, "Transfer Learning," in *Handbook of Research on Machine Learning Applications and Trends*. IGI Global, 2014, pp. 242–264. [Online]. Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60566-766-9.ch011

Appendices

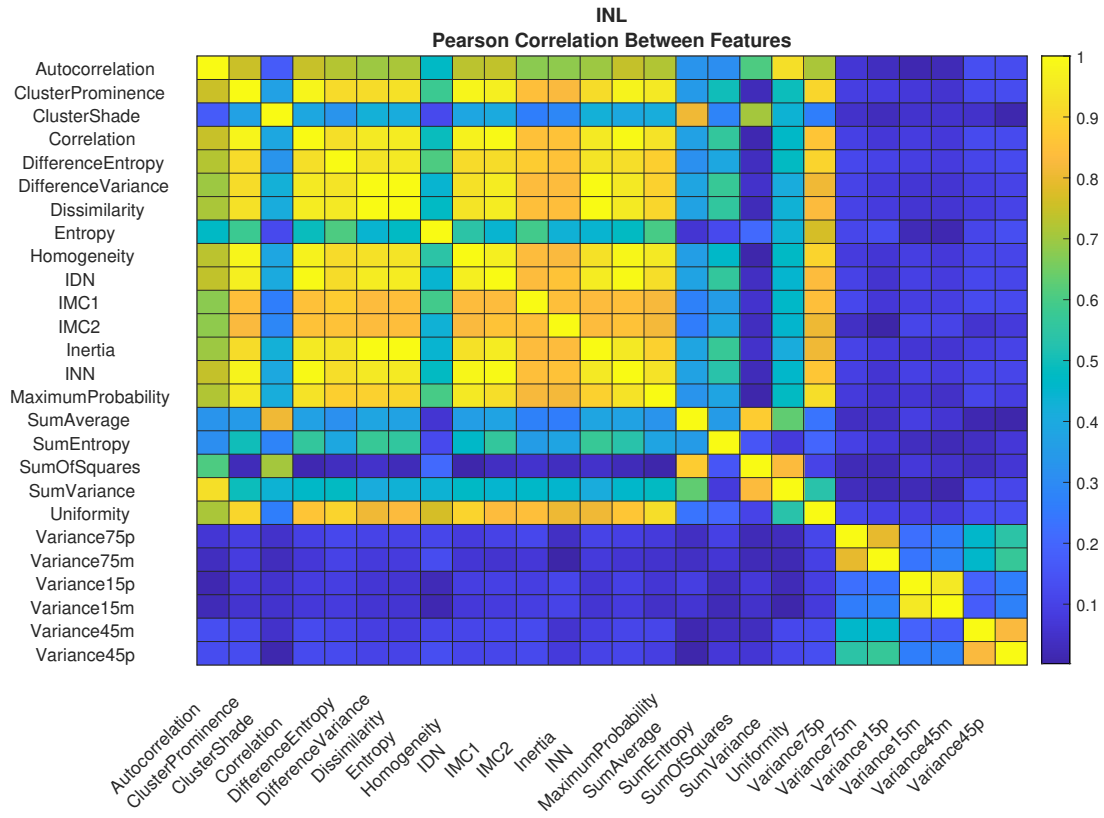
I.1 Appendix A - Heatmaps



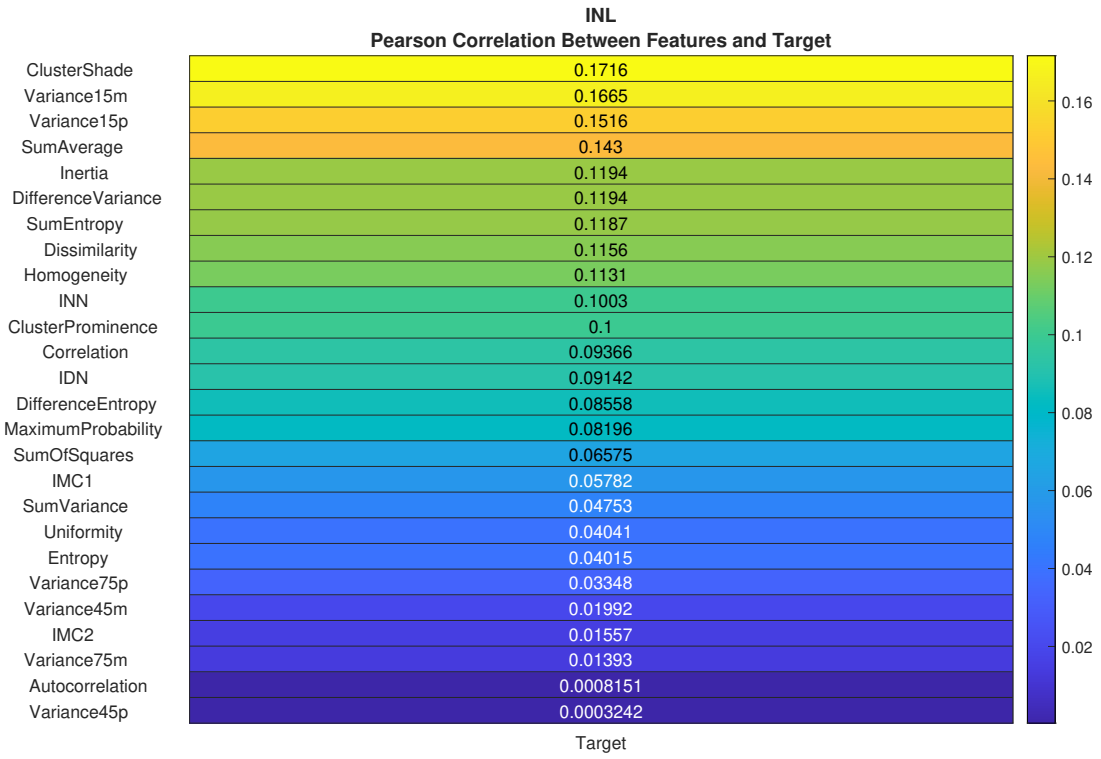
(a) RNFL - Correlation between features.



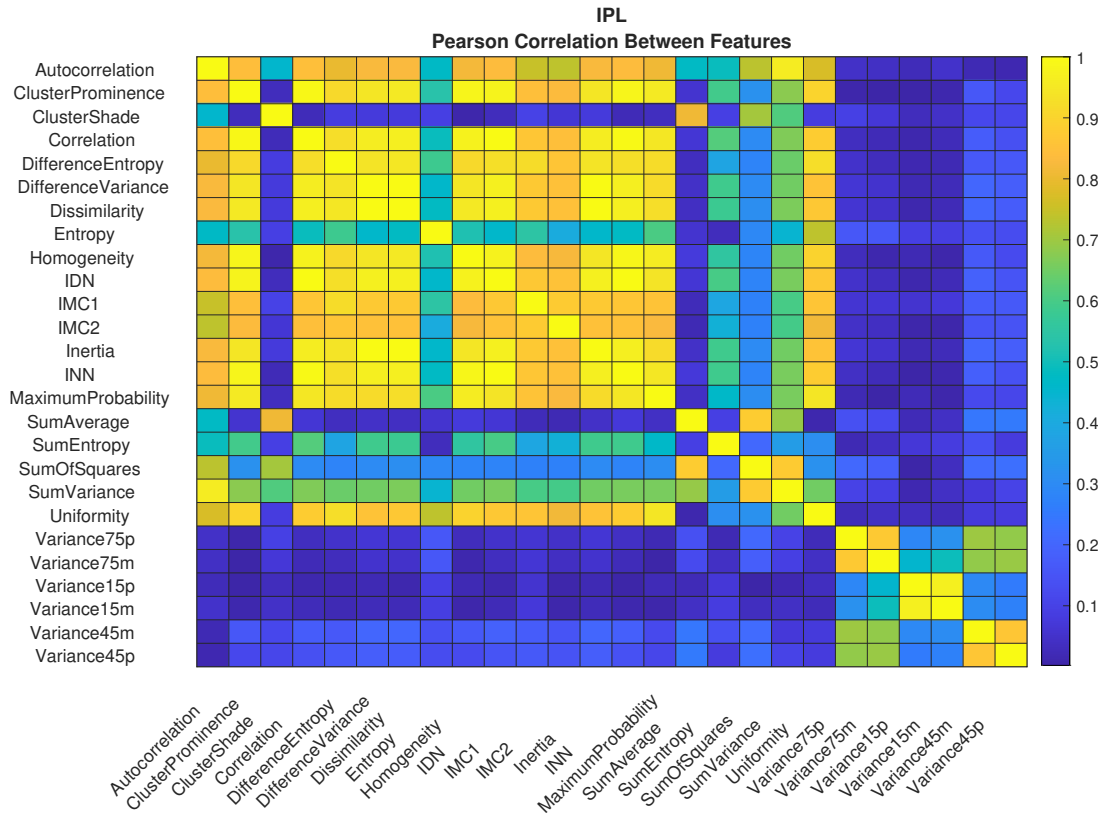
(b) RNFL - Correlation features-target.



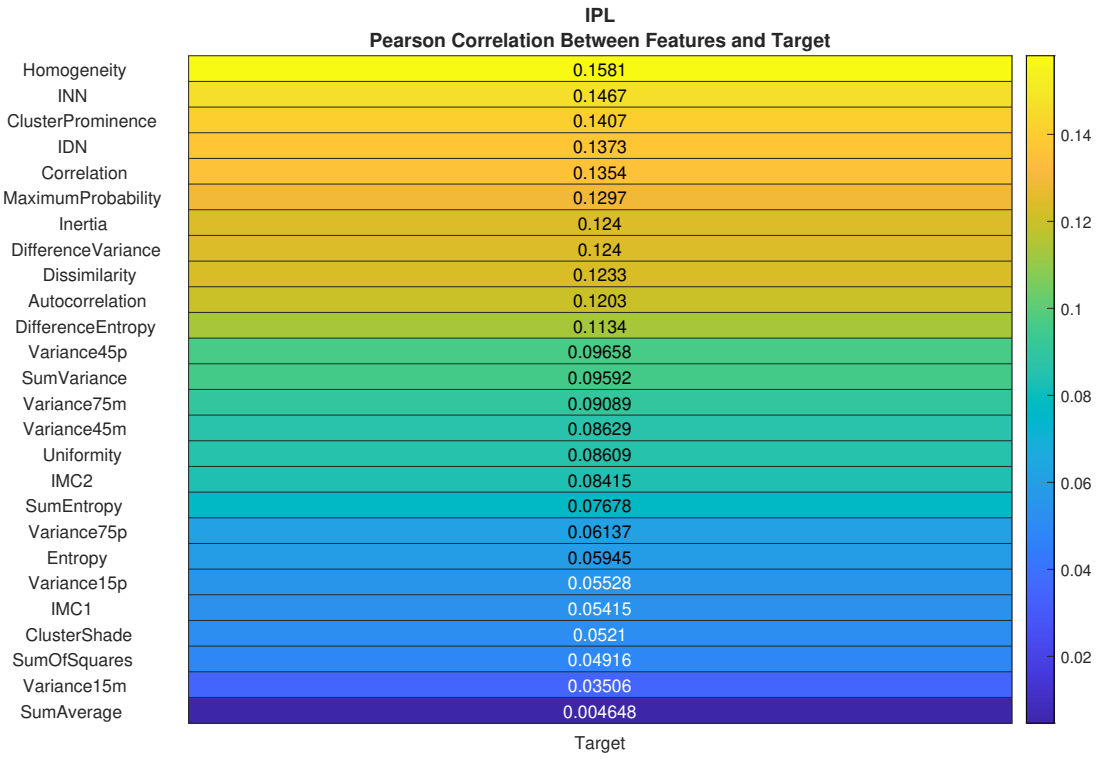
(c) INL - Correlation between features.



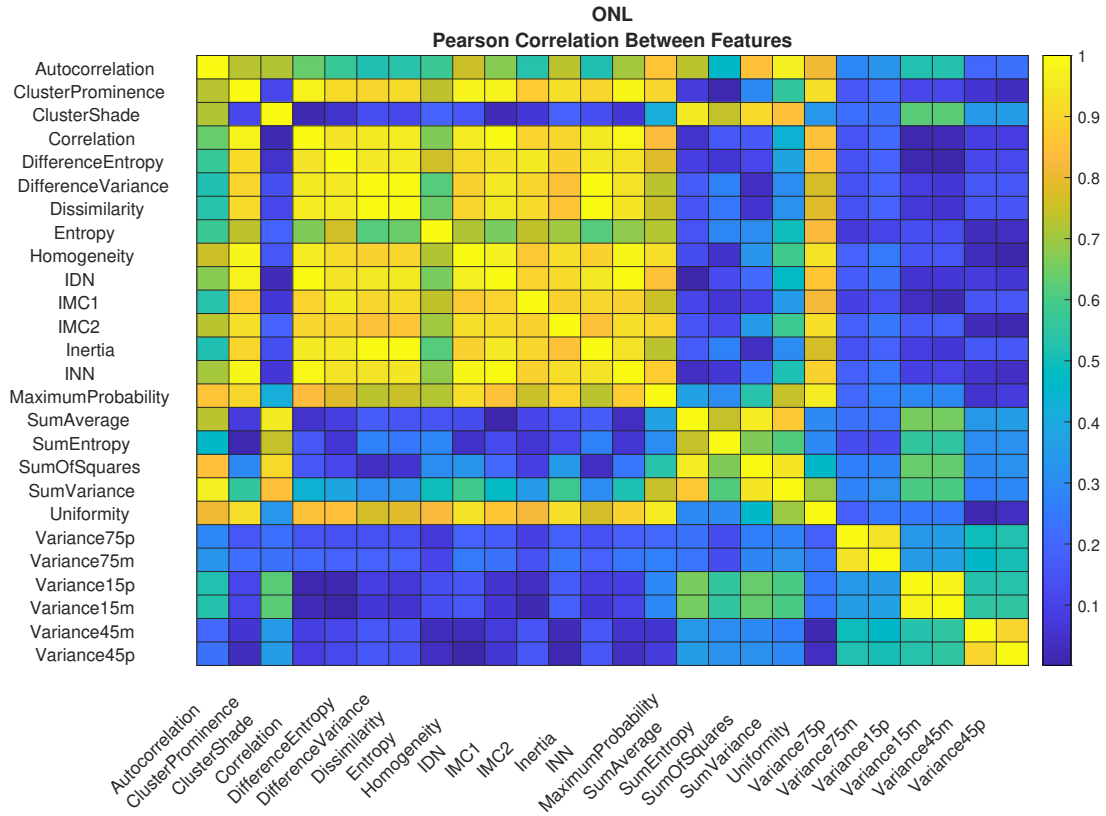
(d) INL - Correlation features-target.



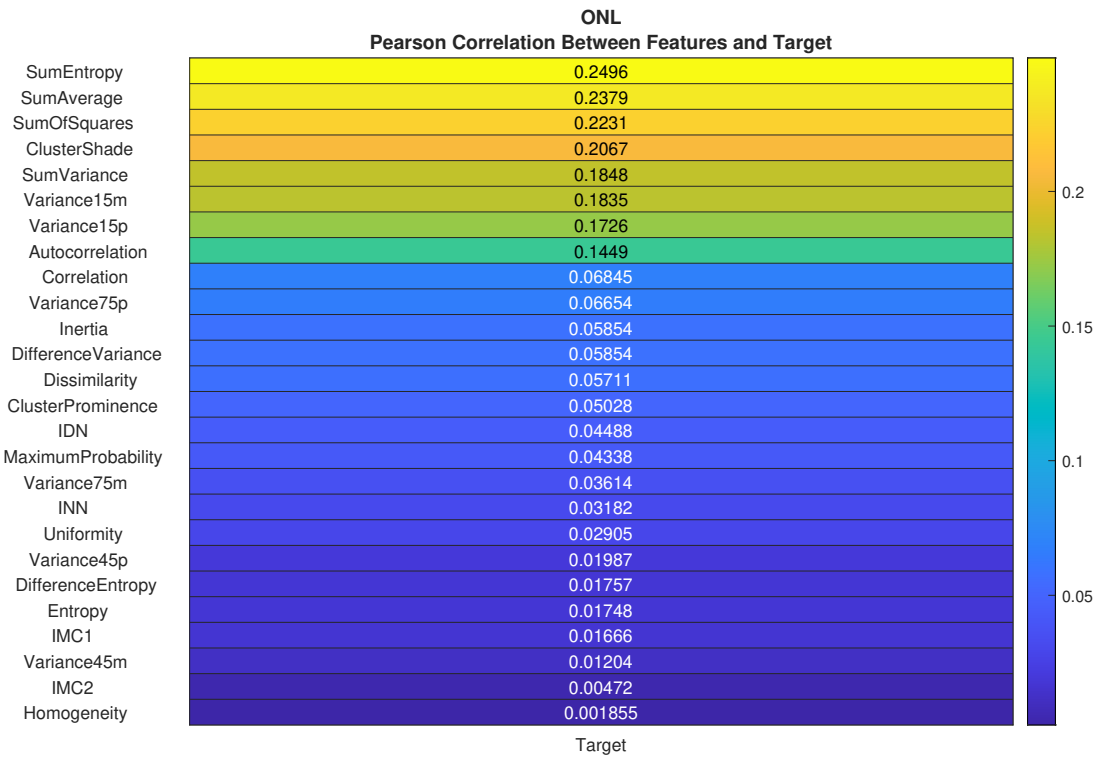
(e) IPL - Correlation between features.



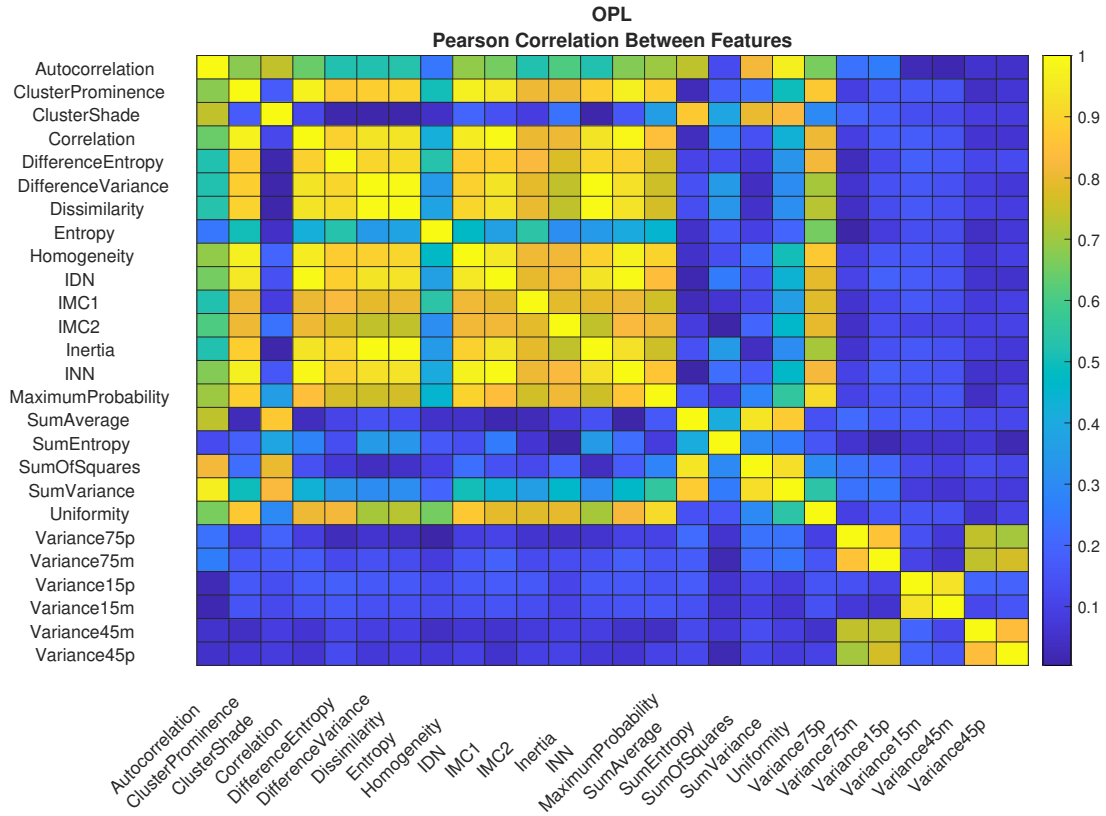
(f) IPL - Correlation features-target.



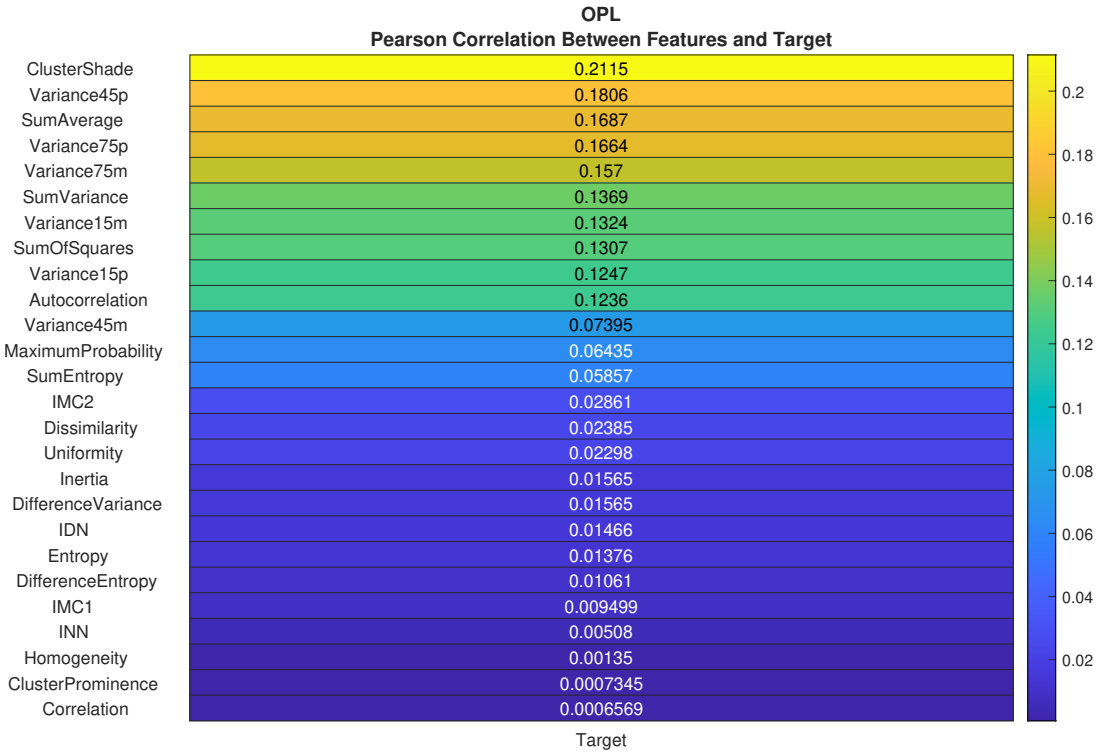
(g) ONL - Correlation between features.



(h) ONL - Correlation features-target.



(i) OPL - Correlation between features.



(j) OPL - Correlation features-target.

Figure I.1.0: Pearson Correlation between features and between features and target, for each retinal layer.